Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of

Doctor of Natural Sciences

presented by
Maximilian Beckers
B.Sc. (Mathematics), B.Sc. M.Sc. (Biochemistry)
born in Laichingen
Oral Examination: 04 March 2020

Statistical Inference of cryo-EM Maps

Referees:
Dr. Martin Beck
Prof. Dr. Robert Russell

" .. orthodoxy is the death of knowledge, since the growth of knowledge depends entirely on the existence of disagreement"

Karl Popper, The Myth of the Framework

Chapter 1

Abstract

1.1 Summary

Investigation of the structure and function of biological matter provides one of the most fascinating aspects of nature: how life with all its complex processes evolves from simple wiggling and jiggling of atoms. Structural Biology is especially interesting, as it involves all three major fields of the natural sciences: the physics of atoms and the scattering of incident radiation, the chemistry of bonding, reactions and interactions and finally the biological interpretation of the chemical structures. However, tools from statistical decision theory are still rare in structural biology compared to their applications in e.g. ”omics”-technologies. Especially when it comes to the interpretation of 3D maps of single particles, more rigorous approaches would be highly desired.

In this thesis I developed new methodology for the statistical inference of three-dimensional molecular maps generated with cryogenic electron microscopy (cryo-EM), which is becoming the method of choice for high-resolution structure determination. The aspects I focused on are very basic topics of the method and were not properly solved so far.

The first part of the dissertation deals with the problem of assigning molecules and molecular parts in noisy 3D densities. Based on multiple hypothesis testing and false discovery rate control, the cryo-EM map is transformed in a map containing detection probabilities, termed confidence map. Confidence maps allow the assignment of map features by means of statistical significance and avoid rather arbitrary thresholds, as they have to be used for the analysis of cryo-EM maps.

Another important quantity that has to be inferred from cryo-EM maps is their resolution. It is probably the most important number regarding the quality of the map and describes up to which spatial frequency we can faithfully interpret the data. However, despite its importance and implications for map processing, resolution estimation remains a highly controversial issue. I describe how the resolution estimation problem can be reformulated into multiple hypothesis testing of Fourier shell correlation coefficients and how this results in a unified threshold criterion applicable to global, local, directional and map-model resolution estimation.

The last chapter provides a high-resolution study of tobacco mosaic virus (TMV). Although structural studies of tobacco mosaic virus by Rosalind Franklin marked one of the birth events of structural biology, the most important biological aspect of controlled release of the viral genome remained a mystery for more than 50 years. Two cryo-EM structures at 1.9 and 2 Å resolution together with the developed confidence maps allowed us to elucidate the structural mechanism of the disassembly switch of TMV. Moreover, it may also serve as a general case study how careful interpretation with state-of-the art methods can be used to deal with densities that are notoriously difficult to model.

1.2 Zusammenfassung

Untersuchungen der Struktur und Funktion belebter Materie stellen einen der faszinierendsten Aspekte des Universums dar: wie sich das Leben mit seiner ungemeinen Vielfalt an orchestrierten Prozessen aus dem einfachen "wackeln" der Atome entwickelt. Insbesondere die Strukturbiologie ist hierbei interessant, da sie alle drei großen Felder der Naturwissenschaften vereint: die Atomphysik und die Streuung von Strahlung, die Chemie der Bindungen, Reaktionen und Interaktionen sowie letztlich die biologische Interpretation der chemischen Strukturen. Trotz der großen Bedeutung der Strukturbiologie sind Verfahren der statistischen Entscheidungstheorie, insbesondere im Vergleich mit deren Anwendungen in den "Omics"-Technologien, immer noch relativ selten in ihr anzutreffen, insbesondere wenn es zur Interpretation der 3D Karten einzelner Partikel kommt.

In dieser Dissertation habe ich neue Verfahren der statistischen Inferenz von dreidimensionalen molekularen Karten, welche mit kryogener Elektronenmikroskopie (Kryo-EM) aufgenommen werden können, entwickelt. Kryo-EM ist mittlerweile die Methode der Wahl wenn es um hoch-aufgelöste Strukturbestimmung geht. Bei den Aspekten, an denen ich mich zu arbeiten entschied, handelt es sich um grundlegende Dinge der Methode, welche ungenügend gelöst waren.

Der erste Teil dieser Dissertation behandelt das Problem in verrauschten 3D Kryo-EM Karten Moleküle und Teile davon zu annotieren. Basierend auf multiplem Testen statistischer Hypothesen und Kontrolle der falsch-positiv Rate wird die Kryo-EM Karte in eine Karte von Detektionswahrscheinlichkeiten transformiert, welche wir als Konfidenzkarten bezeichnen. Diese Konfidenzkarten erlauben die Annotierung von Teilen basierend auf statistischer Signifikanz und umgehen die Wahl beliebieger Niveaus, welche bei der Analyse von einfachen Kryo-EM Karten verwendet werden müssen.

Eine weitere wichtige Kennzahl, die von Kryo-EM Karten geschätzt werden muss, ist deren Auflösung. In Bezug auf die Qualität der Karte handelt es sich hierbei wohl um die wichtigste Größe; sie beschreibt bis zu welcher räumlichen Frequenz wir die Daten sicher interpretieren können. Trotz ihrer Bedeutung und Verwendung in der weiteren Prozessierung der Karten, ist die Schätzung der Auflösung immer noch ein kontrovers diskutierter Prozess. Ich beschreibe ich einen neuen Ansatz, wie das Problem in eines des multiplen Hypothesetestens von Fourier-Hüllen Korrelationskoeffizienten umformuliert werden kann und wie selbes zu einem einheitlichen Kriterium für die Auflösungsbestimmung führt.

Zuletzt beschreibe ich noch eine hoch-aufgelöste Studie von Tabak-Mosaik Virus (TMV). Obwohl Rosalind Franklins Studien zur Struktur von TMV zusammen mit dem Beginn des gesamten Gebiets der Strukturbiologie fielen, blieb der biologisch wichtigste Aspekt, nämlich die kontrollierte Freigabe des viralen Genoms, für mehr als 50 Jahre ein Rätsel, da die beteiligten Reste sehr flexibel sind und darüber hinaus anfällig für Strahlungschäden. Zwei Kryo-EM Strukturen mit Auflösungen von 1.9 und 2 Å, zusammen mit den entwickelten Konfidenzkarten erlaubten uns den Mechanismus im Detail aufzuklären. Darüber

hinaus liefert diese Studie auch ein allgemeines Beispiel, wie sorgfältige Interpretation mit modernen Methoden benutzt werden kann um schwierige Teile von Molekülen zu modellieren.

Contents

List of Abbreviations

ADP Atomic displacement parameters

CCPEM Collaborative computational project for cryo-EM

CDF Cumulative distribution function

CP Coat protein

cryo-EM cryogenic electron microscopy

CTF Contrast transfer function

ECDF Empirical cumulative distribution function

EM Electron microscopy

EMBL European molecular biology laboratory

EMDB Electron microscopy data bank

ER Endoplasmatic Reticulum

FSC Fourier shell correlation

FDR False discovery rate

FFT Fast Fourier transform

FWER Family wise error rate

GUI Graphical user interface

HA Hemagglutinin

HDA High density amorphous ice

IFT Inverse Fourier tranform

kV kilovolt

KS Kolmogorow-Smirnow

LDA Low density amorphous ice

MAP Maximum-a-posteriori

MCMC Markov Chain Monte Carlo

ML Maximum likelihood

MSE Mean squared error

NMR Nuclear magnetic resonance

PC Phosphatidylcholin

PDB Protein data bank

PETG Phenylethyl beta-D-thiogalactopyranoside

pFDR Positive false discovery rate

PPV Positive predictive value

RMSD Root mean square deviation

SNR Signal-to-noise ratio

SPA Single particle analysis

SPoC Statistical processing of cryo-EM maps

SSNR Spectral signal-to-noise ratio

STA Sub-tomogram averaging

TMV Tobacco mosaic virus

TRP1 Transient receptor potential cation channel subfamily V member 1

Chapter 2

Publications Covered in This Thesis

Maximilian Beckers, Colin Palmer and Carsten Sachse. Confidence Maps - Statistical inference of cryo-EM maps. *Acta Cryst. Sect. D* (submitted)

Felix Weis*, Maximilian Beckers*, Iris von der Hocht and Carsten Sachse. Elucidation of the viral disassembly switch of tobacco mosaic virus. *EMBO Reports* (2019) **20**:e48451
*equal contribution

Maximilian Beckers, Arjen Jakobi and Carsten Sachse. Thresholding of cryo-EM maps by false discovery rate control. *IUCrJ* (2019) **6**, 18-33

Chapter 3

Introduction

3.1 Electron Microscopy of Biological Specimens

The three dimensional structure of a protein is the main determinant for its function. With the advent of X-ray crystallography, structural studies of biological molecules became a major discipline in order to understand how these diverse complexes carry out all the basic functions in a living cell. Electron microscopy (EM) was used from quite early on for biological studies [43, 42]. However, compared to X-ray crystallography or NMR, it failed to yield high-resolution information in the beginning, which is a prerequisite for confident three dimensional assignment of individual atom positions. Although limitations of both crystallographic and NMR methods in dealing with big and more flexible macromolecular complexes became obvious, electron microscopy remained a minor technique. Three big topics had to be optimized to make high-resolution studies with EM possible: efficient reconstruction of a 3D map from 2D projections, preserving the biological structures including hydration of the molecules in an electron microscope and improved electron detection at low dose rates.

Although powerful reconstruction algorithms and vitrification protocols of biological specimens were developed quite early, it was until the introduction of direct electron detectors in 2013, which allowed the acquisition of dose fractioned movies, that we faced several near-atomic resolutions structures, not only for highly symmetrical viruses but also for smaller and less symmetric particles. Now commonly known as the "resolution revolution" [78], this marked the point where cryogenic electron micrscopy (cryo-EM) became one of the methods of choice for biological structure determination.

Since then, high-resolution cryo-EM has been used to visualize the 3D structures of long-hunted structures like diverse RNA-polymerase complexes [115, 63], ribosomes [72], spliceosomes [105], membrane proteins [155] and neurodegenerative filaments [41]. Finally in 2017, these developments culminated in the Nobel prize in Chemistry for Richard Henderson, Joachim Frank and Jaques Dubochet for their studies in the development of cryo-EM. At the Electron Microscopy Data Bank (EMDB) this resolution boost is expressed through an exponentially growing number of deposited cryo-EM maps with steady improving resolutions [1]. In 2019, ca. 10000 cryo-EM maps have been submitted to the EMDB with a resolution record of 1.54 Å [1], which highlights the importance cryo-EM gained in the field of structural biology.

3.1.1 Image Formation in an Electron Microscope

The actual image formation in an electron microscope is the result of incident electrons interacting with the sample of interest as well as with the microscope itself. Interactions of the electrons with the specimen can be distinguished in elastic and inelastic. Elastic scattering involves no transfer of energy to the specimen, i.e. the kinetic energy of the incoming electron is conserved, while inelastic scattering involves energy transfer [51]. While elastic scattering is the result of interactions with the electrostatic potential of the atom nuclei screened by the outer electrons in the atom, inelastic scattering is a less localized process due to plasmon and interband excitations [119]. Roughly spoken, elastic interactions are mainly the cause of interactions with the nucleus, which has a much higher mass than the electrons (99.9% of the mass is concentrated in the atom nucleus) and, in concordance with the Born-Oppenheimer approximation [17], can be roughly treated invariant from the electrons. Elastic scattering is then the result of conservation of momentum. In contrast, interactions with other electrons involve interplay between particles of similar mass, which will exchange energy.

Compared to elastic scattering, inelastic scattering is a much less localized process. Thus, the main source of structural information content in the images is delivered from elastic scattering. However, as the inelastically scattered electrons consequently mainly contribute noise, it can be advantageous to filter them out by means of energy filtering [163].

The Lambert-Beer law [11] provides us a simple theoretical framework for electron-specimen interactions, similar to the case of light. Based on this, the probability density function $p(x)$ for a inelastic (elastic) scattering event after travelling a distance x through the specimen is given as a exponential distribution

$$p(x) = \frac{1}{\lambda}\exp(-\frac{x}{\lambda}), \tag{3.1}$$

where λ is the so called mean free path for inelastic (elastic) scattering. Given a specimen thickness d, the probability $F(d)$ of inelastic (elastic) scattering for an electron travelling through this specimen is then given as

$$F(d) = \int_0^d p(x)dx = 1 - \exp(-\frac{\lambda}{d}). \tag{3.2}$$

Calculating the expectation value of this distribution, it can be seen that the mean free paths for elastic and inelastic scattering indeed characterize the expected travelling distance between two consecutive scattering events in the specimen. Therefore, they provide a simple measure for the strength of the electron-specimen interaction. An important feature of the mean free paths is their dependence on the acceleration voltage, which thus forms the main rational behind the specific choice of the used voltages [119].

Amplitude and Phase Contrast

Related to the interactions with the specimen, two main sources of contrast are typically distinguished: amplitude and phase contrast. Amplitude contrast results from attenuation of the intensity of the incoming electron beam, i.e. absorption of electrons in the specimen, scattering outside the objective aperture or as the result of an energy filter [120]. Phase contrast is a more complex process and is a result of the interference of the unscattered with the scattered electrons [163].

Phase contrast itself is a rather old and established concept, originally developed in light microscopy in the first half of the 20th century by Frits Zernike [164]. In cryo-EM of vitrified specimens it represents the main source of contrast in the images, while amplitude contrast usually only plays a minor role.

However, phase contrast needs to be generated, and, in an aberration free microscope in focus, nothing would be visible besides noise. The common way to generate phase contrast is to use a defocus, which was investigated in detail by Otto Scherzer. More sophisticated approaches were developed with Zernike [32] and Volta phase plates [31, 30, 33]. Especially Volta phase plates result in dramatically improved contrast, which was shown to be beneficial for the analysis of small particles [73] and thicker specimens like tomograms [91]. However, for general single particle analysis they do not seem to lead to improved resolutions of the resulting maps.

A very interesting approach termed laser phase plate was developed very recently, where the phases of the electrons are shifted by means of a high-intensity laser beam [130, 99]. This allows direct tuning of the phase shifts and avoids issues of the Volta phase like electrostatic charging and adverse scattering. While this seems to be a very promising direction, it requires a modified microscope architecture with a high-intensity laser including mirrors incorporated in the electron microscope. Thus, wide acceptance might still lay ahead a few years.

A Mathematical Description of Image Formation

Mathematically, the image formation process is usually stated in a wavefunction description [74, 151]. The wavefunction Ψ_{exit} of the incident electron waveform Ψ_{in} after passing through a thin specimen is given in the phase-grating approximation [74] as

$$\Psi_{exit}(x, y) = \exp(i\sigma v(x, y))\Psi_{in}, \tag{3.3}$$

where i is the imaginary unit. The phase shift $\sigma v(x, y)$ at point x, y in the projection is then given as a line integral of the electrostatic potential $V(x, y, z)$ along the beam multiplied by $\sigma = 2\pi me\lambda/h^2$, where m is the relativistic mass, e is the magnitude of the charge on an electron, λ the wavelength of the electron and h Planck's constant. For a beam parallel to the z-axis this becomes

$$v(x, y) = \int_{-\infty}^{\infty} V(x, y, z)dz. \tag{3.4}$$

Amplitude contrast is usually modelled as the imaginary part of the electrostatic potential $V(x, y, z)$. However, biological specimens usually exhibit almost no amplitude contrast and inelastically scattered electrons are filtered out with an energy filter [163], which makes the phase contrast the dominant source of contrast in EM images.

In addition to modulation by the specimen, the exit wavefunction is further modulated by the aberration function of the microscope. The aberration function is easiest to stater in Fourier space, and is in Fourier space as a function of the spatial frequency k given as

$$\chi(k) = \pi\lambda k^2(0.5C_{S3}\lambda^2k^2 - \Delta f), \tag{3.5}$$

with λ the electron wavelength, Δf the defocus and C_{S3} the third-order spherical aberration. The aberration function χ quantifies the phase shift of each spatial frequency due to the microscope. Application of the phase shifts to each pixel in the Fourier transform corresponds to multiplication with

$$\mathrm{CTF}(k) = \exp(-i\chi(k)), \tag{3.6}$$

where CTF is the so called contrast transfer function. The CTF describes how each pixel in the Fourier transform of the image is modified from the microscope. It has to be noted, that, for the sake of simplicity, astigmatism has been ignored in the given description. A depiction of a simulated CTF is shown in Fig. 3.1, where high resolutions are attenuated with a realistic B-factor of 50 Å^2.

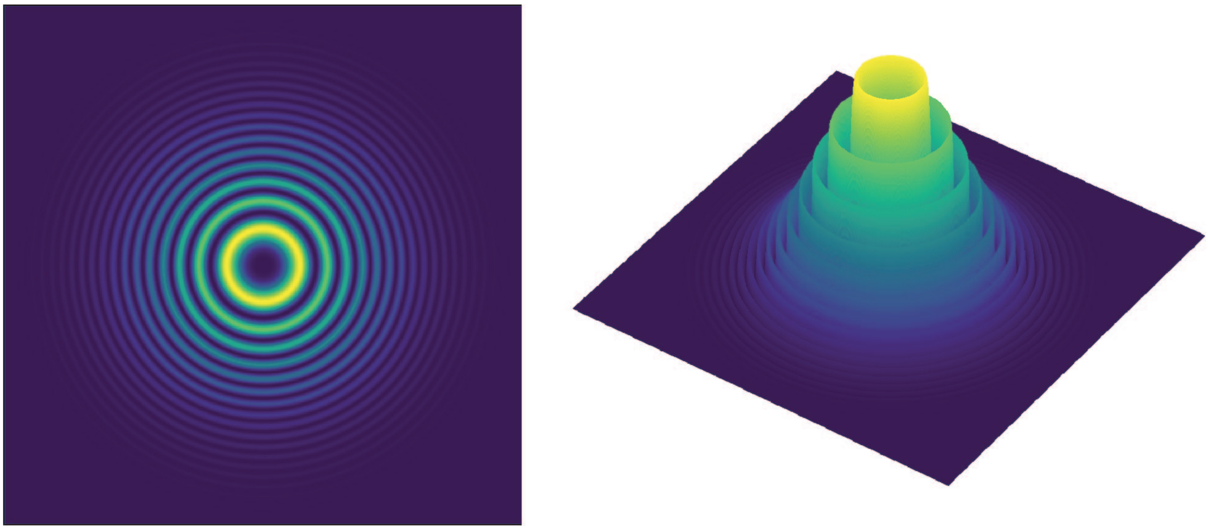


Figure 3.1: Power spectrum of the contrast transfer function. 2D (left) and 3D (right) views of the power spectrum of the CTF with a third-order spherical aberration of 2.7 mm and a defocus of 500 nm at 300 kV. High resolutions are attenuated with a realistic B-factor falloff of 50 Å^2

With this, the final wavefunction Ψ_{final} is given as

$$\Psi_{final}(x, y) = \Psi_{exit}(x, y) \otimes \mathrm{IFT}[\exp(-i\chi(k))], \tag{3.7}$$

where $\otimes$ denotes convolution and IFT the inverse Fourier transform. The probability density function of the image intensities $\mathrm{I}_0(x, y)$ is then given by

$$\mathrm{I}_0(x, y) = |\Psi_{final}(x, y)|^2. \tag{3.8}$$

It has to be noted that the presented image formation model is rather simplistic and ignores aspects like defocus spread, which would further complicate the situation. Moreover, electrons accelerated at 300 kV travel at 77% of the speed of light, which makes it necessary to include relativistic effects. Fortunately, a full relativistic treatment based on the Dirac equation gives similar results as the Schroedinger equation with relativistic corrections [79], which is the framework from which the phase grating approximation can be derived from [152]. However, the presented framework captures the most important aspects and was chosen for the sake of readability. It should be mentioned here as well, that the phase-grating approximation, which is also known as the phase-object or projection assumption, should not be confused with the weak-phase object assumption. The weak-phase object assumption is rather unimportant conceptually, but can be used to simplify the equations I presented.

3.1.2 Cryo-Electron Microscopy of Vitrified Aqueous Specimens

Due to small amplitude contrast and weak phase contrast of biological macromolecules in an electron beam, imaging of biological specimens has for a long time been limited to negatively stained samples. However, biological specimens in their native state are in an aqueous environment, but the visualization of aqueous samples in an electron microscope posed several big challenges: at room temperature the water evaporates in the vacuum needed for imaging with electrons. Moreover, at lower temperatures the sample starts freezing and makes crystalline ice, thereby destroying the native structure of the embedded macromolecules. A workaround was found in 1981 at EMBL Heidelberg, where Alisdair McDowall and Jaques Dubochet managed to vitrify pure water [36], i.e. to freeze water in a non-crystalline state without any long-range order. This opened the way for electron microscopic investigations of vitrified specimens.

The vitrification of water can for example be achieved by plunge freezing a thin layer of aqueous solution in liquid ethane [36]. Due to the rapid plunging process, the freezing happens so fast that no ice crystals can form. Two forms of vitreous ice have been reported: high-density amorphous ice (HDA) and low-density amorphous ice (LDA). While HDA was first obtained by compressing normal ice at liquid nitrogen temperature [97], LDA seems to be the form of ice observed during plunge freezing. Moreover, HDA can be transformed to LDA and reverse [98]. However, the exact nature of the different forms of amorphous ice is not clear. Recent results suggest that HDA is a kinetically arrested state [144], something which might also hold for LDA, i.e. HDA and LDA are not clearly defined states but snapshots of frozen water on its way to some crystalline form of ice.

Once vitrified, the sample can be imaged in a cryo-capable electron microscope. However, vitrification of thicker samples is still problematic. The thicker the sample, the harder complete vitrification becomes and the following build-up of crystalline ice makes the respective regions useless. Moreover, the vitrification process is still not always reproducible and requires further automation in the future.

Vitreous Ice in an Electron Beam

The complete interaction processes of the incident electrons with vitreous water and how it reacts is still not fully understood. Radiation damage of water molecules will generate reactive species that likely react with each other and might form water again [51]. Moreover, this will lead to continuous rearrangement of water molecules. This hypothesis could be supported by means of an analysis of the power spectra from amorphous ice [96] and explains the missing Thon rings at resolutions around the water ring (3-4 Å). Based on this, it can be estimated that water is moved by a mean squared distance of ca. 1 Å for every electron at an acceleration voltage of 300 kV.

However, not only the water molecules show beam-induced motion, but also the complete ice layer including the embedded biological macromolecules. With the advent of direct electron detectors and the possibility to acquire dose-fractionated movies, it became feasible to correct for this movement, with highly improved resolutions of the resulting reconstructions [19, 23, 82, 167, 170]. While specimen doming seems to be a major source of the particle movement [19, 167], additional mechanisms like charging effects and beam deflection are likely to contribute as well [51]. However, the actual source of the specimen doming is still investigated. It is believed that contraction of the grid holes deforms the ice layer [19], which can be supported with the observation that gold grids show substantially less movement compared to carbon [124]. However, expansion of the ice layer due to radiolysis and increased internal pressure [52] as well as relieve of stress, which is accumulated during the rapid freezing process, seem to play a role as well [132]

Towards the perfect Acceleration Voltage

High-end image acquisition is today typically done at acceleration voltages of 300 kV where the mean free path for inelastic scattering of vitreous ice is $\approx$ 400 nm [163, 120], while the mean free path for elastic scattering is $\approx$ 900 nm. Inelastic scattering consequently happens more often than elastic scattering. It can also be observed that for realistic ice-thickness of $\approx$ 30 nm, a big fraction (ca. 85%) of the electrons does not interact with the specimen at all. Recent attempts also showed that lower acceleration voltages of 200kV [61][60] and 100 kV [103] can be used for high-resolution imaging, with possible advantages for smaller particles due to improved contrast and also improved information per unit damage [110]. Additionally, 100 kV and 200 kV microscopes are more cost efficient. However, at higher voltages the scattering cross sections for both elastic and inelastic scattering decrease, which allows to use a larger electron exposure and prevents plural scattering [51]. For thicker samples, where multiple scattering events become limiting factors, 300 kV microscopes still offer important advantages compared to lower acceleration voltages.

3.1.3 3D Reconstruction of cryo-EM Maps

High-resolution structural studies of biological macromolecules by electron microscopy poses the problem of calculating a 3D reconstruction from 2D EM images. Two commonly used approaches can be distinguished: single-particle analysis (SPA) and sub-tomogram averaging (STA). However, before going into more detail of SPA and STA, I will briefly introduce the Fourier slice theorem at this point.

The Fourier Slice Theorem

The Fourier slice theorem, sometimes also referred to as central or projection slice theorem, is a central mathematical theorem that relates the projections from a ray-transform to the Fourier transformation of the projected object and was originally derived by Ronald Bracewell [18]. In the context of cryo-EM, it states that the Fourier transform of a 2D projection (in our case with a simple phase object the projection is a ray-transform) is equal to a 2D slice of the 3D Fourier that stands orthogonal to the projection direction and goes through the origin.

Denoting the 3D and 2D Fourier transforms with F_3 and F_2, respectively, the projection operator with P and the slice operator with S, the theorem can be mathematically stated as

$$F_2 P = S F_3. \tag{3.9}$$

A graphical illustration with a simple duck map can be found in Fig. 3.2. Algorithmically, the theorem provides an easy way to generate the Fourier transformation of projections without ever generating a projection. Moreover, it also provides a powerful theoretical construct that allows a simple understanding of the effects of preferred orientations and other artefacts that come from incomplete sampling like the missing wedge.

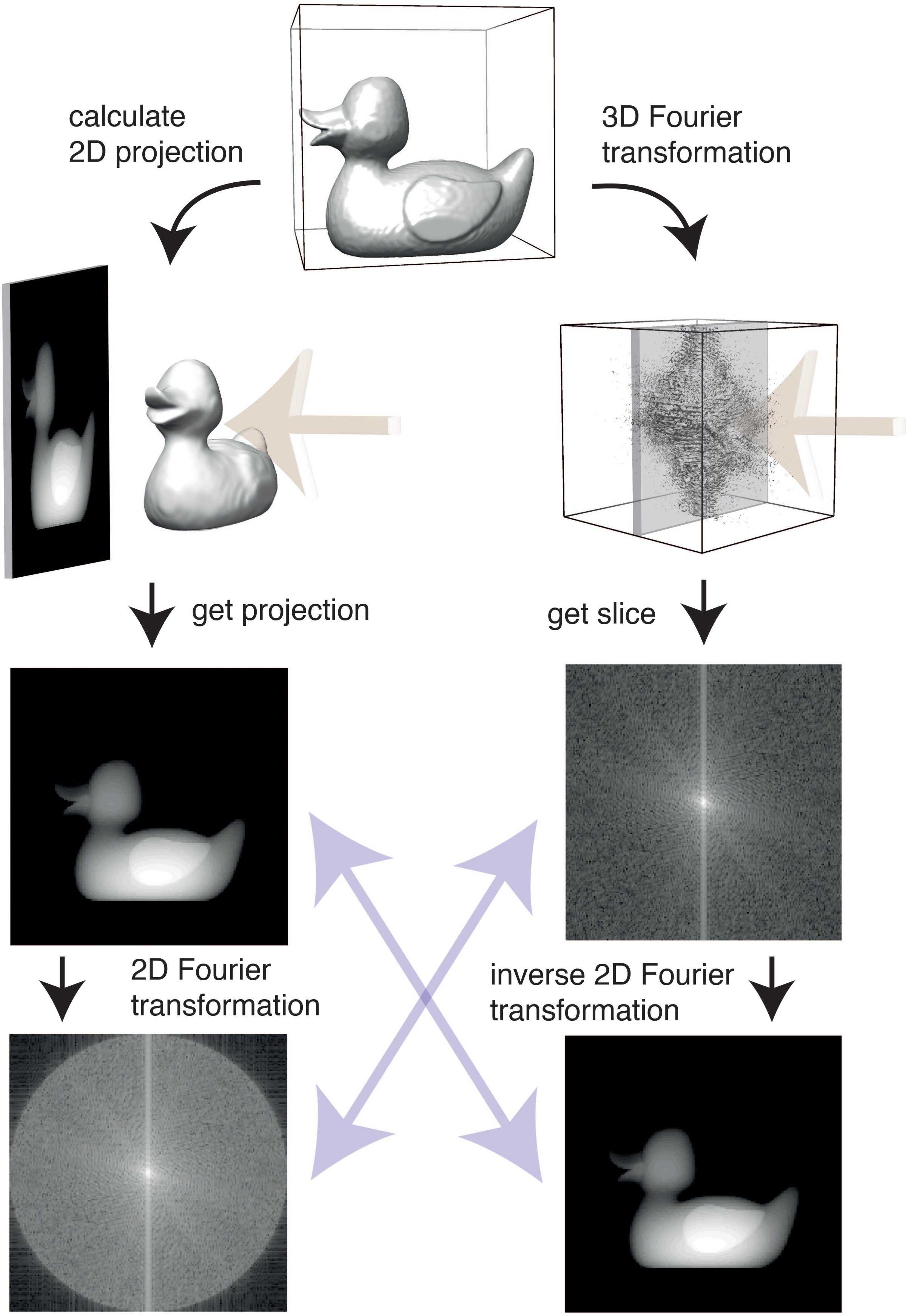


Figure 3.2: Illustration of the Fourier slice theorem with a 3D duck density.

Single Particle Analysis and Sub-Tomogram Averaging

SPA aims to estimate the 3D map from a set of random 2D projections. In addition to statistical estimation of the 3D reconstruction, orientation parameters of the individual projections need to be estimated as well. The basic work flow works by iterating two steps: first comparing projections of the current estimate of the 3D map to the experimental images in order to get estimates of their orientation and second calculating a new 3D map from the newly determined orientations of the images. Some milestones with this respect, only to mention a few, were the introduction of Projection Matching [111], Maximum Likelihood as a target function [133] and simple-to-use software protocols [128, 55, 117, 54].

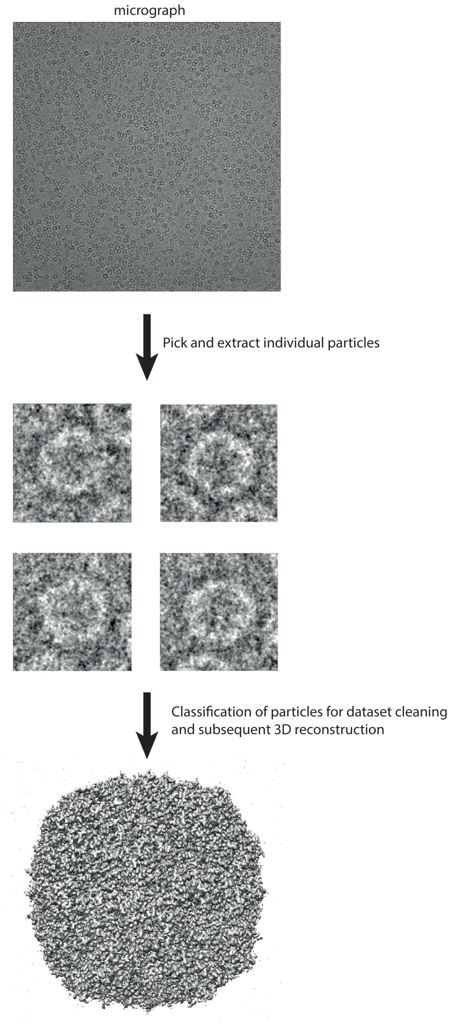


Figure 3.3: Basic work flow of single particle analysis. Particles are picked from the micrographs and extracted as small images for each particle. Multiple rounds of 2D and 3D classification are usually carried out in order to clean the dataset before the high-resolution 3D reconstruction is performed.

While in SPA the final 3D map is directly reconstructed from 2D projections, in STA a tomogram of the sample is first reconstructed. Tilting the sample in the microscope leads to known orientations of the 2D projections that can then be used for reconstruction of the tomogram. Individual particles are subsequently picked in the tomogram and aver-

aged in order to obtain the final 3D map [153].
SPA is the method of choice to obtain high-resolution structures of purified particles. Although high resolutions are in principle possible with STA [145], STA does not yet achieve this goal routinely. However, cryo-electron tomography in combination with focussed ion beam milling and STA offers the possibility to visualize individual macromolecules *in situ* [102, 91]. This is a powerful approach, as it allows the analysis of the organisation and interactions with other structures inside the cell. It is obvious that both SPA and STA methods can be used complementary: rough architectures can be determined with STA *in situ* while high-resolution structures can be determined of sub-complexes by SPA or other methods.

3.1.4 Analysis and Interpretation of cryo-EM Maps

The final result of the 3D reconstruction process for both SPA and STA is a 3D map which contains a description of the Coulomb potential of the structure of interest. However, the ultimate goal is the description of the 3D map by an atomic model with which we want to understand biochemical mechanisms, but deriving meaningful atomic coordinates from the map is a daunting task, especially as model building from cryo-EM maps was impeded by limited resolutions for a long time. In the following, an introduction to processes of resolution estimation, post-processing of the maps and atomic model refinement will be given.

Resolution Estimation

One of the first steps in the analysis work flow is estimation of the resolution. The resolution describes the degree of detail that we are able to see in the reconstruction. This is an important quantity, as it provides a target for sample, data acquisition and reconstruction optimization, it makes different maps comparable and also guides the different post-processing steps, which are necessary to get the most out of the 3D maps.
Resolution estimation is usually done by means of Fourier shell correlation (FSC) curves, which was introduced by van Heel in 1986 [57]. The FSC analysis starts on the reconstruction level by splitting the image data in two halves and reconstructing two maps, so called halfmaps. Although the "gold-standard" approach is often referred to calculating two completely independent reconstructions, diverging orientations of the two halfmaps are usually prevented by keeping low-resolution the same in both maps, which makes them not completely independent [129]. The FSC curve is then calculated by Fourier transforming both halfmaps and correlating each resolution shell R by

$$FSC(R) = \frac{\sum_{r \in R} F_1(r) F_2(r)^*}{\sqrt{\sum_{r \in R} |F_1(r)|^2 |F_2(r)|^2}}, \tag{3.10}$$

where F_1 and F_2 are the Fourier transforms of halfmap 1 and 2 and $*$ denotes the complex conjugation. The FSC values are then plotted against the resolution, giving the so called FSC curve, which should show high correlations at lower resolutions until they drop and finally fluctuate around 0 at higher resolutions. A typical FSC curve shown in Figure 3.4.

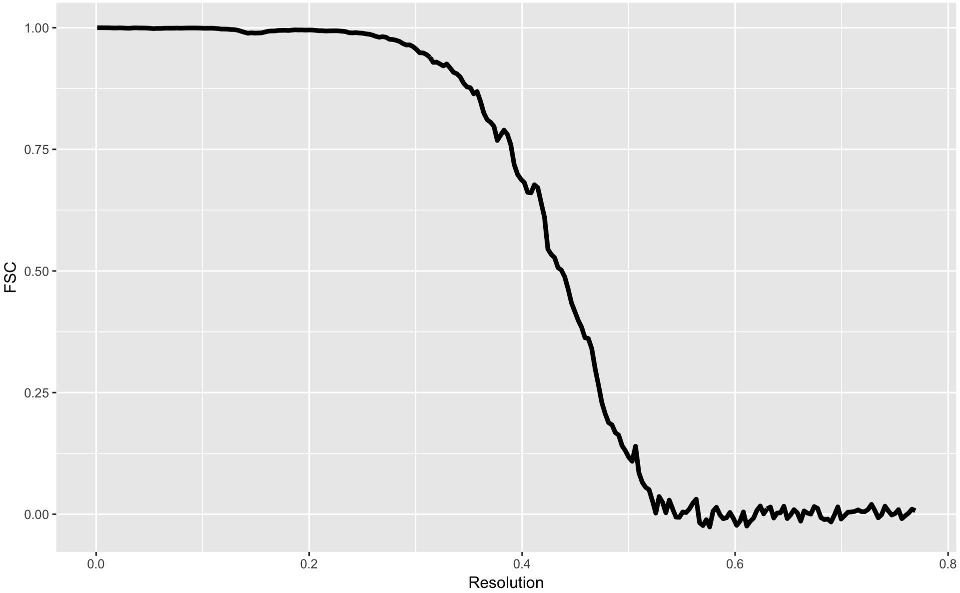


Figure 3.4: Typical FSC curve. The FSC curve has high correlations at lower resolutions and drops until it starts to show statistical fluctuations around 0.

While the basic procedure of calculating FSC curves became standard in cryo-EM and also in super-resolution microscopy, the actual process of reading a resolution from such curves remains a controversial and unresolved problem. The basic idea is to use the resolution at which the FSC drops for the first time below a specified threshold. Several thresholds have been proposed, e.g. 0.143 FSC [123], 0.5 FSC, 3σ [107] and the half-bit criterion [147], but 0.143 FSC is the most widely used. However, it has to be noted that fixed FSC value thresholds like 0.5 and 0.143 ignore important aspects of the FSC and are highly sensitive to masking, which makes them less reproducible. A reproducible threshold criterion has to take into account the size of the respective resolution shell, as this is directly related to the uncertainty of the respective FSC value. Moreover, symmetry and masking introduce dependencies in real space and thus also in Fourier space, which leads to reduced effective sample sizes. Fixed threshold criteria are unable to take these effects into account.

While the 3σ criterion aims to find a point at which the FSC is for the first time indistinguishable from noise and has as such the best theoretical foundation, it tends to give inflated resolution estimates. A unified framework as a possible solution to the problem (correction of the multiple testing problem and no parametric assumptions) is developed in this thesis.

Although a global resolution estimate is an important measure of the data quality, cryo-EM maps usually exhibit local resolution anisotropies. Some parts are more flexible than others and will thus be less well defined. With this respect, quantification of local resolutions has become an important target, especially as this information can be used for local processing of the maps. The simplest, but still most exact, approach uses local FSC calculations by sliding small windows over both halfmaps and in this way calculates a local resolution for each voxel [25]. However, due to small window sizes, usage of rigorous statistical FSC threshold criteria becomes highly important. Although different algorithms, which are not based on FSC calculations, have been proposed, e.g. ResMap [77] and MonoRes [149], the former tends to give unrealistic high resolutions while the latter assigns usually too low resolutions.

3.1.5 Post-Processing of cryo-EM Maps

Cryo-EM maps, as they result from the 3D reconstruction process, usually do not show the map features which would be expected at the estimated resolution. Work from Peter Rosenthal and Richard Henderson in the early 2000s [123] started to quantify this contrast loss at high resolutions. As a result from sample heterogeneity, alignment inaccuracies and the under-determined 3D reconstruction process, the high-resolution information in the final 3D reconstructions, which is stored in the high-frequency amplitudes of the Fourier transform, is under-represented compared to low-resolution information.

They proposed a simple sharpening approach in order to increase the intensities of higher resolutions compared to lower resolutions based on B-factor sharpening. With this respect, each voxel in the Fourier transform is multiplied with a frequency dependent factor given as

$$\exp\Big(-\frac{B_{restore}}{4d^2}\Big), \tag{3.11}$$

where d is the spatial frequency and $B_{restore}$ is the B-factor. This B-factor is dependent on the map at hand and can be reliably estimated from Guinier-plots for structures with resolutions better than 10 Å [123]. In a Guinier plot, the natural logarithm of the amplitudes is plotted against the squared spatial frequency. It is known from the so-called Wilson statistics [159], which describe the amplitude profile of randomly placed atoms, that the amplitudes at resolutions better than 10 Å can be assumed to be constant. However, inspection of cryo-EM maps reveals that this is not the case for experimental maps, where amplitudes show a clear falloff (Fig. 3.5, blue line). Fitting a simple linear function from 10 Å until the resolution limit provides an estimate of this falloff, where the slope corresponds to the B-factor (Fig. 3.5, red line).

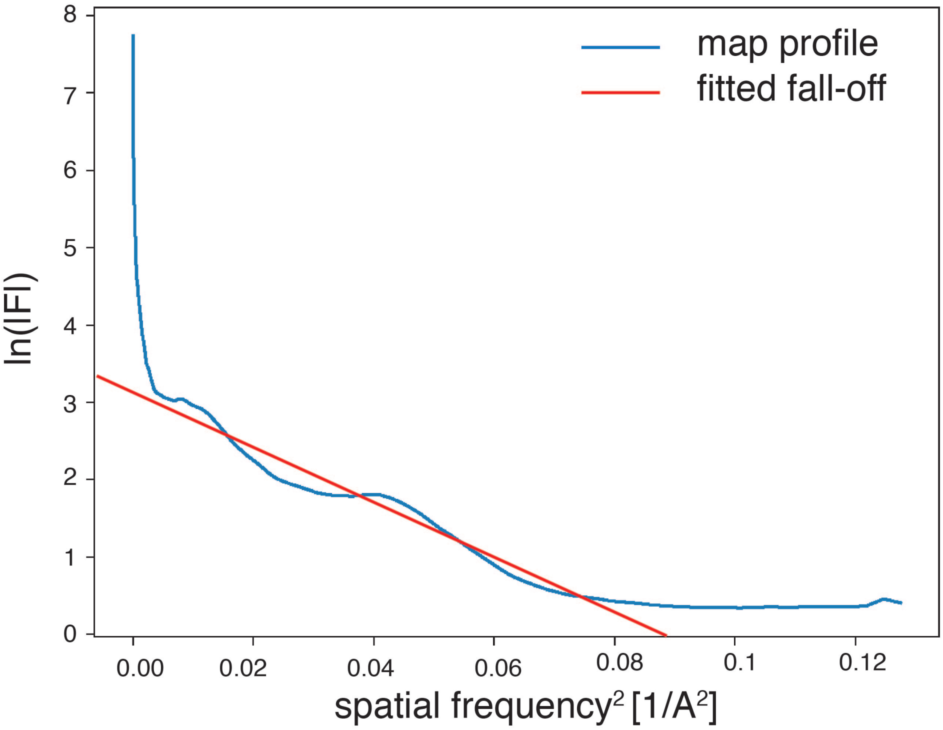


Figure 3.5: Guinier plot and B-factor estimation. The Guinier plot for a γ-secretase map (blue, EMD30161)[4] together with the fitted linear fall-off (red) is shown.

A problem with sharpening is that it also increases background noise. Filtering of the maps based on the estimated resolution reduces the noise levels by suppressing high frequency information beyond the resolution limit of the data. Moreover, a weighting of the Fourier transform of the cryo-EM map by the FSC curve [123] has been proposed and takes account the shape of the FSC. However, the actual shape of the FSC curve is highly dependent on the used mask, the relative box size with respect to the particle and also influenced from noise, which hampers a rigorous justification of the weighting process.

More recent approaches have been developed to take into account local anisotropies of map quality in the post-processing work flow. Filtering of the map at an average resolution estimate leads suppression of features at higher resolutions than the average while lower resolution features remain too noisy. A local filtering approach has been proposed by [25] and shows improved overall representation of the cryo-EM map.

In a similar fashion, sharpening can be done on a local level, taking into account local anisotropies of high-resolution contrast loss. However, local B-factor estimation has been shown to be imprecise [69]. The first approach towards local sharpening was proposed by Jakobi *et al.* [69] and is based on local amplitude scaling using simulated maps from refined atomic models as a reference. However, the need of a refined model still leaves the problem of refining the intial atomic model against the non-locally sharpened map. Different approaches for local sharpening were proposed subsequently [142].

A big problem that remains, especially at initial stages of the atomic modelling, is the assignment of actual molecular map density from background noise. Due to the sharpening process, cryo-EM maps contain excessive amounts of background noise. Moreover, visualization of cryo-EM maps requires the choice of a map threshold, which is especially difficult as different parts show different density levels, e.g. a bound ligand might be represented with much weaker density due to occupancy of binding. Although in x-ray crystallography thresholding of maps by multiples of the standard deviation σ is common, this still ignores the multiple testing problem (see below) and is problematic for cryo-EM maps, as local resolutions lead to substantial amounts of density variation over the map. Moreover, normalization is commonly done by estimating mean and variance over the complete map, thereby ignoring the fact that this also includes the signal component, which will lead to biased background noise estimates.

Many modelling ambiguities have been attributed to wrong choices of thresholds in both cryo-EM as well x-ray crystallography [8], a problem which is sometimes referred to as the "thresholding problem" [90]. With this respect, a statistical framework based on false discovery rate control has been developed in this thesis, that allows interpretation of cryo-EM maps based on detection errors with respect to background noise.

3.2 Statistical Inference

3.2.1 Statistical Inference

The main objective of statistical inference is to estimate or to make probabilistic statements about parameters of the distribution of a random variable based on observed realizations of it. Thus, the goal is to get information about the generative model based on the data. Statistical inference can be seen as the step of upward-reasoning from the data towards the data generating process. Typical questions that arise are: What is the mean effect of a treatment? Or given a specified error, is there any significant effect beyond 0? As such, statistical inference is different from purely descriptive statistics, which aims for visualization and summary of possibly complex and high-dimensional data. In the following, we will give a short introduction to the two main aims of statistical inference: parameter estimation and hypothesis testing. The focus will be on demonstrating the essential concepts and on readability instead of complete mathematical rigour.

Parameter Estimation

Estimating the parameters of a probabilistic model is of pivotal importance and occurs in many different settings. For example, machine learning with deep neural networks works by estimating parameters of the network, during the cryo-EM reconstruction process the 3D reconstruction as well as orientation parameters of the particle images need to be estimated or for many hypothesis tests we need estimates of variances and location parameters. Estimation theory is a big part of mathematical statistics and deals with the behaviour as well as the optimality of different types of estimators.

An estimator for a parameter φ can be understood as a function $\hat{\varphi}(X_1, ..., X_n)$ which maps the data $(X_1, ..., X_n)$ to an estimate of φ. Mathematization of estimation theory then requires the definition of desirable properties that these functions should have. Intuitively, a desirable property would be that the expected output of the estimator is the true value of the parameter, something which is called unbiasedness. Moreover, the estimate should become better when adding more data, a property that is commonly referred to as consistency.

Mathematically, an estimator $\hat{\varphi}$ for φ is called unbiased if the expected value of the respective estimator is the true value of the parameter, i.e.

$$\mathbb{E}(\hat{\varphi}) = \varphi \tag{3.12}$$

where $\mathbb{E}$ denotes the expectation value. As stated above, an other important feature of interest is weak (or strong) consistency, which means convergence in probability (or almost surely) of the estimator to the true value of the parameter for increasing sample sizes, i.e.

$$\hat{\varphi}(X_1, ..., X_n) \xrightarrow{p\ (a.s.)} \varphi, n \to \infty. \tag{3.13}$$

However, while unbiasedness and consistency are important properties, it has to be noted that an estimator is a function of random variables and as such a random variable itself. This means the estimator also has a defined variance (if finite), which is an important additional quantity to be considered. For example, if we have two unbiased estimators at hand, the one with lower variance will tend to give parameter estimates closer to the true value. The Cramer-Rao lower bound gives a lower limit for the variance of an unbiased estimator for a specific parameter and can be used to find optimal estimators [118]. However, the situation can be complicated further, because an estimator with slight bias but reduced variance can yield estimates closer to the true value than its biased counterpart.

A quality measure that takes into account the variance of an estimator is the mean squared error MSE. It is defined as the expected squared difference between the estimate and the true value, i.e.

$$MSE(\hat{\varphi}) = \mathbb{E}[(\hat{\varphi} - \varphi)^2]. \tag{3.14}$$

A well known example is the James-Stein estimator [70], an adjusted estimator for the mean of normally distributed random vectors, which can be shown to yield a lower MSE compared the unbiased least-squares estimates in more than two dimensions. In the context of construction of estimators, the Rao-Blackwell theorem provides a general framework to improve an arbitrary estimator with respect to the MSE.

Several statistical frameworks for the construction of estimators have been proposed, and the most widely used are moment estimators, Maximum-Likelihood (ML) estimators as well as Bayesian Maximum-A-posteriori (MAP) estimators. Especially ML estimators are mathematically well characterized. Under weak conditions, ML estimators have both weak and strong consistency. Moreover, they are asymptotically normal distributed and reach the Cramer-Rao lower bound asymptotically. However, in complex settings the actual ML estimate is not analytically available and needs to be found numerically, which is difficult for non-convex likelihood functions and tends to get stuck in local optima [15, 49]. While Maximum likelihood and moment estimators only assume a parametric model of the data generating process, Bayesian estimators allow incorporation of prior knowledge in order to get more precise estimates, but with the danger of additional bias through the prior distribution. Moreover, in a Bayesian setting the parameters are actually modelled as random variables, and assigning a single estimate to a random variable, as done in MAP estimation, is hardly logically. As such, proper Bayesian treatment requires sampling from the posterior distribution via Markov chain Monte Carlo (MCMC) sampling [9, 39, 20]. In the case of an uniform prior distribution, a MAP estimator becomes a

simple ML estimator. Taken together, the construction of ideal estimators is a complex topic and needs heavy adjustment to the specific situation, for example the development of cryo-EM reconstruction algorithms is still a field of active research.[128, 117]
As already mentioned, one of the biggest problems arising for parameter estimates is the inherent uncertainty. The estimate can be far off from the actual value of the parameter, depending on the uncertainty. In order to quantify the uncertainty, confidence intervals, or credible intervals in a Bayesian setting [9], can be constructed. Naively spoken, confidence intervals are regions of the parameter space that contain the parameter of interest with a certain probability and provide a more detailed description of the actual parameter estimation process compared to single parameter estimates. Moreover, confidence intervals can be used to make probabilistic statements about statistical hypotheses, as we will see in the next sections.

The Concept of Falsification

Falsification builds the basis of our modern theory of science. First thoroughly characterized in "Die Logik der Forschung" by Karl Popper [116] in the 1930s as part of the critical rationalism, the main statement is that a scientific theory can never be proven to be true. A scientific theory and the conclusions we deduce from it will always be incomplete. But what do we mean when talking about scientific facts? What marks the difference between empirical research and religion/ideology and what are the limitations of science? A very simple, but highly powerful requirement of any scientific statement is that it has to be falsifiable. This is a much stronger property than a simple statement in its linguistic sense, which is a sentence that is true or false. For example "God exists." is a sentence that is true or false, with the answer dependent on the person we ask, but it is impossible to be falsified, and thus not scientific. However, despite its rather philosophical character, falsification also lays the foundation for scientific evaluation by statistical significance.
Statistical significance testing provides the methodological framework for falsification based on empirical data. Given the uncertainty inherent to the data, can we reject a specified hypothesis? Although quantification of the uncertainty can be rather difficult, p-values and Bayes factors are standard measures when it comes to judgement of the validity of given hypotheses and are ubiquitous in the scientific literature. In the next section, I will give an introduction to statistical hypothesis testing.

Statistical Hypothesis Testing

Statistical hypothesis testing under the Neyman-Pearson paradigm deals with the situation of testing a hypothesis against an alternative hypothesis based on experimental data with finite sample sizes and the inherent uncertainty [121]. Especially consideration of the uncertainty is important when we have to make sure that the error of having a false-positive decision is limited. For example, when comparing the treatment effect of a drug between a placebo and a treatment group, we have to make sure that there is a significant effect beyond simple statistical fluctuations.

According to Neyman-Pearson, statistical hypothesis testing works by specifying a null hypothesis H_0 and an alternative hypothesis H_1. H_0 contains the events from which we want to find deviations, e.g. for finding a treatment effect ρ different from 0, we would have the hypothesis system $H_0 : \rho = 0$ and $H_1 : \rho \neq 0$. A statistical test is then a mathematical function ϕ which maps the data $(X_1, ..., X_n)$ to 0 if we cannot reject H_0, or 1 if we reject H_0, i.e.

$$\phi(X_1, ..., X_n) = \begin{cases} 0 & T(X_1, ..., X_n) \in H(\alpha) \\ 1 & T(X_1, ..., X_n) \in K(\alpha), \end{cases} \tag{3.15}$$

where H is the non-rejection area, $K(\alpha) = \Theta \backslash H(\alpha)$ is the rejection area, i.e. the complete parameter space Θ excluding $H(\alpha)$, and $T()$ the respective test statistic used.

Construction of the non-rejection area K is done by specifying a region of all possible values of T which contains realizations of T with a specified probability P, considering the distribution of T under the null hypothesis H_0. While the actual specification of the non-rejection area is often not unique, it is usually chosen as the smallest area containing the specified probability mass, which guarantees the biggest rejection area under a specified significance level. The probability mass P of the non-rejection region is then directly related to the significance level α by

$$\alpha = 1 - P. \tag{3.16}$$

α is also referred to as the Type I error and makes a statement about the probability of a erroneous rejection of the null hypothesis, e.g. $\alpha = 0.01$ means that there is a chance of 1% of rejecting H_0 even if the null hypothesis is actually true. Clearly, the rejection and non-rejection regions are dependent on the specific significance level α, which explains our notion of $H(\alpha)$ and $K(\alpha)$.

However, it is not always possible to provide a test that controls the significance level α exactly. For example, if the null hypothesis is of composite structure, as in the one sided test scenario $H_0 : \mu \leq 0$, then the distribution of the test statistic under H_0 is not clearly defined. Using the distribution of T, as it would occur for $\mu = 0$ in this scenario, will give exact control of the significance level only if $\mu = 0$, but more conservative if $\mu < 0$. This motivates the definition, that a test is defined to have level α if it rejects H_0 with a probability less or equal than α under the null hypothesis, i.e.

$$\mathbb{P}(\phi = 1 | H_0) \leq \alpha. \tag{3.17}$$

Related to the Type I error is the Type II or β error, which is the probability of not-rejecting H_0 when H_0 is actually wrong, i.e. the probability of a false negative decision. Unfortunately, Type I and Type II errors can usually not be controlled simultaneously when the sample size is fixed. Although the Type II error is important as well, control of Type I errors is considered as more important. Interpretation of not-rejecting H_0 should therefore be regarded as a failure of rejecting H_0 at the specified significance level and does not necessarily demonstrate the correctness of H_0.

However, when comparing two tests at the same significance level for the same hypotheses system, the test that will reject H_0 more often obviously has a lower Type II error and will be more powerful, something which is reflected in the definition of statistical power as $1 - \beta$. Moreover, both Type I and Type II errors influence each other: The smaller α is set, the higher β will be.

Another widely used quantity used in hypothesis testing is the p-value. The p-value is defined as the probability under the null hypothesis of having a more extreme event than the observed one. As such the actual mathematical definition is dependent on the specific test scenario and can be quite complicated.

In an abstract notion and assuming nested rejection regions $K(\alpha)$, the p-value is given as

$$p = \inf\{\alpha : T(x_1, ..., x_n) \in K(\alpha)\} \tag{3.18}$$

where $T(x_1, ..., x_n)$ denotes the value of the test statistic for the observed data $x_1, ..., x_n$. One of the most useful features of p-values is their direct relation to the significance level. If the p-value is smaller than α, we can reject the null hypothesis at a significance level α. Moreover, the p-value shows how small α could in principle be chosen in order to reject H_0 and reporting p-values instead of binary test results has become scientific practice by now.

Although most of statistical test theory is already quite old, the fundamental Neyman-Pearson lemma dates back to 1933 [104], it is still an active field of research. Emergence of modern non-parametric procedures [21], often based on heavy computations via resampling, as well as high-dimensional inference problems provide new fields of application and new challenges, one of them being large scale simultaneous testing of millions of hypotheses [81]. Moreover, hypothesis testing in a Bayesian framework based on Bayes-factors, their relation to p-values [59] and the Bayesian calibration of p-values provides new insights into statistical significance.

t-Test

In order to illustrate the rather abstract mathematics behind statistical hypothesis testing, an example of the famous t-test will be given now. Assume we acquired a dataset $(x_1, ..., x_n)$ with sample size n, which follows normal distribution and we are interested if the mean μ is significantly deviating from 0. This gives the hypothesis system

$$H_0 : \mu = 0 \text{ and } H_1 : \mu \neq 0. \tag{3.19}$$

Taking the test statistic

$$T(X_1, ..., X_n) = \sqrt{n}\frac{\overline{X_n} - \mu}{\hat{\sigma}}, \tag{3.20}$$

where $\overline{X_n}$ is the sample mean and $\hat{\sigma}$ the estimator of the standard deviation of the sample, it can be shown that under the null hypothesis $T(X_1, ..., X_n)$ follows a student-t distribution with $n-1$ degrees of freedom.
If T' denotes the observed value of the test-statistic T from the respective sample, the p-value in this case can be calculated as

$$p = 2 * \min\{F(T'), 1 - F(T')\}, \tag{3.21}$$

where F is the cumulative distribution function of the student-t distribution with $n-1$ degrees of freedom. A graphical depiction of the t-test is given in Fig. 3.6. The distribution of the test statistic T with a sample size of 31 is shown. Given a value of ~1.9 for the test statistic observed with the data, the p-value would then be the area of the region framed with red.

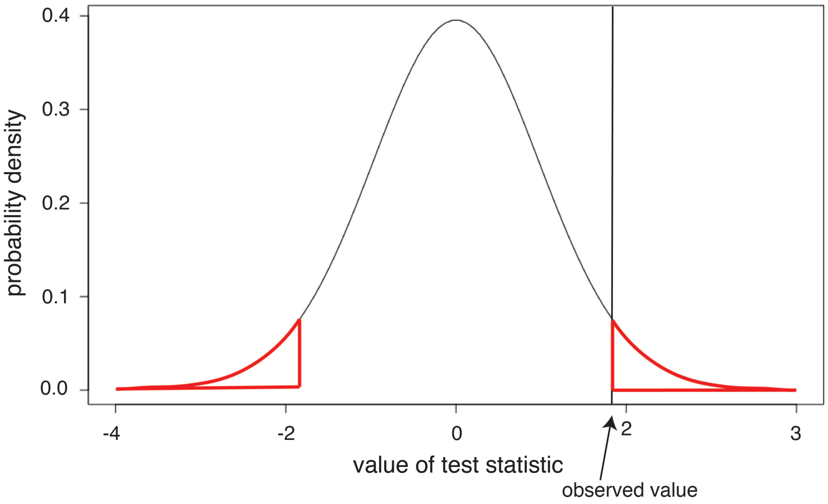


Figure 3.6: Illustration of the t-test. The probability density of a t-distribution with 30 degrees of freedom is shown, which corresponds to a t-test with a sample size of 31. In the two sided test setting, as described in the example, the p-value for the observed value of the test statistic ~1.9 would then be the area of the region which is framed with red lines.

3.2.2 Multiple Hypothesis Testing

Testing multiple hypotheses simultaneously results in a multiple testing problem. The main problem emerging is the multiplicity: Not just every single hypothesis test can lead to a false positive decision, but also multiple tests can have false positive results simultaneously. Moreover, multiple false positive hits can occur in any combination of single tests, leading to a combinatorial explosion and subsequently to a loss of error control, which is often referred to as the α-error inflation [81].

The basic problem can be easily illustrated for the case of a simultaneous inference of n independent hypothesis tests with significance level α. The probability p of having at least one false positive result can then be calculated as the result of a binomial event by $p = 1 - f_{\text{Bin}(n,\alpha)}(x)$, where f_{Bin} is the density function of the binomial distribution. For $n = 100$ tests and $\alpha = 0.01$ this gives already a probability of more than 99% for at least one false positive, which shows the problem we need to correct for.

Family Wise Error Rate

Dealing with multiple hypothesis tests requires the use of specific error criteria that allow the interpretation of multiple inferences simultaneously. The oldest and most widely used criterion is the family wise error rate ($FWER$). The $FWER$ is defined as the probability of having at least one false positive, i.e.

$$FWER := \mathbb{P}(V > 1), \tag{3.22}$$

where we denote with V the number of false positive hypotheses. In order to control the $FWER$ for n simultaneous statistical tests, the significance level α for each test needs to be adjusted accordingly. We say that the $FWER$ is controlled at level α if the true probability of any false positives is smaller than α, i.e.

$$\mathbb{P}(V > 1) \leq \alpha. \tag{3.23}$$

The simplest approach for $FWER$ control is the Bonferroni procedure, where the adjusted significance level for each test is set to $\alpha_{adj} = \frac{\alpha}{n}$ [81]. Therefore, in order to control the $FWER$ at level α, each of the n tests is conducted at level $\frac{\alpha}{n}$. Although, the Bonferroni procedure is quite simple and works under arbitrary dependencies between the individual tests, it is a rather conservative procedure, i.e. the actual error is well below α. Especially for a large number of tests it becomes very restrictive and more powerful procedures have been developed.

Most $FWER$ controlling approaches result from the construction of a so called closed testing procedure, which was proposed by Marcus, Peritz and Gabriel in 1976 [92], and provides a general mathematical framework for the control of $FWER$. The probably most widely used procedure nowadays is Bonferroni-Holm [64], which has the advantage that it controls the $FWER$ under arbitrary test dependencies similar to the Bonferroni approach. Even more powerful are the Hochberg [62], Simes [134] and the Hommel [65] procedures. However, these approaches make assumptions about the dependency structure, e.g. the Hochberg procedure assumes non-negative dependencies, and are therefore restricted in applicability.

A big problem for $FWER$ control is the limited statistical power, especially for large number of tests. However, a very recent approach using harmonic mean p-values in a multilevel test procedure demonstrated that $FWER$ control [160] can still be improved substantially, but further validation and mathematical characterization of the proposed framework is required. It has to be noted at this place that the multiple testing correction usually works on the level of p-values and not on the individual significance levels as I showed it for the Bonferroni example, i.e. after calculation of the p-value for all tests the p-values are adjusted. This has the advantage that we do not have to repeat the whole inference when changing the significance level.

False Discovery Rate

While the $FWER$ is a useful criterion for small or medium sized number of hypothesis tests, it becomes very conservative for large scale inference in the order of several millions of hypotheses, as they occur in images or genomic studies [50][86]. Instead of controlling the error of having any false positives at all, it makes sense in many settings to tolerate a few false positives in order to reduce the amount of false negatives. Such a screening approach has been developed with the false discovery rate (FDR), which specifies the expected ratio of false positives in the set of rejected hypotheses [12]. Mathematically, the FDR is defined as

$$FDR := \begin{cases} \mathbb{E}(\frac{V}{V+R}) & V+R \neq 0 \\ 0 & V+R = 0, \end{cases} \tag{3.24}$$

where V is the number of false positives and R the amount true positives. We say that the FDR is controlled at level α, if the true FDR is smaller than α, i.e.

$$FDR \leq \alpha. \tag{3.25}$$

The first approach to control the FDR has been proposed by Benjamini and Hochberg in 1995 [12] and assumes a condition termed "positive dependency" between the p-values. In 2001 Benjamini and Yekutieli provided a modified version of their initial approach [13], which allows control of the FDR under arbitrary dependencies. However, compared to

Benjamini-Hochberg the Benjamini-Yekutieli approach is more conservative. Due to the importance of the Benjamini-Yekutieli procedure for this thesis, we will give an overview of the algorithm at this point.

As input the Benjamini-Yekutieli approach takes the p-values $p_i, i = 1, ..., m$ of m simultaneous statistical tests. The first step is sorting of the p-values from small to large which leads to the sorted p-values $p_{(i)}, i = 1, ..., m$. The FDR-adjusted p-values, often referred to as q-values, are then calculated from the p-values by

$$q_{(i)} = \min_{i \leq k \leq m} p_{(k)} \frac{m}{k} \gamma, \tag{3.26}$$

where m is the number of hypotheses, k is a running index and $\gamma = \sum_{l=1}^{m} \frac{1}{l}$. The q-values are then sorted back to the original order and assigned to each hypothesis. It should be mentioned at this stage that for $\gamma = 1$ the Benjamini-Yekutieli approach becomes the Benjamini-Hochberg procedure.

While both the Benjamini-Hochberg and the Benjamini-Yekutieli procedures are classical frequentist approaches, the concept of false discovery rates was also developed further in a Bayesian setting. Storey introduced in the early 2000s the positive false discovery rate ($pFDR$) and showed how it can be derived from a Bayesian point of view [137, 138]. Moreover, Bradley Efron developed the concept of local false discovery rates, which gives for each hypothesis and estimate of the probability of being a false positive in an empirical Bayesian framework [38, 139], and beyond that methods have been developed recently to incorporate additional information for the hypotheses [166, 68].

Although false discovery rate control became an accepted concept, it has to be mentioned that it only has a screening character and does not provide a stringent error rate like $FWER$. Depending on the number of hypotheses or the amount of positive hits, 1% FDR can mean different things. For example, addition of tests which that likely reject H_0 will allow more significant findings at 1% FDR. A possibility for manipulation that needs to be treated with caution.

Chapter 4

Confidence Maps - Thresholding of cryo-EM Maps by False Discovery Rate Control

The first part of this chapter is based on the published article Beckers *et al.* in *IUCrJ* [10] and some of the figures from the publication, which I use in this chapter, are reproduced with permission of the International Union of Crystallography. All figures adapted from the publication are marked as such. The project was conceptualized based on ideas from a lecture on multiple testing theory that I attended at Ulm University and was a good fit to recent efforts in the lab to infer more information from the cryo-EM maps. I performed all analyses, programming and preparation of all the figures and the paper was written together with Carsten Sachse. The article was highlighted by Peter Rosenthal [122] and selected for oral presentations at several international conferences, including the Gordon Research Conference on 3D electron microscopy and the CCPEM Spring Symposium. My talk at the CPPEM Spring symposium is also available on YouTube at *www.youtube.com/watch?v=xOX-5EzmjwI.* Some of the recent extensions and the implementation in the CCPEM software package, which are presented at the end of this chapter, were submitted as a research article to *Acta Crystallographica Section D* as part of the proceedings of the CCPEM Spring Symposium.

4.1 Abstract

The interpretation of cryo-EM maps by atomic models is the ultimate goal of the structure determination process with cryo-electron microscopy. However, the actual results of the 3D reconstruction process are subject to several problems: Contrast loss at high resolutions, high levels of background noise, radiation damage and local resolution variations make extensive post-processing of the maps necessary.

Annotation of density at initial steps of the atomic model building process is done by visualizing the map at isosurfaces, which requires the choice of appropriate thresholds.

Especially corruption of the signal with noise makes the assignment of weak densities close to background noise a difficult task. However, many interesting features are expected to be very weak, for example due to flexibility of a mechanistic switch or because a ligand might not be bound to all particles.

In order to facilitate this process, I developed a new type of map, termed confidence map, which contains for each voxel a probabilistic quantity for deviations from background noise. Confidence maps can be thresholded by specifying the upper bound for the expected false discovery rate in the visible volume and allow rigorous density interpretation by means of statistical significance. Confidence maps avoid the choice of arbitrary and subjective thresholds and show drastically reduced intensity variations that aids in the map interpretation work flow.

4.2 Methods

In order to introduce statistical significance as an additional measure for validation and interpretation in the cryo-EM data processing workflow, I applied multiple hypothesis testing using false discovery rate (FDR) control to cryo-EM maps. A rough outline of the developed statistical procedure is as follows: The distribution from the background fluctuations is estimated from a sharpened cryo-EM map, subsequent statistical hypothesis testing is applied for each voxel and finally the resulting multiple testing problem is accounted for by controlling the FDR (Fig. 4.1 a). Additionally, I developed and investigated methodology to improve the signal detection by incorporating additional information like local resolutions. In the first section of this chapter I will provide a detailed mathematical description of the developed algorithms.

4.2.1 Mathematical Framework

For each voxel in the cryo-EM map, which will be indexed with i, j, k, the corresponding map values $X_{i,j,k}$ are modelled as the random variable

$$X_{i,j,k} = \mu_{i,j,k} + \epsilon_{i,j,k}, \tag{4.1}$$

with $\epsilon_{i,j,k}$ a real-valued random variable representing the background noise with mean $\mu_{0,i,j,k}$ and variance $\sigma^2_{i,j,k}$ and with $\mu_{i,j,k}$ the true map value without background noise. Analysis of the maps is usually done under a positive density assumption, i.e. the object of interest are positive map values that deviate form the background noise, which gives the following hypothesis system to be inferred:

$$H_0 : \mu_{i,j,k} \leq 0 \tag{4.2}$$

$$H_1 : \mu_{i,j,k} > 0. \tag{4.3}$$

In order to test this hypothesis system, background noise is estimated outside the particle (Fig. 4.1 b), giving its mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$. Subsequently, p-values are calculated by

$$p_{i,j,k} = \begin{cases} \mathbb{P}(X_{i,j,k} \geq x_{i,j,k} | H_0) = 1 - F(x_{i,j,k}) & x_{i,j,k} \geq \tilde{\mu} \\ 1 & x_{i,j,k} < \tilde{\mu} \end{cases}, \tag{4.4}$$

where $X_{i,j,k}$ denotes the random variable of the map intensities as above, $x_{i,j,k}$ the respective realizations observed in the map and $F()$ the cumulative distribution function of the background noise. The background noise distribution can either be approximated with a normal distribution, leading to a standard normal distribution after subtraction of $\tilde{\mu}$ and division by $\tilde{\sigma}$, or by replacing the cumulative distribution function (CDF) with the empirical CDF $\hat{F}()$, given as

$$\hat{F}(t) = \frac{\text{number of elements in the sample} \leq t}{\text{total number of elements in the sample}}, t \in \mathbb{R}. \tag{4.5}$$

The p-values are then subsequently corrected for the multiple testing by controlling the false discovery rate, as described already in the introduction. In a last step, the FDR-adjusted p-values, $q_{i,j,k}$, are then inverted by

$$q'_{i,j,k} = 1 - q_{i,j,k}, \tag{4.6}$$

in order to allow visualization with standard software tools. The confidence map is then composed of the inverted and FDR-adjusted p-values $q'_{i,j,k}$ for each voxel and can thresholded by specifying a FDR, i.e. a threshold of 0.99 means an expected FDR of less than 1%.

In order to account for the complex dependencies apparent after sharpening and filtering of cryo-EM maps, the Benjamini-Yekutieli procedure has been chosen as the default. However, Benjamini-Hochberg FDR-control, Bonferroni-Holm $FWER$-control and $localFDR$ have been implemented as well. If not stated differently, all following analyses have been done with Benjamini-Yekutieli FDR-control.

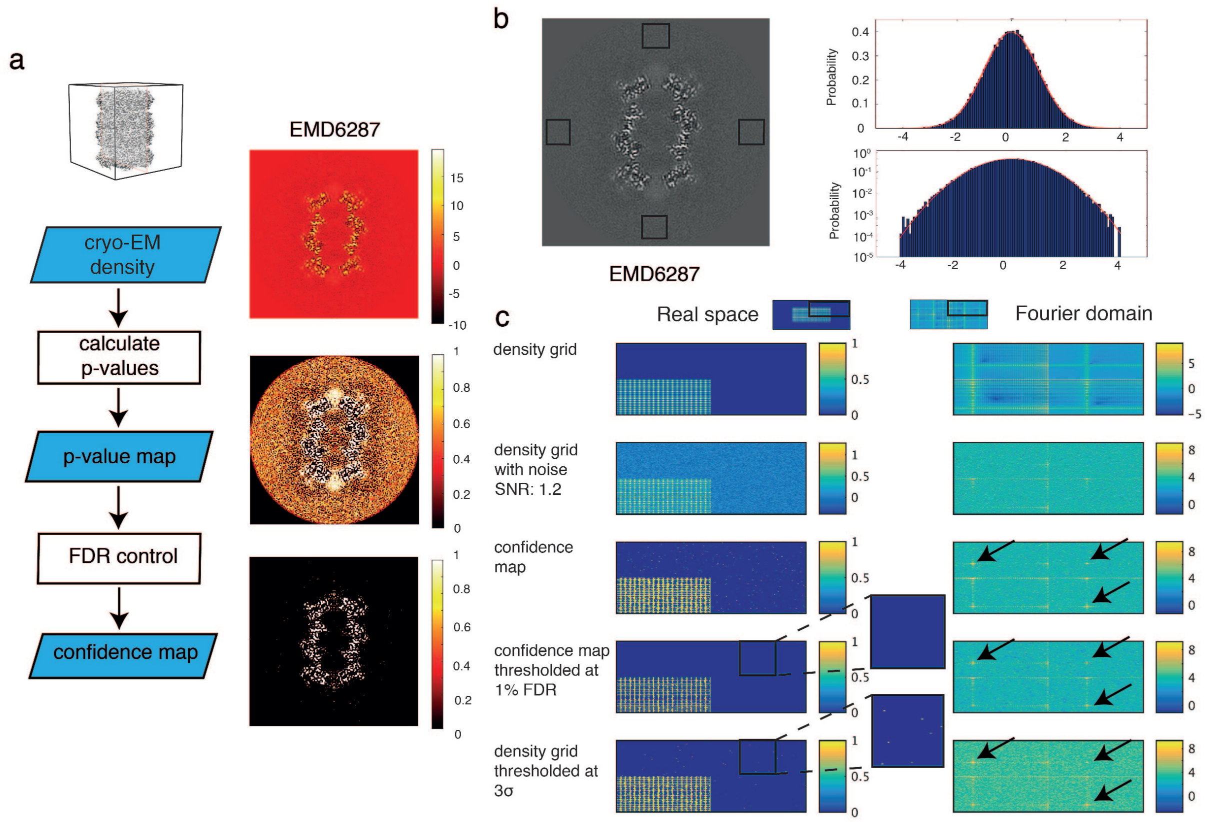


Figure 4.1: False discovery rate (FDR) control of cryo-EM maps. a) Left: flowchart of confidence map generation. The cryo-EM map is converted to p-values and finally FDR-controlled. Right: slice views through a cryo-EM map of 20S proteasome (EMD-6287 [24]) depicted at the respective stages of the algorithm (blue boxes) on the left. Note the strong increase in contrast when the sharpened map is converted to the confidence map. b) Left: estimation of the background noise from windows (red) outside the particle. Right: histograms (top, probability on a linear scale; bottom, probability on a log scale) of the background window together with the probability density function of the estimated Gaussian distribution. c) Evaluation of the algorithm on a simulated two-dimensional density grid. The upper right quadrant of images in real space (left column) together with the corresponding power spectrum in the Fourier domain (right column) are displayed. Addition of noise leads to a loss of contrast at high resolution. Confidence maps recapitulate these high-resolution features (arrows), showing that high-resolution signal is detected with high sensitivity. FDR thresholding at 1% recovers a similar binary grid in comparison with 3σ thresholding while controlling noise contributions and minimizing detected noise (enlarged insets). This figure is taken from [10] and was prepared by myself.

In the following proposition I show mathematically that replacing voxel-wise background noise estimates with upper bounds for mean and variance will only lead to more conservative FDR estimates, i.e. the FDR is still controlled in these cases. This is an important statement, as from the map itself we can only estimate the background outside the particle, which is assumed to have higher background noise levels than the particle itself due to relative ice thickness.

Proposition 1. *Consider Gaussian-distributed random variables representing the background noise at all voxels i, j, k in the three-dimensional map with true mean* $\mu_{0,i,j,k} \in \mathbb{R}$ *and variance* $\sigma^2_{i,j,k} \in \mathbb{R}_{>0}$. *Moreover, let* $\tilde{\mu} \geq \mu_{0,i,j,k}$ *and* $\tilde{\sigma}^2 \geq \sigma^2_{i,j,k}, \tilde{\mu} \in \mathbb{R}, \tilde{\sigma}^2 \in \mathbb{R}_{>0}$ *for all i,j,k the overestimated background noise parameters. Then*

$$\tilde{q}_{i,j,k} \geq q_{i,j,k}, \tag{4.7}$$

where $\tilde{q}_{i,j,k}$ corresponds to the q-value as defined above and calculated with the overestimated background noise parameters and $q_{i,j,k}$ corresponds to the q-value as obtained with true parameters $\mu_{0,i,j,k}$ and $\sigma^2_{i,j,k}$.

Proof. In order to prove the statement, I will now recapitulate the algorithm and prove the inequality at all necessary steps. I start showing that the true p-value at voxel i, j, k, $p_{i,j,k}$, is smaller when compared with the p-value $\tilde{p}_{i,j,k}$, as calculated from the overestimated background noise parameters. In other words, we want to show that $p_{i,j,k} \leq \tilde{p}_{i,j,k}$ or, equivalent to this, $\tilde{p}_{i,j,k} - p_{i,j,k} \geq 0$.
If $x_{i,j,k} < \tilde{\mu}$, then the statement is trivial, because $\tilde{p}_{i,j,k} = 1$ and $p_{i,j,k} \leq 1$, which is a general property of p-values.

For $x_{i,j,k} \geq \tilde{\mu}$, it follows that

$$\tilde{p}_{i,j,k} - p_{i,j,k} = 1 - 0.5\Big[1 + \mathrm{erf}\Big(\frac{x_{i,j,k} - \tilde{\mu}}{\sqrt{2}\tilde{\sigma}}\Big)\Big] - 1 + 0.5\Big[1 + \mathrm{erf}\Big(\frac{x_{i,j,k} - \mu_{0,i,j,k}}{\sqrt{2}\sigma_{i,j,k}}\Big)\Big] \tag{4.8}$$

$$= -0.5\mathrm{erf}\Big(\frac{x_{i,j,k} - \tilde{\mu}}{\sqrt{2}\tilde{\sigma}}\Big) + 0.5\mathrm{erf}\Big(\frac{x_{i,j,k} - \mu_{0,i,j,k}}{\sqrt{2}\sigma_{i,j,k}}\Big) \tag{4.9}$$

As the error function erf() is monotonically increasing, it is sufficient to show that

$$\frac{x_{i,j,k} - \mu_{0,i,j,k}}{\sqrt{2}\sigma_{i,j,k}} \geq \frac{x_{i,j,k} - \tilde{\mu}}{\sqrt{2}\tilde{\sigma}}. \tag{4.10}$$

Because $x_{i,j,k} - \tilde{\mu} \geq 0$ and thus also $x_{i,j,k} - \mu_{0,i,j,k} \geq 0$, as well as $\tilde{\sigma} \geq \sigma_{i,j,k}$, we have

$$\frac{x_{i,j,k} - \mu_{0,i,j,k}}{\sqrt{2}\sigma_{i,j,k}} - \frac{x_{i,j,k} - \tilde{\mu}}{\sqrt{2}\tilde{\sigma}} \tag{4.11}$$

$$= \frac{(x_{i,j,k} - \mu_{0,i,j,k})\tilde{\sigma} - (x_{i,j,k} - \tilde{\mu})\sigma_{i,j,k}}{\sqrt{(2)}\tilde{\sigma}\sigma_{i,j,k}} \tag{4.12}$$

$$\geq \frac{(x_{i,j,k} - \mu_{0,i,j,k})\sigma_{i,j,k} - (x_{i,j,k} - \tilde{\mu})\sigma_{i,j,k}}{\sqrt{(2)}\tilde{\sigma}\sigma_{i,j,k}} \tag{4.13}$$

$$= \frac{(-\mu_{0,i,j,k} + \tilde{\mu})\sigma_{i,j,k}}{\sqrt{2}\tilde{\sigma}\sigma_{i,j,k}} \geq 0, \tag{4.14}$$

where in the last equation I used that $\tilde{\mu} \geq \mu_{0,i,j,k}$ and $\tilde{\sigma} \geq \sigma_{i,j,k} > 0$. This gives $\tilde{p}_{i,j,k} \geq p_{i,j,k}$.
Recapitulating the calculation of q-values together with the conversion of the three-dimensional volume to a one-dimensional sequence, it follows that

$$q_{(a)} = \min_{a \leq k \leq m} \Big[p_{(k)}\frac{m}{k}\gamma\Big] \leq \min_{a \leq k \leq m} \Big[\tilde{p_k}\frac{m}{k}\gamma\Big], a = 1, ..., m, \tag{4.15}$$

where m is the number of hypotheses, k is a running index and $\gamma = \sum_{l=1}^{m} \frac{1}{l}$. This proves the statement. $\square$

4.2.2 Incorporation of Local Filtering and Local Sharpening

Local sharpening and local filtering have become powerful techniques in order to deal with local resolution variation. Consideration of local spatial information allows to filter out additional noise and to adjust the sharpening accordingly and thus leads to improved local signal-to-noise ratios (SNR). It is beneficial in terms of statistical power to incorporate this additional information also in the calculation of confidence maps.

Conceptually, local filtering and local sharpening are straightforward to incorporate. For each local processing step, the same operation (filtering or sharpening) needs to be applied to the box of background noise as well. For example in local filtering the complete map is filtered for each local resolution, which allows the estimation of the background outside the particle similar to the simple setting above. The respective local background noise levels are then saved to their corresponding locations in the 3D map. Subsequently, p-values are calculated as above, with the only difference that for each voxel the corresponding local estimates of the background noise distribution have to be used. All following steps of the algorithm remain the same.

4.2.3 Simulation of Density Grids

To test the proposed framework under controlled conditions, I prepared simulations of two-dimensional images of a density grid that resembles the behaviour of cryo-EM density maps with both strong and very weak features. The simulated images were generated with a size of 400 x 400 pixels. The scaled grid was generated by adding two orthogonal two-dimensional cosine waves with a period of five pixels, where all values <0 were set to 0, and multiplying the sum by a factor of 0.5 in order to set the maximum to 1. The scaled grid had a size of 200 x 200 pixels and was embedded in the center of the 400 x 400 image. Gaussian-distributed noise with mean 0 and a variance of 0.01 (Fig. 4.1) was added to the grid image. The mean and variance for the multiple testing procedure were estimated outside the scaled grid and the FDR procedure was carried out as described. This simulation was implemented in MATLAB (MathWorks).

4.2.4 Implementation

The confidence map tools are all implemented in Python3 based on mrcfile, NumPy and Matplotlib libraries. The software termed *FDRthresholding* is available on gitlab at *git.embl.de/mbeckers/FDRthresholding/*.

4.2.5 Figure Preparation

FSC and ADP graphics were visualized with ggplot2 in R [141] [158]. Chimera [114] was used for the figure preparation of the molecular maps and atomic models.

4.3 Results

4.3.1 False Discovery Rate Control of a Simulated Density Grid

Application to a simulated two-dimensional density grid, which I prepared as a linear combination of two orthogonal cosine waves in the two-dimensional plane (Fig. 4.1 c), shows that the generation of a confidence map itself already improves the detectability of weak features compared to the noisy grid, as visible in the presence of high-resolution peaks in the Fourier transform. Moreover, thresholding based on FDR-control allows to control the expected maximal amount of detected false positive signal, which is not possible by simple thresholding of the grid.

One dimensional snippets of the grids show that, beyond statistical control of false positives, also the peak detection is improved in the confidence maps and that the signal is flattened during the transformation to false discovery rates (Fig. 4.2), i.e. variations of the peak heights, which are only due to statistical fluctuations in this case, disappear to a large extent. This is especially useful as it avoids the over-interpretation of artificial and only statistical fluctuations. (Fig. 4.2).

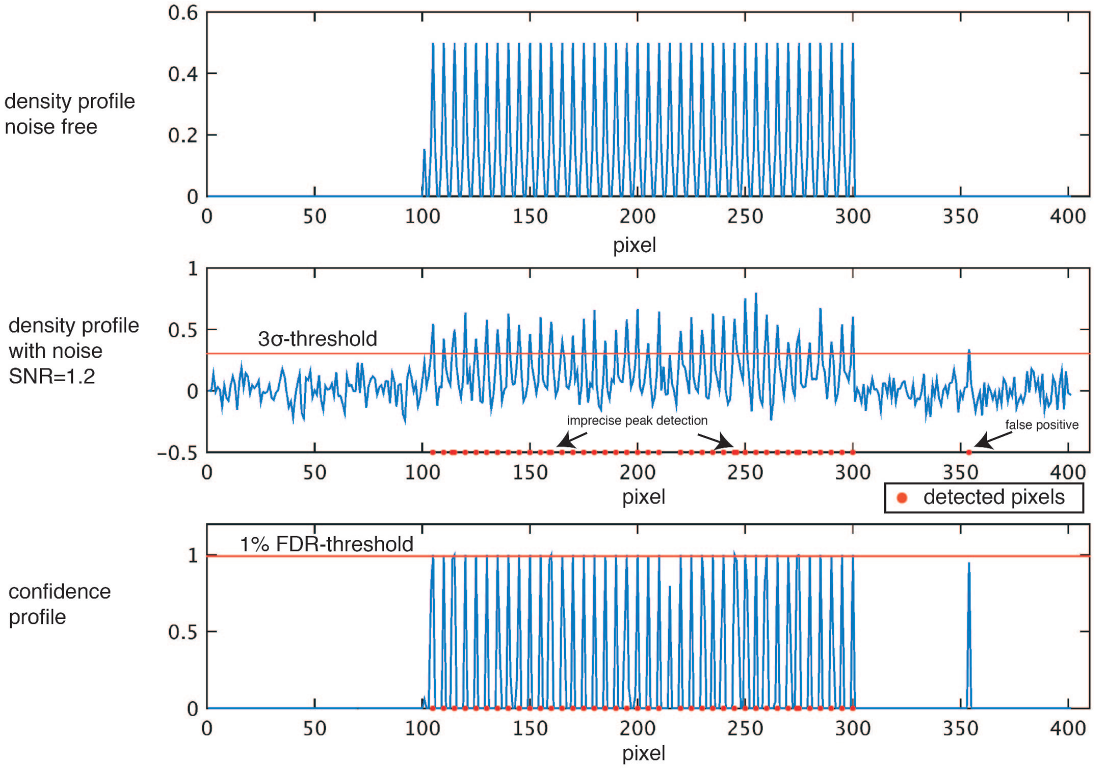


Figure 4.2: Effect of σ and FDR thresholding on 1D density profiles. One-dimensional stacked plots of grid density with noise-free original (top), at signal-to-noise ratio of 1.2 (center) and confidence map (bottom). The noisy density grid is thresholded at 3σ and the confidence map is thresholded at 1% FDR. Conventional 3σ-thresholding yields higher rates of false positives and some imprecise peak positions (arrows). This figure is taken from [10] and was prepared by myself.

4.3.2 Statistical Properties of cryo-EM Maps

Cryo-EM maps are subject to several sources of noise. Background noise is mainly the result from shot noise, i.e. electron counting statistics, as well as interactions with the amorphous ice. While the first can be assumed uniform over the images, the latter will be higher outside the particle, due to lower relative ice thickness in the region of the ice embedded particles.

Beyond background, additional sources of inaccuracy influence the map values: mainly flexibility, stoichiometry of binding, radiation damage and computational inaccuracies. Although these additional sources of noise add additional variation to the signal, they are irrelevant for the task of detecting significant deviations from the background noise. To calculate false discovery rates of deviations from background noise, we only need to be able to get estimates of the background noise distribution. In this section I provide an analysis which investigates the basic statistical properties of cryo-EM maps: the distribution of background noise and the expected behaviour of the signal of the embedded particles.

Distribution of Background Noise

In order to analyse the distributions of background noise in cryo-EM map, I analysed 32 deposited cryo-EM maps in a resolution range from 2 to 8 Å and compared the empirical cumulative density functions (ECDF) with the estimated Gaussian CDFs (Fig. 4.3 a). Clearly all of the analysed maps follow the ideal Gaussian CDF closely. Moreover, for each map normality has been assessed by Anderson–Darling testing [3] and showed that 75% and 87.5% of the maps do not significantly deviate from normality when thresholds corresponding to 1% and 0.1% family-wise error rates ($FWER$) are chosen (Fig. 4.3 b). An explanation for deviations from a Gaussian distribution is given by the three-dimensional reconstruction procedure. When correctly aligned images containing white Gaussian noise are combined by linear inversion, the obtained three-dimensional volume will also have a Gaussian distribution. However, when the images need to be aligned, alignment of noise might result in heavy-tailed distributions. Moreover, three-dimensional reconstructions will always contain local correlations. In order to investigate the resulting noise of three-dimensional reconstructions in more detail, I generated from pure noise images with even angular sampling. Inspection of the resulting amplitude spectrum reveals that it differs from pure white noise as it shows correlations between adjacent pixels (Fig. 4.3 c left). Moreover, the estimated variances for each voxel from 900 reconstructions show that they can be approximated uniform over the central sphere (Fig. 4.3 c right).

To further investigate the effects of non-normality I calculated confidence maps of the map EMD-6287, which shows a significant deviation from normality according to the Anderson–Darling test, using both Gaussian and empirical CDF. Inspection of the confidence maps (Fig. 4.3 d) showed that the visual agreement between the two maps is remarkably high and a difference map between the two confidence maps shows no systematic deviations.

All in all, background noise in cryo-EM maps can be approximated sufficiently well by a normal distribution. However, deviations from normality are monitored during confidence map calculation by Anderson-Darling testing as well as comparison of the empirical CDF with the Gaussian CDF. If the assumption of Gaussian noise is violated, the non-parametric approach should be used in these cases.

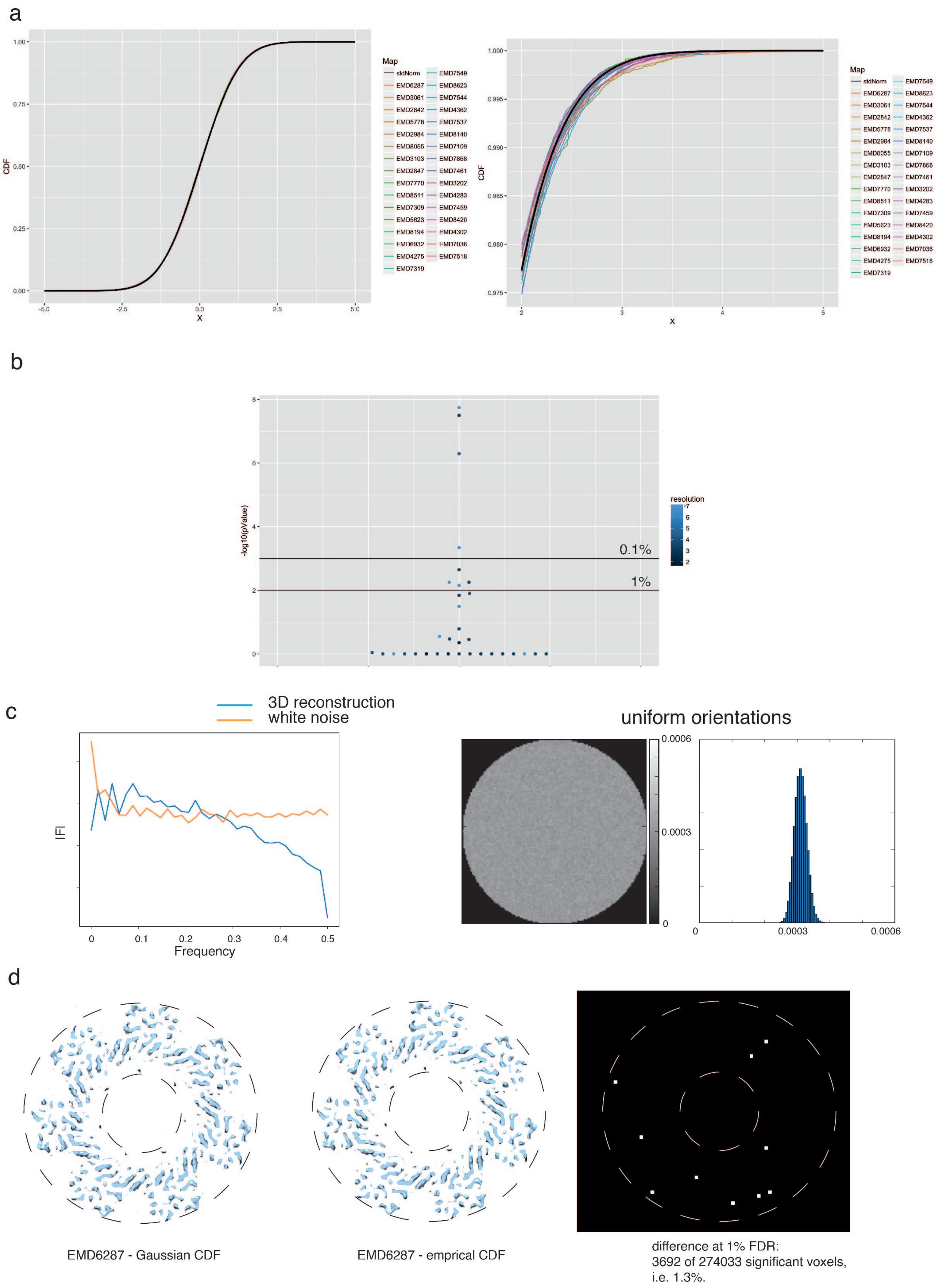


Figure 4.3: Analysis of normality of cryo-EM densities. a) Left. Overlay of 32 cumulative density functions (CDF) derived from the above EMDB entries with ideal Gaussian CDF in black. Right. Zoomed inset to better highlight small differences. b) 32 map entries are assessed with respect to normality according to the Anderson-Darling test, significance thresholds are displayed 1.0 and 0.1% $FWER$ respectively. c) Left. Rotational power spectrum of a 3D reconstruction of white noise images in comparison with pure white noise spectrum. Right. Slice through 3D volume of variances estimated from 900 independent reconstructions from Gaussian white noise images with similar uniform orientations together with a histogram of the estimated variances, showing that background noise can be assumed uniform over the central sphere in the reconstructed volume. d) Cross-sectional view of confidence maps generated of EMD-6287 using Gaussian and empirical CDF. Difference map between 1% FDR binarized confidence maps in the respective image slice. This figure is taken from [10] and was prepared by myself.

Positivity of Signal in cryo-EM Maps

Analysis of cryo-EM maps in standard visualization software is inherently based on the assumption that protein gives rise to positive map values. In the statistical framework this is incorporated via testing for positive values beyond background (see Methods). However, recent results also suggested the presence of some negative signal [66] in cryo-EM maps.

The statistical framework of confidence maps can be easily adjusted to negative values by only inverting the sign of the input maps, which results in testing for negative values beyond background. Visualization of confidence maps from negative densities indeed reveals significant signal. It mainly occurs in regions between positive protein density, often in a spatially complementary way (Fig. 4.4 a, left). To test if the significant negative densities coincide with atomic coordinates, I compared the overlap with negative densities using the independently determined X-ray structure of the 20S proteasome (PDB entry 1pma; [88]). However, the significant negative density has only a very small 2.5% overlap with atoms, which is close to the used false discovery rate of 1% (Fig. 4.4 b). Using positive density, the same analysis gives 60% of the atomic coordinates in the 1% FDR-contoured confidence map and that 10% of this volume is occupied by modelled atoms. Taken together, there is significant negative signal beyond background noise in cryo-EM maps, but only a small fraction of it is occupied by atoms. Biggest amounts of negative density can be found embedded between strands of positive density, most likely owing to the fact that the molecular density there is lower compared to the surrounding solvent area, which will lead to reduced scattering and thus to weaker intensities.

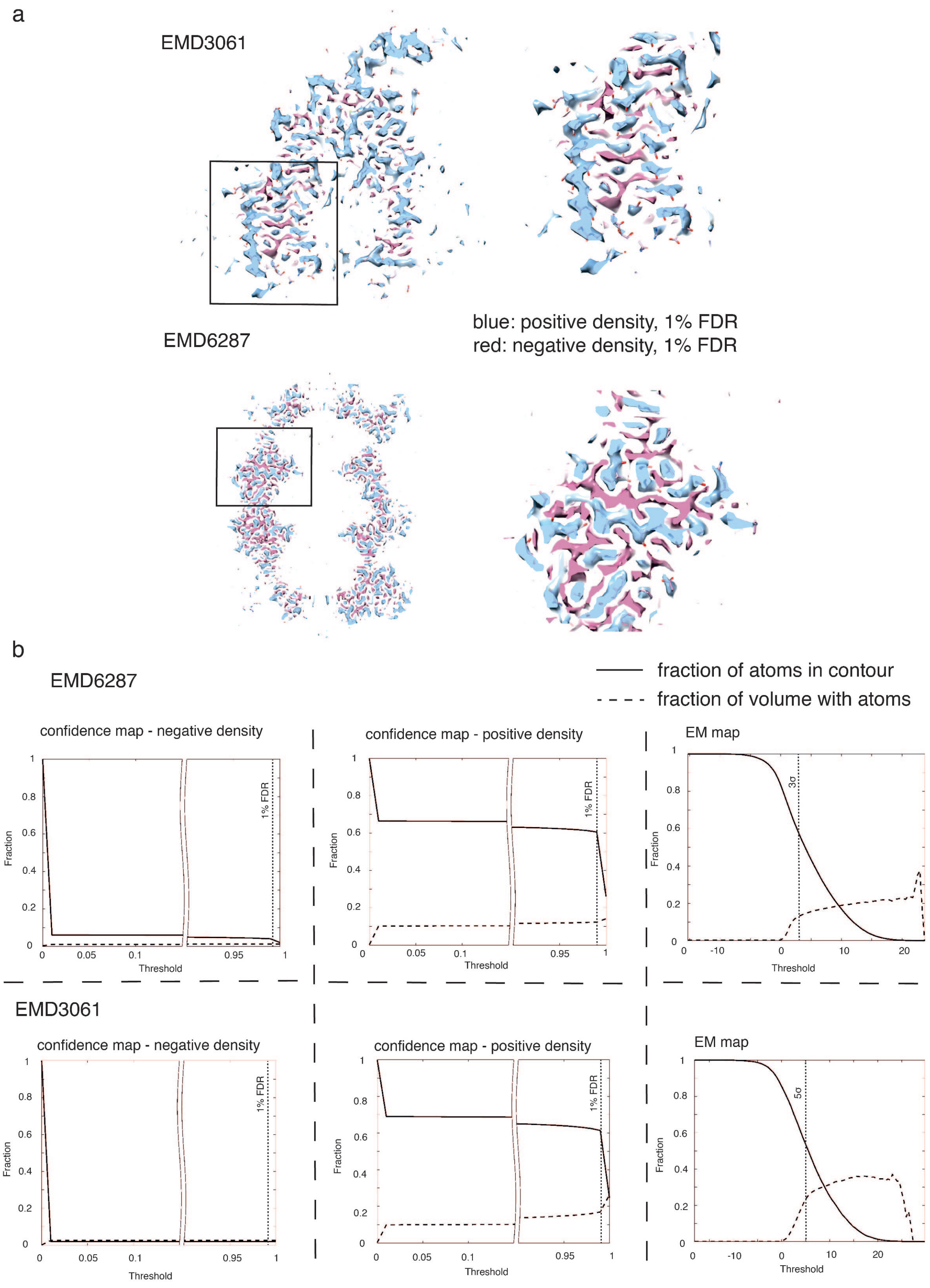


Figure 4.4: Analysis of positive and negative densities using confidence maps. a) Overlay of 1% FDR positive (blue) and negative (red) confidence maps from original and inverted densities of EMD-3061 (top) and EMD-6287 (bottom) respectively. b) Comparison of detected signal with corresponding atomic models by determining the fraction of overlap of atoms with volume and fraction of volume with atoms as a function of threshold for negative (left), positive (center) confidence maps and cryo-EM maps (right), respectively. This figure is taken from [10] and was prepared by myself.

4.3.3 False Discovery Rate Control of cryo-EM Maps

Basic properties of confidence maps can be illustrated with the 3.3 Å cryo-EM map of *Tobacco Mosaic virus* (TMV, EMD-2842). Analysis of a slice through the cryo-EM map of TMV and comparison of the confidence map with the cryo-EM density (Figs. 4.5 a and b) reveals backbone traces that contain values close to 1 in confidence maps, while the cryo-EM maps exhibits a continuum of map values. Confidence maps show strong contrast with respect to background noise, which is suppressed towards values of 0. The histograms of the confidence map revealed a obvious peak beyond 0.99 PPV (1% FDR), thresholding 5.7% of all voxels within the density. In the case of the deposited cryo-EM map, the recommended 1.2σ threshold also yielded a recognizable outline of helical pitch contours. However, only detecting 3.7% of voxels from the map. Similar to isosurface-rendered cryo-EM densities, the confidence map exhibits recognizable details, such as the helical pitch and many side chains of the central helices (Fig. 4.5 c). When going to a very low FDR of 0.01%, some density becomes discontinuous and smaller features disappear. At higher FDR thresholds such as 10%, obvious noise starts to appear. At a 1% FDR threshold, the appearance of noise is minimal in the confidence maps. For reasonable FDR thresholds around 1% the features are robust towards the actual choice of FDR threshold. This is in contrast to cryo-EM maps, where the appearance of noise is very sensitive to small changes in the threshold level, especially close to background noise levels.

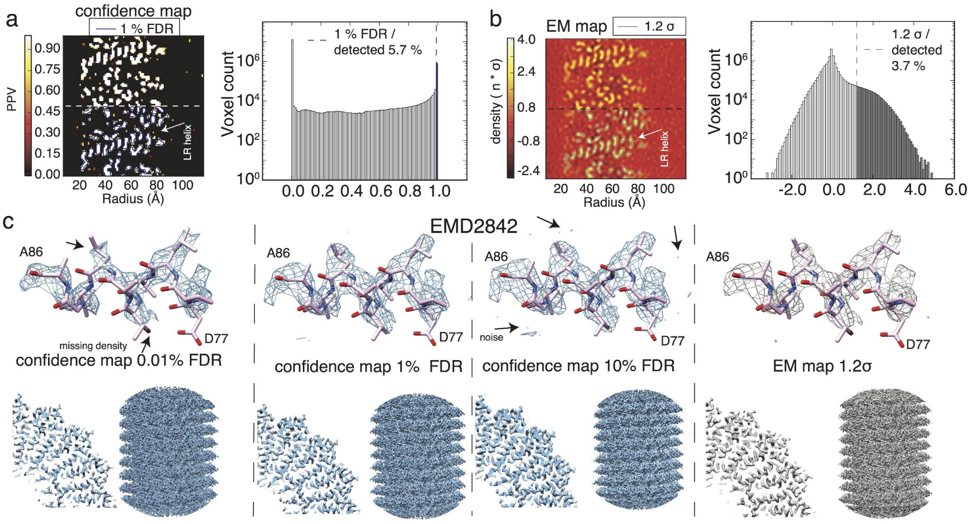


Figure 4.5: Confidence maps separate signal from noise for molecular-density interpretation. a) Left: confidence map with a longitudinal section through the TMV coat protein displayed, indicating the α-helical pitch of the LR helix. The lower half shows the chosen contour at 1% FDR in blue with 5.7% of voxels detected. Right, the corresponding histogram of the confidence map with signal separated above 0.99 PPV (1% FDR). b) Left: the same section as in a) from cryo-EM density and the recommended threshold contoured at 1.2σ in gray with 3.7% of voxels detected. Right: the corresponding histogram of the cryo-EM density with thresholded values displayed in gray. c) Isosurface-rendered thresholded confidence maps at 0.01%, 1% and 10% FDR (left, center left and center right, respectively) shown in blue and sharpened cryo-EM density with a 1.2σ threshold (right) in gray from TMV (EMD-2842 [46]). Shown are helix Ala86–Asp77 (top), a quarter cross-section (bottom left) and a side view (bottom right) of the TMV map. This figure is taken from [10] and was prepared by myself.

Additionally, I also examined generated confidence maps from EMDB model challenge targets, which were determined at near-atomic resolution: 20S proteasome [24] and γ-secretase [4] (Figs. 4.8 a and b). They confirm the previous observation that at rather low FDR levels of 1% they provide structural details at near-atomic resolution while effectively controlling background noise.

To evaluate high-resolution molecular details of the confidence map, I inspected more ambiguous density features of the TMV map. Peripheral density at lower and higher radii of the virus was very difficult to interpret in previous work [46, 125, 101]. For these regions, we found that well defined features are present in the 1% FDR confidence maps. The densities of the coat protein for the loops Gln97–Thr103 located at the inner radius and Thr153–Gly155 at the outer radius are not present in the respective EM map, but can be clearly traced in the 1% FDR confidence map (Fig. 4.6, center). In addition, side-chain density for Lys53 facing the adjacent subunit was found to be clearly significant, while being discontinuous in the original map (Fig. 4.6, bottom left). Based on confidence maps, the readjustment of side-chain rotamers was possible, as illustrated for example by significant density for Arg61, which suggests a realignment of Arg61 to form stabilizing interactions with the aromatic residue Trp152 (Fig. 4.6, bottom right). The presented examples of TMV show that confidence maps represent a fundamentally new way for density visualization and interpretation. Although adjustments of thresholds of cryo-EM maps can also help model interpretation in ambiguous regions and enhance weak density features, they also amplify noise and increase the risk of noise interpretation. Validation by means of statistical significance is essential in such cases.

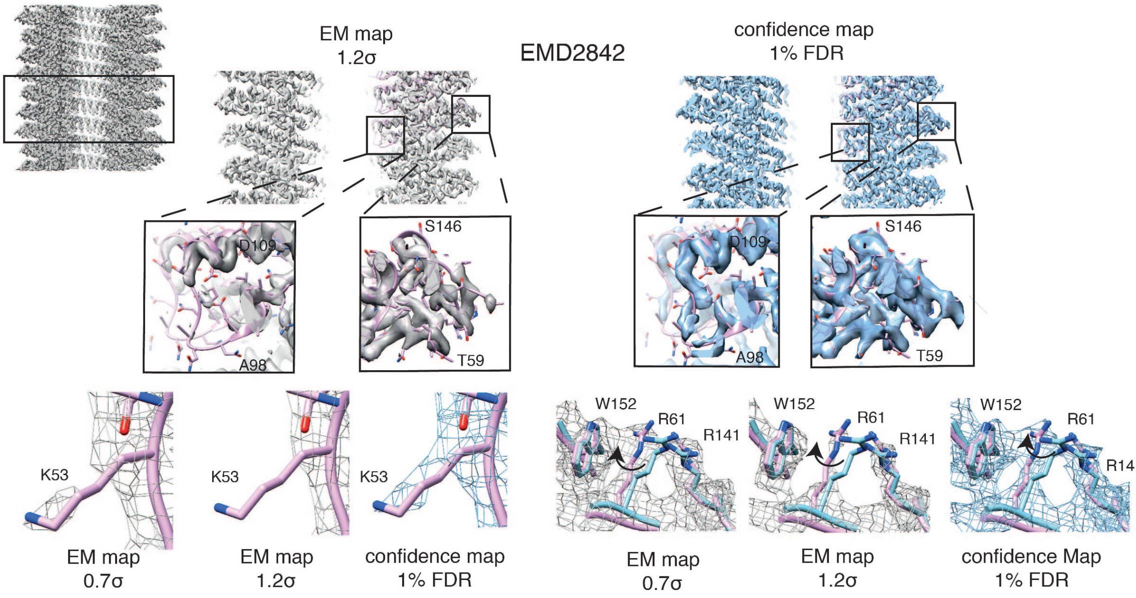


Figure 4.6: Confidence maps facilitate the detection of weak density features. Detailed comparison of TMV density and the corresponding confidence map. A slice view through the TMV rod with enlarged insets for inner and outer radii density (top). Lys53 side-chain density (left) and the molecular environment of Arg61 side chains (right) are shown at 0.7σ and 1.2σ thresholds and in a 1% FDR confidence map. This figure is taken from [10] and was prepared by myself.

4.3.4 Robustness of FDR-controlled Density Transformation

To test the influence of the actual box sizes used for background noise estimation on the resulting noise statistics, I compared the background noise variances for box sizes from 5 to 30 voxels. Results show that the variances are rather robust towards the specific box size (Fig. 4.7). Moreover, in order to test the effect of inaccurate background noise estimates on the resulting confidence maps, I calculated confidence maps by using half or three quarters of the determined variance of the 20S proteasome densities (Fig. 4.8 c). The resulting confidence maps displayed at 1%FDR revealed declaration of background as signal, which poses a principal risk of over-interpretation. As the variance is underestimated, p-values will be underestimated as well and as such also the FDR. This risk, however, is less relevant when the variance measurements outside the particle proposed here are used, as we tend to overestimate noise as a consequence of the specific imaging process (see section Statistical properties of cryo-EM maps).

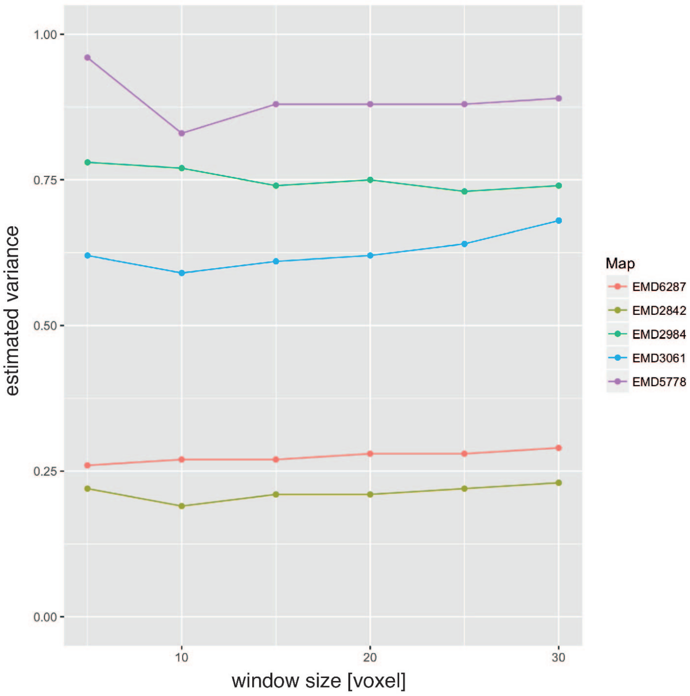


Figure 4.7: Effect of window size on estimated variance. Estimated variance is stable with increasing window size from 5 to 30 voxels for a series of EMDB entries. This figure is taken from [10] and was prepared by myself.

Therefore, I tested the effect of overestimation of the variance by 1.25-fold, 2-fold and 8-fold and calculated the respective confidence maps. We know from *Propositon 1* that this will lead to a more conservative procedure, i.e. less voxels will be declared as signal. However, obvious map features at 1% FDR only start to disappear when the variance is heavily overestimated by a factor of 8; for smaller errors in the estimates the effect is hardly visible in the appearance of the map.

The post-processing of the cryo-EM map is another important noise-related aspect before the FDR-control, namely the degree of sharpening and filtering. Application of a series of B factors from 0 to 250 Å to the 20S proteasome map and subsequent calculation of confidence maps showed that with increasingly negative B factors the corresponding confidence maps at 1% FDR declare less signal as significant. Compared to cryo-EM maps, which become severely oversharpened and the density features are dominated by noise (Fig. 4.8 d), the increase in background noise will be taken into account in the confidence maps. As such, confidence maps inherently avoid the enhancement of noise features that could be mistakenly interpreted as signal and can be used to optimize the sharpening level.

Using undersharpened maps for the calculation of confidence maps, the resulting confidence maps will contain only low-resolution features lacking high-resolution detail at the respective significance level, in analogy to cryo-EM densities. A *vice-versa* behaviour is apparent for low-pass filtering: filtering at lower resolutions will lead to suppression of high-resolution features. On the other hand, filtering at too high resolutions leads to higher noise levels.

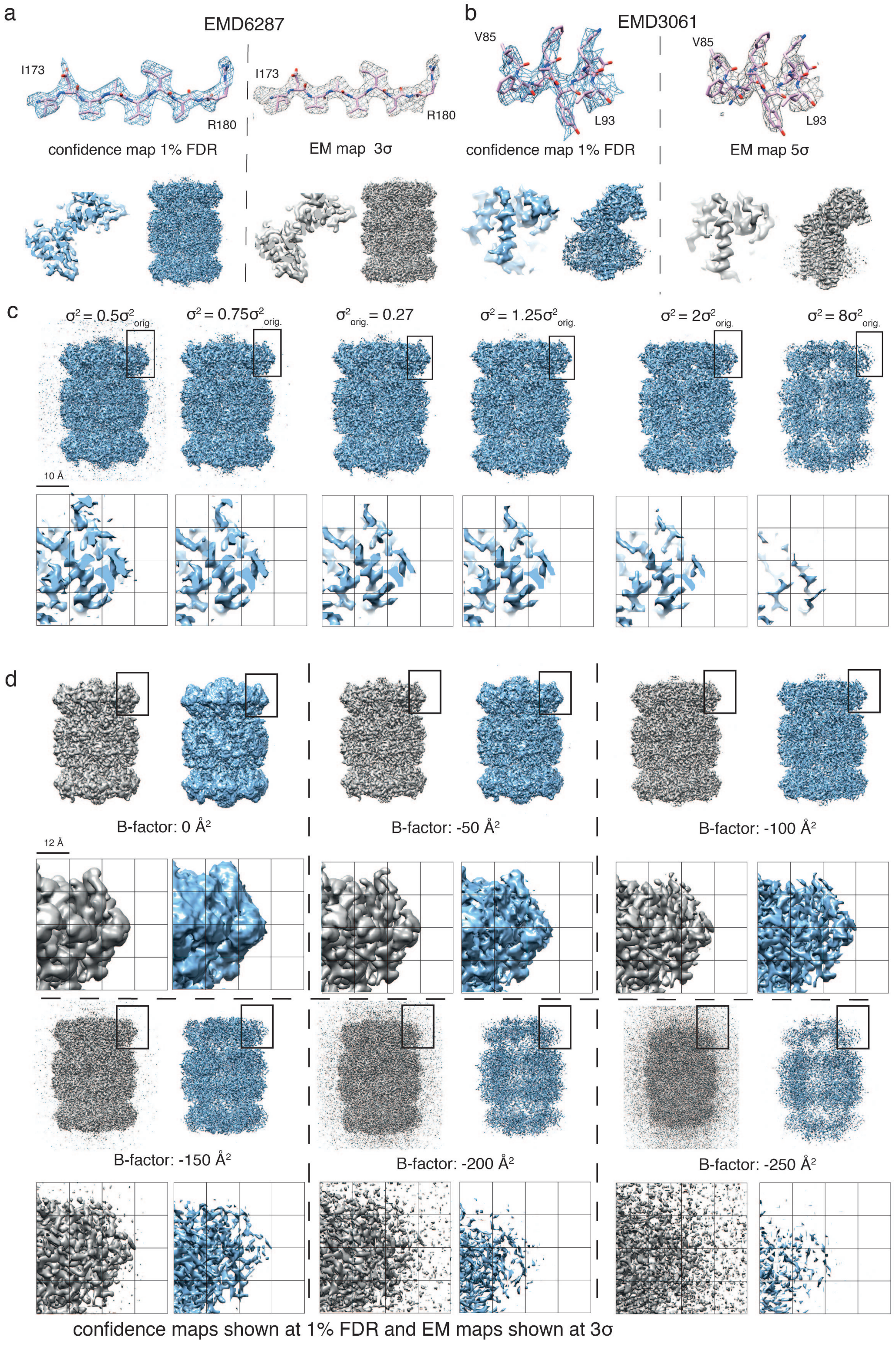


Figure 4.8: Confidence maps and effect of incorrect noise estimation. a) 20S proteasome map (EMD-6287) comparison of 1% *FDR* density (left) and 3σ-thresholded map (right). Shown are molecular details from I173 – R180 (top), slice view (bottom left) and side view (bottom right) of density. b) γ-secretase map (EMD-3061) comparison of 1% *FDR* confidence map and 5σ-thresholded map. c) Six confidence maps of 20S proteasome (EMD-6287) including magnified inset based on incorrect variance estimation: 1st and 2nd left noise is underestimated by 0.5 and 0.75 times the variance (σ^2). In comparison with the correctly estimated noise (3rd), they show excessive noise features declared as signal at 1% *FDR*. When noise is overestimated, which is more likely for cryo-EM maps, confidence maps are quite insensitive to changes in map appearance. For multiples like $1.25\sigma^2$ and $2\sigma^2$ no apparent density changes become visible (4th and 5th) unless strong overestimation like $8\sigma^2$ (6th) leads to disappearance of map features at a 1% *FDR* threshold. d) When applying a series of B-factors to the 3D reconstruction of the 20S proteasome map, we see that with higher B-factors, sharpened EM densities become dominated by noise whereas corresponding confidence maps displayed at 1% *FDR* show disappearance of significant features thereby avoids over-interpretation of noise features. This figure is taken from [10] and was prepared by myself.

4.3.5 Confidence Maps from Subtomogram Averages

In addition to high-resolution structures, the interpretation of lower resolution maps also clearly benefits from confidence maps. with this in mind, I analysed a subtomogram average of the HeLa nuclear pore complex reconstructed from eight pore particles at 90 Å resolution [91], which were determined *in-situ*. Continuous densities for the cytoplasmic and inner rings can be clearly found in the deposited map. However, the map is noisy when visualized at a threshold of 2.0σ (Fig. 4.9 a). The confidence map at 1% FDR shows continuous features for the ring structure with controlled background noise and allows unambiguous interpretation.

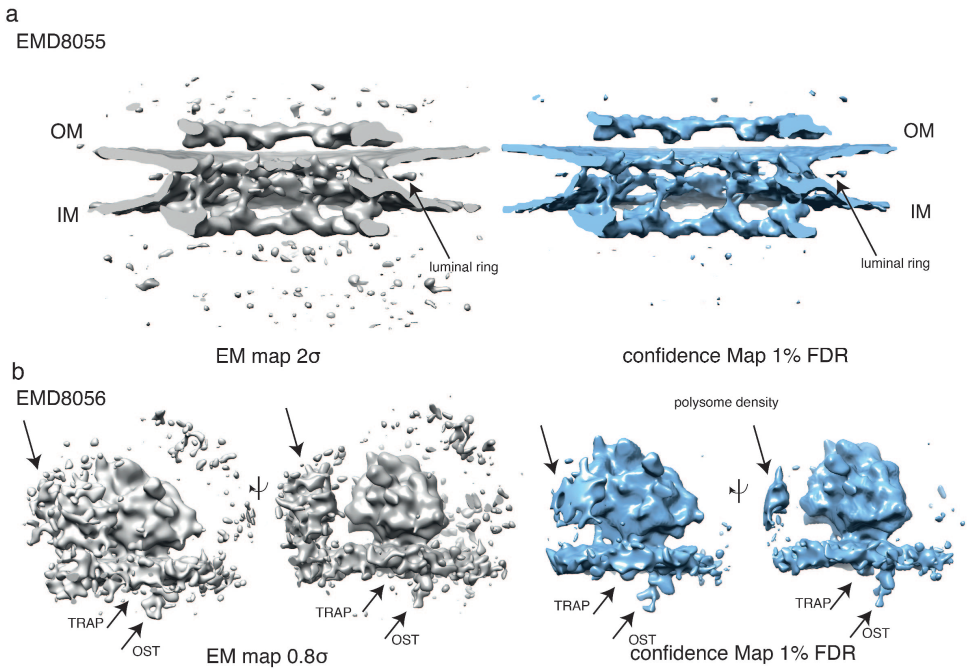


Figure 4.9: Confidence maps from subtomogram averages. a) Nuclear pore structure at 90 Å(EMD-8055) from eight pore particles: cryo-EM map at 2.0σ threshold (left, gray) and confidence map at 1% FDR threshold (right, blue). Note that the confidence map minimizes the appearance of noise. b) ER-associated ribosome structure at 35 Å resolution (EMD-8056) in two side views at a 0.8σ threshold (left) and 1% FDR confidence map (right). Note that in confidence maps weaker densities assigned to the peripheral protein complexes TRAP and OST (arrows) can easily be visualized in the absence of noise. This figure is taken from [10] and was prepared by myself.

From the same tomograms a subtomogram average of ER-associated ribosomes could be produced as well. The map could be reconstructed at a resolution of 35 Å at the membrane, with weak density below the membrane that was interpreted as a translocon-associated protein complex and an oligosaccharyltransferase. However, the corresponding densities can only be visualized at lower thresholds corresponding to 0.8σ, which increases the amount of background noise and prevents faithful molecular interpretation (Fig. 4.9 b). The 1% FDR confidence maps are able to display the additional protein complexes under controlled noise levels. However, in order to calculate a confidence map from a subtomogram average, care must be taken to identify areas of noise without any signal in order to estimate the noise variance reliably (Fig. 4.10 a,b, and c), which usually requires manual specification of the noise boxes.

Moreover, I analysed the deposited maps of the 23 Å resolution nuclear pore structure and the corresponding confidence maps, again determined by subtomogram averaging ([150]; Fig. 4.10 d). Focusing on the ambiguous density assignment of the linker region of Nup133, the presence of density in the 1% FDR confidence maps confirms the interpretation of the authors of placing the Nup133 linker region connecting the N-terminal-propeller and C-terminal-helical domain (Fig. 4.10 d, top right). Additional densities in the connecting regions between the inner and nuclear ring as well as between the inner and the cytoplasmic ring can be found in the confidence maps (Fig. 4.10 d, bottom). Both densities are not found at the recommended threshold of 2.1σ, but significant at 1% FDR.

It should be mentioned here, that in contrast to clearly defined features in high-resolution protein structures (like secondary structure and side chains), we generally do not know how the expected density features of lower resolution structures should look like, which further complicates manual thresholding as well as the validation of additional densities.

In conclusion we can say that also for lower resolution maps as subtomogram averages, confidence maps facilitate density interpretation and help in the assignment of ambiguous map features.

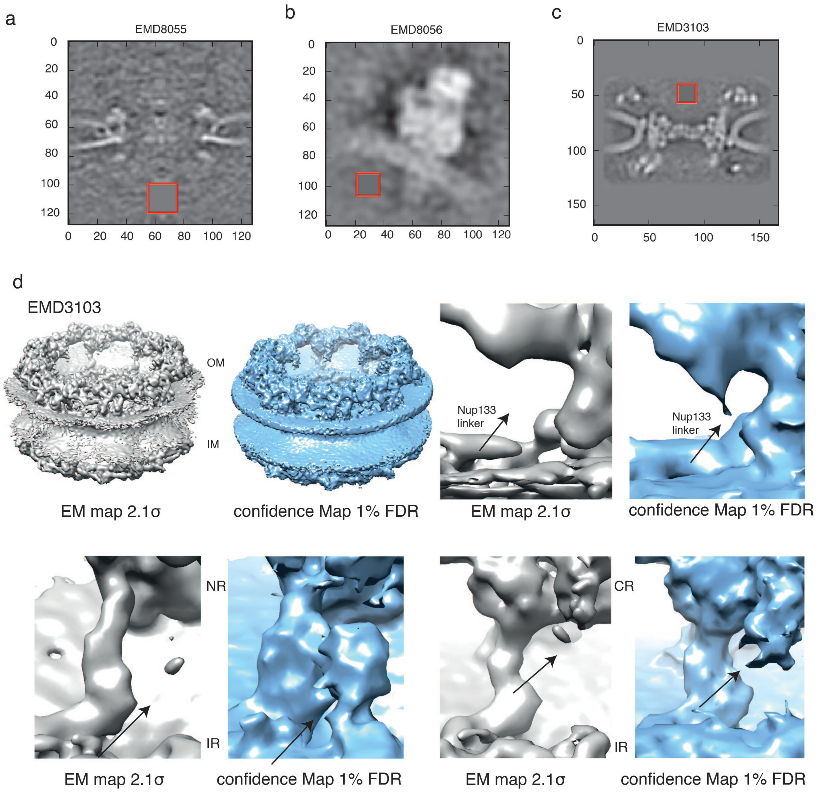


Figure 4.10: Noise estimation in subtomogram averages. Gray-scale density slices with red windows for the voxel region used for variance estimation: a) EMD-8055: nuclear pore from HeLa cells by FIB-SEM, b) EMD-8056: ER-associated ribosomes, c) EMD-3103: 23 Å resolution nuclear pore subtomogram average. d) Nuclear pore structure at 23 Å resolution (EMD-3103) comparing cryo-EM map at 2.1σ threshold (left) and 1% FDR confidence map (right). Comparison of map pairs for Nup133 linker density (top right), densities located between inner and nuclear ring (bottom left) and inner and cytoplasmic ring (bottom right). In contrast to sharpened cryo-EM maps at 2.1σ threshold, confidence maps show consistently densities at the connections between the inner and outer rings at 1% FDR threshold (arrows). This figure is taken from [10] and was prepared by myself.

4.3.6 Incorporation of Prior Information

As described in the methods section, additional information based on local resolutions can be incorporated via local filtering and sharpening to increase local signal-to-noise levels. Incorporation of these post-processing steps and consideration of the resulting local background noise levels will lead to improved detection of low-resolution features due to improved signal-to-noise ratios in the map.

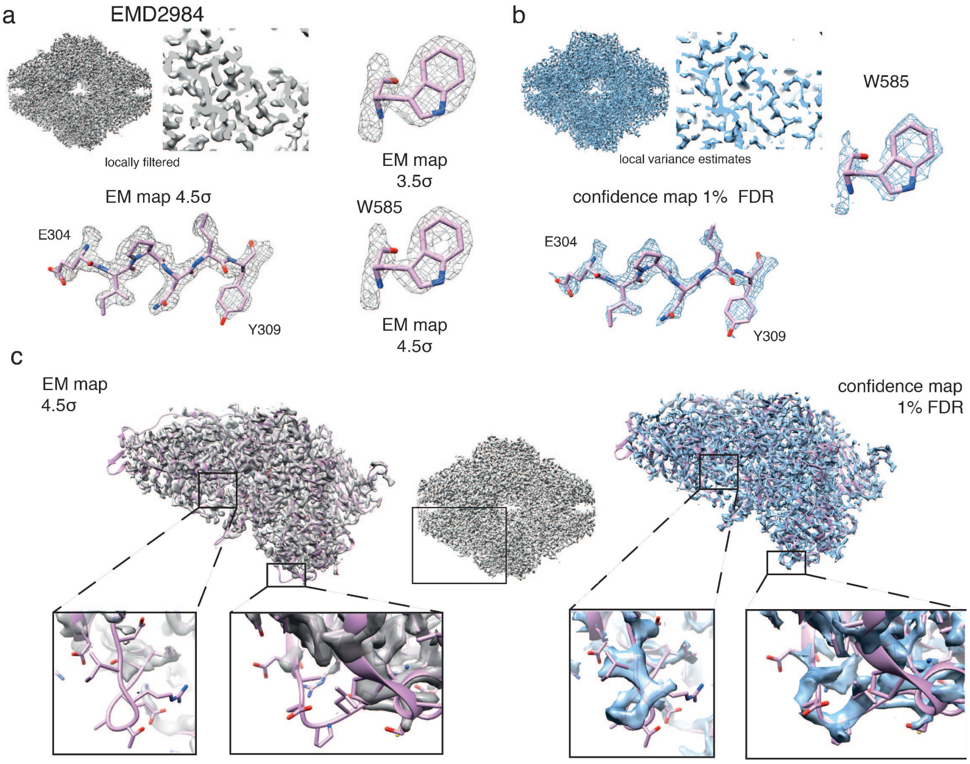


Figure 4.11: Confidence maps benefit from prior information based on local resolutions. a) Locally filtered β-galactosidase (EMD-2984) cryo-EM map (gray) displayed at a 4.5σ threshold (left) and b) confidence map (blue) including prior information based on local resolution at a 1% FDR threshold (right) in side view and cross-section. High-resolution features such as Glu304–Glu398 and holes in the aromatic rings of Trp585 in the 3.5/4.5σ - thresholded cryo-EM map (a) in comparison with the 1% FDR confidence map b). c) Comparison of density features from peripheral loop regions not covered by density in the locally filtered cryo-EM map (left) compared with the 1% FDR confidence map that shows densities for the respective loops. This figure is taken from [10] and was prepared by myself.

Analysis of the very high-resolution map (2.2 Å resolution) of β-galactosidase (EMD-2984; [8]) showed that the applied sharpening levels, which were necessary to reveal the high-resolution details in the center of the map, resulted in heavily oversharpened peripheral parts of the protein, mainly due to resolution ranges from <2 Å in the center to 3.8 Å in the periphery. Calculating a confidence map from the deposited map confirmed this observation, with almost no significant signal at 1% FDR in the periphery of the protein complex (Fig. 4.13). However, incorporation of local resolutions followed by FDR control resulted in a evenly distributed 1% FDR confidence map including the β-gal periphery (Figs. 4.11 a and b, top). Moreover, side-chain details such as holes in aromatic rings can be resolved at the same significance level, as shown for Trp585 (Figs. 4.11 a and b, bottom). Inspection of the cryo-EM map shows that density for the peripheral loops of the β-gal complex can not be found at a 4.5σ threshold, but obvious significant continuous loop density at a FDR of 1% in the resolution-compensated confidence map (Fig. 4.11 c, left and right).

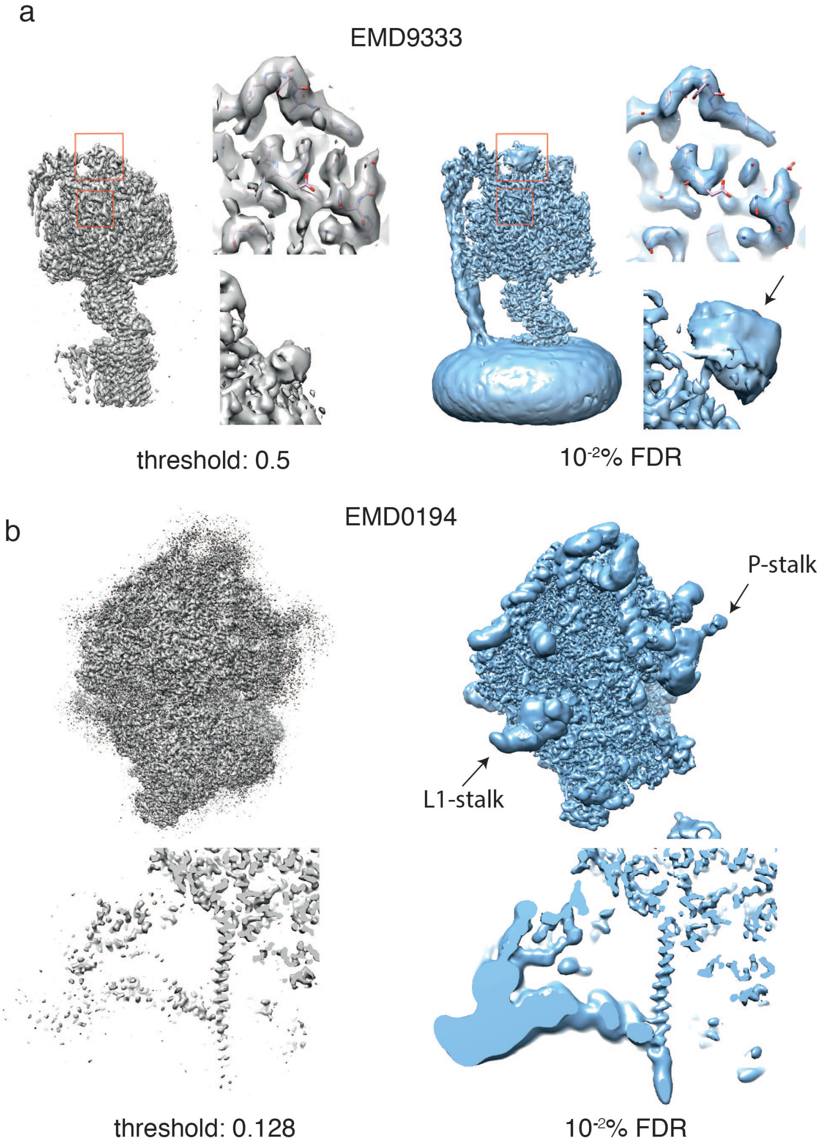


Figure 4.12: Confidence maps facilitate the overall interpretation cryo-EM maps. a) Locally filtered cryo-EM map (EMD9333) of a Bacterial ATP synthase (left) with the corresponding confidence map (right) and magnified views. The arrow shows the density that likely correspond to a His-tag. b) Cryo-EM map of a eukaryotic ribosome (EMD0194, left) together with the corresponding confidence map (right) and magnified views.

Beyond the β-gal example, local resolution variations occur for many maps, and often also in a much more extreme way. As many maps contain disordered protein parts when compared to the rest of the complex. For example, comparison of the map of a bacterial ATP synthase (EMD9333)[56], which I locally filtered in this case, to the corresponding confidence map clearly shows that even in the locally filtered map, low resolution parts like the stalk domain are still missing at rather low thresholds at which high-resolution parts already become noisy. In contrast, the confidence maps allows interpretation of the complete complex at a rather low FDR of 0.01%, even with significant low-resolution density that likely corresponds to a 10x His-tag (Fig. 4.12 a). Eukaryotic ribosomes have parts like the expansion segments and the ribosomal stalks, which are usually much worse resolved due to flexibility (Fig. 4.12 b) (EMD0194) [71]. In the deposited cryo-EM map, the respective parts are heavily oversharpened and appear noise-like. Calculating the confidence map with local resolution information clearly depicts the respective domains, with both high and low-resolution features visible at a single threshold of 0.01% FDR. Taken together, the statistical power of the procedure can be improved, i.e. the power with which we can detect signal, while still controlling the FDR by the incorporation of local resolution information. Moreover, similar improvements can be observed with local-sharpening and combination of both local-resolution filtering and sharpening, as illustrated in Figure 4.13. However, for the the peripheral loops of the β-gal complex, which were already examined above, we can see that the incorporation of local resolution via local filtering has a stronger impact. While local sharpening via LocScale definitely leads to improvements in the confidence maps, the loops are not as well defined in the confidence maps as they are when local filtering is incorporated. The reason is that for the task of signal detection we are only interested in good signal-to-noise ratios, which is optimized by local resolution filtering. In contrast, local sharpening aims to scale the amplitude profile in Fourier space, which has itself a less strong effect on the signal-to-noise ratio. However, both local resolution filtering and local sharpening can be combined, but only with minor improvements on the confidence maps compared to simple incorporation of local filtering.

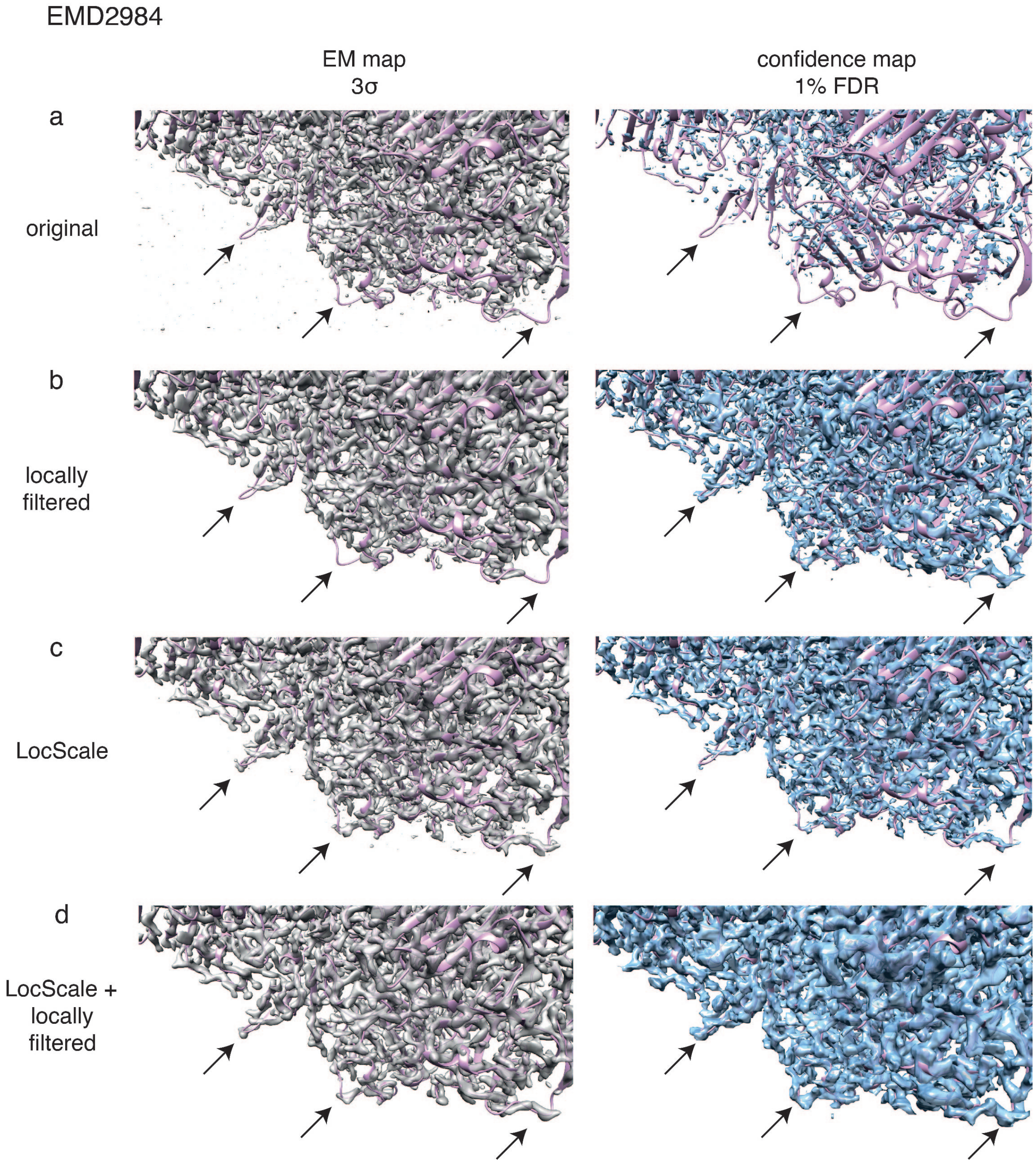


Figure 4.13: Effect of local variance adjustments on confidence maps. β-galactosidase (EMD-2984) cryo-EM map at 3.0σ threshold (left, gray) and 1% FDR confidence map based on different post-processing methods (right, blue). Global sharpening with uniform filtering, local filtering based on local resolution measurements, local sharpening and the combination of local sharpening with local filtering were compared. Confidence maps were generated with local noise estimate based on local resolution measurement, locally scaled window from a model reference structure and the combination of both, which in this case shows the best preservation of molecular density with respect to confidence. This figure is taken from [10] and was prepared by myself.

4.3.7 Confidence Maps for the Annotation of Bound Molecules

Most cryo-EM maps have resolutions between 3 and 4.5 Å. In these resolution regimes most main-chain and large side-chain densities can usually be modelled. However, for residues with smaller side chains and non-protein components such as water molecules and ions, maps at near-atomic resolutions pose the risk of modelling atoms into noise. In order to investigate whether confidence maps can help to mitigate this problem I inspected a putative Mg^{2+} site coordinated by Glu416, Glu461, His418 and three additional H_2O molecules. I performed a rigid body docking of the Mg^{2+} ion together with coordinated water molecules based on a 1.6 Å resolution X-ray crystal structure ([157]; PDB entry 4ttg) into the deposited cryo-EM density map and looked at the overlay. The map at the lower 3.5σ threshold shows convincing density for only two water molecules (Fig. 4.14 a, top left), while the confidence map reveals distinct densities for all three proposed water molecules at 1% FDR (Fig. 4.14 a, top right).

Moreover, the β-galactosidase structure has been acquired in the presence of the small-molecule drug PETG. Finding ligands and modelling their conformation is challenging, as both the ligands as well as the ligand binging sites usually exhibit flexibility and the binding might not occur for all particles used in the 3D reconstruction (Fig. 4.14 a, bottom left), which will lead to worse resolutions and weaker map densities. Looking at the confidence map and comparing the densities with the cryo-EM maps, ligand placement is clearly facilitated using confidence maps, with density being well resolved for the complete inhibitor (Fig. 4.14 a, bottom right).

To investigate whether the detection of smaller ions can be facilitated by confidence maps as well, I took the TRPV1 ion channel [83] and inspected the density in the selectivity filter of the ion channel, which is mainly made of Gly643. In the deposited map a clear density peak can be found in the symmetry center that is compatible with a small ion. In support, the confidence map also shows a density peak at the same position, supporting the presence of an ion at a significance level of 1% FDR (Fig. 4.14 b, bottom right).

By inspection of the γ-secretase complex, density for membrane-embedded phosphatidylcholine (PC) lipid molecules can be found. However, the deposited EM map requires thresholding at different σ-levels between 4 and 5 in order to visualize the two PC acyl chains, (Fig. 4.14 c). In contrast, the corresponding confidence map at 1% FDR contains most of the density of the two acyl chains without any additional threshold adjustments.

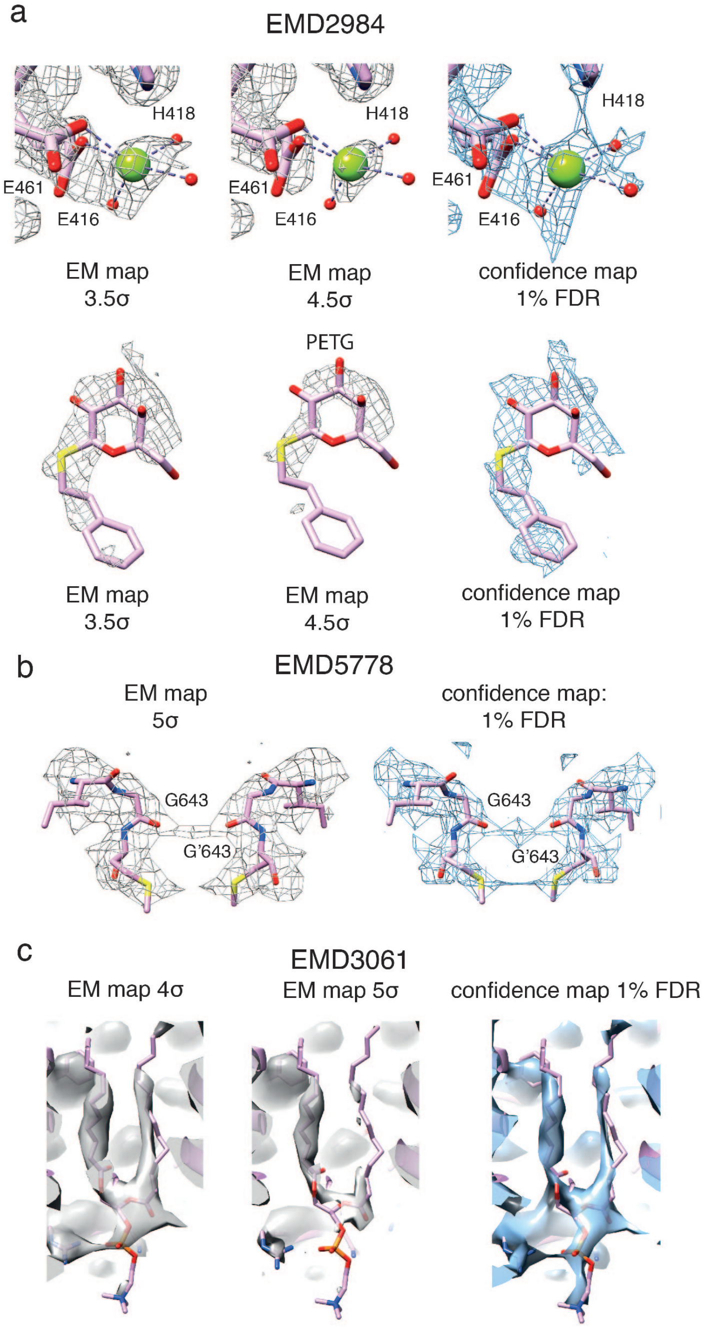


Figure 4.14: Confidence maps confirm the localization of nonprotein components. a) β-Galactosidase (EMD-2984) with 3.5/4.5σ-thresholded cryo-EM maps (left and center, gray) and a 1% FDR-thresholded confidence map (right, blue). Top: the Mg^{2+} ion is coordinated by Glu461, Glu416, His418 and three H2O molecules. Bottom: density of bound PETG ligand in 3.5/4.5σ-thresholded cryo-EM maps and the 1% FDR confidence map. b) TRPV1 channel (EMD-5778) with a 5σ-thresholded cryo-EM map (left) and a 1% FDR-thresholded confidence map (right): the selectivity filter formed by the carbonyls of symmetry-related Gly643 residues. The presence of a putative ion is supported by the confidence map. c) γ-Secretase (EMD-3061) with 4σ- and 5σ-thresholded cryo-EM maps (left) and a 1% FDR-thresholded confidence map (right). The confidence map reveals density for both acyl chains of phosphatidylcholine at a single threshold. This figure is taken from [10] and was prepared by myself.

4.4 Further Developments

Beyond the published work, the concept of statistical signal detection in cryo-EM maps has subsequently been extended towards other error measures ($FWER$ and *local*FDR) and other FDR-controlling procedures in addition to basic Benjamini-Yekutieli FDR. The Benjamini-Hochberg procedure has shown to be more powerful than the Benjamini-Yekutieli procedure. However, it makes assumptions about the dependency structure of the p-values. A similar situation applies to the Hochberg and Bonferroni-Holm procedures, with the former more powerful but also more restrictive in dependency structure. The question arises whether cryo-EM maps fulfil assumptions about certain dependency structures. Moreover the application of the framework to single tomograms has been explored, which will be discussed in this chapter as well.

4.4.1 Beyond FDR - Comparison to $FWER$ and *local*FDR

Multiple testing is a major field of research in statistical inference [160, 166, 68] and, in addition to false discovery rate control, several other error rates have been proposed to account for the multiple testing problem. The Family-wise-error rate ($FWER$) as well as *local*FDR have been implemented in the *FDRthresholding*-software and can be chosen as additional options. Moreover, several approaches for controlling FDR and $FWER$ exist, where in both cases more powerful procedures can be obtained when making rather weak assumptions about the dependency between p-values.

LocalFDR can be seen as a local pendant to the FDR and, as such, quantifies the probability for each voxel to be a false positive. This gives error rates which are readily interpretable in contrast to FDR-adjusted p-values, which only give a global measure over all tested hypotheses. However, it does not deliver information about the global amount of false positives and is more difficult to calculate, as it is a Bayesian posterior probability and requires estimation of the probability densities of both the background noise and the signal distribution. Most useful application of *local*FDR is therefore in conjunction with basic FDR in order to judge the individual local false discovery rates for signal detected at a specified global FDR.

$FWER$ specifies the probability of having false positives after all. In contrast to FDR and *local*FDR, $FWER$ is a rather strict and more conservative error criterion, but in the context of cryo-EM map interpretation $FWER$ is able to deliver important additional information. Especially modelling of water molecules and small ions requires interpretation of single voxels, which can be easily confused with noise features. Thus, the cost of false-positives can be severe. Although controlling the FDR already drastically facilitates this process, we still expect false positives, namely up to as many as is specified by the FDR threshold. Considering 100000 significant voxels, 1% FDR corresponds to a expected maximum of 1000 false positives. Controlling the $FWER$ instead of FDR would as such be desirable. However, the biggest problem of $FWER$ control of cryo-EM maps remains the low statistical power, i.e. a high false-negative rate, which is basically the result of the vast amount of hypotheses that need to be tested in a complete 3D map. In order to analyse the behaviour of $FWER$, FDR and $localFDR$ applied to cryo-EM maps, confidence maps have been calculated from the 20S proteasome map (EMD-6287, [24]) for all three error criteria. Both FDR and $localFDR$ show remarkably similar results (Fig. 4.15 middle and right), with $localFDR$ having a bit more signal. In contrast, $FWER$ clearly annotates less significant signal compared to FDR and $localFDR$. While several noise peaks might be correctly assigned as not significant signal (Fig. 4.15 middle row), also possibly true signal from the protein seems to be missed. For a more quantitative assessment I refer to the next section, when I will also provide a detailed comparison between the different error controlling procedures.

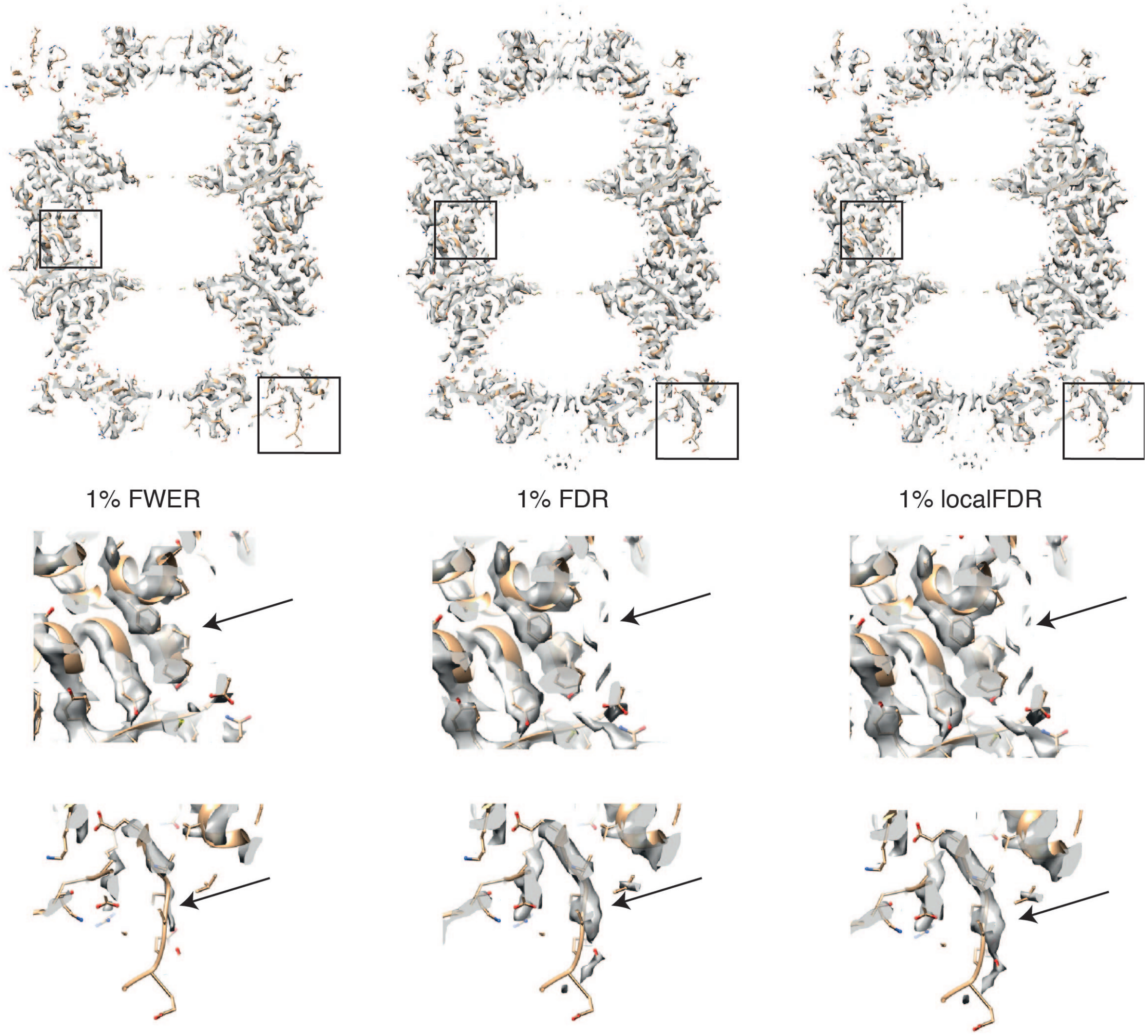


Figure 4.15: Comparison of different error rates for confidence map calculation. Confidence maps for the 20S proteasome (EMD-6287) are shown for FWER (left column), FDR (middle column) and localFDR (right column) in presence of the corresponding atomic model (pdb entry 6bdf). Slices through the maps (top row) as well as two zoomed views (middle and bottom row) are presented.

Choice of the Error Controlling Procedure

Generic black-box methods for the control of $FWER$ and FDR are given by the Bonferroni-Holm and the Benjamini-Yekutieli approach. It can be shown that both procedures control the respective error rates under arbitrary dependencies between the p-values. However, with the Hochberg procedure for $FWER$ control, and Benjamini-Hochberg for FDR control, respectively, alternatives exist that have been shown to posses lower Type II errors, under the restriction of certain allowed dependencies. The Benjamini-Hochberg procedure assumes a so called "positive regression dependency", which is further specified in their paper [12]. The Hochberg $FWER$ procedure is based on the Simes inequality [134], and also assumes positively dependent distributions. Under negatively dependent distributions, both methods fail to control the error level [16].

Cryo-EM maps undergo extensive post-processing with both sharpening and low-pass filtering of the Fourier transform. Low-pass filtering will lead to positive dependencies, sharpening will lead to negative dependencies. As such, the actual dependency pattern between voxels can be assumed to be very complex. While voxels close to each other will be highly positively dependent due to the low-pass filter cut-off, at longer distances the sharpening might lead to negative dependencies. However, if no or very careful sharpening is applied, positive dependency might be a good approximation.

In order to investigate the difference of the multiple testing approaches in the context of cryo-EM maps, we compared the respective approaches to each other at hand a 3.4 Å cryo-EM map (EMD3061) of γ-secretase [4]. As expected, at 1% $FWER$ less density is declared significant compared to 1% FDR, no matter what approach is used (Fig. 4.16). While several noise peaks might be correctly assigned as not significant signal, also likely true signal from the complex seems to be missed, for example for the head of the embedded lipid (Fig. 4.16 right column) appears smaller at 1% $FWER$.

In order to quantify the performance of the different error rates, I applied them to a simulated noisy map of 4194 water molecules (taken from pdb 6cvm [6]). I compared the detected false positive map peaks that cannot be attributed to water molecules and also the amounts of missed water molecules for the different procedures at 1% FDR or $FWER$, respectively. The simulated map was generated with the Chimera molmap function and we added Gaussian white noise with a standard deviation of 0.5, which results in a SNR of 1.75 for the density peaks. While at 1 % $FWER$ we do not have any false positives, at 1% FDR we have, as expected, some false positives hits that could be mistakenly interpreted as water (Tab. 4.1), with the Benjamini-Hochberg procedure having more false positives that almost reaches the controlled FDR of 1 %. However, the decreased amount of false positives in the case of $FWER$ control comes for the price of some missed water molecules.

Moreover, the Hochberg $FWER$ seems not make a obvious difference compared to Holm $FWER$, but the Benjamini-Hochberg FDR-control declares more significant signal compared to Benjamini-Yekutieli. However, as described above, the actual dependency pattern between voxels can assumed to be very complex in a cryo-EM map, and the Benjamini-Yekutieli for FDR and Holm for $FWER$ might be the safer approaches.

All in all, additional error rates like $FWER$ and $localFDR$ can be useful to use beyond basic FDR-controlled confidence maps. Both have their own limitations (low statistical power for $FWER$ and missing global information for $localFDR$), which need to be considered and do not make any of them clearly superior to FDR control. However, with respect to the analysis of cryo-EM maps $FWER$ provides the error criterion that would be the most useful for the interpretation of weak and isolated signal, as it occurs especially for water molecules and bound ligands and is only hampered due to low statistical power.

Table 4.1: Comparison of detected false positive voxels and false negative water molecules for a simulated map of 4194 water molecules (taken from pdb 6cvm). The number of false negative water molecules and false positive voxels together with the true FDR this corresponds to.

Controlling procedure	False positive voxels	False negative water molecules
Holm $FWER$ 1%	0 / FDR: 0.00%	18 of 4194
Hochberg $FWER$ 1%	0 / FDR: 0.00%	18 of 4194
Benjamini-Yekutieli FDR 1%	7 / FDR: 0.05%	0 of 4194
Benjamini-Hochberg FDR 1%	147 / FDR: 0.88%	0 of 4194

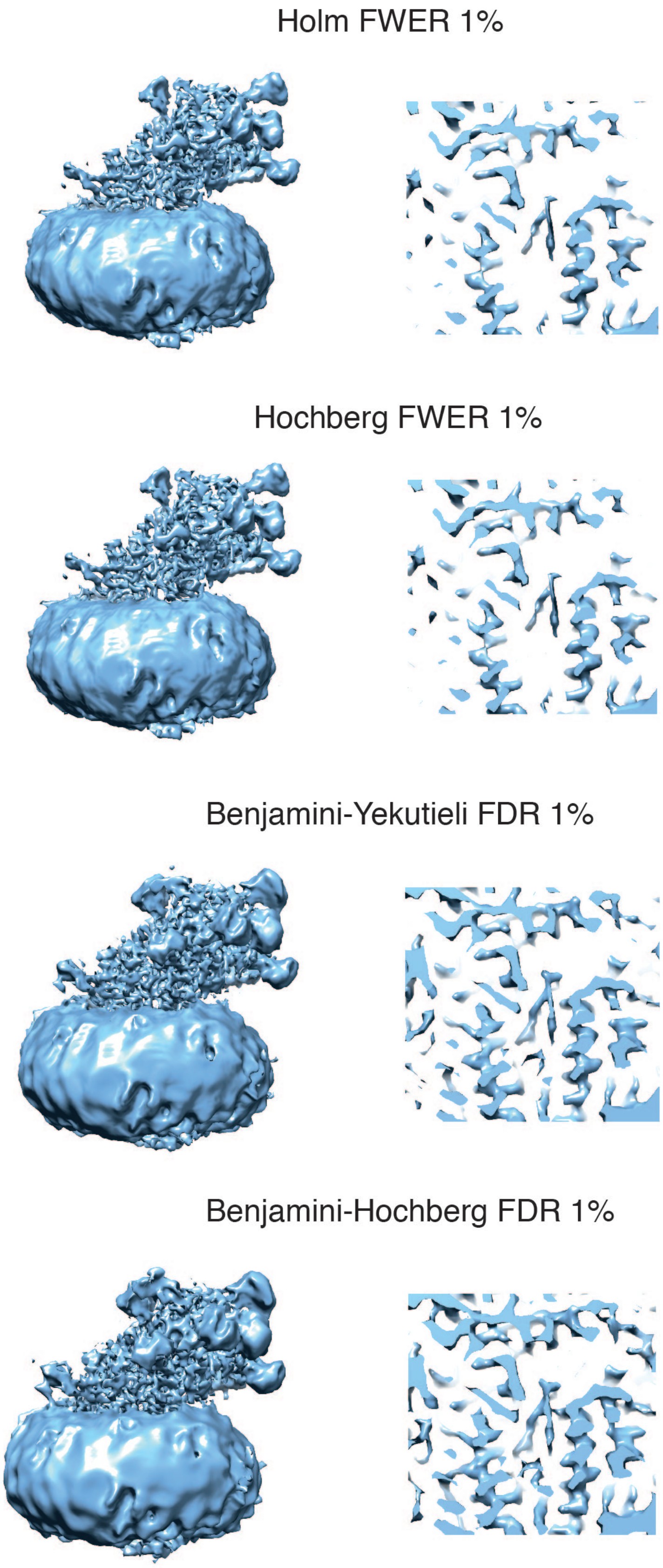


Figure 4.16: Comparison of different $FWER$ and FDR controlling procedures. Confidence maps based on Holm $FWER$, Hochberg $FWER$, Benjamini-Yekutieli FDR and Benjamini-Hochberg FDR are compared with a γ-secretase map (EMD3061). Respective zooms in the inner of the transmembrane domain are sown in the right column.

4.4.2 Application to Tomograms

Cryo-electron tomography is an emerging technique for the analysis of biological samples *in situ*. While the analysis of subtomogram maps by means of multiple statistical hypothesis testing is quite similar to high-resolution maps obtained by single-particle analysis, raw tomograms have a much lower signal-to-noise ratios compared to averaged cryo-EM maps. In principle, the calculation of confidence maps from raw tomograms is possible as well. Electron counting noise will be uniform over the tomogram; however, noise from interactions with the amorphous ice will be different over the tomogram, depending on the sample thickness in the respective regions (an effect that can be ignored when looking at picked particles). If the tomogram has uniform ice thickness, background noise will be roughly uniform as well (with the exception that embedded particles will have lower background from ice due to relative ice thickness as is the case for averaged maps, see above).

Application to a tomogram of the *Chlamydomonas reinhardtii* Golgi apparatus (EMD-3977, [22]) shows that cellular details can be clearly recognized in the statistically significant signal (Fig. 4.17) with clear separation of the Golgi cisternae. However, application of the statistical framework to tomograms is of limited use, as the signal in tomograms is usually not interpreted itself, but rather used for particle picking with subsequent averaging or rough morphological analyses. In both cases, approaches based on supervised learning will be more appropriate, as they can incorporate additional semantic information about the expected specific shapes.

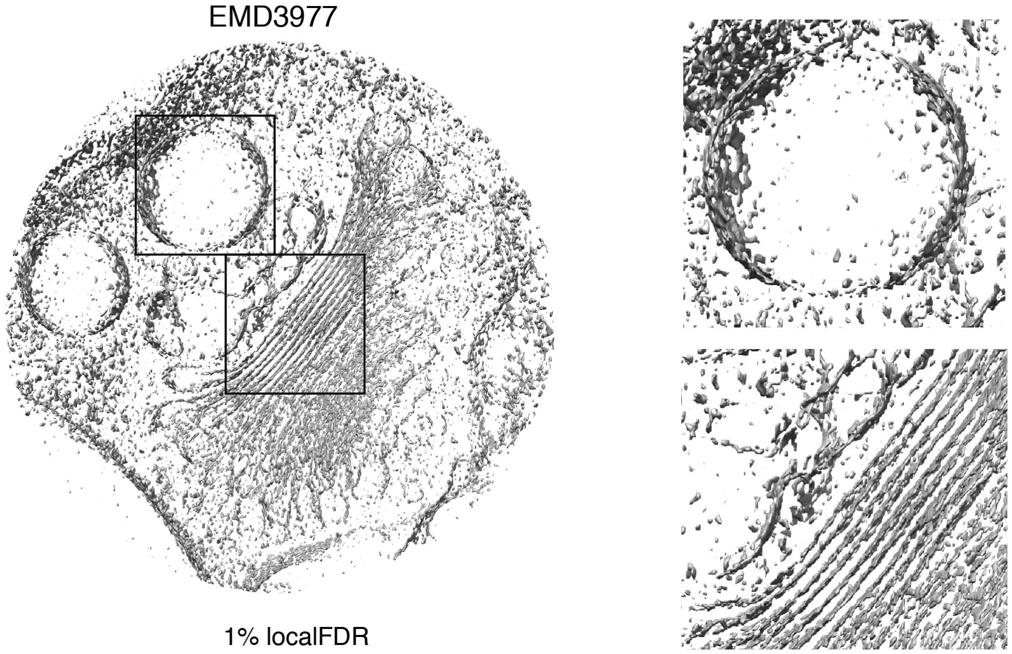


Figure 4.17: Confidence tomograms. A confidence map of a tomogram from the Golgi apparatus of *Chlamydomonas reinhardtii* (EMD-3977). Zoomed views focusing on a vesicle and on the Golgi cisternae shown on the right.

4.5 How to Interpret a Confidence Map

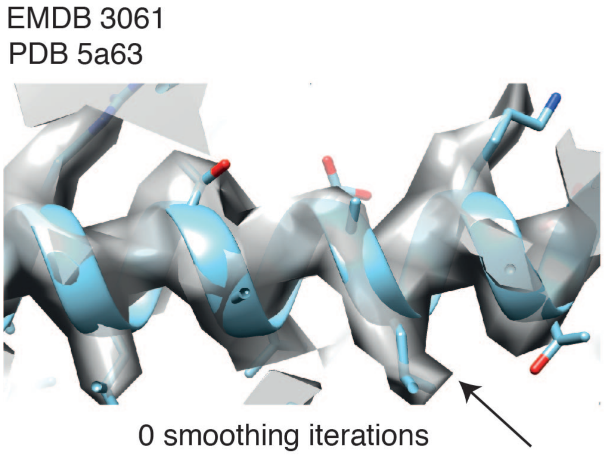
Confidence maps are based on multiple hypothesis testing, and, as such, similar interpretation of the results applies. No significant signal at a given FDR does not imply that there is no signal at the respective voxels. It means that the uncertainty is too high to make a decision against the background noise hypothesis at the specified significance level. Controlling a global error rate in a multiple testing setting is a difficult task, as complex dependencies between the individual tests have to be taken into account. Thus, control of the error rates can only be done rather conservative, which means that a confidence map at a FDR of 1% will have a true FDR below 1%. However, we described above how local resolutions can be incorporated in order to increase the statistical power, i.e. how we can reduce the amount of false negative voxels.

When choosing an appropriate threshold of the confidence maps, the FDR provides an error criterion that is directly interpretable: e.g. choosing a FDR of 1% means that up to 1% of the visible density can be actually background noise. Choice of the FDR is also related to the properties of the interpreted features. While for continuous density, as they appear for low-resolution structures, the occurrence of single false positive voxels is less problematic, the cost of false positive hits can be substantially higher when interpreting water and ion densities in high-resolution maps. In these cases, usually more restrictive thresholds are necessary. Alternatively, $FWER$ instead of FDR control can be chosen in such situations. In conclusion, the lower the error level can be chosen, the more confident statements can be made about the respective densities.

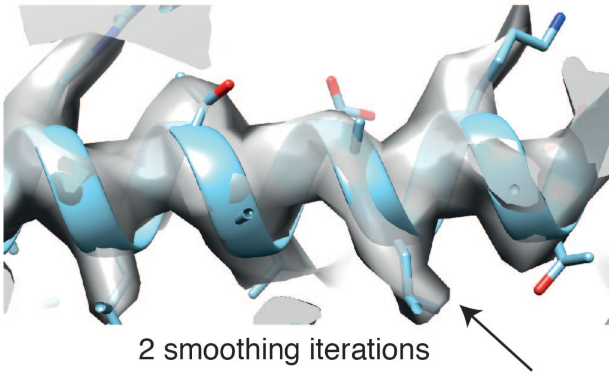
Although the threshold of a confidence map still allows adjustment, it has to be noted that it differs substantially from the threshold of cryo-EM maps. The FDR-threshold provides an error criterion that is directly interpretable and gives feedback about the validity of the visible density in a probabilistic way. For cryo-EM maps, however, the threshold is difficult to interpret in detail and the specific choice remains arbitrary due to plenty of different reasons that can in principle influence the actual map values, as described.

Moreover, as soon as density can be clearly distinguished from background noise, they will have values very close to 1 in a confidence map. This means that lots of variations of the map values are flattened in confidence maps. This effect becomes obvious for example in membrane proteins, where the detergent micelle of course also contributes true signal. In order to interpret the transmembrane domain in confidence map, it needs to be sliced. In a cryo-EM map simply the threshold can be increased until the detergent is not visible any more. This is possible, as the the detergent micelle usually has lower map values compared to the protein parts. But such a threshold adjustment is problematic, as of course also density will be ignored that is real: not only the detergent disappears, but also other weaker density like ligands.

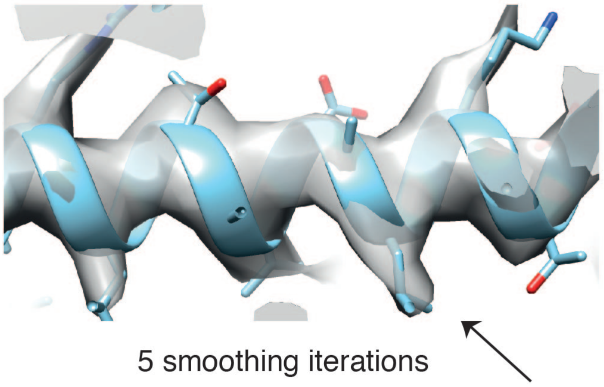
When visualizing confidence maps, for example in Chimera [114], they look different from common cryo-EM maps. Confidence maps appear very sharp and have obvious edges, which is the result of the strong contrast between signal and background voxels, with values close to 1, or 0, respectively. However, this visualization problem can be easily solved by oversampling and smoothing the surface with the respective visualization program, in order to make them look more EM like (Fig. 4.18).



0 smoothing iterations



2 smoothing iterations



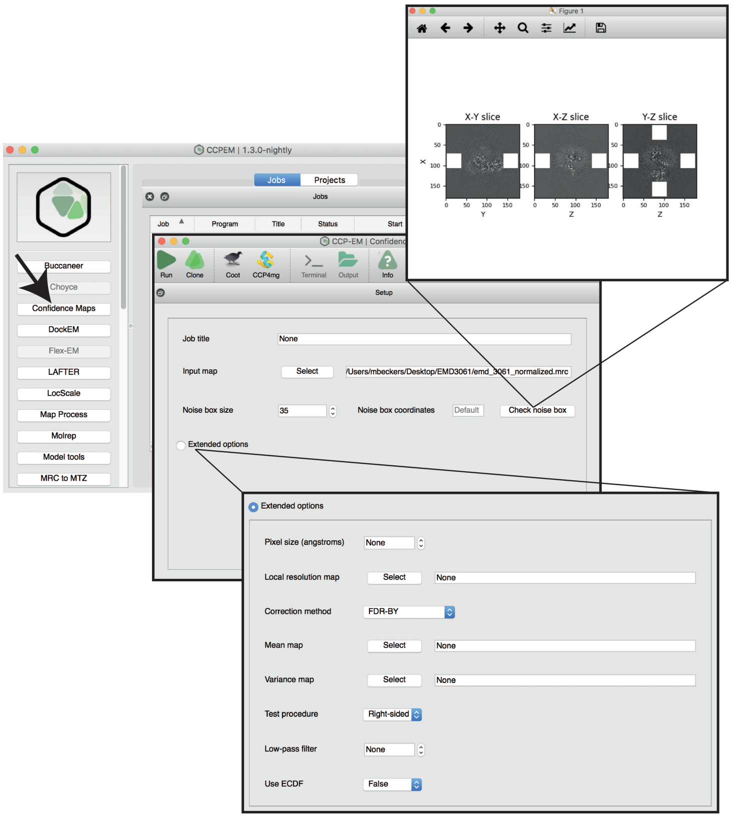
5 smoothing iterations

Figure 4.18: Confidence maps have very high contrast, which results in sharp edges (top), as shown for the map EMD3061. For example in Chimera, additional surface smoothing can be used to make the appearance smoother and look more like a normal cryo-EM map (middle, bottom).

Confidence maps aim to detect signal from background noise. If significant signal is detected, it means that, up to the specified confidence level, it is neither background from the amorphous ice nor detection noise. As such, signal can be everything that contains contributions beyond background noise. However, situations can be imagined, where this might result in uninterpretable confidence maps, even if the background estimation was performed correct. For example, substantial amounts of misaligned particles will of course contribute true signal, but without any clear structural meaning. Similar effects will occur with highly preferred orientations and overfitting of noise. As confidence maps are able to detect very weak signal, especially when local resolutions are provided, such unwanted signal will be more prominent in confidence maps compared to normal cryo-EM maps. Although this can be seen as a conceptual problem of the signal detection approach, it is also indicative of problems in the reconstructions.

4.6 Implementation of Confidence Maps in the CCPEM Software Suite

Together with Colin Palmer from CCPEM, the confidence map tools have been integrated into the CCPEM software suite v1.4 with an easy-to-use graphical user interface (Fig. 4.19). Upon clicking the "Check noise box" button, a pop-up window appears, which shows three slice-views through the map together with the locations of the used noise boxes. In this way, noise boxes can be easily inspected and adjusted accordingly. Moreover, additional error rates and additional input of a local resolutions can be selected as extended options.



Job title: None

Input map: /Users/mbeckers/Desktop/EMD3061/emd_3061_normalized.mrc

Noise box size: 35

Noise box coordinates: Default

Pixel size (angstroms): None

Local resolution map: None

Correction method: FDR-BY

Mean map: None

Variance map: None

Test procedure: Right-sided

Low-pass filter: None

Use ECDF: False

Figure 4.19: Confidence map implementation in CCPEM v1.4. Overview of the graphical user interface in CCPEM. A simple "Check noise box" button allows direct visualization of the noise boxes.

4.7 Limitations and Future Perspectives

A major limitation of the approach at this stage is the background noise estimation process. Estimation in the solvent area outside the particle and subsequent usage of this estimates for the complete map assumes that we can get information from the solvent background about the background affecting the particle itself. If the background noise is homogeneous over the whole map, then this will result in accurate background estimates for all voxels in the map.

This is of course an approximation to reality. From the image formation process with lower relative ice-thickness in the region of the particles compared to the solvent surrounding them, one might argue that background noise is higher in the solvent area outside the particle. However, we showed in a mathematical proof that the FDR is still controlled with overestimated background noise levels, although more conservative. Moreover, it seems to be rather robust towards small inaccuracies of the background estimates. We argue that the presented approach is robust, especially if we recapitulate that the major problem is the correction of the multiplicity from the multiple testing, which can only be done conservatively. Improved background estimation approaches that could deliver voxel-wise parameter estimates, however, would of course be beneficial to the approach in order to increase the statistical power. Another limiting point is the choice of the boxes used for background noise estimation, which adds a level of subjectivity, especially for the case of sub-tomogram averages. Although guidelines can be given on how to choose the parameters (as big as possible and as small as necessary), this should become a fully automated process in the future in order to get a fully unsupervised procedure.

In addition to cryo-EM maps, a similar thresholding problem also applies to maps from x-ray crystallography. However, it remains an open question if the concept is transferable, as the estimation of background noise might be more complicated due to the phasing process. But the data is in general interpretable, and, as such, there must at least be rough hints about the expected uncertainty in the maps.

To generalize the concept even further, thresholding is also inherent to other imaging techniques that deal with high amounts of background noise and where one has to make a decision if something is still background or already a significant hit. While a simple inference on individual pixels ignores the multiple testing problem, it is essential to correct for it, especially for large images. Multiple testing issues are often overlooked and lead to inflated amounts of positive hits. However, they can be corrected quite powerful. While tools from high-dimensional statistical inference like multiple testing are widely used in all kinds of *omics* technologies, they have not found their way to image analysis and processing in the life sciences, especially as visibility to statisticians has been low. The presented work shows that statistical inference can be a valid and highly useful tool for the analysis and inference of imaging data, and that similar approaches might be influential tools beyond cryo-EM.

Chapter 5

A Non-Parametric Permutation Approach for Adaptive Thresholding of Fourier Shell Correlation Curves

The work presented in this chapter is based on the observation that thresholding of Fourier shell correlation (FSC) curves poses a multiple testing problem as well. Together with a novel non-parametric permutation algorithm to learn the null distribution from the data itself, I could apply similar methodology as used for the confidence map approach. Chronologically, this was the last work done for this thesis and remains unpublished so far. However, work on the paper is ongoing. I prepared all figures and the programming was done by myself.

5.1 Abstract

Fourier shell correlation (FSC) curves have become the standard quantity used for resolution estimation in both cryo-EM and super-resolution microscopy. However, the actual resolution determination process requires the selection of a specific threshold as well as processing of the maps beforehand (masking), which remains a controversial and subjective issue [148]. In this chapter I describe a novel statistical framework based on modern non-parametric permutation sampling and multiple testing correction which allows fully automated and mask-free resolution determination based on thresholding of Fourier shell correlation curves. I demonstrate applicability for global, local, directional and map-model resolution estimation and show that the developed criterion termed $FDR-FSC$ gives accurate resolution estimates without any user intervention. The algorithms are implemented in a user-friendly GUI based software termed SPoC.

5.2 Resolution Estimation

In the last decades, novel imaging approaches greatly improved our understanding of biological matter. With the advent of cryo-EM and super-resolution microscopy, to mention but a few, microscopy with both electrons and photons has undergone major improvements, allowing to see complex biological mechanisms at high detail. At the core of all these developments are improvements in the resolutions that can be achieved. The resolution is a highly important number, as it makes different methods comparable and, more importantly, gives insights up to which level of detail we can faithfully interpret our data. FSC curves are nowadays the standard tool for resolution estimation. FSCs were originally introduced to the field of cryo-electron microscopy [57] and gradually proliferated into super-resolution microscopy [106, 5, 34, 76]. The FSC measures the correlation between Fourier coefficients in the respective resolution shells and typically shows high correlations at low resolutions and starts to drop until it oscillates around zero at higher resolutions where only noise is present. In order to get a resolution number from FSC curves, a threshold value needs to be chosen. The most widely used value is the 0.143 cutoff [123] for resolutions better than 10 Å. For lower resolutions and for local FSCs, as used for local resolution estimation [25], a 0.5 cutoff is typically used. However, fixed value thresholds ignore the uncertainty inherent to FSCs, which are dependent on the effective number of Fourier coefficients in the respective resolution shell. This effect becomes especially important at lower resolutions and for small sample sizes, as they occur for local resolution estimation [25] and directional FSCs [168]. Other criteria like the 2σ and 3σ [127, 107] as well as the half-bit criterion [147] have been proposed that take into account this effects.

The FSC is directly related to the spectral-signal-to-noise ratio (SSNR) [146] and thresholds like the 0.143/0.5 or half-bit criterion choose the cutoff to have a least information content present in the highest resolution shell. However, in order to get correct estimates of the SSNR of the molecular density, noise needs to be masked out of the volume, as big parts of the volume are only noise and thus highly bias the SSNR towards lower values. Masking however poses the danger of introducing artificial correlations, an effect which needs to be further corrected by heuristics [27]. Typically, several masks are tried and the user decides about a reasonable resolution based on reproducibility as well as justification by molecular features. Although approaches exist for the calculation of corrected FSC curves considering the molecular mass [135], this still requires a particular threshold criterion. In summary, it remains a highly subjective task.

The σ-thresholds were developed in order to give cut-offs when the FSC exceeds the random correlations of pure noise. Compared to the above-mentioned criteria, testing for signal in the resolution shells has the advantage that the resolution shells that contain signal also contain signal in presence of noise, i.e. the quantity we are interested in is unbiased by the noise. Only the detectability is influenced, which makes it necessary to use powerful statistical methods with high specificity. Additionally, inference of statistically significant signal in the resolution shells only requires knowledge of the random correlations of pure noise and avoids the consideration of complicated correlations between signal and noise. However, statistics based on simple σ-thresholds suffer from several drawbacks. Connecting the σ-levels to probabilities requires strict assumptions about the underlying probability distribution, which is usually unknown. Furthermore, statistical inference on all resolution shells poses an inherent multiple testing problem. Both problems have not been considered in the context of resolution estimation so far and neglecting them could lead to erroneous resolution estimates.

5.3 Methods

In order to circumvent these problems, I developed a new probabilistic procedure for thresholding of FSC curves based on non-parametric permutation sampling and multiple comparisons correction. Permutation tests, or sometimes also referred to as exact tests, are powerful statistical procedures for learning the null-distribution from the data itself and thus do not require any assumptions about the underlying distributions [81]. These methods are still subject of active research [109, 35]. The idea for permutation testing of correlation coefficients is straightforward: under the null-hypothesis the two resolution shells are independent and we can generate new samples by permutation of one of them. This allows to sample the null distribution of the FSC coefficients of each resolution shell and subsequently calculation of p-values (Fig. 5.1). In order to account for the multiple testing problem, p-values are then corrected by means of false discovery rate (FDR) [12] control, similar to the case of cryo-EM map thresholding [10]. Effects of specific symmetry and windowing functions on the effective sample sizes are incorporated by subsampling before permutation of the Fourier coefficients. In the following section, I will give a detailed outline of the mathematical framework.

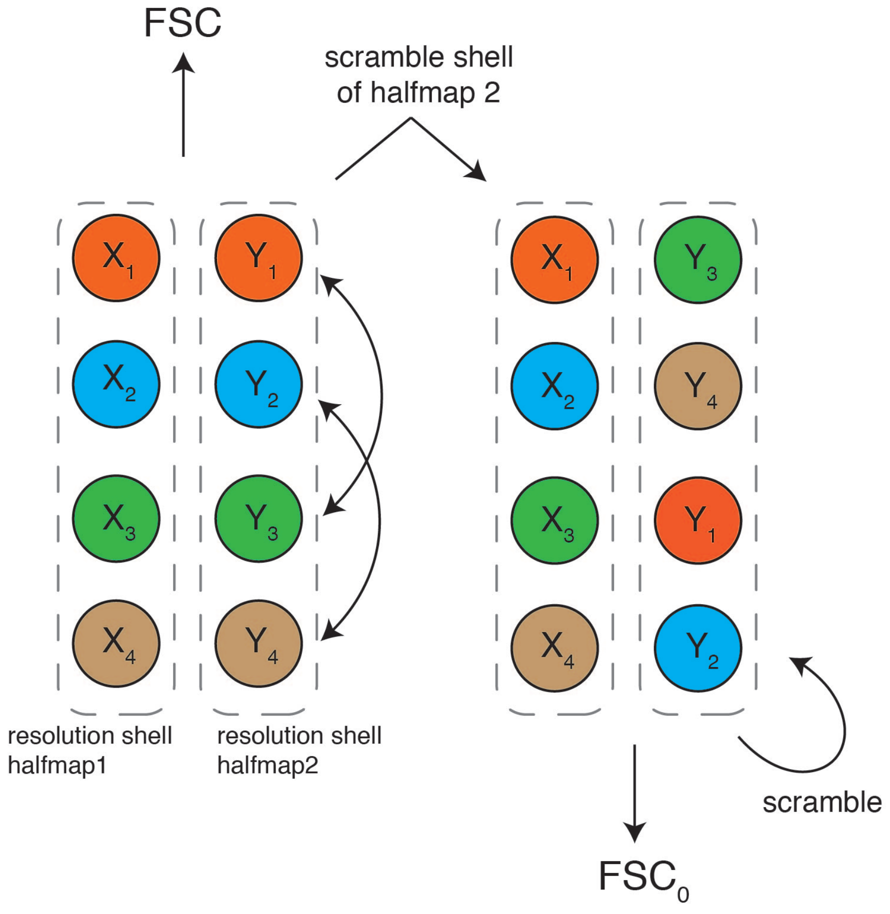


Figure 5.1: Samples FSC_0 under the null hypothesis are generated by permutations of the paired Fourier coefficients, leading to new paired Fourier coefficients.

5.3.1 A Permutation Test for Fourier sell Correlation Coefficients

Let X_{r_i} and Y_{r_i} complex random variables, with which we denote the Fourier coefficients at the specific location $r_i, i = 1, ..., N$ in resolution shell r of halfmap 1 and 2, where $N \in \mathbb{N}$ is the number of Fourier coefficients in the respective shell. Due to the Friedel symmetry inherent to real Fourier transforms, we restrict ourselves now to one half of all locations r_i, which contains one partner of each symmetry pair. The other is than simply given as the complex conjugate. We denote the Friedel symmetry corrected sample size with n. In the case of two independent reconstructions, dependence between X_{r_i} and Y_{r_i} is introduced through an effect of the signal at position r_i, which we denote with $S(r_i) \in \mathbb{C}$ and which is the same in both volumes. Thus, Fourier coefficients are usually modelled as

$$X_{r_i} = S(r_i) + N_X(r_i) \text{ and } Y_{r_i} = S(r_i) + N_Y(r_i), \tag{5.1}$$

where $N_X(r_i)$ and $N_Y(r_i)$ are complex valued noise variables [147]. We do not impose any assumptions about the specific distribution of the noise in resolution shell r. Furthermore, we denote with $X_R = (X_{r_1}, ..., X_{r_n})$ and similarly $Y_R = (Y_{r_1}, ..., Y_{r_n})$ the complete resolution shells.

We are interested if the Fourier coefficients in the respective resolution shell are dependent with respect to the location r_i, i.e. that groupings (X_{r_i}, Y_{r_i}) show statistical dependencies. In statistical terms, the null hypothesis H_0 is that X_{r_i} and Y_{r_i} are independent of each other, and the alternative H_1 is that there are dependencies.

In order to test such a hypothesis system, correlation coefficients can be used [81]. As a test-statistic we use the Fourier shell correlation. Using the Friedel symmetry inherent to Fourier transforms of real valued data and denoting complex conjugation with *, we can write the FSC as

$$FSC(X_R, Y_R) = \frac{\sum_{i=1}^{n} X_{r_i} Y_{r_i}^* + Y_{r_i} X_{r_i}^*}{\sqrt{\sum_{i=1}^{n} 2|X_{r_i}|^2 \sum_{i=1}^{n} 2|Y_{r_i}|^2}}, \tag{5.2}$$

where it becomes clear that $FSC(X_R, Y_R)$ is a real valued random variable. Statistically, $FSC(X_R, Y_R)$ is the estimator for the true value of the Fourier shell correlation, which we denote from now on with FSC. Thus, the values that we calculate with $FSC(X_R, Y_R)$ are only noisy estimates for the truth.

It is important to note here that we are testing for dependencies by using a correlation coefficient; the null hypothesis is not FSC$= 0$. The FSC can be zero and there can still be dependencies. Independence is a stronger property than correlation, i.e. correlation implies dependence while the reverse is in general not true.

In order to test H_0 we calculate a p-value as follows. With fsc we denote the estimated FSC value for resolution shell r. Moreover, with FSC we denote the random variable describing the FSC as above. The p-value p of this observation is then given as the probability that FSC is bigger than fsc under the null hypothesis, i.e.

$$p = \mathbb{P}(FSC > fsc|H_0), \tag{5.3}$$

where $\mathbb{P}$ is the true probability measure under the null hypothesis. In case for correlation coefficients, the true probability measure $\mathbb{P}$ is rather complicated and usually unknown. Although analysis can be done based on Fisher's z-transformation [112], this assumes a normal distribution of the underlying sample. However, non-parametric statistical methods based on resampling got attractive in the recent years due to increased computational power and correlation coefficients like the Fourier shell correlations can be easily resampled by permutation.

A permutation test is built as follows. Under H_0 we have that the Fourier coefficients X_{r_i} and Y_{r_i} are independent for $i = 1, ..., n$. New paired samples of Fourier coefficients can thus be generated by permutation of Y_R. This is where the null hypothesis of independence becomes important, because under the null hypothesis of no correlations there could still be dependencies and simple permutations would not be allowed. Thus, the sampled null distribution is the distribution of FSC from independent half maps. For a detailed discussion regarding this topic we refer to [35]. From each generated sample, we then calculate the Fourier shell correlation, which results in a sample of FSC under the null hypothesis. Denoting with S_n the set of all permutations π of $1, ..., n$, which is known in abstract algebra as the symmetric group, we can calculate the p-value p by

$$p = \frac{1}{n!} \sum_{\pi \in S_n} \mathbb{I}(FSC(X_R, Y_{\pi(R)}) \geq fsc), \tag{5.4}$$

where $\mathbb{I}$ denotes the indicator function. As the number of possible permutations grows very rapidly with the sample size, technically often just a random subset of $H \subset S_n$ of all $n!$ possible permutations are used. Thus, p is replaced by its Monte-Carlo estimator $\hat{p}_{MC}$, given as

$$\hat{p}_{MC} = \frac{1}{|H|} \sum_{\pi \in H} \mathbb{I}(FSC(X_R, Y_{\pi(R)}) \geq fsc), \tag{5.5}$$

where $|H|$ is the cardinality of the set H, i.e. the number of permutations selected for the Monte-Carlo estimate.

It is important to note, that the null distribution of the FSC estimated by the permutation approach is not necessarily the distribution of the FSC in absence of any signal. As we are permuting in presence of possible signal, which adds additional variation, the permutation distribution of the FSC will have bigger tail probabilities than the distribution of the FSC without any signal. This could give rise to more conservative resolution estimates. However, simulations in presence of signal at high and low signal-to-noise ratios showed that this effect seems to have less practical relevance (Fig. 5.2 a). Moreover, the amount of signal in the most important resolution shells close to the actual resolution is rather low and will thus have limited influence on the actual distribution. As a compromise between computational efficiency and statistical accuracy, we restrict the number of permutations to a maximum of 1000. Moreover, permutations are only done for resolution shells with an effective sample size >10, which allows for more than 1000 permutations. Even in the case of high symmetries this usually only excludes the first two resolution shells. A 0.9 threshold is used in these cases.

5.3.2 Multiple Testing Correction

Testing of all resolution shells results in a multiple testing problem. In order to correct for the α-error inflation due to the multiple testing problem, estimated p-values are subsequently adjusted for the false discovery rate (FDR) of detected resolution shells, i.e. the amount of false positive resolution shells. To account for arbitrary dependencies between the p-values, the Benjamini-Yekutieli [13] procedure has been selected for FDR control.

5.3.3 Effective Sample Size Corrections

Correction of the sample size has been found to be an important factor in presence of symmetry and masking [147], as this leads to dependencies between Fourier coefficients. Simple heuristics have been proposed in order to get estimates the effective sample size n_{eff}. The effect of imposed symmetry during reconstruction can be conservatively corrected by

$$n_{eff} = \frac{n}{n_{as}}, \tag{5.6}$$

where n is the number of Fourier coefficients in the respective resolution shell and n_{as} is the number of asymmetric units under the respective symmetry. Effective sample sizes are incorporated in the permutation framework by sub-sampling of Fourier coefficients.

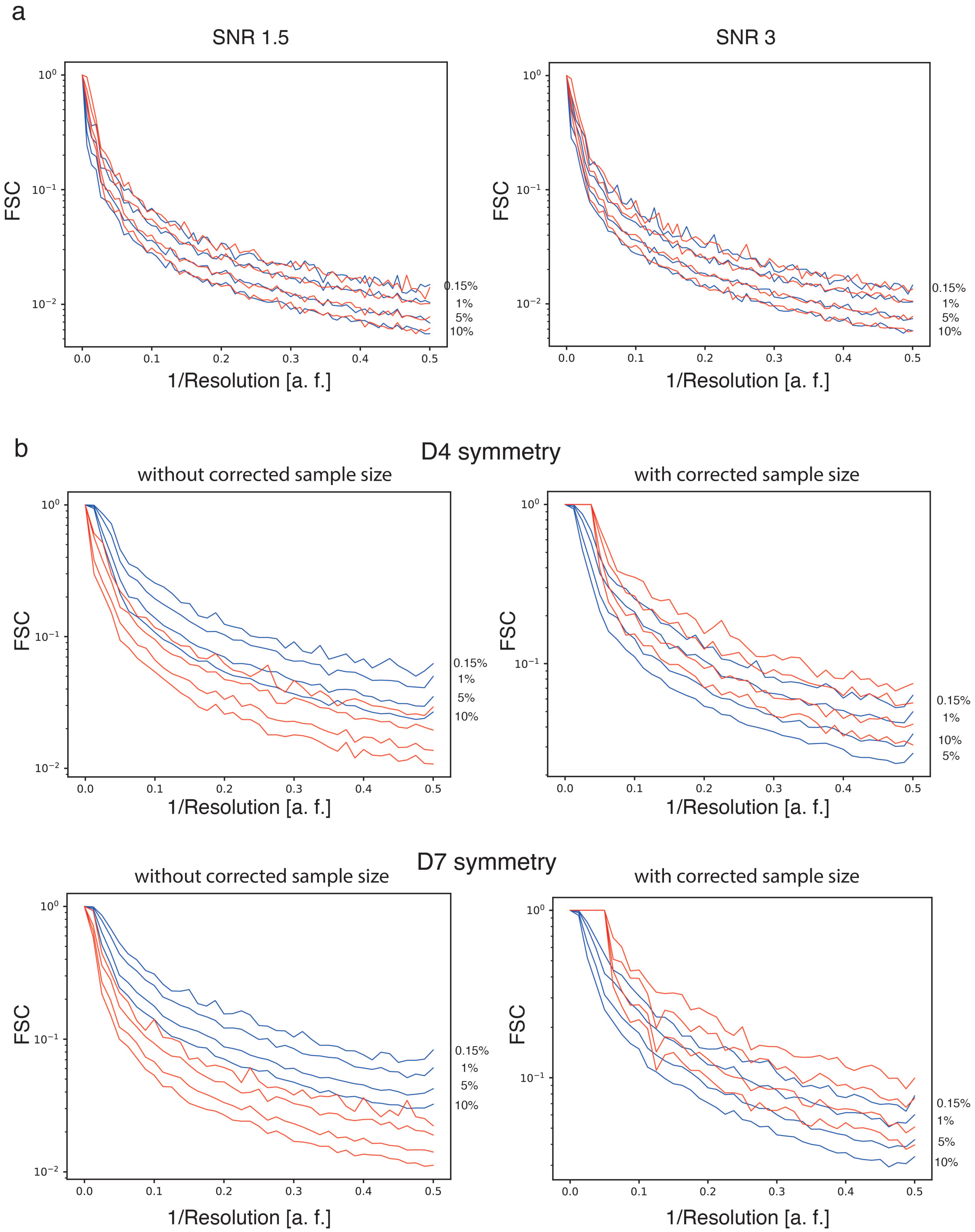


Figure 5.2: a) Comparison of right-sided 10%, 5%, 1% and 0.15% percentiles, as estimated by permutation sampling (red lines) from two simulated half maps of β-galactosidase (pdb 5a1a) at SNRs of 1.5 (left) or 3 (right), with the true FSC distribution (blue lines), as it can be obtained with 5000 simulations of the respective half maps. The y-axis corresponding to the FSC values is shown in a logarithmic scale. b) Again, comparison of the respective percentiles between the permutation approach and simulation of the true FSC distribution, but in this case both halfmaps were noise volumes and symmetrized with D4 (top) and D7 (bottom) symmetries. Plots on the left side show that the presence of symmetry reduces the effective sample size, which leads to underestimated tail probabilities, and can be conservatively corrected (right).

Effects of masking on the effective sample size are more complicated and depend on the specific shape and volume of the mask. We propose a novel procedure for the estimation of the effective sample size $n_{eff} = \alpha * n$ by finding the factor $\alpha \in [0, 1]$ which minimizes the mean Kolmogorov-Smirnov distance $\overline{D}_{\alpha}$ over the resolution shells.

The two-sample Kolmogorov-Smirnov statistic [93] is a measure of similarity of two empirical cumulative distribution functions (ECDF) and is defined as

$$D_{\alpha} = \sup_{x \in [-1,1]} |F_{1,\alpha}(x) - F_{2,sim}(x)|, \tag{5.7}$$

where in our setting $F_{1,\alpha}(x)$ denotes the ECDF of the permutation approach, which is estimated with an effective sample size $n_{eff} = \alpha * n$, and $F_{2,sim}(x)$ can be obtained by repetitive simulation of two masked noise maps and subsequent FSC calculation.

D_{α} is then calculated as the mean of the Kolmogorow-Smirnov statistics $D_{\alpha,r}$ over all resolution shells r, i.e.

$$\overline{D}_{\alpha} = \frac{1}{m} \sum_{r=1}^{m} D_{\alpha,r}, \tag{5.8}$$

where m is the number of resolution shells. An estimate $\hat{\alpha}$ for α is then given by

$$\hat{\alpha} = \text{argmin}_{\alpha \in (0,1]} \overline{D}_{\alpha}. \tag{5.9}$$

In the presented algorithm, we apply a soft circular mask for global resolution estimation by default, which effects can be corrected with an effective sample size of $n_{eff} \approx 0.7n$, i.e. $\overline{D}_{\alpha}$ is minimized for $\alpha \approx 0.7$ (Fig. 5.3a), which allows accurate calculation of tail probabilities for various box sizes (Fig. 5.4).

Application of windowing functions, as used for local resolution estimation (see below), also lead to reduced effective sample sizes. In a similar way as for the soft circular mask, we found that a Hann-window leads to an effective sample size $n_{eff} \approx 0.2n$ (Fig. 5.3b, 5.5).

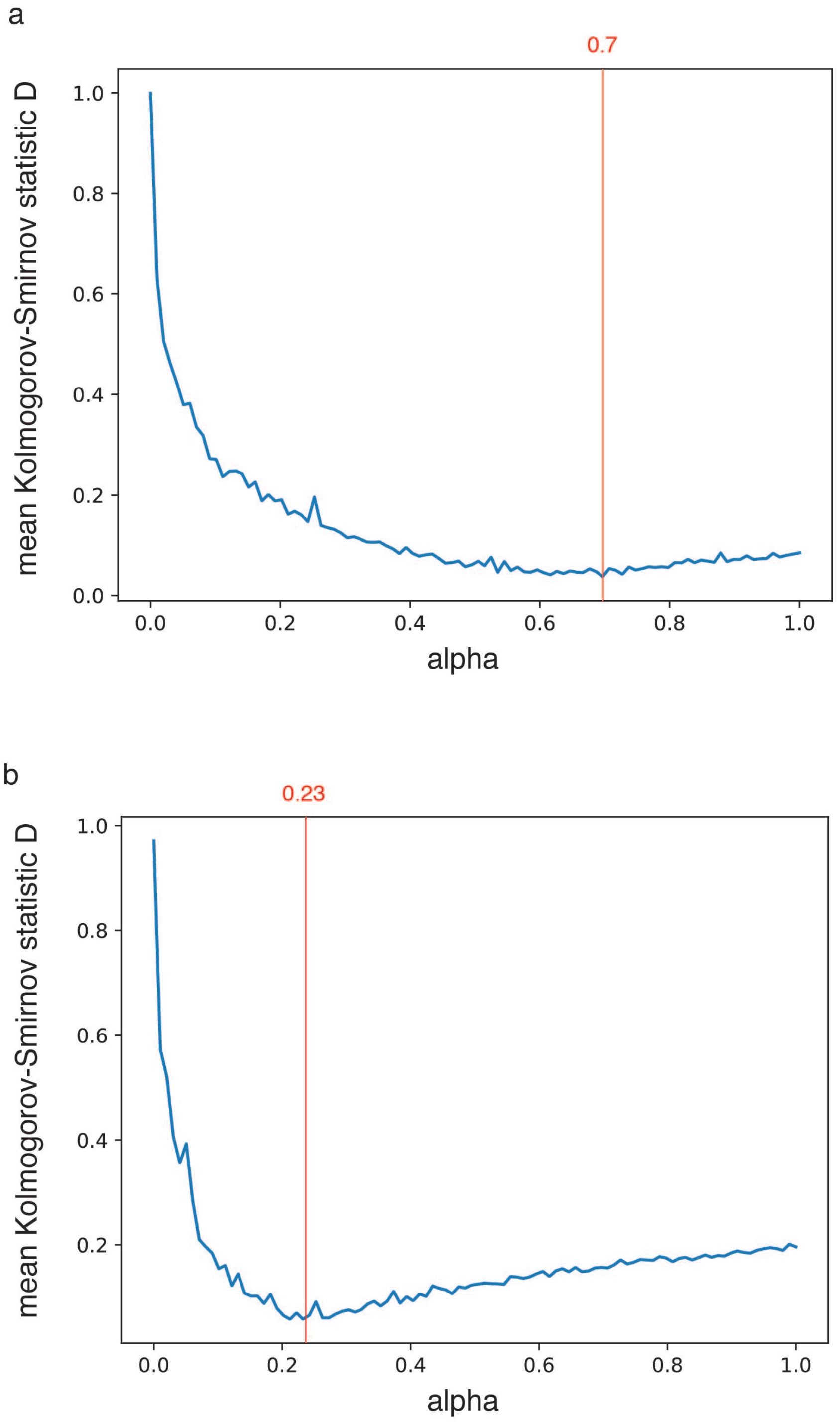


Figure 5.3: Sample size correction factor plotted against the mean Kolmogorov-Smirnov statistic of the resolution shells. a) A soft circular mask leads to an optimal value for α of ≈ 0.7 . b) A Hann window leads to an optimal value for α of ≈ 0.23.

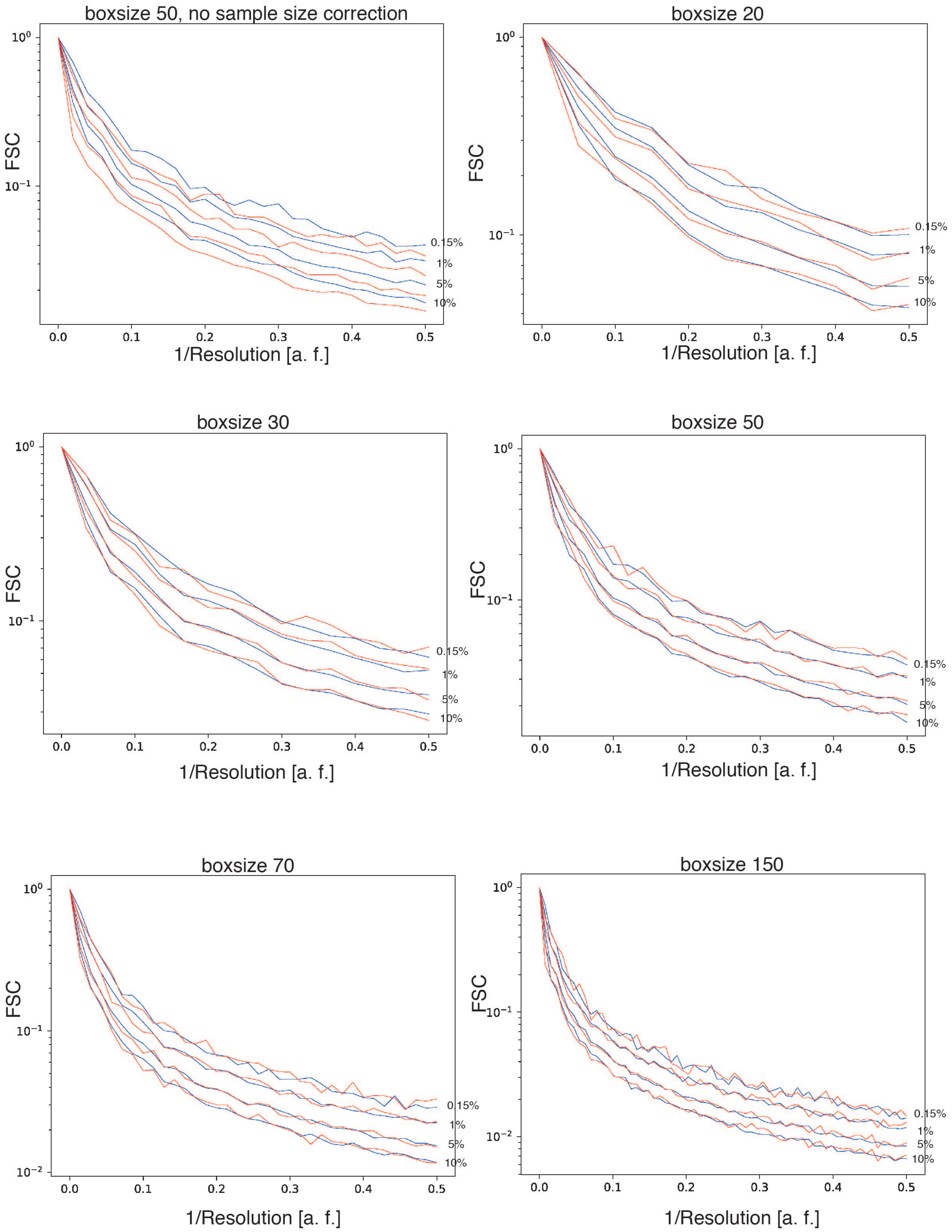


Figure 5.4: Effective sample size after circular masking. Comparison of right-sided 10%, 5%, 1% and 0.15% percentiles, as estimated by permutation sampling (red lines), with the true FSC distribution (blue lines), as estimated from 5000 simulations of the respective noise half maps, which were masked with a soft circular mask. Without correction for the masking effect, tail probabilities are underestimated (top left), but this can be efficiently corrected with an effective sample of 0.7 for different box sizes. The y-axis corresponding to the FSC values is shown in a logarithmic scale.

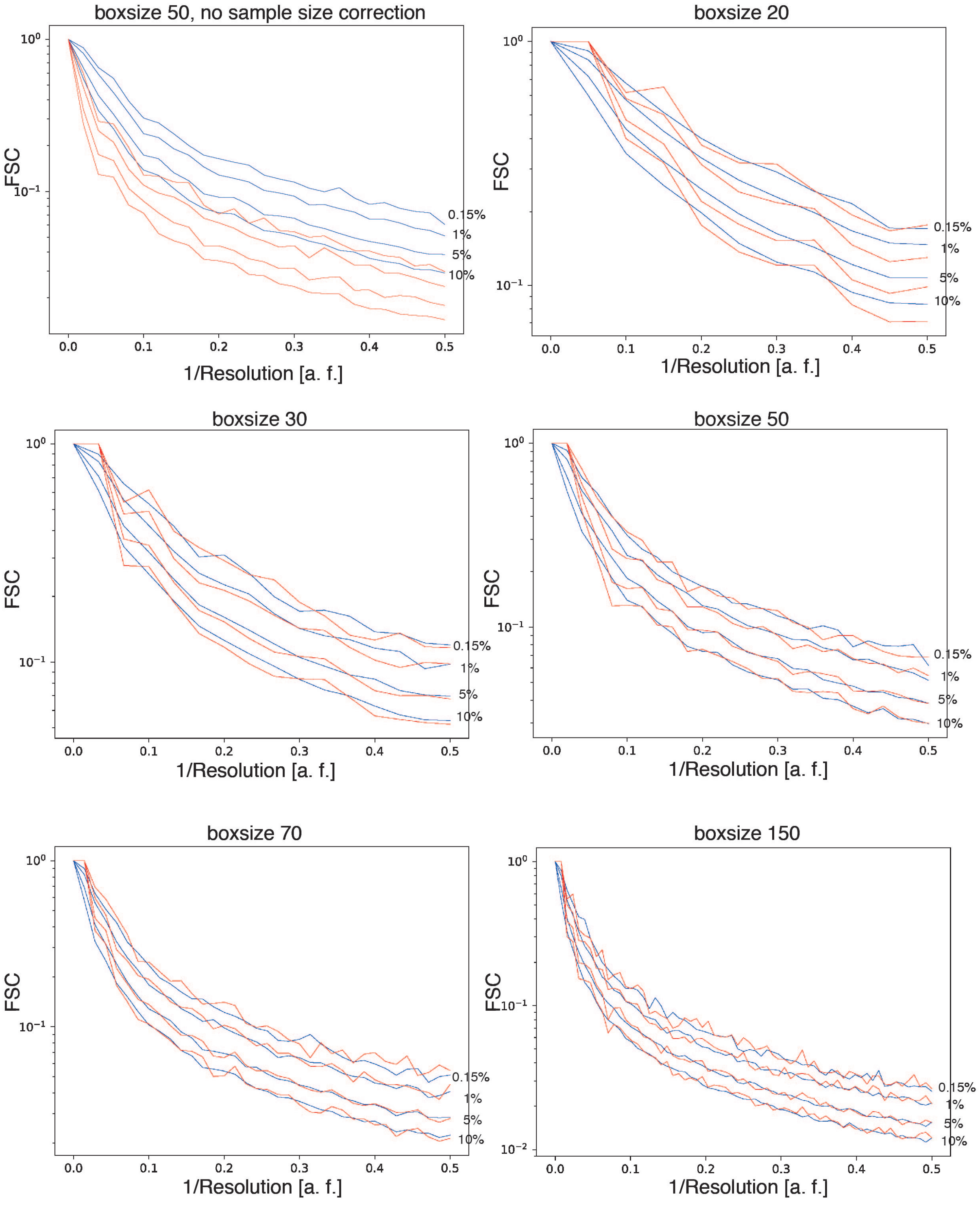


Figure 5.5: Effective sample size after application of a Hann window. Comparison of right-sided 10%, 5%, 1% and 0.15% percentiles, as estimated by permutation (red lines), with the true FSC distribution (blue lines), as obtained from 5000 simulations of the respective noise half maps, which were masked with a Hann window. Without correction for the masking effect, tail probabilities are underestimated (top left), but this can be efficiently corrected with an effective sample of 0.23 for different box sizes. The y-axis corresponding to the FSC values is shown in a logarithmic scale.

5.3.4 Local Resolution Estimation

Local resolutions are estimated by sliding a window over both half maps and subsequent calculation of local FSC thresholds [25]. The described permutation procedure with subsequent multiple testing correction can be straightforward applied to local FSC curves as well. In order to account for high-resolution artefacts introduced through spectral leakage, a Hann-window is used as a windowing function [25]. The masking effects, which are introduced from the Hann-window, are corrected as described above with a correction factor $\alpha = 0.23$. Moreover, effects of symmetry do not have any influence on the effective sample sizes if the size of the sliding windows is smaller than the size of the asymmetric units in the map, as it is commonly the case.

In order to speed-up the calculations, a step-size option is implemented which allows movement of the sliding-window of more than a single voxel. Moreover, in order to avoid repeating permutations on the same map, the permutations are only done on 10 random locations of the sliding window. The resulting samples of the null-distribution are then merged and subsequently used for p-value calculation at all locations.

The size of the sliding window needs to be chosen as a compromise between locality and resolution in Fourier space (Fig. A.1, A.2). Too small box sizes tend to give overestimated resolutions due to coarse sampling of low resolutions. Too large windows loose locality and thus tend to converge to the global resolution measurement, which will result in overestimated local resolutions for low resolution parts, while for high-resolution parts still high-resolution signal will be detected. The characteristic peak shape of the median resolution plots (Fig. A.1) is a consequence of the locality: if the box becomes large enough to contain both high and low-resolution features, the FSC will show significant correlations up to high frequencies. That's why the median resolution curve is expected to fall with increasing box sizes. In order to estimate low-resolution features accurately, we need reasonably big boxes, which thus explains the increase of the median resolution curves at the beginning. The peak thus marks the box size that gives the best separation of local resolutions and can be used to identify a optimal box size. An advantage of the $FDR - FSC$ criterion compared to 0.5 FSC is the clearer maximum.Box sizes around 20-30 pixel present a reasonable compromise between locality and resolution in Fourier space for most cases. All experiments were done with a box size of 25.

5.3.5 Directional Resolution Estimation

The implementation of the directional resolution estimation was done in analogy to Lyumkis and colleagues [168]. For each direction, the FSC is calculated by taking all voxels from each halfmap that deviate in direction no more than a specified angle from the desired direction. This results in rotating an inverse cone over one half of the 3D Fourier transforms, thereby accounting for the Friedel symmetry, and calculating the FSC only from samples inside the inverse cone. Similar to [168], an angle of 20° was used for all experiments, which gives a good compromise between the number of Fourier coefficients per shell and the preservation of local directionality. Sample size corrections need to be done as well for directional resolutions. Similar to global resolution estimation, this results in a correction factor for soft spherical masking. Moreover, as for local resolutions, locality of the directional resolution makes the correction of symmetry effects unnecessary and, again similar to local resolution estimation, it is possible to accelerate the algorithm by calculating the resolutions only for a few directions and interpolating the results in order to avoid repetitive FSC calculations of overlapping cones.

5.3.6 Model-Map Resolution Estimation

Estimation of global and local model-map resolutions requires similar effective sample size corrections as described above for global halfmap-halmap resolutions and their local counterparts. Although the simulated map does not contain any noise, it has to be noted that this does not generate any artificial similarities between the experimental and the model map and thus also no artificial correlations (Fig. 5.12 a). It only reduces the noise levels of the FSC, which is implicitly considered with the permutation sampling. Moreover, the effective sample size will be reduced due to symmetry if present. This needs to be corrected for the calculation of global map-model resolutions, as described already above for the halfmap-halfmap resolution estimation.

5.3.7 Figure preparation

Plots were visualized with ggplot2 in R [141] [158]. Chimera [114] was used for the figure preparation of the molecular maps and atomic models.

5.4 Results and Discussion

5.4.1 Global Resolution Estimation with the $FDR-FSC$ Criterion

In order to illustrate the accuracy of the permutation approach, we compared right-sided 10%, 5%, 1% and 0.15% percentiles of the FSC distribution, as it is sampled using our permutation approach, with the true percentiles, which can be estimated by iterative simulation of random noise half maps with subsequent FSC calculation (Fig. 5.6 a). We focus on the tail probabilities here, as they are most important for the accurate calculation of small p-values. The lines follow each other closely for all tested quantiles, emphasizing the accuracy of the permutation approach. Under a normal distribution, thresholding at 3σ corresponds to a 0.15% percentile. Direct comparisons of the simple 3σ curve [107] shows that it gives higher estimates of the percentiles. A modified version [147], marked with modified 3σ, gives rise to more accurate percentiles, in this case though with an unrealistic effective sample size higher than the number of Fourier coefficients in the resolution shell. Comparison of the empirical cumulative distribution functions (ECDFs) of the sampled FSC distributions for a single resolution shell confirm the observation, that the permutation approach is able to accurately represent the FSC distribution under the null hypothesis and outperforms the modified 3σ and the 3σ criteria (Fig. 5.6 b). While the ECDFs for the simulation and the permutation approach are highly similar, it can be seen that the distributions which result from the σ criterion tend to overestimate the tail probabilities. All in all, σ-curves suffer from substantial inaccuracies.

Based on our approach, resolution is assigned as the highest spatial frequency which contains significant signal at 1% FDR before the first non-significant resolution shell occurs. Although a 1% FDR threshold is somewhat arbitrary, the method is robust towards reasonable error thresholds between 0.1 and 10% (Fig. 5.6 c).

Comparison with uncorrected voxel-wise p-value thresholds for a γ-secretase map [4] shows that the multiple testing correction can be necessary to avoid overestimated resolutions (Fig. 5.7 a left). To test the sensitivity in the presence of noise, we reduced the box size of the volume of the same γ-secretase map in steps of 20 voxels, only cutting away noise. While the 0.143 FSC threshold falls from 4.1 Å to 3.7 Å (Fig. 5.7 a middle, right), the 1% FDR threshold only fluctuates at the second decimal place (Fig. 5.7 a right), which shows that the 1% FDR threshold, from now on referred to as 1% $FDR-FSC$, is robust towards masking out of noise, as only detectability is influenced. Moreover, the actual shape of the FSC is also highly dependent on the actual map volume or the specific mask (Fig. 5.7 a middle), making it difficult to read any reproducible information from the shape of the FSC curve. To further the validate the robustness towards different noise levels, we simulated a map of β-galactosidase (pdb 5a1a [8]) at 2.5 Å resolution and added different amounts of white Gaussian noise. Similar to above we find that the 1% $FDR-FSC$ threshold is highly robust towards the actual noise levels while the resolutions at 0.143

FSC are highly biased from noise (Fig. A.3 left).

A benchmark of our algorithm to the reported resolutions of 77 maps from the EMDB achieves a median deviation of only -0.02 Å for maps better than 5 Å with a median absolute deviation of 0.15, which presents the accuracy of 0.1 Å to which resolutions are reported (Fig. 5.7 b left, middle, Table A.1). Comparison with the 0.143 FSC threshold without masking, which has a clear bias towards lower resolutions with a median deviation of 0.53 Å (Fig. 5.7 b left, right), shows that, especially at lower resolutions, the 1% $FDR-FSC$ resolution is highly accurate without any masking (Fig. 5.7 b middle). Moreover, the consideration of the number of Fourier coefficients in each resolution shell avoids the need to switch to a more conservative 0.5 FSC threshold at lower resolutions. However, deposited halfmaps are rare for low-resolution structures and resolutions are harder to judge based on the visible features, which makes a detailed low-resolution benchmark difficult. Results on a few low-resolution examples are shown in Table A.2.

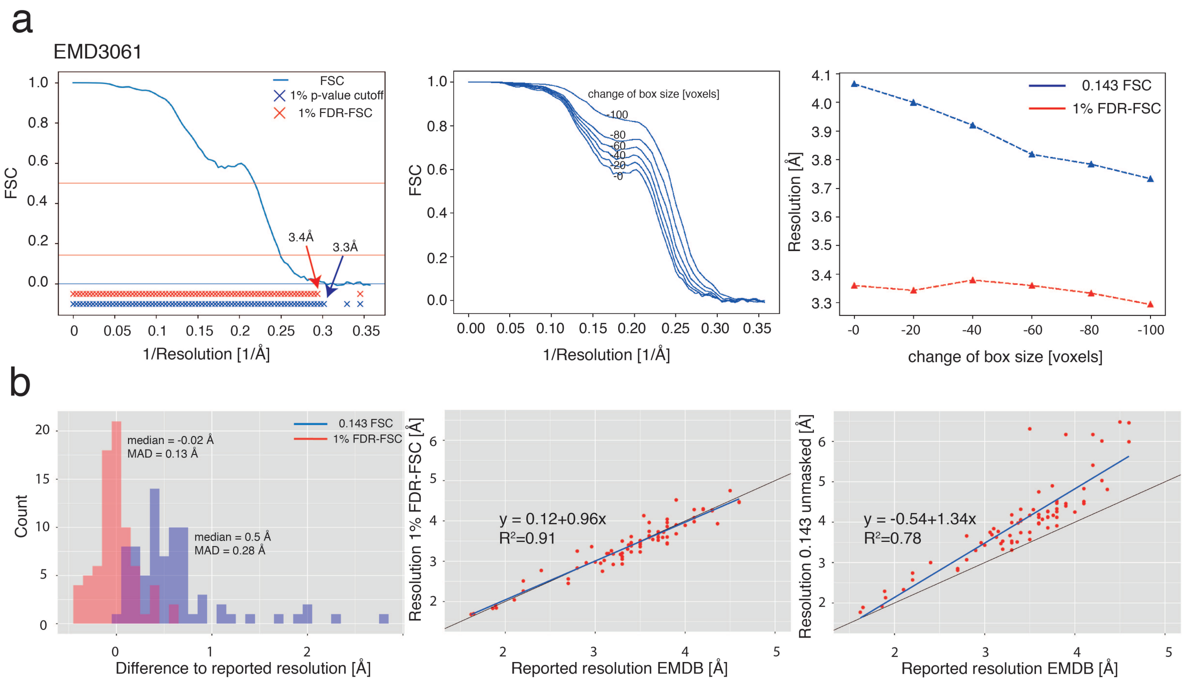


Figure 5.7: a) left: Example of a FSC curve for EMDB entry 3061. Resolutions shells with significant correlations beyond random fluctuations at 1% $FDR-FSC$ are marked with red crosses. Significant resolution shells at a 1% p-value threshold without multiple testing correction are marked with blue crosses. Respective resolutions are depicted above the arrows, showing that neglecting the multiple testing problem is necessary for accurate resolution estimation. Middle: Effect of noise on FSC curve shape for decreasing box sizes in steps of 20 voxels, which only excludes noise. Right: Effect of noise on resolution estimates at 0.143 FSC (blue) compared to 1% $FDR-FSC$ (red). b) left: Histograms of deviations from the reported resolutions for the 1% $FDR-FSC$ criterion (red) and unmasked 0.143 FSC criterion. middle: Reported resolutions against 1% $FDR-FSC$. right: Reported resolutions against unmasked 0.143 FSC. Fitted linear models are shown as blue lines together with the fitted function, diagonals shown in black.

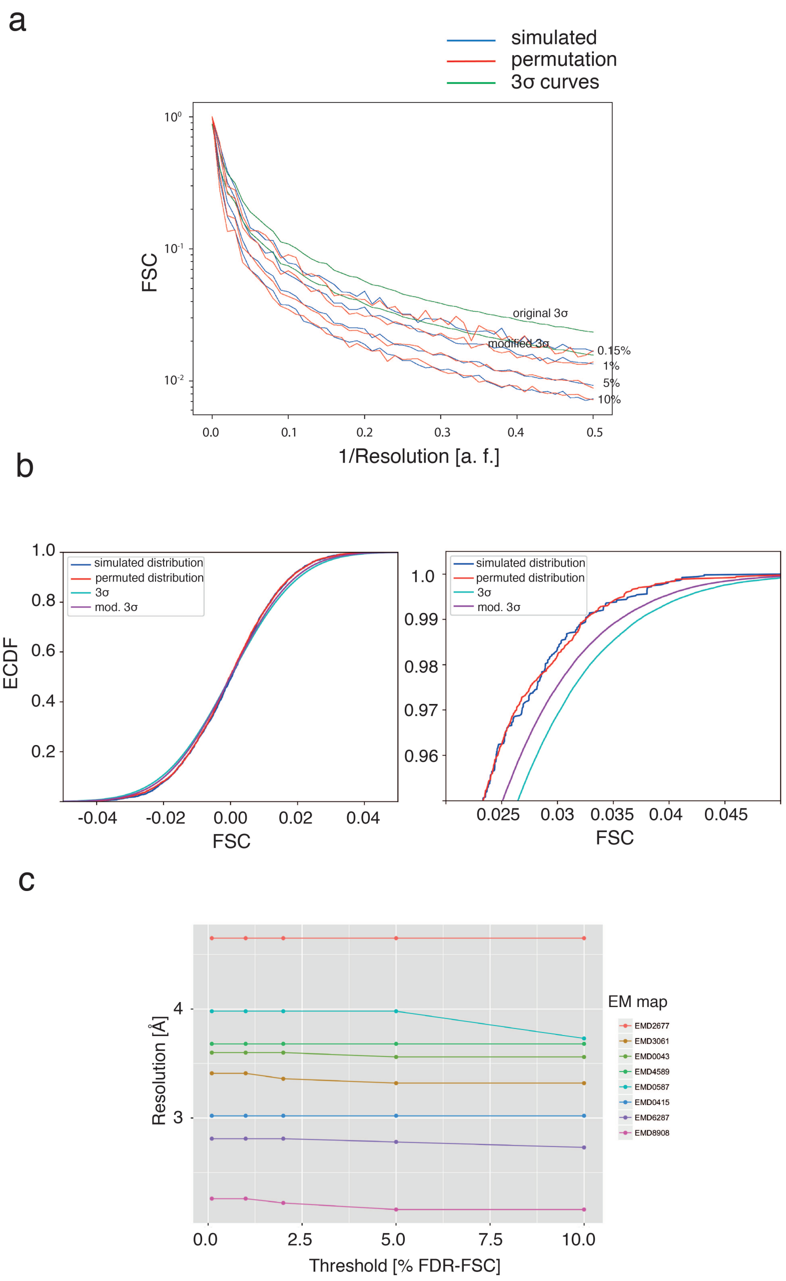


Figure 5.6: a) Right-sided 10%, 5%, 1% and 0.15% percentiles for the permutation approach (red), for 5000 simulated FSCs from random noise maps (blue) and both the original as well as the modified 3σ curves (green), marked with original and modified, respectively. The y-axis corresponding to the FSC values is shown in a logarithmic scale. b) Comparison of the empirical cumulative distribution functions (ECDFs) of the FSC for a single resolution shell (left) with zoomed view (right). The simulated (blue) and the permutation based distributions (red) follow each other closely, especially for the tail probabilities (right). Moreover, it can be seen that the distribution from the modified 3σ criterion is more accurate than the original 3σ c) Resolutions at different FDR thresholds for 8 maps between 2.2 Å and 4.6 Å resolution show that the FDR threshold is stable towards the specific significance level for reasonable errors levels between 0.1 and 10%.

5.4.2 Application to Local Resolution Estimation

EM maps usually exhibit local variations of resolutions and estimating these local resolutions has become an important task in recent years [25, 77, 149]. Based on local FSC curves resolutions are usually estimated with a 0.5 FSC threshold in order to account for small samples sizes in the resolution shells of the sliding windows. However, the 0.5 FSC threshold tends to underestimate resolution in higher resolution shells and might be too optimistic in lower resolution parts. Adaptive thresholds should provide superior estimates in these situations. Extension of our approach to local FSC curves is straightforward and we tested it on challenging test cases with large resolution variations.

In order to test the approach on simulated data, we generated a map of β-galactosidase (pdb 5a1a [8]) an filtered the subunits at different resolutions of 2,3,4 and 5 Å and added white Gaussian noise. Based on local 1% $FDR-FSC$ thresholding, local resolutions can be accurately determined (Fig. A.3).

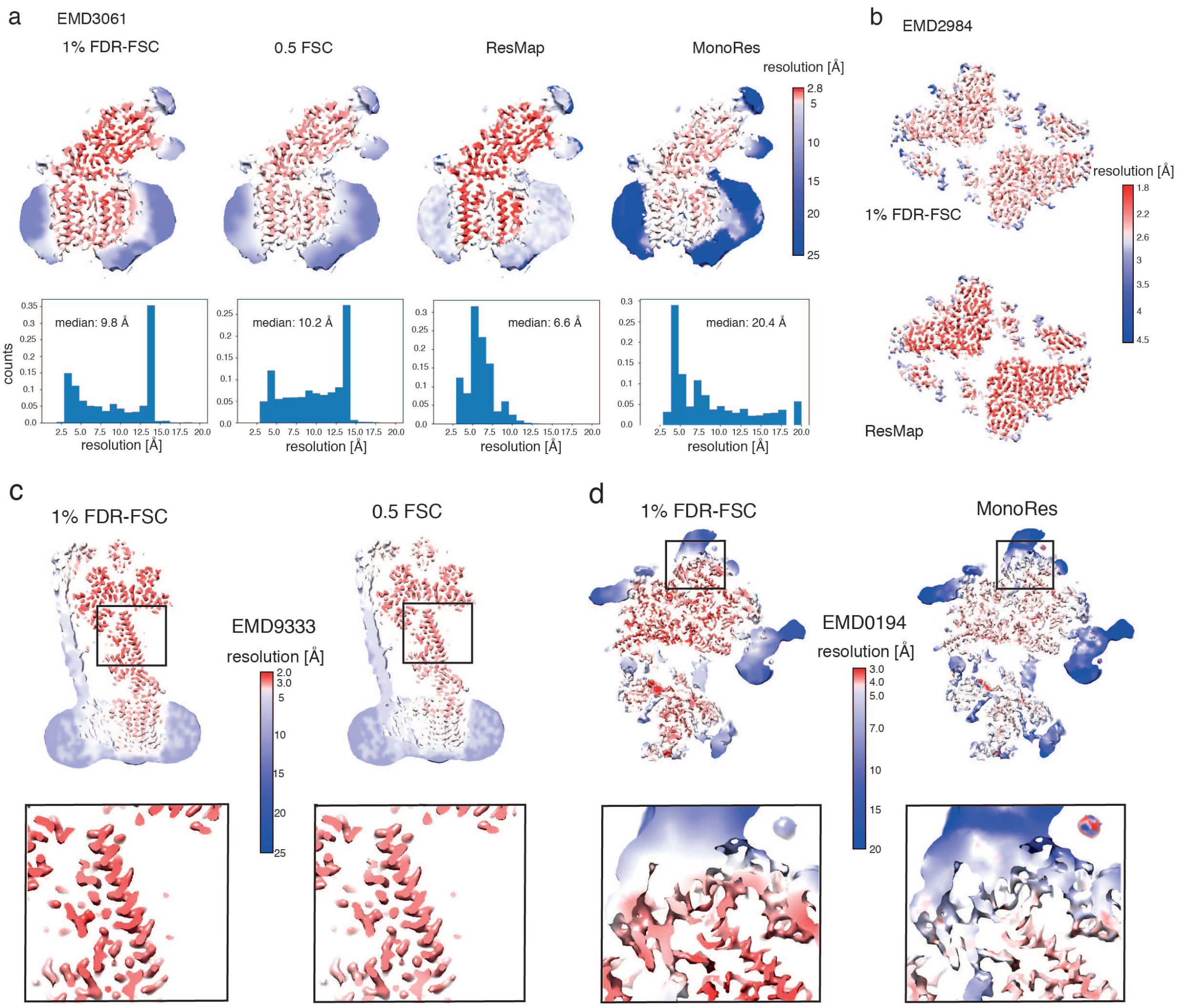
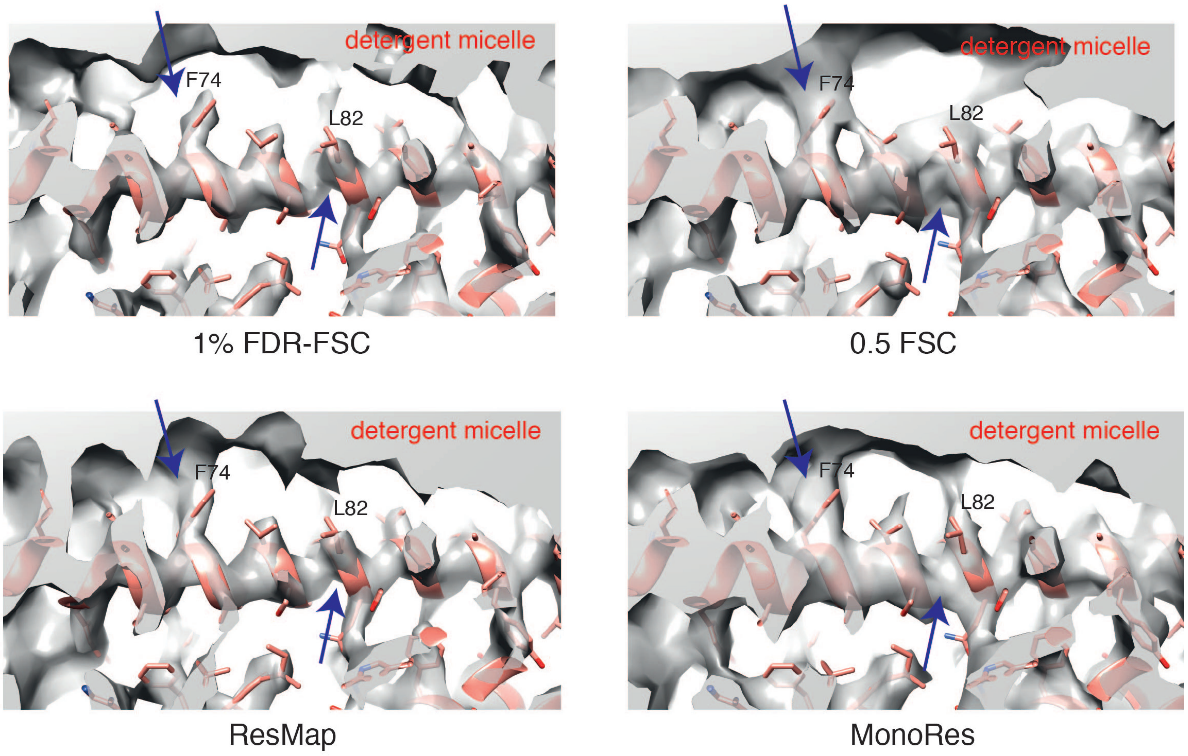


Figure 5.8: a) Comparison of local resolutions estimated for the γ-secretase map EMD3061 for 1% $FDR-FSC$, 0.5 FSC, ResMap and MonoRes (top) together with resolution histograms. b) Local resolutions at 1% FDR for the high-resolution β-galactosidase EMD2984 map compared to ResMap. c) Local resolutions at 1% $FDR-FSC$ of a 3.9 Å map of a bacterial ATP synthase (EMD9333) compared to 0.5 FSC. d) Local resolutions at 1% $FDR-FSC$ of a 3.8 Å map of a eukaryotic ribosome (EMD0194) compared to MonoRes.

Application to a 3.4 Å γ-secretase map [4](EMD3061) shows that the 1% $FDR-FSC$ criterion assigns higher resolutions to the resolved protein parts compared to 0.5 FSC, while at the same time, slightly lower resolutions around 10-15 Å to the detergent and gycosylations (Fig. 5.8 a). ResMap [77] assigns resolutions around the Nyquist frequency to the protein parts and 5 Å to the disordered parts, while MonoRes [149] gives estimates even more conservative than the 0.5 FSC criterion. Analysis of a 2.2 Å β-galactosidase [8](EMD2984) and 3.9 Å ATP synthase [56] (EMD9333) map show that the 1% $FDR-FSC$ criterion assigns higher resolutions to high-resolution parts compared to 0.5 FSC and, at the same time, avoids overestimation of low-resolution parts (Fig. 5.8 b and c). Comparison of the 1% $FDR-FSC$ criterion with MonoRes on a eukaryotic ribosome map [71](EMD0194) shows similar results as for the γ-secretase map: low resolution features like the L1-stalk domain are assigned lower resolutions >7 Å for both methods while the higher resolutions seem to be underestimated in MonoRes (4-5 Å) compared to 1% $FDR-FSC$ (3-4 Å), as side-chain density is clearly visible for the complete α-helical segment (Fig. 5.8 d).

One of the main applications of local resolutions is local resolution filtering, which requires accurate local resolution estimates. Underestimation of resolutions will directly lead to disappearance of high-resolution features while overestimation will leave the map too noisy. Comparison of local filtering with our method to 0.5 FSC, ResMap and MonoRes shows again that the 1% $FDR-FSC$ criterion avoids too conservative as well as inflated resolution estimates and outperforms the other approaches (Fig. 5.9). Both 0.5 FSC as well as MonoRes lead to smeared densities for a critical region close to the detergent micelle in the 3.4 Å γ-secretase map around residues F74 and L82, which indicates too heavy low-pass filtering in this region. Both ResMap and 1% $FDR-FSC$ give rise to higher resolution features, with 1% $FDR-FSC$ the only method to resolve residue F74.



1% FDR-FSC

0.5 FSC

ResMap

MonoRes

Figure 5.9: Effect of different local resolution estimates on resolvable map features after local filtering. Shown is a α-helix next to the detergent micelle, which poses a big local resolution gradient. A 0.5 FSC threshold (top right) and MonoRes (bottom right) tend to underestimate local resolutions for the well resolved parts, which leads to disappearance of high-resolution features. Local resolutions estimated at a 1% $FDR-FSC$ threshold (top left) and with ResMap (bottom left) give rise to similar higher resolution features. However, ResMap tends to overestimate resolutions (c.f. Fig. 5.8). All maps are contoured at a map threshold of 1% FDR.

5.4.3 Application to Directional Resolution Estimation

Complementary to local resolutions, directional resolutions have been developed recently in order to assess the effect of preferred particle orientations in the ice layer on the resulting 3D maps [168]. This can be realized by calculating the FSC curve from a subset of the 3D Fourier transform, where the subset is built from all voxels that deviate from the respective direction up to a specified angle, i.e. inside a cone with specified opening. Due to the limited size of Fourier coefficients inside the cone, directional FSCs, like local FSCs, suffer from substantial statistical uncertainties and the $FDR-FSC$ approach should provide similar benefits.

I tested the approach in more detail with three different cases: a map of the soluble portion of the small influenza hemagglutinin (HA) trimer, which exhibits highly preferred orientations (EMD8731, Fig. 5.10 a top) [168], a highly symmetric apo-ferritin map (EMD0144, Fig. 5.10 a middle) [169] and an asymmetric map of γ-secretase (EMD3061, Fig. 5.10 a bottom) [4].

First of all, Inspection of the directional resolution plots reveals that the 0.143 FSC-threshold tends to give lower resolution estimates than the 1% $FDR-FSC$ and again would require tight masking. Moreover, the 1% $FDR-FSC$ criterion gives resolution ranges much closer to the reported resolutions.

A more detailed analysis of the HA structure shows obvious directional resolution differences, with lower resolutions up to 8 Å in the vertical direction and higher resolutions up to 4.2 Å in the horizontal direction (Fig. 5.10a top and b). Due to highly preferred orientations, such a result is expected in this case [168]. On the other hand, the high-resolution map of apo-ferritin at 1.6 Å resolution does not exhibit such clear directional resolution differences (Fig. 5.10 a middle), due to high-degree of symmetry and more homogenous particle orientations. Similarly, the asymmetric map of γ-secretase seems not to exhibit strong directional resolutions (Fig. 5.10 a bottom).

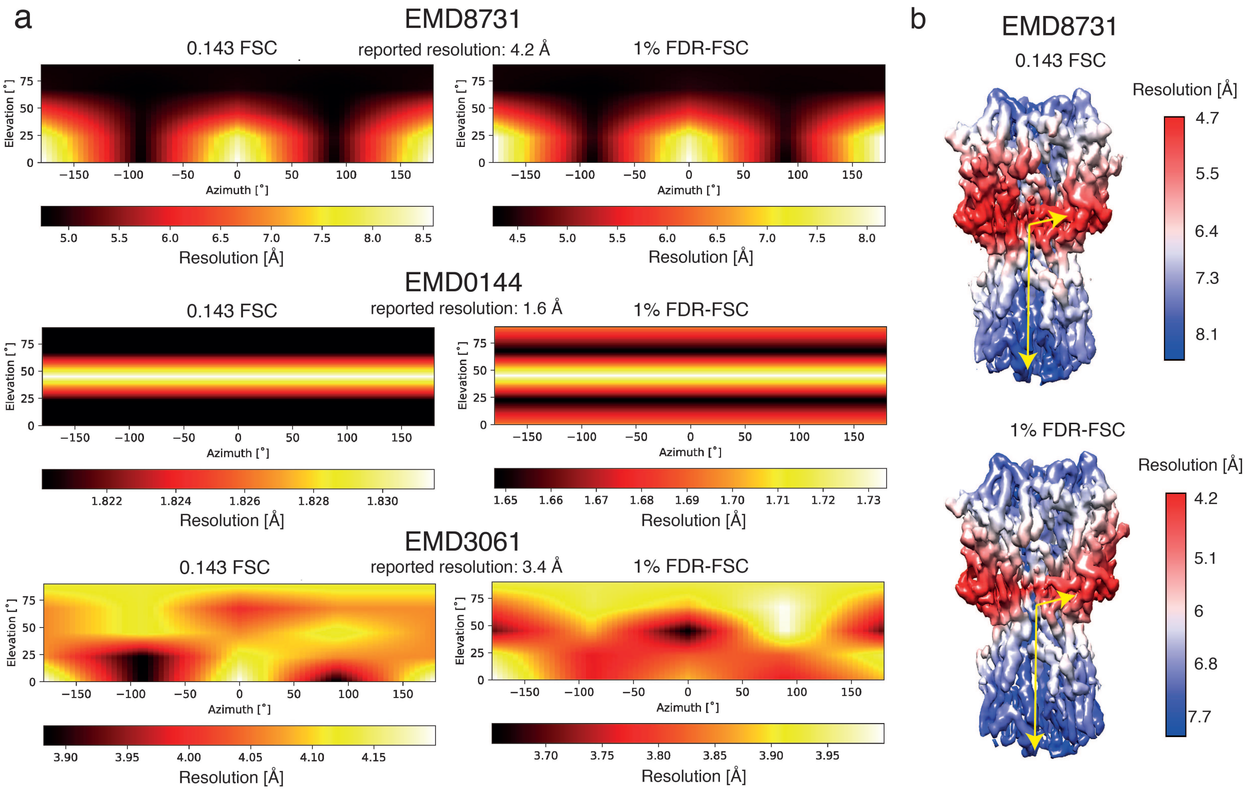


Figure 5.10: a) Comparison of directional resolution plots of EMDB entries 8731, 0144 and 3061 for 0.143 FSC (left) and 1% $FDR-FSC$ thresholds (right). Resolutions are shown in colors for the respective directions as characterized by azimuth and elevation. b) Directional resolutions mapped on the surface of EMDB entry 8731. The resolution at each voxel specifies the resolution in the direction given by the vector from the center to the respective voxel (c.f. yellow arrows). Clear directional resolution differences are visible with lower resolutions in the vertical direction.

5.4.4 Application to Model-Map Resolution Estimation

Fourier shell correlations also became important in assessment of overfitting of the atomic model to the EM map. However, although global model-map resolutions can provide useful hints about overfitting, they also suffer from substantial disadvantages. The averaged EM map has lower noise levels compared to the two half maps used for averaging. Moreover, many maps contain low-resolution parts which cannot be modelled and will thus be missed in the map-model FSC. As a result, the map-model resolution is generally expected to be better than the halfmap-halfmap resolution. Although masking and threshold adjustments can in principle be used adjust for these factors, these is again highly subjective. The 1% $FDR-FSC$ criterion extends to the case of model-map FSCs as well (Fig. 5.12 a) and has the advantage (as shown above) that it is very robust to the noise levels. Although it will still be biased from unmodelled features, overfitting can thus be assessed by comparing the model-map resolution to the model-halfmap resolutions, which should be the same if no overfitting occurs. As the atomic models are usually refined against the averaged maps, overfitting will be visible as a better map-model resolution compared to the model-halfmap resolutions.

Application to a map of β-galactosidase (EMD2984) shows that the model–map resolution of 2.0 Å is substantially better than the model-halfmap resolutions, which also differ for the two halfmaps (Fig. 5.12 b). Such an obvious difference to the model-halfmap resolutions indicates that the model was refined against information that is different between the two halfmaps, which is likely noise. In contrast, the 20S proteasome map EMD6287 and the γ-secretase map EMD3061 with their corresponding atomic models give similar results for the two halfmaps and the averaged map. In all cases, however, is the model-map resolution better than the halfmap-halfmap resolution, showing the bias explained above.

More useful information can in principle be obtained with local model-map resolutions and these can be used to identify problematic regions in the models. However, this is difficult with existing threshold criteria, as global model-map resolutions already require a 0.5 FSC threshold and going to smaller local windows would make an even more conservative threshold necessary. The 1% $FDR-FSC$ provides a simple solution to this problem, as it is implicitly taking into account the statistical uncertainties from smaller windows. Local model-map resolutions can be used together with local map resolutions to assess their local differences and, with this, local over and underfitting (Fig. 5.12 c and d). For example, for the β-galactosidase map we find that most parts of the model are not affected by overfitting, but some overfitting can be identified, for example at residue E136 (Fig. 5.12 c top). Vice versa, calculating local model-map resolutions for the same map against the atomic model from lower resolution data at 3.2 Å (pdb 3j7h) [7], underfitting can be identified and less overfitting (Fig. 5.12 c bottom). Investigation of the map features around the respective residues validated the findings: E980 and R961 from pdb 3j7h do not properly fit the map compared to pdb 5a1a (Fig. 5.11 top). Moreover, the carboxylate group of E136 in pdb 5a1a is fitted in two map peaks that only appear at a very low threshold (Fig. 5.11 bottom left), which are likely noise.

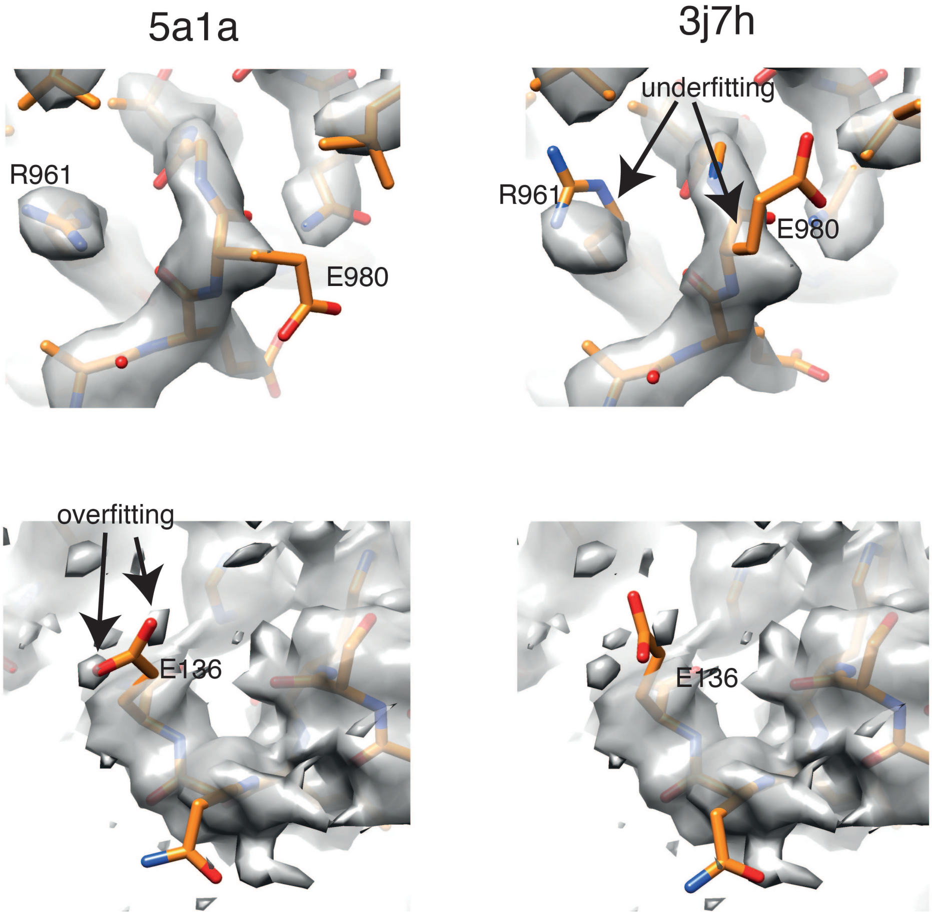


Figure 5.11: *FDR − FSC* applied to model-map Fourier shell correlations. Cryo-EM maps are shown for the residues that were identified by local map-model resolutions. It can be seen that residues R961 and E980 from the pdb 3j7h do not correspond to the maps (top right) and were thus identified as underfitted, while pdb 5a1a shows nice fits (top left). However, residue E136 from pdb 5a1a was identified as overfitted, which can be confirmed with the map. The two oxygen atoms from the carboxylate group seem to be fitted into two noise peaks (bottom left), which only appear at low thresholds. In comparison, E136 from pdb 3j7h cannot be related to any noise peak.

A similar analysis for the γ-secretase map shows that overfitting can be found at the glycosylation sites and at residues interacting with the membrane (Fig. 5.12 d left). Moreover, comparison to local model-map resolutions obtained with an atomic model determined from a different map (pdb 6idf) [162] reveals again substantial underfitting, but also disappearance of overfitting (Fig. 5.12).

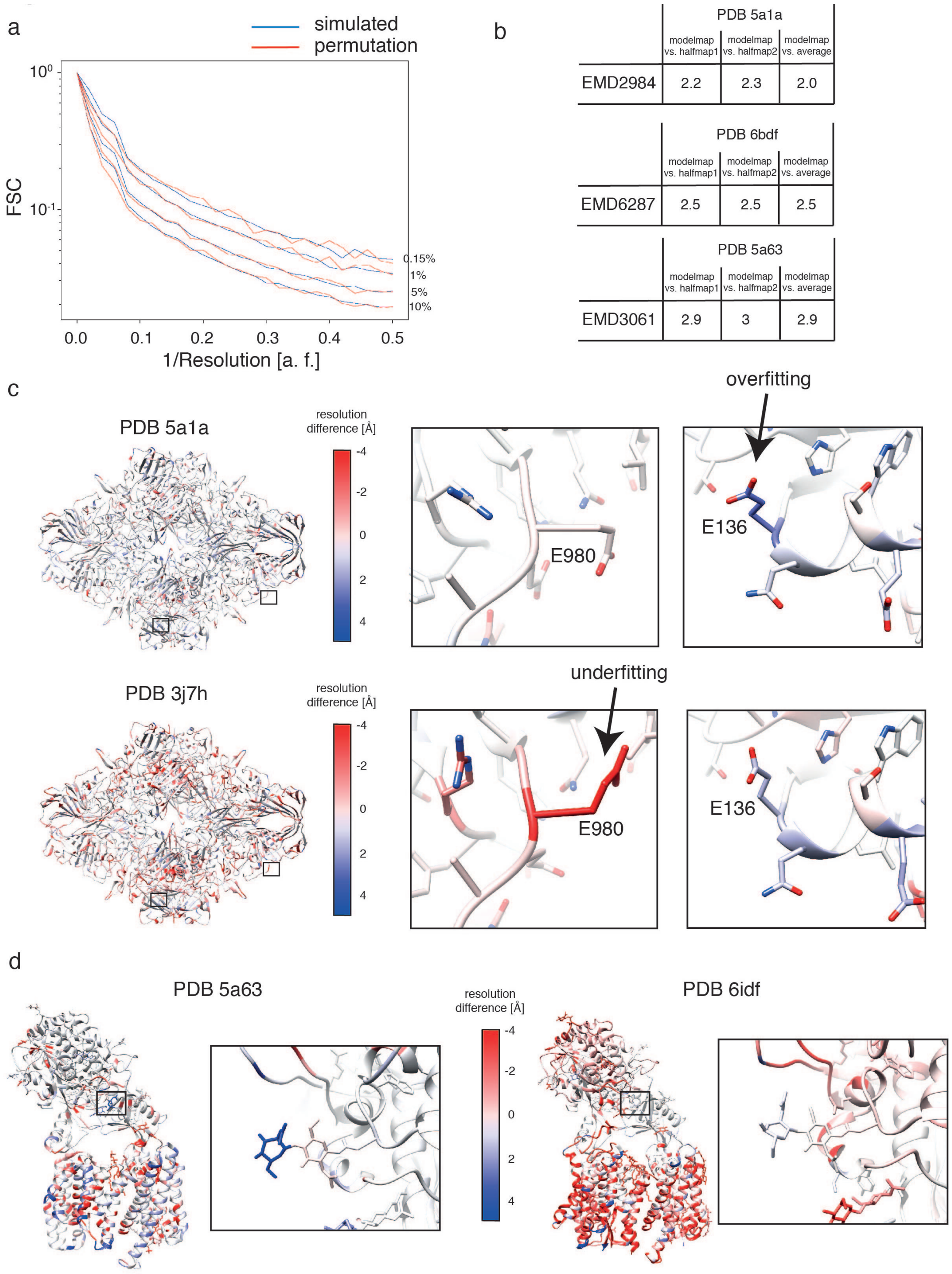
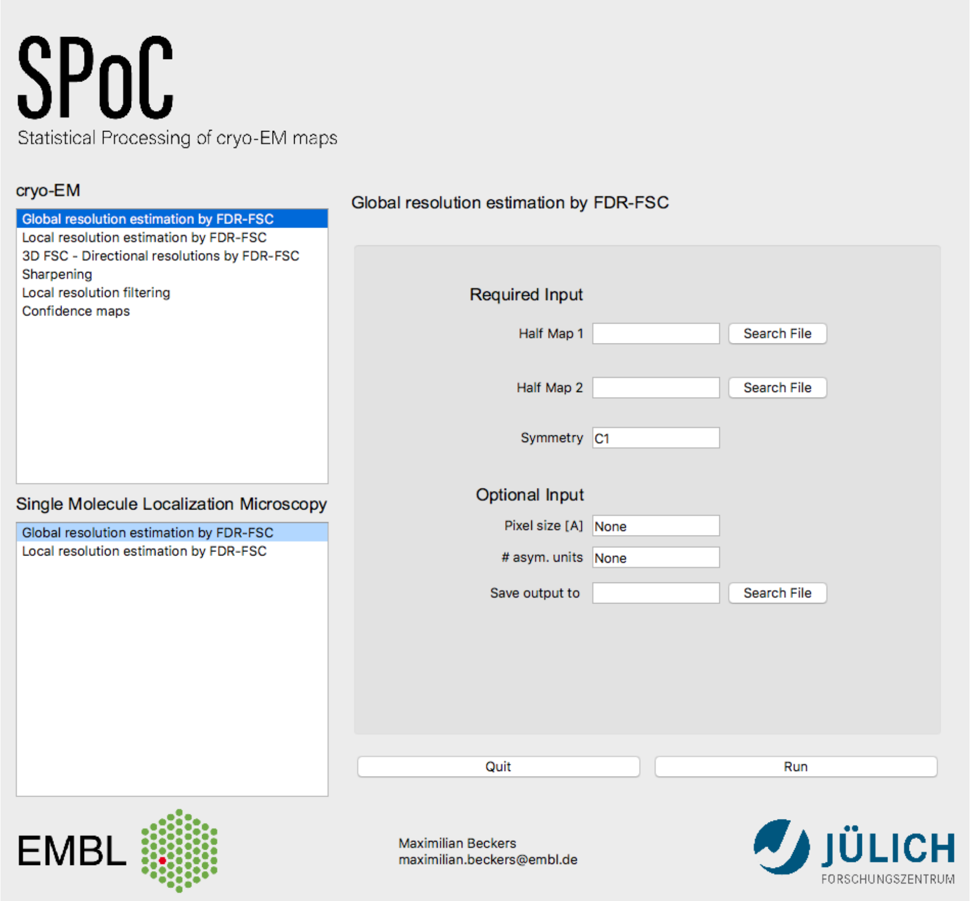


Figure 5.12: a) Right-sided 10%, 5%, 1% and 0.15% percentiles from the permutation approach (red) and from 5000 simulated FSCs from random noise maps against a model map simulated from the β-galactosidase structure pdb 5a1a (blue). b) Model-map resolutions at 1% $FDR-FSC$ for EMD2984 and EMD6287. Resolutions for both halfmaps as well as the averaged maps are shown. c) Differences between local model-map resolutions and local map resolutions mapped on the atomic models of β-galactosidase. EMD2984 was used as cryo-EM map and atomic models were pdb 5a1a (top) and pdb 3j7h (bottom). Overfitting is shown in blue and underfitting in red. d) Similar to c), but for γ-secretase. EMD3061 was used as cryo-EM map and atomic models were pdb 5a63 (left) and pdb 6idf (right).

5.5 Implementation

All algorithms for the analysis of cryo-EM maps that I developed in this thesis have been implemented in an easy-to-use and GUI based software termed SPoC (Statistical processing of cryo-EM maps). Besides of all FSC-based tools, SPoC includes confidence maps as well as sharpening tools and can thus be used a stand-alone software for the complete analysis of cryo-EM maps, starting only with the two halfmaps, as they commonly result from any reconstruction software. The incorporated confidence map tools also contain an interactive noise estimation button similar to the CCPEM implementation.The GUI is implemented with PyQt5 and the whole software is available as a pre-compiled executable file.



Symmetry C1

Pixel size [A] None

asym. units None

Figure 5.13: GUI of SPoC. SPoC is a stand-alone software that allows complete processing and analysis of cryo-EM maps starting with the two halfmaps form the 3D reconstruction.

5.6 Summary

Taken together, the presented results show that thresholding of FSC curves by multiple hypothesis testing and false discovery rate control allows completely automated resolution determination. The approach does not have any free parameters, only the symmetry used for reconstruction needs to be provided. Moreover, it allows objective resolution estimation without masking and thus avoids the specific choice of the mask. Although it can be argued that FSC curves from masked half maps are still necessary for FSC-weighting of maps, FSC-weighting does not correct for local effects and is thus inferior to local resolution filtering, which is independent of the specific shape of the FSC. Morevoer, the presented algorithm is computationally fast: a few seconds for small maps (<200 pixels box size) and a few minutes for big maps (>400 pixels box size).

Application to local resolution estimation showed that the adaptive 1% $FDR-FSC$ threshold corrects for the known underestimation of high-resolution features of fixed FSC thresholds while avoiding too optimistic estimates. Moreover, our criterion can be used with model-map FSC curves as well and we applied it to local model-map resolutions in order to assess local over and underfitting, which allows identification of problematic residues. Complementary to local resolutions, directional FSCs can be analyzed by $FDR-FSC$ as well and resulted in resolutions ranges closer to the reported overall resolutions of the maps.

Chapter 6

Elucidation of the Viral Disassembly Switch of Tobacco Mosaic Virus

The last chapter of this thesis is a direct application of the methods I developed, applied to two high-resolution maps from tobacco mosaic virus (TMV). This work was a collaboration with Felix Weis and was published as shared first-authors in EMBO Reports [156]. Some figures from the paper are reproduced here as permitted under the Creative Commons Attribution 4.0 Unported Licence (CC BY). Data acquisition and reconstructions of the 3D maps were done by Felix Weis. I processed and analyzed the cryo-EM maps, built and interpreted the atomic models and proposed the biochemical mechanism. Moreover, I wrote the initial version of the paper and prepared all the figures, which were then refined with constant feedback from Felix Weis and Carsten Sachse. Infection experiments were done by Iris von der Hocht.

6.1 Abstract

Capsids of RNA viruses protect the viral RNA from degradation in the extracellular environment. However, when the virus enters the cell in an enveloped state, the virion has to be disassembled in order to allow access of the viral genome. This requires a molecular mechanism that is capable of sensing the environmental differences that occur upon viral entry and modulate disassembly. Biochemically, this process has been extensively studied with tobacco mosaic virus (TMV) and carboxylate interactions have been proposed to play a major role. However, due to lacking high-resolution information, a detailed structural mechanism remains elusive. Here I describe two cryo-EM structures of the helical TMV assembly at 1.9 and 2.0 Å resolution in conditions mimicking the intracellular and extracellular environment. At the time of writing this thesis, these structures represent the best resolved helical reconstructions obtained with cryo-EM so far. Confidence maps and local post-processing of cryo-EM maps allowed the assignment and interpretation of the flexible inner loop region, which was found to contain the residues of the disassembly switch. Based on the atomic models, I was able identify the conformational details of the

switch mechanism: in high Ca^{2+}/acidic pH environment the virion is stabilized between neighbouring subunits through carboxyl groups E95 and E97 in close proximity to a Ca^{2+} binding site that is shared between two monomers. When the pH increases and the Ca^{2+} concentration drops, mutual repulsion of E95 and E97 and Ca^{2+} removal destabilize the network of interactions between neighbouring monomers and release the switch for disassembly.

6.2 The Disassembly of TMV

RNA viruses require a capsid structure that ensures the protection of the viral RNA from degradation in the extracellular environment. Especially the capsid of tobacco mosaic virus is highly thermostable in ambient temperatures and resistant to degradation across a wide range around neutral pH [80]. The TMV capsid is composed of a 17kDa coat protein (CP) that forms a helical filament and encloses the viral RNA between the CPs [45, 44, 101, 100]. During infection, TMV enters the plant cell via mechanical lesions that allow passing of the outer membrane [84]. Inside the host cell, the capsid undergoes a controlled destabilization which then allows co-translational virion disassembly by the replisome [85]. However, such a controlled destabilization requires a molecular mechanism that is able to sense the differences between the distinct environments, and then initiates the destabilization of the virion and further further facilitates the following replication inside the host cell. Ca^{2+} and pH dependent disassembly processes [37] have been found for a series of plant viruses, which exploit the lower Ca^{2+} and H^+ concentrations inside the plant cell compared with the extracellular environment [101, 101, 37].

Biochemically, the disassembly behaviour of TMV has been extensively studied, where it was found from titration experiments that TMV contains groups that titrate with pKa values between 7 and 8 [26]. This motivated Don Caspar to postulate the hypothesis that a carboxylate cluster in the TMV capsid may bind protons with high affinity. The respective residues, which were subsequently referred to as "Caspar-Carboxylates" [154], are assumed to drive the disassembly process through mutual repulsion upon entering the intracellular environment, as the lower proton concentration will lead to deprotonation of the carboxylates and the resulting negative charges repel each other. Additionally, it has been shown that TMV binds Ca^{2+} [87, 47].

The TMV capsid protein is arranged in a four-helix bundle, which is conventionally distinguished between three regions corresponding to the distance from the helical axis [125]: lower radius 20 – 40 Å, middle radius 40 – 60 Å and higher radius 60 – 90 Å. Studies of site-directed mutagenesis of the CP allowed identification the important residues [29, 89] involved in the disassembly process: Glu50 and Asp77 at middle radius have been hypothesized to be involved in axial carboxylate interactions, whereas Glu95, Glu97 and Glu106 at lower radius contribute lateral carboxylate and possible Ca^{2+} interactions [101].

Furthermore, experiments by single molecule force spectroscopy demonstrated that upon decrease of Ca^{2+} concentrations the 5' end of the capsid becomes exposed and RNA-coat protein interactions are weakened at the remainder of TMV [85], which likely helps to advance cotranslational disassembly by the replication machinery.

However, despite the plethora of experimental data at hand, a thorough structural understanding of the exact conformational switch sensing the environmental changes is still lacking. Although TMV was the subject of pioneering structural studies and also shaped the whole field of structural biology [45, 44, 101, 100, 125, 46, 75, 140], resolution of the helical rod was limited to 2.9 Å when determined by X-ray fiber diffraction [100] and to 3.3 Å by electron cryo-microscopy (cryo-EM) [125, 46]. More recently a resolution up to 2.3 Å could be achieved[136]. In these resolution regimes, the maps were sufficiently clear for the annotation of the architecture of the CP and assignment of bulky side-chains, whereas specific information about individual conformations for smaller and more flexible side-chains is lacking. To make things worse, the proposed Caspar carboxylate residues Glu95, Glu97, Glu106 and the calcium binding site are found in a rather flexible part of the capsid protein at lower radius with high B-factors, as it is expected for such a metastable switch. Within the disk assembly of TMV in absence of RNA, the respective residues were not detectable in high resolution maps and thought to be disordered [14]. Moreover, negatively charged amino acid residues suffer from fast radiation damage when visualized with an electron beam [46][53], which results in weak and more ambiguous density features and makes it hard to assign their exact conformations. Taken together, the exact structural mechanism of the viral disassembly switch of TMV remains to be discovered.

In order to investigate this open question of virus disassembly, we exploited latest developments of high-resolution imaging by cryo-EM, data processing and map interpretation methods and determined two structures of TMV at $\leq$2 Å resolution. Based on these maps and and the corresponding atomic models we find that the metastable switch relies on a Ca^{2+} sensitive network of carboxylate and iminocarboxylate residues at lower radius, which becomes destabilized by Ca^{2+} release and carboxylate repulsion at higher pHs. The maps captured the virion at different conformational states of the interaction network and allow direct visualization the structural rearrangements that drive the viral disassembly switch.

6.3 Methods

This section is taken from the corresponding publication [156] and was written by all authors.

6.3.1 Sample Preparation

TMV sample was isolated as described in [125] and stored in 0.1 M Tris-HCl pH 7.0, 0.02% NaN_3 (w/v) at a concentration of ca. 33 mg/ml at 4 °C. A total of 50 μl of virus stock solution was dialyzed for 1 hour at room temperature against 50 ml of 0.1 M NaOAc pH 5.2, 20 mM $CaCl_2$ and 50 ml of MilliQ H_2O, respectively. Before plunge-freezing sample concentration was adjusted to 22 mg/ml and 1.1 mg/ml for the Ca^{2+}/acidic pH and the water condition, respectively. A total of 3.6 μl were applied on holey carbon grids (C-flat 300 mesh R2/2, Protochips) that had been glow discharged in an EasyGlow (Pelco) device. Grids were plunge-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific) with a blotting time of 2 s at 10°C and 100% humidity.

6.3.2 Electron Microscopy

Data acquisition was performed on a Titan Krios microscope (Thermo Fisher Scientific) operated at 300 kV, through a Gatan Quantum 967 LS energy filter using a 20 eV slit width in zero-loss mode. The dataset was recorded on a Gatan K2-Summit direct electron detector operated in super-resolution mode, at a calibrated magnification of 215,000 (resulting in a super-resolution pixel size of 0.319 Å on the object scale) with a defocus range of 0.15 – 0.35 μm. For the TMV in water, a total of 20 frames were recorded in movies of 5 s exposure at a dose rate of ca. 2.6 e^-/physical pix/s, accumulating a total dose of 30.8 $e^-/\text{Å}^2$ at the sample level. For the TMV in Ca^{2+}/acidic pH conditions, a total of 40 frames were recorded in movies of 4 s exposure with a dose rate of ca. 3.7 e^-/physical pix/s, accumulating a total dose of 41.3 $e^-/\text{Å}^2$ at the sample level. For both samples, data collection was performed on a single grid using SerialEM [94].

6.3.3 Image Processing

After inspection of the micrographs, 62 images for the TMV in water, and 197 images for the TMV in Ca^{2+}/acidic pH conditions, were selected and both datasets were processed in the same way. Movie frames were aligned and dose-compensated with MotionCor2 [167] using patch-based alignment (5 x 5) followed by 1/2 cropping in the Fourier domain, resulting in 2x lower pixel sampling and a pixel size of 0.638 Å. Contrast transfer function parameters for the micrographs were estimated using Gctf [165]. Helix coordinates were determined automatically using MicHelixTrace [67], resulting in ca. 20,000 segments for each sample. Complete 2D and 3D classifications and refinements were performed using RELION implementation of single-particle based helical reconstruction [58], including per-particle refinement of CTF parameters, correction of estimated beam tilt and "Bayesian polishing" [170]. Helical symmetry parameters were refined to a helical rise/rotation of 1.405 Å/22.036° and 1.406 Å/22.038° for the Ca^{2+}/acidic and the water structure, respectively. The reported overall resolutions for TMV of 2.0 Å in Ca^{2+}/acidic pH and 1.9 Å in water conditions, were calculated using the Fourier shell correlation (FSC) 0.143 criterion. The final maps were corrected for the modulation transfer function of the detector and sharpened by applying a negative B factor that was estimated using automated procedures[32] (-41 $Å^2$ for the TMV in water and -42 $Å^2$ for the TMV in Ca^{2+}/acidic pH conditions). Local resolution maps were calculated with BlocRes [25] at a 0.5 FSC cutoff and the maps were subsequently local resolution filtered. To annotate significant molecular map features in the 3D reconstruction and to control false positive voxels, confidence maps using local resolution information were generated [10].

6.3.4 Atomic Model Building and Refinement

Atomic models were built and refined as 9-mers in order to account for inter-subunit interactions. PDB 4udv [46] was used as starting model and rigid body fitted into the processed maps using Chimera [114]. Additional H_2O and Mg^{2+}/Ca^{2+} ions were placed in the maps where biochemically appropriate and confirmed using the 1% FDR thresholded confidence maps. Mg^{2+} was placed in proximity of the RNA and justified by the known tendency of RNA to be stabilized by Mg^{2+} ions. The Ca^{2+} ion was identified by the combination of high map intensities, octahedral coordination and lower B-factors in the respective region, which distinguishes it clearly from water molecules. The Ca^{2+} ion was only found in the cryo-EM structure with high Ca^{2+} concentrations. Several rounds of real-space refinement with phenix.real_space_refine [2] using electron scattering factors and manual rebuilding with Coot [40] were done to obtain the presented models. Refinement was performed using rotamer, Ramachandran and C-β restraints in addition to standard restraints of bond-lengths, angles, etc. Real-space refinement was carried out with global minimization and local grid search options activated. Atomic coordinates and B-factors were refined against the sharpened and locally filtered maps. Residues 154 – 158 were not modelled as the corresponding map features were not sufficiently well resolved. An additional significant map feature at 1% FDR at the N-terminus was modelled as a N-terminal acetylation, as reported in [143]. Grouped atomic displacement factors (ADP) were refined with phenix.real_space_refine. Validation scores were calculated with phenix.molprobity [28] and DipCheck [113]. To assess overfitting of the refinement, we introduced random coordinate shifts into the final models using the program phenix.pdbtools with the shake option and a mean error of 0.5 Å, followed by refinement against the first unfiltered halfmap (halfmap 1) with the same parameters as above. Comparisons of FSC curves of the randomized model refined against halfmap1 versus halfmap 1 and the FSC curve of the same model map versus halfmap 2 do not indicate overfitting.

6.3.5 Infection of Nicotiana tabacum with TMV

The Bel B and Bel W3 variants of Nicotiana tabacum were chosen for infection experiments as they are known to be sensitive to TMV. Seeds distributed on a soil surface were watered and placed in a greenhouse for germination and cultivation at the following growth conditions: length of day: 14h, day: 28 °C/ night: 22 °C, relative humidity: 70%. Seedlings were piqued after 16 days, repotted after 31 days for the first time and repotted after 58 days for the second time. Before infection, the plants reached a height of 65 – 85 cm. On Day 60, one plant of both variants was infected with the TMV whereas the second plant was cultivated free of virus as a control. They were grown under ambient room temperature conditions and 16 hours of neon light. For infection, 25 µl of tobacco mosaic virus stock (33 mg/ml) was diluted into 10 ml PBS pH 7.5 and mixed with 105 mg Silicon carbide (SiC, 200-450 mesh, Sigma-Aldrich) in a porcelain mortar. SiC was used as an abrasive to cause small wounds and lesions supporting virus entry [108] [131]. The pestle was dipped into the virus/SiC suspension and rubbed gently onto the top surface of each plant leaf [126] [95]. Ten days after the infection event, first symptoms of TMV replication were visible. The variety Bel B developed deformed leaves, yellowish spots and new leaves were unusually light green and showed stunted growth. The variety Bel W3 showed lesions, necrotic spots on the leaves and stunted growth. After 35 days of infection, the plants possessed heights of 100 cm and 124 cm of Bel B and of Bel W3, respectively whereas the corresponding non-infected control plants were of 168 cm and 167 cm heights.

6.3.6 Figure Preparation

FSC and ADP graphics were visualized with ggplot2 in R [141] [158]. Chimera [114] was used for the figure preparation of the molecular maps and atomic models.

6.4 Results and Discussion

6.4.1 Cryo-EM Maps of TMV at 1.9 and 2.0 Å Resolution

In order to obtain different structural states of the virion, TMV has been prepared in two different conditions: in presence of 20 mM $CaCl_2$ at a pH of 5.2 (from now on referred to as Ca^{2+}/acidic pH) and the second sample was prepared in pure water in the absence of any ions. Both samples were plunge-frozen and imaged using a 300 kV electron microscope equipped with a GIF Quantum K2 camera (Fig. 6.1), and resulted in 2.0 and 1.9 Å resolution structures of TMV in a Ca^{2+}/acidic pH condition and in pure water (FSC curves are shown in Fig. B.1).

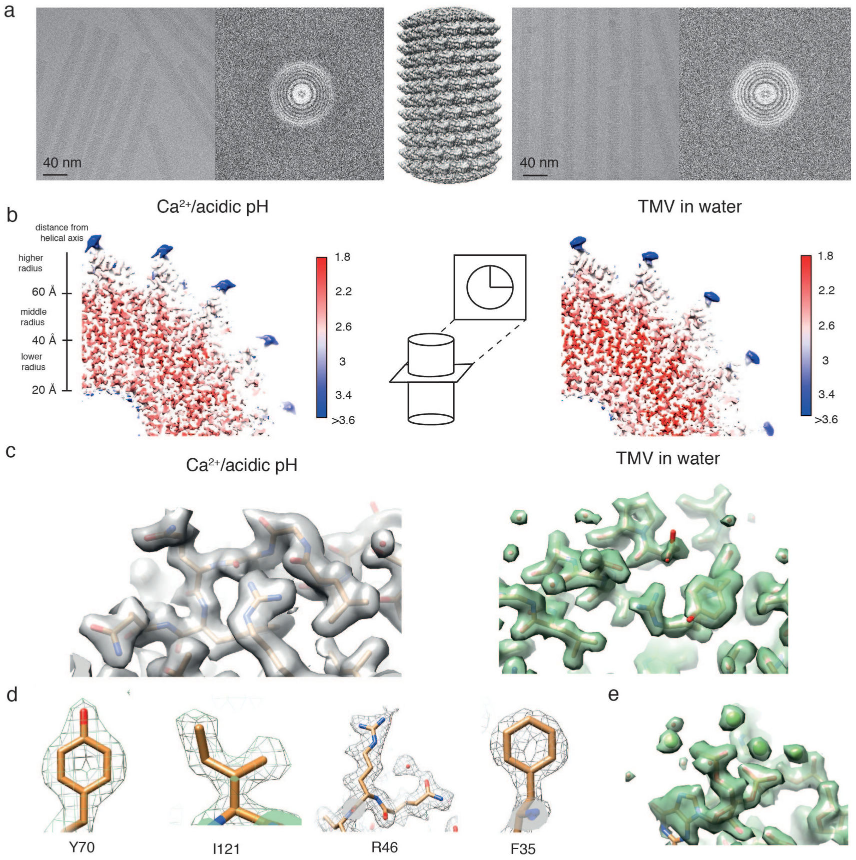


Figure 6.1: High-resolution cryo-EM structures of tobacco mosaic virus in conditions of Ca^{2+}/acidic pH (grey, left) and water (green, right). a) Characteristic micrographs for both data sets, respectively (left and right). b) Local resolutions mapped on the respective 3D reconstructions. The center of the coat protein is resolved up to ca. 1.8 Å with some features of approx. 5 Å resolution at the C-terminus. c) Both maps show well-resolved protein features including water molecules (left, right). (d) Side-chain features of Tyr70, Ile121, Arg46 and Phe35 residues at Ca^{2+}/acidic pH.(e) Snapshot of RNA map features with Mg^{2+} ions in water conditions. This figure is taken from [156] and was prepared by myself. Experimental data was acquired by Felix Weis.

Local resolutions for the best resolved parts of the CP reach 1.8 Å and ca. 5 Å for the disordered C-terminal tail (Fig. 6.1 b, Fig. B.2 a and c). Map details such as defined carbonyl oxygens of the protein backbone and holes in aromatic rings are clearly visible, as it is expected for parts with resolutions better than 2 Å. Moreover, a total of 92 water molecules for TMV in water and 71 water molecules under Ca^{2+}/acidic pH have been modelled. Donor-acceptor hydrogen bond lengths fall in the expected range between 2 and 5 Å and have a peak at ca. 2.8 Å (Fig. B.2 d). In addition, 4 Mg^{2+}-ions bound to RNA as well as well-defined side-chain conformers per CP in both conditions (Fig. 6.1 c, d, e) have been found. A single Ca^{2+} ion could be located in the Ca^{2+}/acidic pH structure in the inner loop at lower radius. However, no Ca^{2+} binding could be identified for TMV in water. Moreover, the proposed Ca^{2+} site at the RNA [100][48] was found to contain no Ca^{2+} in both maps (Fig. B.3 a).

Biological activity of the used virus batch was shown in infection experiments in tobacco plants. Infected plants exhibited typical symptoms of a TMV infection (stunted growth and necrotic lesions 35 days after infection (Fig. 6.2)).

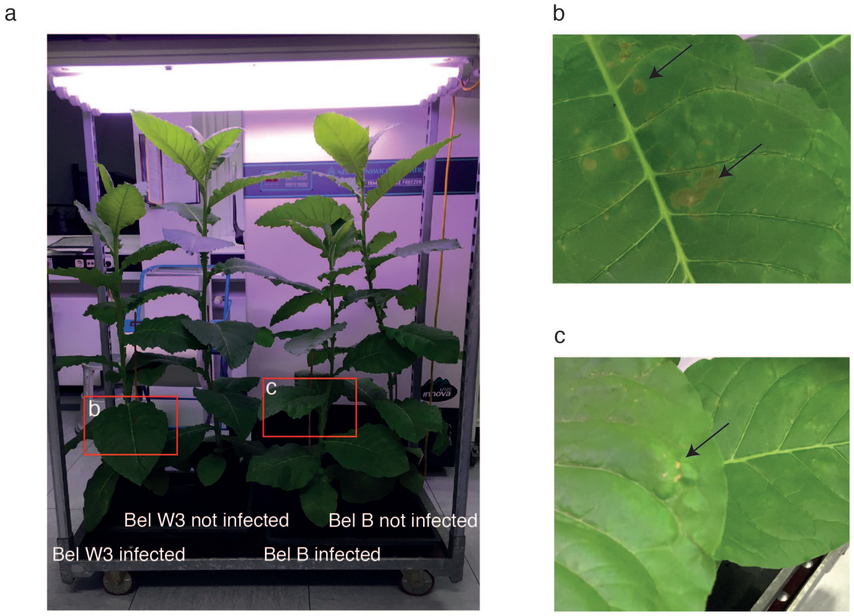


Figure 6.2: Symptoms of TMV infection on tobacco plants. a) Four tobacco plants from left to right: variant Bel W3 infected, Bel W3 not infected, variant Bel B infected, Bel B not infected. Infected plants are significantly reduced in height in comparison with the non-infected control plants. b) Leaf of variant Bel W3 with necrotic lesions (arrows). c) Leaf of variant Bel B with light green spots (arrow) and bulges. This figure is taken from [156] and was prepared by myself. The photos were taken by Iris von der Hocht, who also performed the infection experiments.

6.4.2 Structures at Lower Radius

Based on the statistical framework of confidence maps for the annotation of cryo-EM densities [10], that I developed and allows statistical control of the amount false positive signal in the interpreted volumes (see Chapter 4), I was able to interpret the complete maps from the Ca^{2+}/acidic pH and water samples including the lower radius region (Fig. 6.3 a). Although for most parts of the CP no obvious differences are apparent, the analysis also revealed that the lower radius regions shows significant differences at 1% FDR (Fig. 6.3 a right).

Comparison with a recently determined cryo-EM map (EMD2842) [46](Fig. B.2 b) revealed that this part of the protein was only poorly resolved. The main difficulties arising for the interpretation of this region is flexibility and radiation damage. Although previous studies on the layer aggregate of TMV assumed this region to be disordered [14], it was possible to locate statistically significant signal in the confidence maps and subsequently to compare the respective parts between the two confidence maps. Detailed comparison of the lower radius regions revealed that the protein backbone follows a different path in the Ca^{2+}/acidic pH and water structure (Fig. 6.3 b top): the Ca^{2+}/acidic pH structure adopts a more densely folded conformation.

For Ca^{2+}/acidic pH map I could built a single atomic model explaining the complete density (Fig. 6.3 b left). However, for the water structure 3 co-existing models in the residue range 97 – 100 at lower radius were necessary to describe the significant density, e.g. the map of the water structure is consistent with multiple conformations of E97 (Fig. 6.3 b right). It should be noted that the cryo-EM density for these three models is probably not discrete but compatible with a continuous ensemble of models.

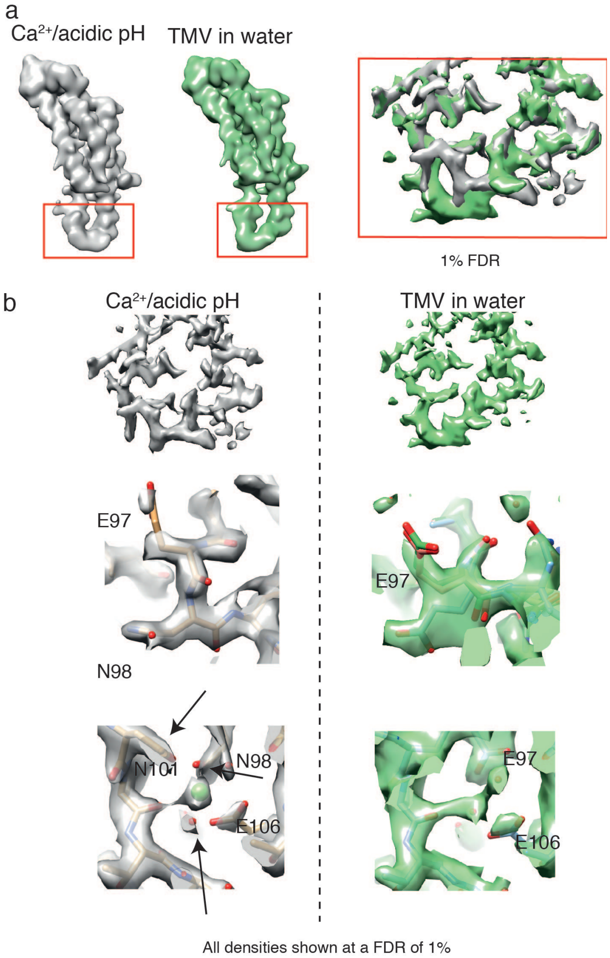


Figure 6.3: Confidence maps thresholded at 1% FDR from cryo-EM maps of Ca^{2+}/acidic pH and water structure. a) Low-pass filtered monomer map features of Ca^{2+}/acidic pH (grey) and water structure (green) with differences at lower radius (left). Zoomed inset (right) with maps displayed at a false discovery rate (FDR) threshold of 1% showing significant differences. b) Detailed map comparison at lower-radius region of Ca^{2+}/acidic pH (grey, left column) with the water condition (green, right column). A total of 3 atomic models (model 1: cyan, model 2: pink, model 3: green) describe the map of TMV in water whereas the Ca2+/acidic pH map could be modelled with a single atomic model (center). Located Ca^{2+} ion in the Ca^{2+}/acidic pH map with different conformations in the water structure (bottom). This figure is taken from [156] and was prepared by myself.

In order to quantify the structural differences, I calculated RMSDs between the models for the different regions. As anticipated from visual inspection of the models, the differences between the Ca^{2+}/acidic pH and water structures in the region outside the 90 – 110 region are very low with RMSDs <0.5 Å, whereas substantially higher with RMSDs between 2.4 and 2.8 Å inside the lower radius region. The different models placed for the water structure in the lower radius region show less strong deviations with RMSDs around 1.0 Å (Tab. 6.1).

Table 6.1: RMSD values for model comparisons. Combined main and side chain RMSD values for residues 90-110 and $\Delta(90-110)$ are shown in Tables a) and b). The respective side-chain only RMSD values are shown in c) and d).

a) Main and side-chain RMSD between models of TMV (90 – 110).

	Model 1	Model 2	Model 3
Ca^{2+}/acidic pH	2.80 Å	2.53 Å	2.38 Å
Model 1	-	0.94 Å	1.28 Å
Model 2	-	-	0.75 Å

b) Main and side-chain RMSD between models of TMV $\Delta(90-110)$.

	Model 1	Model 2	Model 3
Ca^{2+}/acidic pH	0.33 Å	0.40 Å	0.40 Å
Model 1	-	0.24 Å	0.24 Å
Model 2	-	-	0.01 Å

c) Main-chain RMSD between models of TMV (90 – 110).

	Model 1	Model 2	Model 3
Ca^{2+}/acidic pH	2.21 Å	1.93 Å	1.92 Å
Model 1	-	0.43 Å	0.46 Å
Model 2	-	-	0.24 Å

c) Main-chain RMSD between models of TMV $\Delta(90-110)$.

	Model 1	Model 2	Model 3
Ca^{2+}/acidic pH	0.09 Å	0.09 Å	0.09 Å
Model 1	-	0.01 Å	0.01 Å
Model 2	-	-	0.01 Å

Comparison of the two maps at the lower radius interface between neighbouring subunits uncovers additional differences: in the Ca^{2+}/acidic pH structure, densities for a bound ion are present. The map suggests coordination by Glu106, Asn101, Asn98 and a backbone carbonyl oxygen. In the water structure such a complete coordination is not present and asparagines Asn101 and Asn98 are facing away from the central density. In order to confirm the observed differences on the level of the cryo-EM maps, I compared the respective loop regions by calculating the difference of the two maps, a so called difference map (Fig. 6.4 a and b). The difference map clearly shows the rearrangement of the α-helical segment including the presence and coordination of the Ca^{2+} ion (Fig. 6.4 b). The specific assignment of the density by a Ca^{2+} ion is based on the strong density at the respective site and obvious octahedral coordination (Fig. 6.4 c). Local resolutions plots of the cryo-EM maps justify the placement of side chains and highlight again the stabilizing effect of the Ca^{2+}/acidic pH condition (Fig. 6.4 d), which leads to improved resolutions of the respective loop. This being said, it can be concluded that under Ca^{2+}/acidic pH conditions this subunit interface is stabilized by Ca^{2+} interactions, whereas in pure water the respective site is interacting with water molecules.

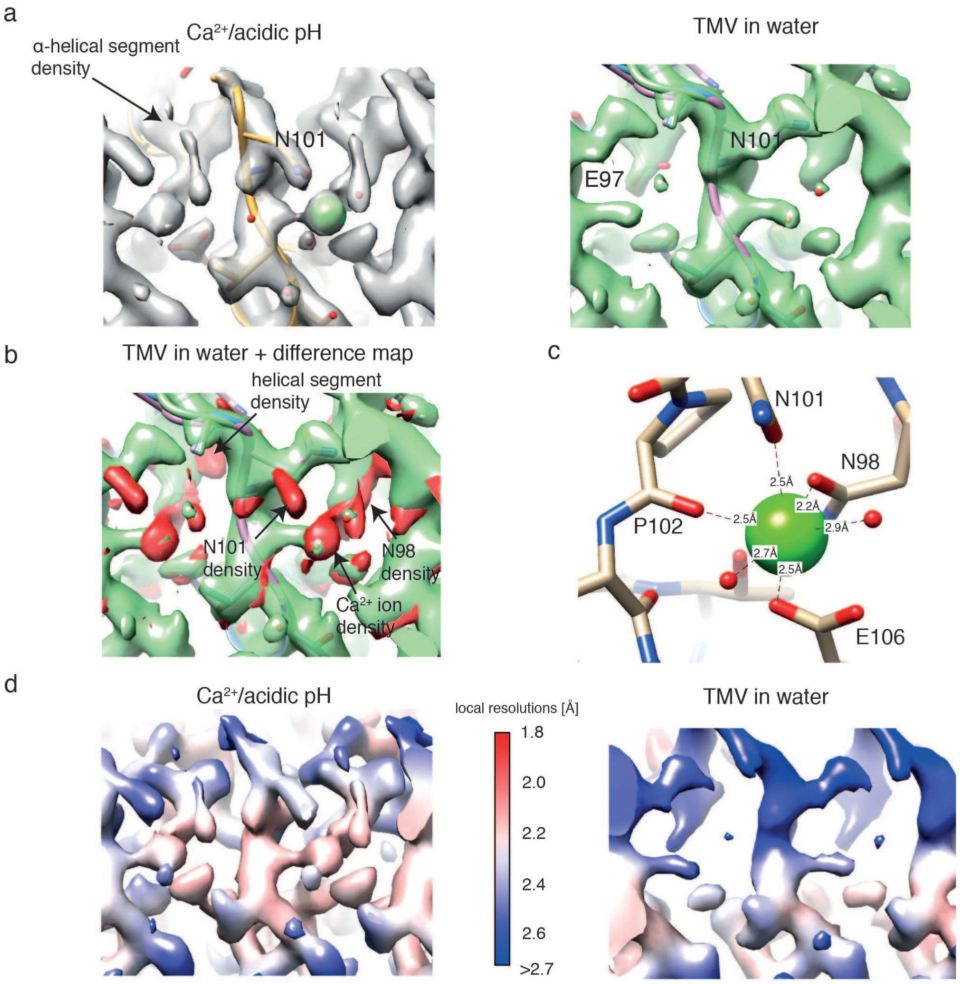


Figure 6.4: Visualization of structural differences between Ca^{2+}/acidic pH and water cryo-EM maps of determined states. a) TMV at lower radius in the cryo-EM maps under Ca^{2+}/ acidic pH condition (left) and of TMV in water (right). b) The corresponding view of the water cryo-EM structure (green) together with the difference map (red). Map features corresponding to the Ca^{2+} ion as well as for the coordinating residues and the rearranged helical segment are visible in the difference map. c) Detailed depiction of the Ca^{2+} ion including coordination distances with neighbouring residues. d) Local resolution plots of the lower radius region for the Ca^{2+}/ acidic pH map (left) and for TMV in water (right). This figure is taken from [156] and was prepared by myself.

6.4.3 Interactions in the Metastable Switch

For further in-depth analysis of the interactions within the two structures, I inspected the atomic coordinates which were refined by a common real-space optimization approach [2]. Comparison of the refined atomic models shows an extended helix of the short α-helical segment in the Ca^{2+}/acidic pH model by residue N98 at lower radius (Fig. 6.6 a). Moreover, visualizing the refined atomic B-factors clearly shows a drop from 42 to 25 Å^2 in the lower radius region for the Ca^{2+}/acidic pH condition. This confirms the finding that Ca^{2+} stabilizes the assembly structure once more (Fig. 6.5).

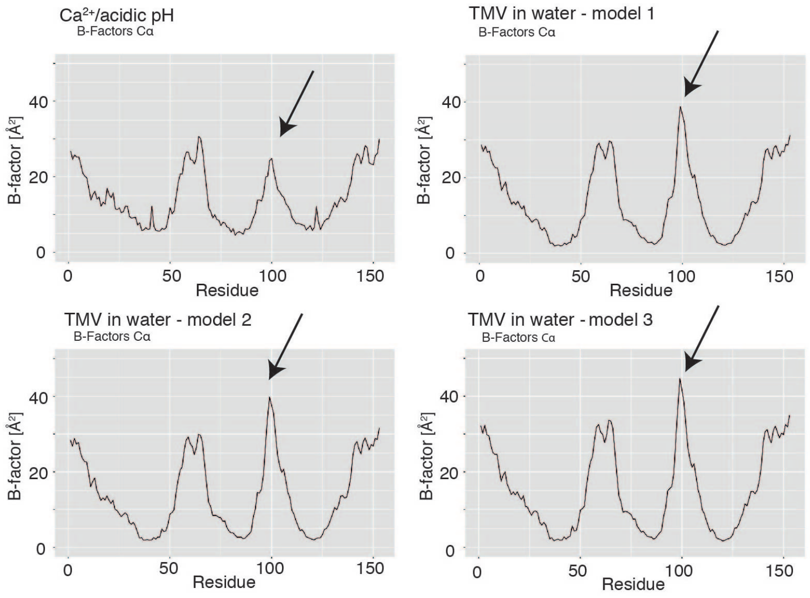


Figure 6.5: Plot of Cα B-factors with corresponding residue number. Plot of 4 determined models from this study are shown. The peak for residues 90 to 110 at the lower radius region (highlighted with an arrow) shows lower B-factors in the Ca^{2+}/acidic pH condition. For the remaining residues, the overall profile is very similar. This figure is taken from [156] and was prepared by myself.

I further analysed the carboxylate residues that were previously found to be important in the disassembly process (Glu50, Asp77, Glu95, Glu97 and Glu106). Glu50 and Glu77 contribute to tight axial carboxylate contacts at medium radius with a distance of 3.0 Å. These residues did not exhibit any differences between the two conditions. Moreover, glutamates Glu97 and Glu95 make up tight inter-subunit interactions at lower radius in the Ca^{2+}/acidic pH structure with a distance between the carboxylate groups of ~2.5 Å (Fig. 6.6 b left, top). Glu106 is not found to participate in carboxylate interactions. However, it contributes to the coordination of the Ca^{2+} ion in the respective Ca^{2+} site, which is shared between the neighbouring subunits. The ion itself is coordinated by Glu106 and Asn98 from one CP monomer and Asn101 as well as the backbone carbonyl oxygen of Pro102 from the adjacent monomer (Fig. 6.6 b left, bottom). These close inter-subunit interactions between neighbouring coat proteins stabilize the virion.

In contrast to the Ca^{2+}/acidic pH map, the water structure is lacking these close inter-subunit carboxylate contacts. Glu97 assumes two different conformations in the water structures, one towards Glu106 (model 1) and one towards Glu95 (model 2), both with longer distances of 3.9 Å and 5.4 Å to the closest carboxylate residues (Fig. 6.6 b right, top). Asn101 and Asn98 exhibit different orientations in the water structures as well(Fig. 6.6 b right, bottom), but contribute in the coordination of Ca^{2+} in the Ca^{2+}/acidic pH condition. Based on the atomic models, I proposed the following mechanism of the viral disassembly switch: Due to the loss of Ca^{2+} coordination as well as the build-up of carboxylate repulsion at higher pH, the residue network in proximity of Glu106, Asn101 and Asn98 is destabilized and changes its conformation (Fig. 6.7).

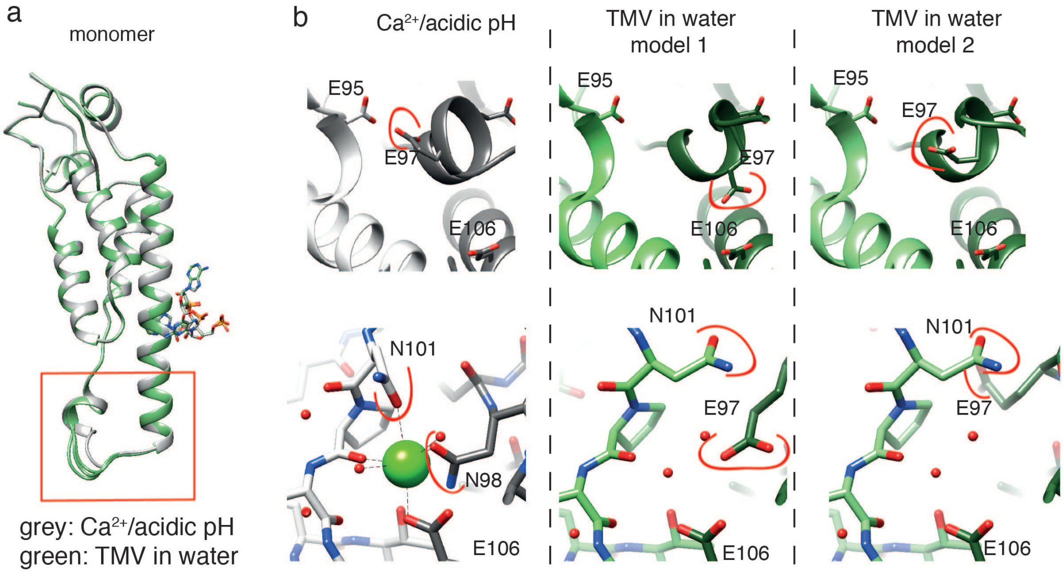


Figure 6.6: Model comparison of Ca^{2+}/acidic pH and water structural states. a) Superposition of the monomer structures with the lower radius region highlighted in the red box. The Ca^{2+}/acidic pH state (grey) shows an additional α-helical segment when compared with the water models (green). b) Comparison of the Glu95-Glu97 interaction (top row) and of the Ca^{2+} binding site in the Ca^{2+}/acidic pH model (bottom row). Close proximity of Glu95 and Glu97 in the Ca^{2+}/acidic pH state whereas in water Glu97 flips towards Glu106 and is rather flexible. No Ca^{2+} ion and corresponding coordination is evident in the water model. Residues that change conformation are marked in red. Adjacent subunit models are displayed in lighter shade of the main color. Model 3 is not shown due to high structural similarity with model 2 in the displayed region. This figure is taken from [156] and was prepared by myself.

6.4.4 Charge Build-Up at Carboxylates Destabilizes the Ca^{2+} Site

Two high-resolution cryo-EM structures of TMV with and without Ca^{2+} allowed us to reconstruct the structural details of the viral disassembly switch. At acidic pH and in presence of high levels of Ca^{2+}, inter-subunit interactions mediated by Ca^{2+} coordination and attractive carboxylate interactions between Glu95 and Glu97 stabilize the virion. Under these conditions, carboxyl groups can bind protons with high affinity and are able interact via hydrogen bonding between the carboxyl groups. Proton binding also neutralizes their negative charges, which attenuates the repulsive force between Glu95 and Glu97 (Fig. 6.7).

Entering the cell leads to an increase of the pH and carboxyl groups tend become deprotonated, which will result in the mentioned repulsive forces between them. The subsequent tendency of Glu97 to move away from Glu95, together with the correlated motion of Asn98 and Asn101, destabilize the ion binding site and further promote Ca^{2+} removal, especially in a low Ca^{2+} environment. All these concerted conformational transitions attenuate the inter-subunit interactions which originally stabilize the helical assembly and thus release the switch to disassemble the virus. Moreover, the participation of further residues such as Asn101 and Asn98 in the coordination of the Ca^{2+} ion suggests a more complex conformational network responsible for rearrangements beyond the previously postulated mechanism based on simple carboxylate repulsion. This observation is in concordance with results from a series of mutation studies [89], where additional residues like Ans98 were shown to be important.

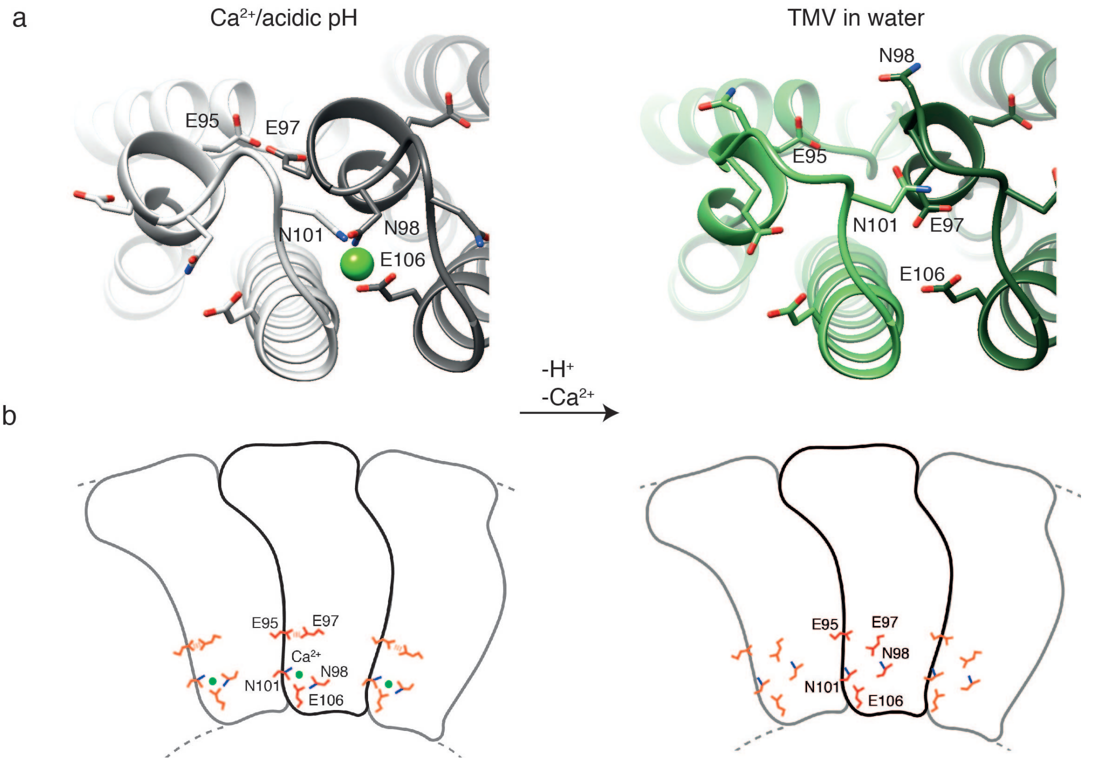


Figure 6.7: Towards a disassembly mechanism based on the Ca^{2+}/acidic pH and water states. a) The close proximity of the E95-E97 interaction and tight coordination of Ca^{2+} in the Ca^{2+}/acidic pH state site suggests a mechanism in which, upon pH change and Ca^{2+} removal, repulsive forces between charged carboxylates destabilize the network of interactions at lower radius releasing the switch for viral disassembly. b) Schematic presentation of coat protein with neighbouring subunits including main residues of TMV in the Ca^{2+}/acidic pH state (left) and in water (right) responsible for the metastable disassembly switch. This figure is taken from [156] and was prepared by myself.

An additional Ca^{2+} site that interacts with the RNA backbone has been suggested previously [100] [48], which could, however, not be located in the two structures presented here (Fig. B.3 a). More precisely, the conformations around the RNA in all the structures of this study correspond to what has been referred to as the low Ca^{2+} state [48], with Glu116, Arg92 and Arg90 involved in RNA binding [125] and a direct interaction of Arg92 with Asp116. Taking this observation into account, the proposed second Ca^{2+} site and the altered conformation around the RNA [100] [48] may not be necessary for TMV stabilization. Moreover, Glu109 was neither found to assume different conformations in the two structures, nor to form interactions with one of the before mentioned residues (Fig. B.3 b), although it was thought to be important for the disassembly process.

How far these observed differences in the maps reflect distinct sample preparations or experimental uncertainties of previous datasets is not easy to resolve. In order to make sure that the used batch of TMV is a biologically active virus, we showed in infection experiments that the used sample is capable of infecting tobacco plants (Fig. 6.2) and although the local resolutions at lower radius of the structures of this work drop to ~2.5 Å, the maps are sufficiently clear to locate all the mentioned side chains with high confidence (Fig. 6.3).

It has to be noted that the structural destabilization mechanism provides the opportunity for a a cooperative disassembly switch between the subunits: removal of a Ca^{2+} from the coordination site at lower radius is going to result in an instantaneous effect on the neighbouring Ca^{2+} sites, which are located in close distance of only 10 Å to each other and share residues. This has important implications especially on the 5' and 3' ends of the virion, as the end subunits can only have weak interactions at the shared Ca^{2+} coordination site in the Ca^{2+}/acidic pH condition, because they lack one of the stabilizing neighbouring subunits. When the Ca^{2+} concentration decreases, the ends will be even further destabilized. This makes the ends more easily accessible to the pulling replisome machinery.

To sum up, the shared Ca^{2+} site gives a direct explanation for the preferential capsid opening at the virion ends and cooperative weakening of RNA-coat protein interactions, thereby facilitating co-translational virion disassembly by the replisome [85]. I want to mention again, that, although the lower radius region of the virion is destabilized under low Ca^{2+} and high pH conditions, most parts of the CP are not affected by these changes of the environmental conditions. This is a highly important aspect of the CP: only a subtle destabilization of the metastable switch is necessary to trigger cotranslational disassembly [161]. At the same time, the virion is still stable enough to allow spontaneous reassembly of the CP in the helical assembly after viral replication [85].

Chapter 7

Summary and Conclusions

Statistical method development can be referred to as a classical field of research and dates back to the beginning of scientific activities. It illustrates our persistent challenge as scientists to draw objective conclusions from noisy observations and, despite this age, it didn't forfeit importance. Continuous development of new experimental techniques requires new statistical tools for validation, exploration and discovery based on the acquired data. One such technique is cryogenic electron microscopy, or simply cryo-EM, which has become the method of choice for high-resolution imaging of biological macromolecular complexes in recent years.

Throughout this thesis, I presented new methodology for the analysis of three dimensional maps determined with cryo-EM. The problems I focused on are at the very basis of cryo-EM map interpretation and were not properly dealt with so far. A major challenge with this respect was not only providing appropriate solutions, but rather giving a rigorous formulation of what the problem actually is and how we can quantify it.

The thresholding problem for cryo-EM map visualization, as well as the resolution estimation by means of Fourier shell correlation curves, could both be phrased as a signal detection problem and were addressed in a high-dimensional statistical inference framework based on multiple hypothesis testing and non-parametric statistical learning via permutation sampling. While the third big part of the thesis was a rather biological case study of the developed methodology applied to a high-resolution map of tobacco mosaic virus (TMV), TMV can also be seen as one of the reference samples of cryo-EM method development and structural biology in general. Given this, I hope that the thesis at hand could provide some new insights into the respective problems.

Chapter 8

Acknowledgements

I want to express my big gratitude to all people which contributed to this work and provided me with almost endless support during the time of my PhD. A big thanks goes to my supervisor, Carsten Sachse, for giving me the possibility to do this dissertation, for providing plenty of freedom to pursue my own ideas and for all the support, chances and conference trips I was enabled. Although big parts of this thesis fell into the time of the lab moving to Juelich, he always found time for discussions and to publish our work. All the advice towards more comprehensible writing, figure preparation and structured research have truly helped me to become a better scientist.

I also want to thank all the people in the Sachse lab and at EMBL that provided me with big support. Especially Arjen Jakobi and Simon Fromm, who introduced me to field of cryo-electron microscopy when I started at EMBL, which was a great help. Felix Weis was a great collaborator for the work on the TMV structures and a great office neighbour. Colin Palmer always provided very constructive and helpful comments on my software and invested quite some time in order to implement it in the CCPEM software suite.

Finally I want to thank all the people that accompanied me my whole life so far: My friends, many of whom I already got to know in the kindergarten, my brothers and of course my parents. You always supported me and where always there when needed. Growing up in a home, where education always had highest priority and thinking with the herd was frowned upon, truly helped me to develop my own way of thinking and still drives me to work on my own ideas.

Last, but certainly not least, I want express my deep gratitude and love to my wife Vera. Especially our move to Heidelberg, sharing life together and marrying me made my PhD an amazing time. All your understanding and patience when my Mac was opened again in the evening and at the weekend, when slight resentment about ridiculous reviewer comments came up again and also all your advice to my work will be forever remembered.

Bibliography

[1] https://www.ebi.ac.uk/pdbe/emdb/statistics_main.html/, 2019.

[2] Pavel V. Afonine, Billy K. Poon, Randy J. Read, Oleg V. Sobolev, Thomas C. Terwilliger, Alexandre Urzhumtsev, and Paul D. Adams. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology*, 74(6):531–544, 2018.

[3] T. W. Anderson and D. A. Darling. A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49(268):765–769, 1954.

[4] Xiao Chen Bai, Chuangye Yan, Guanghui Yang, Peilong Lu, Dan Ma, Linfeng Sun, Rui Zhou, Sjors H.W. Scheres, and Yigong Shi. An atomic structure of human γ-secretase. *Nature*, 525(7568):212–217, 2015.

[5] Niccolò Banterle, Khanh Huy Bui, Edward A. Lemke, and Martin Beck. Fourier ring correlation as a resolution criterion for super-resolution microscopy. *Journal of Structural Biology*, 183(3):363–367, 2013.

[6] Alberto Bartesaghi, Cecilia Aguerrebere, Veronica Falconieri, Soojay Banerjee, Lesley A. Earl, Xing Zhu, Nikolaus Grigorieff, Jacqueline L.S. Milne, Guillermo Sapiro, Xiongwu Wu, and Sriram Subramaniam. Atomic Resolution Cryo-EM Structure of β-Galactosidase. *Structure*, 26(6):848–856.e3, 2018.

[7] Alberto Bartesaghi, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam. Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(32):11709–11714, 2014.

[8] Alberto Bartesaghi, Alan Merk, Soojay Banerjee, Doreen Matthies, Xiongwu Wu, Jacqueline L.S. Milne, and Sriram Subramaniam. 2.2 å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science*, 348(6239):1147–1151, 2015.

[9] Maximilian Beckers, Florian Drechsler, Tobias Eilert, Julia Nagy, and Jens Michaelis. Quantitative structural information from single-molecule FRET. *Faraday Discuss.*, 184:117–129, 2015.

[10] Maximilian Beckers, Arjen J. Jakobi, and Carsten Sachse. Thresholding of cryo-EM density maps by false discovery rate control. *IUCrJ*, 6(1):18–33, 2019.

[11] Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik*, 86(5):78–88, 1852.

[12] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.

[13] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188, 2001.

[14] Balaji Bhyravbhatla, Stanley J. Watowich, and Donald L.D. Caspar. Refined atomic model of the four-layer aggregate of the tobacco mosaic virus coat protein at 2.4-Å resolution. *Biophysical Journal*, 74(1):604–615, 1998.

[15] Peter J. Bickel and Kjell A. Doksum. *Mathematical statistics: Basic ideas and selected topics, second edition.* Chapman & Hall, 2015.

[16] Henry W. Block, Thomas H. Savits, and Jie Wang. Negative dependence and the Simes inequality. *Journal of Statistical Planning and Inference*, 138(12):4107–4110, 2008.

[17] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–483, 1927.

[18] RN Bracewell. Strip Integration in Radio Astronomy. *Australian Journal of Physics*, 9(2):198, 1956.

[19] Axel F. Brilot, James Z. Chen, Anchi Cheng, Junhua Pan, Stephen C. Harrison, Clinton S. Potter, Bridget Carragher, Richard Henderson, and Nikolaus Grigorieff. Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology*, 177(3):630–637, 2012.

[20] Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao Li Meng. *Handbook of Markov Chain Monte Carlo.* Chapman & Hall, 1 edition, 2011.

[21] Edgar Brunner and Ullrich Munzel. *Nichtparametrische Datenanalyse.* 2002.

[22] Yury S. Bykov, Miroslava Schaffer, Svetlana O. Dodonova, Sahradha Albert, Jürgen M. Plitzko, Wolfgang Baumeister, Benjamin D. Engel, and John Ag Briggs. The structure of the COPI coat determined within the cell. *eLife*, 6, 2017.

[23] Melody G. Campbell, Anchi Cheng, Axel F. Brilot, Arne Moeller, Dmitry Lyumkis, David Veesler, Junhua Pan, Stephen C. Harrison, Clinton S. Potter, Bridget Carragher, and Nikolaus Grigorieff. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure*, 20(11):1823–1828, 2012.

[24] Melody G Campbell, David Veesler, Anchi Cheng, Clinton S Potter, and Bridget Carragher. 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. *eLife*, 4:e06380, mar 2015.

[25] Giovanni Cardone, J. Bernard Heymann, and Alasdair C. Steven. One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *Journal of Structural Biology*, 184(2):226–236, 2013.

[26] D. L.D. Caspar. Assembly and Stability of the Tobacco Mosaic Virus Particle. *Advances in Protein Chemistry*, 18:37–121, 1964.

[27] Shaoxia Chen, Greg McMullan, Abdul R. Faruqi, Garib N. Murshudov, Judith M. Short, Sjors H.W. Scheres, and Richard Henderson. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, 135:24–35, 2013.

[28] Vincent B. Chen, W. Bryan Arendall, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1):12–21, 2010.

[29] James N. Culver, William O. Dawson, Kelly Plonk, and Gerald Stubbs. Site-directed mutagenesis confirms the involvement of carboxylate groups in the disassembly of tobacco mosaic virus. *Virology*, 206(1):724–730, 1995.

[30] Radostin Danev and Wolfgang Baumeister. Cryo-EM single particle analysis with the volta phase plate. *eLife*, page 5:e13046, 2016.

[31] Radostin Danev, Bart Buijsse, Maryam Khoshouei, Jürgen M. Plitzko, and Wolfgang Baumeister. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(44):15635–15640, 2014.

[32] Radostin Danev and Kuniaki Nagayama. Transmission electron microscopy with Zernike phase plate. *Ultramicroscopy*, 88(4):243–252, 2001.

[33] Radostin Danev, Dimitry Tegunov, and Wolfgang Baumeister. Using the volta phase plate with defocus for cryo-em single particle analysis. *eLife*, page 6:e23006, 2017.

[34] A. Descloux, K. S. Grußmayer, and A. Radenovic. Parameter-free image resolution estimation based on decorrelation analysis. *Nature Methods*, 16:918–924, 2019.

[35] Cyrus J. DiCiccio and Joseph P. Romano. Robust Permutation Tests For Correlation And Regression Coefficients. *Journal of the American Statistical Association*, 112(519):1211–1220, 2017.

[36] J. Dubochet and A. W. McDowall. Vitrification of pure water for electron microscopy. *Journal of Microscopy*, 124(3):3–4, 1981.

[37] Anthony C H Durham, Donald A. Hendry, and M. Barbara Von Wechmar. Does calcium ion binding control plant virus disassembly? *Virology*, 77(2):524–533, 1977.

[38] Bradley Efron. Size, power and false discovery rates. *Annals of Statistics*, 35(4):1351–1377, 2007.

[39] T. Eilert, M. Beckers, F. Drechsler, and J. Michaelis. Fast-NPS—A Markov Chain Monte Carlo-based analysis tool to obtain structural information from single-molecule FRET measurements. *Computer Physics Communications*, 219:377–389, 2017.

[40] Paul Emsley and Kevin Cowtan. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12):2126–2132, 2004.

[41] Anthony W.P. Fitzpatrick, Benjamin Falcon, Shaoda He, Alexey G. Murzin, Garib Murshudov, Holly J. Garringer, R. Anthony Crowther, Bernardino Ghetti, Michel Goedert, and Sjors H.W. Scheres. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature*, 547:185–190, 2017.

[42] Joachim Frank. The ribosome - Structure and functional ligand-binding experiments using cryo-electron microscopy. *Journal of Structural Biology*, 124:142–150, 1998.

[43] Joachim Frank, Jun Zhu, Pawel Penczek, Yanhong Li, Suman Srivastava, Adriana Verschoor, Michael Radermacher, Robert Grassucci, Ramani K. Lata, and Rajendra K. Agrawal. A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature*, 376:441–444, 1995.

[44] R E Franklin and A Klug. The splitting of layer lines in X-ray fibre diagrams of helical structures: application to tobacco mosaic virus. *Acta Crystallographica*, 8(12):777–780, dec 1955.

[45] Rosalind E. Franklin. Structure of tobacco mosaic virus. *Nature*, 175(4452):379–381, 1955.

[46] Simon A. Fromm, Tanmay A M Bharat, Arjen J. Jakobi, Wim J H Hagen, and Carsten Sachse. Seeing tobacco mosaic virus through direct electron detectors. *Journal of Structural Biology*, 189(2):87–97, 2015.

[47] Warren H. Gallagher and Max A. Lauffer. Calcium ion binding by isolated tobacco mosaic virus coat protein. *Journal of Molecular Biology*, 170(4):921–929, 1983.

[48] P. Ge and Z. H. Zhou. Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. *Proceedings of the National Academy of Sciences*, 108(23):9637–9642, 2011.

[49] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis.* Chapman & Hall, 3 edition, 2013.

[50] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.

[51] R. M. Glaeser. Specimen Behavior in the Electron Beam. In *Methods in Enzymology*, pages 19–50. 2016.

[52] Robert M. Glaeser. Retrospective: Radiation damage and its associated "Information Limitations". *Journal of Structural Biology*, 163:271–276, 2008.

[53] Timothy Grant and Nikolaus Grigorieff. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife*, 4:e06980, 2015.

[54] Timothy Grant, Alexis Rohou, and Nikolaus Grigorieff. cisTEM, user-friendly software for single-particle image processing. *eLife*, 7:e35383, mar 2018.

[55] Nikolaus Grigorieff. FREALIGN: High-resolution refinement of single particle structures. *Journal of Structural Biology*, 157(1):117–125, 2007.

[56] Hui Guo, Toshiharu Suzuki, and John L Rubinstein. Structure of a bacterial ATP synthase. *eLife*, 8:e43128, 2019.

[57] George Harauz and Marin Van Heel. Exact filters for general geometry three dimensional reconstruction. *Optik*, 73:146–156, 1986.

[58] Shaoda He and Sjors H.W. Scheres. Helical reconstruction in RELION. *Journal of Structural Biology*, 198(3):163–176, 2017.

[59] Leonhard Held and Manuela Ott. On p -Values and Bayes Factors. *Annual Review of Statistics and Its Application*, 5:393–419, 2018.

[60] Mark A. Herzik, Mengyu Wu, and Gabriel C. Lander. Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV. *Nature Methods*, 14:1075–1078, 2017.

[61] Mark A. Herzik, Mengyu Wu, and Gabriel C. Lander. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nature Communications*, 10:1032, 2019.

[62] Yosef Hochberg. A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802, 1988.

[63] Niklas A. Hoffmann, Arjen J. Jakobi, María Moreno-Morcillo, Sebastian Glatt, Jan Kosinski, Wim J.H. Hagen, Carsten Sachse, and Christoph W. Müller. Molecular structures of unbound and transcribing RNA polymerase III. *Nature*, 528:231–236, 2015.

[64] S Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2):65–70, 1979.

[65] G. Hommel. A stagewise rejective multiple test procedure based on a modified bonferroni test. *Biometrika*, 75(2):383–386, 1988.

[66] Corey F. Hryc, Dong-Hua Chen, Pavel V. Afonine, Joanita Jakana, Zhao Wang, Cameron Haase-Pettingell, Wen Jiang, Paul D. Adams, Jonathan A. King, Michael F. Schmid, and Wah Chiu. Accurate model annotation of a near-atomic resolution cryo-EM map. *Proceedings of the National Academy of Sciences*, 114(12):3103–3108, 2017.

[67] Stefan T. Huber, Tanja Kuhm, and Carsten Sachse. Automated tracing of helical assemblies from electron cryo-micrographs. *Journal of Structural Biology*, 202(1):1–12, 2018.

[68] Nikolaos Ignatiadis, Bernd Klaus, Judith B. Zaugg, and Wolfgang Huber. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods*, 13(7):577–580, 2016.

[69] Arjen J. Jakobi, Matthias Wilmanns, and Carsten Sachse. Model-based local density sharpening of cryo-EM maps. *eLife*, 6:e27131, 2017.

[70] W. James and Charles Stein. Estimation with Quadratic Loss. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, 1:361–379, 1961.

[71] Szymon Juszkiewicz, Viswanathan Chandrasekaran, Zhewang Lin, Sebastian Kraatz, V. Ramakrishnan, and Ramanujan S. Hegde. ZNF598 Is a Quality Control Sensor of Collided Ribosomes. *Molecular Cell*, 72(3):469–481, 2018.

[72] Heena Khatter, Alexander G. Myasnikov, S. Kundhavai Natchiar, and Bruno P. Klaholz. Structure of the human 80S ribosome. *Nature*, 520:640–645, 2015.

[73] Maryam Khoshouei, Mazdak Radjainia, Wolfgang Baumeister, and Radostin Danev. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nature Communications*, 8:16099, 2017.

[74] Earl J. Kirkland. Computation in electron microscopy. *Acta Crystallographica Section A: Foundations and Advances*, 71(1):1–27, 2016.

[75] A. Klug. The tobacco mosaic virus particle: Structure and assembly. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1383):531–535, 1999.

[76] Sami Koho, Giorgio Tortarolo, Marco Castello, Takahiro Deguchi, Alberto Diaspro, and Giuseppe Vicidomini. Fourier ring correlation simplifies image restoration in fluorescence microscopy. *Nature Communications*, 10:3103, 2019.

[77] Alp Kucukelbir, Fred J. Sigworth, and Hemant D. Tagare. Quantifying the local resolution of cryo-EM density maps. *Nature Methods*, 11:63–65, 2014.

[78] Werner Kühlbrandt, A. Amunts, M. Liao, E. Cao, D. Julius, Y. Cheng, M. Allegretti, X. Li, N. Ban, B. T. Wimberly, A. R. Faruqi, R. Henderson, B. Daum, A. Walter, A. Horst, H. D. Osiewacz, W. Kühlbrandt, F. K. Schur, and S. H. Scheres. Biochemistry. The resolution revolution. *Science (New York, N.Y.)*, 343(6178):1443–1444, 2014.

[79] Fujiwara Kunio. Relativistic Dynamical Theory of Electron Diffraction. *Journal of the Physical Society of Japan*, 16(11):2226–2238, 1961.

[80] Max A Lauffer and W C Price. Thermal denaturation of Tobacco mosaic virus. *Journal of Biological Chemistry*, 133:1–15, 1940.

[81] E. Lehmann and J. Romano. *Testing Statistical Hypotheses*. 2005.

[82] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R Booth, Michael B Braunfeld, Sander Gubbens, David A Agard, and Yifan Cheng. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods*, 10(6):584–90, 2013.

[83] Maofu Liao, Erhu Cao, David Julius, and Yifan Cheng. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107–112, 2013.

[84] Chengke Liu and Richard S. Nelson. The cell biology of Tobacco mosaic virus replication and movement. *Frontiers in Plant Science*, 4:12, 2013.

[85] Ningning Liu, Ying Chen, Bo Peng, Yuan Lin, Qian Wang, Zhaohui Su, Wenke Zhang, Hongbin Li, and Jiacong Shen. Single-molecule force spectroscopy study on the mechanism of RNA disassembly in tobacco mosaic virus. *Biophysical Journal*, 105(12):2790–2800, 2013.

[86] Gabriele Lohmann, Johannes Stelzer, Eric Lacosse, Vinod J. Kumar, Karsten Mueller, Esther Kuehn, Wolfgang Grodd, and Klaus Scheffler. LISA improves statistical analysis for fMRI. *Nature Communications*, 9:4104, 2018.

[87] Hubert S Loring, Yasuo Fujimoto, and Anthony T Tu. Tobacco mosaic virus—A calcium-magnesium coordination complex. *Virology*, 16(1):30–40, 1962.

[88] Jan Löwe, Daniela Stock, Bing Jap, Peter Zwickl, Wolfgang Baumeister, and Robert Huber. Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 Å resolution. *Science*, 268(5210):533–539, 1995.

[89] Bin Lu, Gerald Stubbs, and James N. Culver. Carboxylate interactions involved in the disassembly of tobacco mosaic tobamovirus. *Virology*, 225(1):11–20, 1996.

[90] Dmitry Lyumkis. Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry*, 294:5181–5197, 2019.

[91] Julia Mahamid, Stefan Pfeffer, Miroslava Schaffer, Elizabeth Villa, Radostin Danev, Luis Kuhn Cuellar, Friedrich Förster, Anthony A. Hyman, Jürgen M. Plitzko, and Wolfgang Baumeister. Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science*, 351(6276):969–972, 2016.

[92] Ruth Marcus, Peritz Eric, and K. R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976.

[93] Frank J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[94] David N. Mastronarde. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*, 152(1):36–51, 2005.

[95] Larry McDaniel, Marina Maratos, and Joan Farabaugh. Infection of Plants by Tobacco Mosaic Virus. *American Biology Teacher*, 60(6):434–439, 1998.

[96] G. McMullan, K. R. Vinothkumar, and R. Henderson. Thon rings from amorphous ice and implications of beam-induced Brownian motion in single particle electron cryo-microscopy. *Ultramicroscopy*, 158:26–32, 2015.

[97] O. Mishima, L. D. Calvert, and E. Whalley. 'Melting ice' i at 77 K and 10 kbar: A new method of making amorphous solids. *Nature*, 310:393–395, 1984.

[98] O. Mishima, L. D. Calvert, and E. Whalley. An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature*, 314:76–78, 1985.

[99] H. Müller, Jian Jin, R. Danev, J. Spence, H. Padmore, and R. M. Glaeser. Design of an electron microscope phase plate using a focused continuous-wave laser. *New Journal of Physics*, 12:073011, 2010.

[100] Keiichi Namba, Rekha Pattanayek, and Gerald Stubbs. Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction. *Journal of Molecular Biology*, 208(2):307–325, 1989.

[101] Keiichi Namba and Gerald Stubbs. Structure of tobacco mosaic virus at 3.6 Å resolution: Implications for assembly. *Science*, 231(4744):1401–1406, 1986.

[102] Kedar Narayan and Sriram Subramaniam. Focused ion beams in biology. *Nature Methods*, 12(11):1021–1031, 2015.

[103] K Naydenova, G McMullan, M J Peet, Y Lee, P C Edwards, S Chen, E Leahy, S Scotcher, R Henderson, and C J Russo. CryoEM at 100keV: a demonstration and prospects. *IUCrJ*, 6(6):1086–1098, nov 2019.

[104] J. Neyman and E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 231:289–337, 1933.

[105] Thi Hoang Duong Nguyen, Wojciech P. Galej, Xiao Chen Bai, Christos G. Savva, Andrew J. Newman, Sjors H.W. Scheres, and Kiyoshi Nagai. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, 523:47–52, 2015.

[106] Robert P.J. Nieuwenhuizen, Keith A. Lidke, Mark Bates, Daniela Leyton Puig, David Grünwald, Sjoerd Stallinga, and Bernd Rieger. Measuring image resolution in optical nanoscopy. *Nature Methods*, 10:557–562, 2013.

[107] Elena V. Orlova, Prakash Dube, J. Robin Harris, Erich Beckman, Friedrich Zemlin, Jürgen Markl, and Marin Van Heel. Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *Journal of Molecular Biology*, 271(3):417–437, 1997.

[108] Peter Palukaitis and Milton Zaitlin. *Tobacco Mosaic Virus Infectivity and Replication*, pages 105–131. Springer US, Boston, MA, 1986.

[109] Markus Pauly, Edgar Brunner, and Frank Konietschke. Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 77(2):461–473, 2015.

[110] Mathew J. Peet, Richard Henderson, and Christopher J. Russo. The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules. *Ultramicroscopy*, 203:125–131, 2019.

[111] Pawel Penczek, Michael Radermacher, and Joachim Frank. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, 40(1):33–53, 1992.

[112] Pawel A. Penczek. Resolution measures in molecular electron microscopy. In *Methods in Enzymology*, pages 73–100. 2010.

[113] Joana Pereira and Victor S. Lamzin. A distance geometry-based description and validation of protein main-chain conformation. *IUCrJ*, 4(5):657–670, 2017.

[114] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[115] C. Plaschka, L. Larivière, L. Wenzeck, M. Seizl, M. Hemann, D. Tegunov, E. V. Petrotchenko, C. H. Borchers, W. Baumeister, F. Herzog, E. Villa, and P. Cramer. Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, 518:376–380, 2015.

[116] Karl Popper. *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft.* Verlag Julius Springer, 1934.

[117] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14:290–296, 2017.

[118] C. Radhakrishna Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91, 1945.

[119] Ludwig Reimer and Helmut Kohl. *Transmission Electron Microscopy Physics of Image Formation.* 2008.

[120] William J. Rice, Anchi Cheng, Alex J. Noble, Edward T. Eng, Laura Y. Kim, Bridget Carragher, and Clinton S. Potter. Routine determination of ice thickness for cryo-EM grids. *Journal of Structural Biology*, 204(1):38–44, 2018.

[121] Joseph P. Romano, Azeem M. Shaikh, and Michael Wolf. Hypothesis Testing in Econometrics. *Annual Review of Economics*, 2010.

[122] Peter B. Rosenthal. Interpreting the cryo-EM map. *IUCrJ*, 6(1):3–4, 2019.

[123] Peter B. Rosenthal and Richard Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, 2003.

[124] Christopher J. Russo and Lori A. Passmore. Ultrastable gold substrates for electron cryomicroscopy. *Science*, 346(6215):1377–1380, 2014.

[125] Carsten Sachse, James Z. Chen, Pierre Damien Coureux, M. Elizabeth Stroupe, Marcus Fändrich, and Nikolaus Grigorieff. High-resolution Electron Microscopy of Helical Specimens: A Fresh Look at Tobacco Mosaic Virus. *Journal of Molecular Biology*, 371(3):812–835, 2007.

[126] Vincent Santilli, J. Piacitelli, and Jia Hsi Wu. The effect of tobacco mosaic virus protein on virus incubation period and infectivity. *Virology*, 14(1):109–123, 1961.

[127] W. O. Saxton and W. Baumeister. The correlation averaging of a regularly arranged bacterial cell envelope protein. *Journal of Microscopy*, 127(2):127–138, 1982.

[128] Sjors H.W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012.

[129] Sjors H.W. Scheres and Shaoxia Chen. Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, 9:853–854, 2012.

[130] Osip Schwartz, Jeremy J Axelrod, Sara L Campbell, Carter Turnbaugh, Robert M Glaeser, and Holger Müller. Laser phase plate for transmission electron microscopy. *Nature Methods*, 16(10):1016–1020, 2019.

[131] John G. Shaw. Tobacco mosaic virus and the study of early events in virus infections. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1383):603–611, 1999.

[132] Huigang Shi, Wuchun Ling, Dongjie Zhu, and Xinzheng Zhang. Increasing vitrification temperature improves the cryo-electron microscopy reconstruction. *bioRxiv*, 2019.

[133] F. J. Sigworth. A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*, 12(3):328–339, 1998.

[134] R. J. Simes. An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.

[135] Charles V. Sindelar and Nikolaus Grigorieff. Optimal noise reduction in 3D reconstructions of single particles using a volume-normalized filter. *Journal of Structural Biology*, 180(1):26–38, 2012.

[136] Boyuan Song, Julian Lenhart, Vanessa Judith Flegler, Cihan Makbul, Tim Rasmussen, and Bettina Böttcher. Capabilities of the Falcon III detector for single-particle structure determination. *Ultramicroscopy*, 203:145–154, 2019.

[137] John D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(3):479–498, 2002.

[138] John D. Storey. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, 31(6):2013–2035, 2003.

[139] Korbinian Strimmer. A unified approach to false discovery rate estimation. *BMC Bioinformatics*, 9(1):303, 2008.

[140] Gerald Stubbs. Tobacco mosaic virus particle structure and the initiation of disassembly. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1383):551–557, 1999.

[141] R Development Core Team and R R Development Core Team. R: A Language and Environment for Statistical Computing, 2019.

[142] Thomas C. Terwilliger, Oleg V. Sobolev, Pavel V. Afonine, and Paul D. Adams. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallographica Section D: Structural Biology*, 74(6):545–559, 2018.

[143] A Tsugita, D T Gish, J Young, H Fraenkel-Conrat, C A Knight, and W M Stanley. The complete amino acid sequence of the protein of tobacco mosaic virus. *Proceedings of the National Academy of Sciences of the United States of America*, 46(11):1463–1469, nov 1960.

[144] Chris A. Tulk, Jamie J. Molaison, Adam R. Makhluf, Craig E. Manning, and Dennis D. Klug. Absence of amorphous forms when ice is compressed at low temperature. *Nature*, 569:542–545, 2019.

[145] Beata Turoňová, Florian K.M. Schur, William Wan, and John A.G. Briggs. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *Journal of Structural Biology*, 199(3):187–195, 2017.

[146] Michael Unser, Benes L. Trus, and Alasdair C. Steven. A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy*, 23(1):39–51, 1987.

[147] Marin Van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of Structural Biology*, 151(3):250–262, 2005.

[148] Marin van Heel and Michael Schatz. Reassessing the Revolutions Resolutions. *bioRxiv*, 2017.

[149] Jose Luis Vilas, Josué Gómez-Blanco, Pablo Conesa, Roberto Melero, José Miguel de la Rosa-Trevín, Joaquin Otón, Jesús Cuenca, Roberto Marabini, José María Carazo, Javier Vargas, and Carlos Oscar S. Sorzano. MonoRes: Automatic and Accurate Estimation of Local Resolution for Electron Microscopy Maps. *Structure*, 26(2):337–344, 2018.

[150] Alexander Von Appen, Jan Kosinski, Lenore Sparks, Alessandro Ori, Amanda L. DiGuilio, Benjamin Vollmer, Marie Therese Mackmull, Niccolo Banterle, Luca Parca, Panagiotis Kastritis, Katarzyna Buczak, Shyamal Mosalaganti, Wim Hagen, Amparo Andres-Pons, Edward A. Lemke, Peer Bork, Wolfram Antonin, Joseph S. Glavy, Khanh Huy Bui, and Martin Beck. In situ structural analysis of the human nuclear pore complex. *Nature*, 526(7571):140–143, 2015.

[151] Miloš Vulović, Raimond B.G. Ravelli, Lucas J. van Vliet, Abraham J. Koster, Ivan Lazić, Uwe Lücken, Hans Rullgård, Ozan Öktem, and Bernd Rieger. Image formation modeling in cryo-electron microscopy. *Journal of Structural Biology*, 183(1):19–32, 2013.

[152] Miloš Vulović, Lenard M. Voortman, Lucas J. Van Vliet, and Bernd Rieger. When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM. *Ultramicroscopy*, 136:61–66, 2014.

[153] W. Wan and J. A.G. Briggs. Cryo-Electron Tomography and Subtomogram Averaging. In *Methods in Enzymology*, pages 329–367. 2016.

[154] Hong Wang, Antonio Planchart, and Gerald Stubbs. Caspar carboxylates: The structural basis of tobamovirus disassembly. *Biophysical Journal*, 74(1):633–638, 1998.

[155] Weiwei Wang and Roderick MacKinnon. Cryo-EM Structure of the Open Human Ether-à-go-go-Related K+ Channel hERG. *Cell*, 169(3):422–430, 2017.

[156] Felix Weis, Maximilian Beckers, Iris Hocht, and Carsten Sachse. Elucidation of the viral disassembly switch of tobacco mosaic virus. *EMBO reports*, 2019.

[157] Robert W. Wheatley, Douglas H. Juers, Bogdan B. Lev, Reuben E. Huber, and Sergei Yu Noskov. Elucidating factors important for monovalent cation selectivity in enzymes: E. coli β-galactosidase as a model. *Physical Chemistry Chemical Physics*, 17:10899–10909, 2015.

[158] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2011.

[159] A. J.C. Wilson. Determination of absolute from relative X-ray intensity data. *Nature*, 150:152, 1942.

[160] Daniel J. Wilson. The harmonic mean p -value for combining dependent tests. *Proceedings of the National Academy of Sciences*, 116(4):1195–1200, 2019.

[161] T. M.A. Wilson. Cotranslational disassembly of tobacco mosaic virus in vitro. *Virology*, 137(2):255–265, 1984.

[162] Guanghui Yang, Rui Zhou, Qiang Zhou, Xuefei Guo, Chuangye Yan, Meng Ke, Jianlin Lei, and Yigong Shi. Structural basis of Notch recognition by human γ-secretase. *Nature*, 565(7738):192–197, 2019.

[163] Koji Yonekura, Michael B. Braunfeld, Saori Maki-Yonekura, and David A. Agard. Electron energy filtering significantly improves amplitude contrast of frozen-hydrated protein at 300 kV. *Journal of Structural Biology*, 156(3):524–536, 2006.

[164] F. Zernike. Phase contrast, a new method for the microscopic observation of transparent objects. *Physica*, 9(7):686–698, 1942.

[165] Kai Zhang. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, 193(1):1–12, 2016.

[166] Martin J. Zhang, Fei Xia, and James Zou. Fast and covariate-adaptive method amplifies detection power in large-scale multiple hypothesis testing. *Nature Communications*, 10, 2019.

[167] Shawn Q Zheng, Eugene Palovcak, Jean-Paul Armache, Kliment A Verba, Yifan Cheng, and David A Agard. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4):331–332, 2017.

[168] Yong Zi Tan, Philip R. Baldwin, Joseph H. Davis, James R. Williamson, Clinton S. Potter, Bridget Carragher, and Dmitry Lyumkis. Addressing preferred specimen orientation in single-particle cryo-EMthrough tilting. *Nature Methods*, 14:793–796, 2017.

[169] Jasenko Zivanov, Takanori Nakane, Björn O. Forsberg, Dari Kimanius, Wim J.H. Hagen, Erik Lindahl, and Sjors H.W. Scheres. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, page e42166, 2018.

[170] Jasenko Zivanov, Takanori Nakane, and Sjors H. W. Scheres. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ*, 6(1):5–17, 2019.

Appendices

Appendix A

A Non-Parametric Permutation Approach for Adaptive Thresholding of Fourier Shell Correlation Curves

Table A.1: 1% FDR-FSC results on maps from the EMDB with reported resolutions better than 5 Å

Map	Symmetry	reported resolution [Å]	0.143 unmasked [Å]	1% FDR-FSC [Å]
EMD2677	C1	4.5	6.48	4.75
EMD3061	C1	3.4	4.06	3.41
EMD2984	D2	2.2	2.74	2.51
EMD5778	C4	3.3	3.99	3.5
EMD6422	D7	4.1	4.51	4.28
EMD6287	D7	2.8	3.07	2.83
EMD5623	D7	3.3	3.31	3.05
EMD6000	I	3.8	4.12	4.04
EMD2847	C1	2.9	3.3	2.96
EMD0144	O	1.65	1.89	1.7
EMD0408	C2	3.2	3.33	2.92
EMD6479	C1	3.5	3.57	3.23
EMD2764	C1	3.75	4.9	3.59
EMD0468	C1	3.9	6.17	4.52
EMD0480	C1	3.54	4.31	4.13
EMD9384	C2	2.96	3.49	3.36
EMD0043	C1	3.7	3.86	3.6
EMD0501	C1	3.5	4.17	3.47
EMD0500	C1	3.4	4.03	3.38
EMD0489	C4	4.3	5.03	4.25
EMD0487	C4	4.0	4.25	3.89
EMD9333	C1	3.0	3.43	2.98

EMD0488	C4	3.4	3.74	3.46
EMD0415	C1	3.1	3.84	3.02
EMD0416	C1	3.6	3.96	3.59
EMD0417	C1	3.7	4.13	3.73
EMD0418	C1	3.8	4.52	3.96
EMD0419	C1	3.3	4.18	3.2
EMD0420	C1	3.4	4.32	3.23
EMD0425	C1	3.5	6.31	3.42
EMD4587	C2	3.8	4.26	3.86
EMD4588	C2	3.6	3.98	3.59
EMD4589	C2	3.7	4.12	3.68
EMD4592	C2	3.6	4.12	3.63
EMD4593	C2	3.7	4.26	3.52
EMD4594	C2	3.6	4.19	3.74
EMD4611	C2	3.2	3.53	3.18
EMD4612	C2	3.8	3.91	3.43
EMD4613	C2	3.5	3.75	3.36
EMD4614	C2	3.3	3.53	2.93
EMD9599	O	1.62	1.77	1.68
EMD9012	I	1.86	1.92	1.83
EMD0153	D2	1.89	2.29	1.91
EMD9203	D3	2.1	2.33	2.04
EMD7541	D2	2.4	3	2.77
EMD8908	D2	2.2	2.57	2.26
EMD4129	C1	3.06	3.75	2.95
EMD0194	C1	3.8	4.33	3.45
EMD0407	C2	2.8	3.66	3.25
EMD0452	C1	3.7	4.27	3.59
EMD0587	C1	4.1	4.83	3.98
EMD0588	C1	4.3	6.01	4.27
EMD0589	C1	3.9	4.46	3.68
EMD0590	C2	3.13	3.63	3.19
EMD0591	C2	3.37	3.98	3.4
EMD0592	C2	3.15	3.77	3.19
EMD9241	C1	3.8	4.47	3.6
EMD0182	I	3.36	3.51	3.3
EMD0596	C1	3.08	3.69	2.75
EMD0183	I	3.18	3.47	3.26
EMD7335	C1	3.5	4.17	3.51
EMD7334	C1	3.9	4.4	3.9

EMD7337	C1	4.6	5.99	4.47
EMD7770	D2	1.9	2.13	1.84
EMD0264	C5	4.6	6.46	4.46
EMD9013	C1	3.4	4.04	3.61
EMD9014	C1	3.3	3.49	3.11
EMD8731	C3	4.2	5.41	4.3
EMD0341	C2	3.6	4.75	3.96
EMD0342	C2	4.2	6.17	4.28
EMD9696	C1	3.76	4.17	3.72
EMD9695	C1	3.64	4.07	3.54
EMD0498	C6	2.7	2.85	2.45
EMD0499	C6	2.7	2.81	2.56
EMD20074	C3	4.1	4.46	4.08
EMD20101	C1	4.36	4.81	3.93
EMD20100	C1	3.28	3.67	3.17

Table A.2: 1% FDR-FSC results on maps from the EMDB with reported resolutions worse than 5 Å

Map	Symmetry	reported resolution [Å]	0.143 unmasked [Å]	1% FDR-FSC [Å]
EMD9625	C1	6.78	7.71	6.48
EMD9306	C1	7.5	9.39	8.05
EMD9779	C1	16	16.87	11.03
EMD4748	C2	5.14	6.07	5.0
EMD0074	C1	18	31.44	21.68
EMD0075	C1	21	31.44	21.68
EMD0086	C1	14	20.96	13.1
EMD0087	C1	16	32.25	23.29

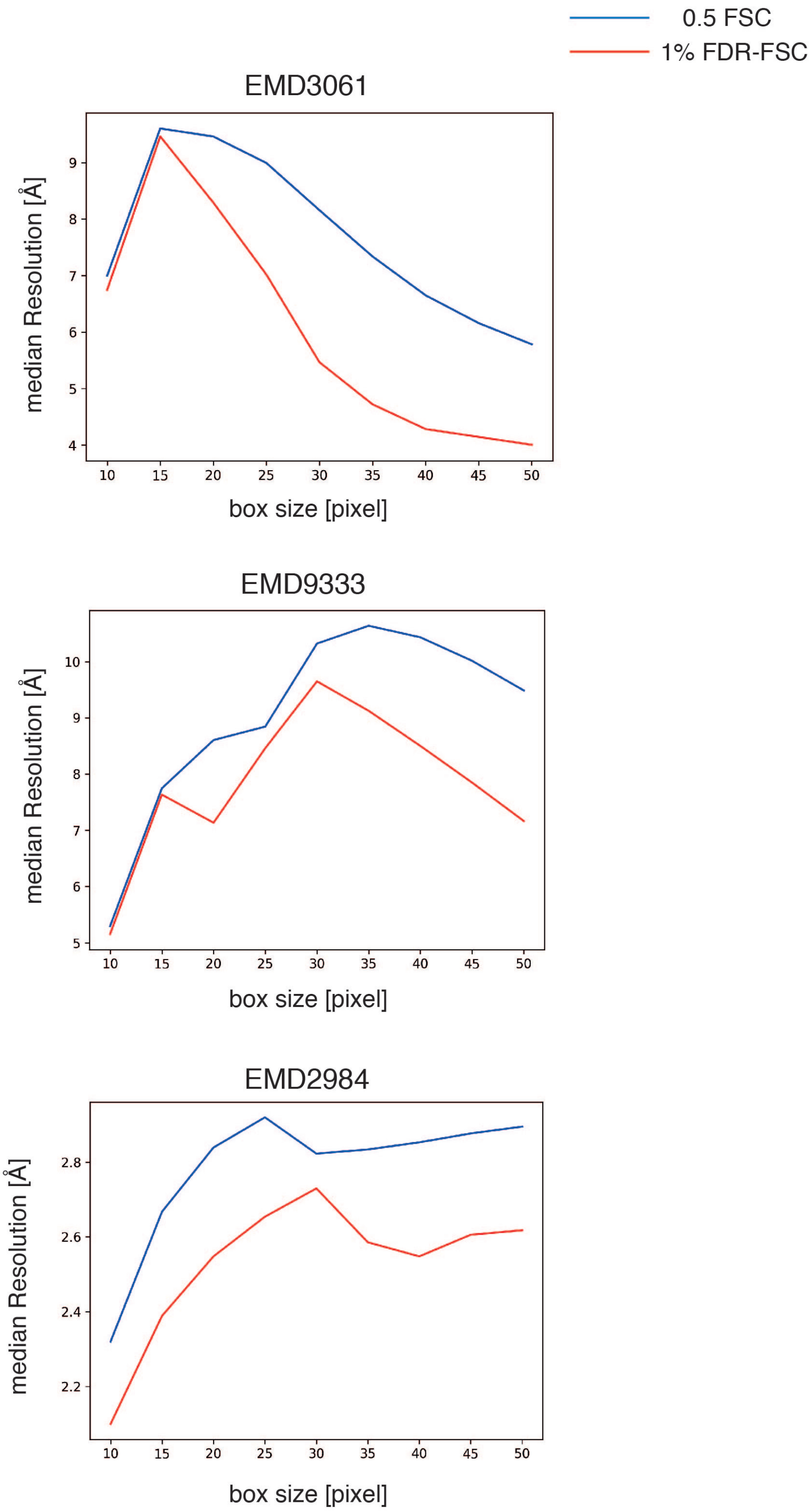


Figure A.1: Influence of the box size on local resolution estimates. The box size of the sliding window is plotted against the respective median local resolutions over the complete map for both the 0.5 FSC (blue) and 1% FDR-FSC (red).

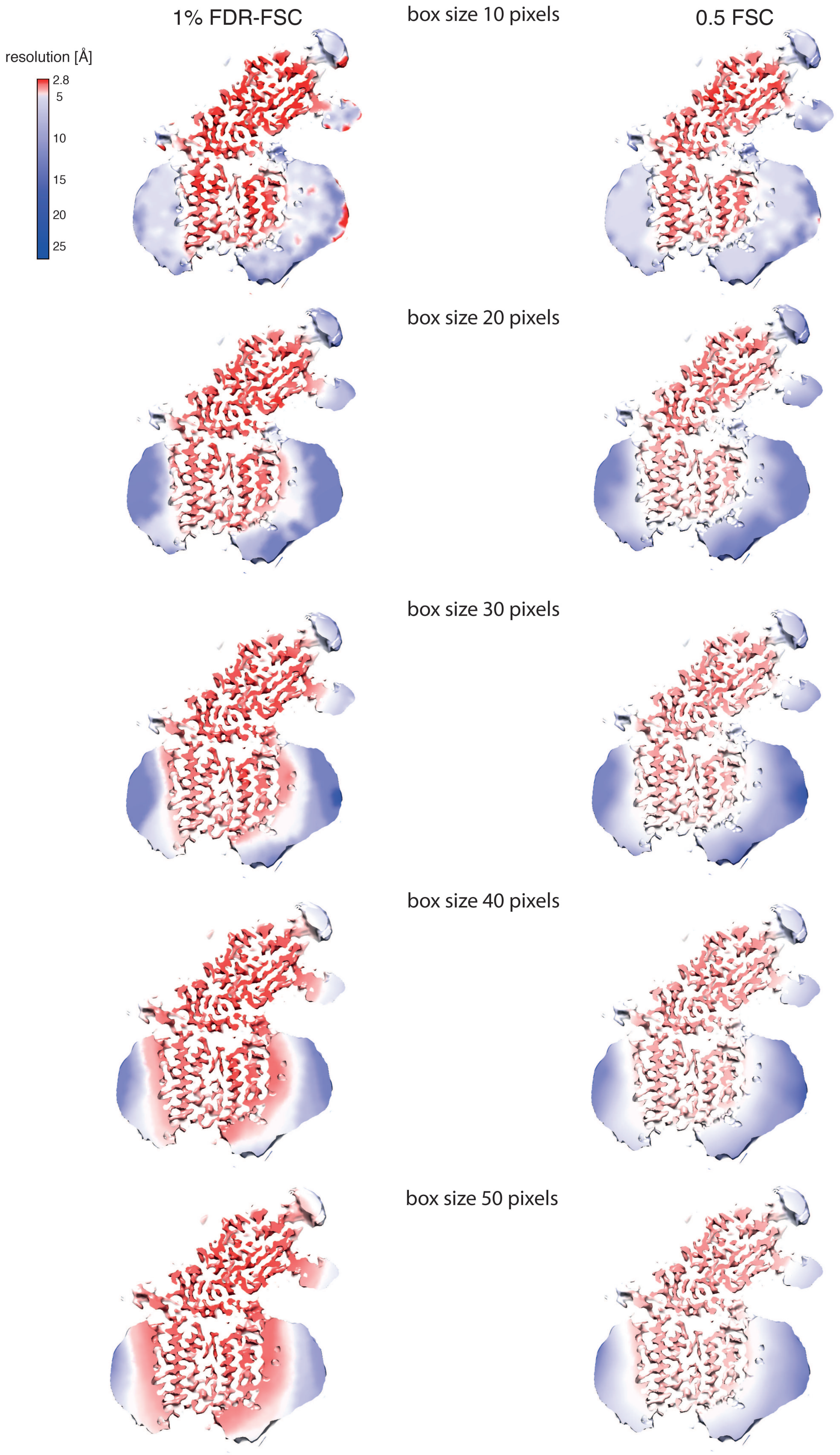


Figure A.2: Influence of the box size on local resolutions for the cryo-EM map EMD3061. The local resolutions are mapped on the surface in the respective colours for both 0.5 FSC thresholding and 1% FDR-FSC.

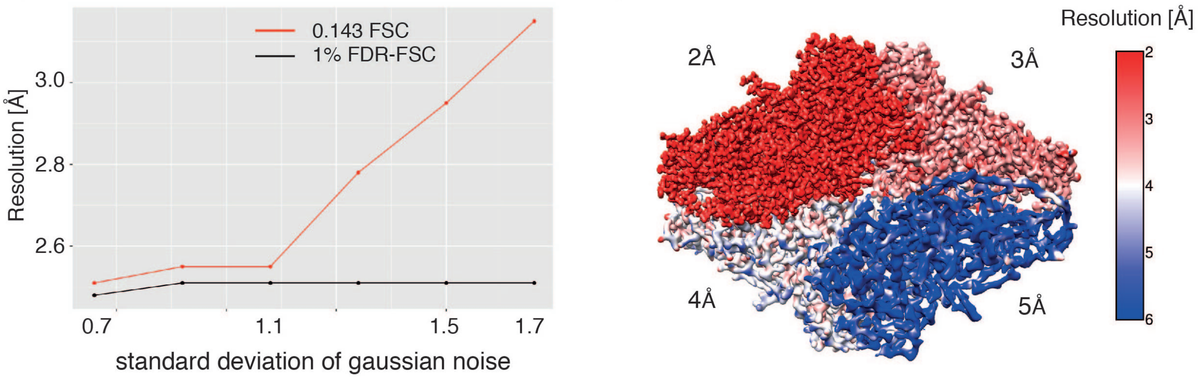


Figure A.3: (left) Resolution estimates of both 0.143 FSC and 1% FDR-FSC thresholds for a simulated map of β-galactosidase (pdb 5a1a) at 2.5 Å resolutions and with different levels of background noise. (right) Local resolution estimates of the 1% FDR-FSC threshold for a simulated map of β-galactosidase (pdb 5a1a). The 4 subunits were filtered to different resolutions of 2,3,4 and 5 Å, respectively.

Appendix B

Elucidation of the Viral Disassembly Switch of Tobacco Mosaic Virus

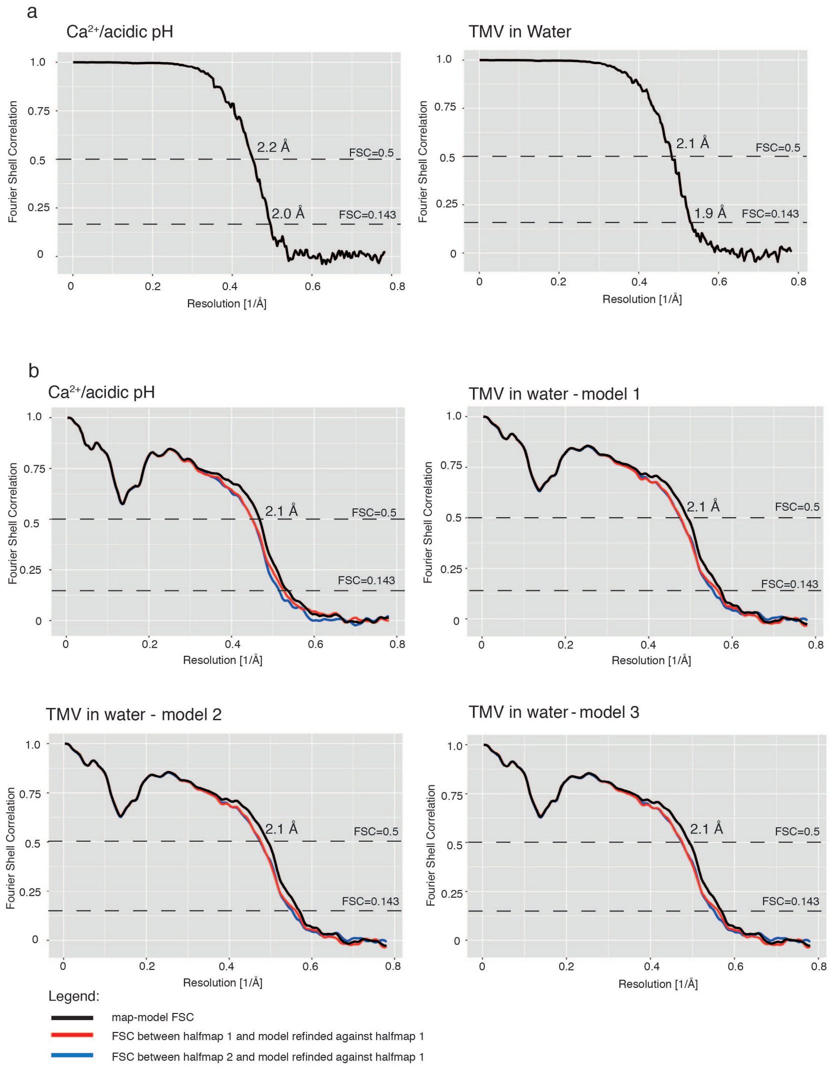


Figure B.1: Resolution assessment using Fourier shell correlation (FSC). (a) Comparison of FSC curves between two half-maps for the Ca^{2+}/acidic pH (left) and water structure (right). (b) FSC curves between map and model (black), between half map 1 and a perturbed model refined against half map 1 (red) as well as between half map 2 and a perturbed model refined against half map 1 (blue) for the 4 determined atomic models. This figure is taken from [156] and was prepared by myself.

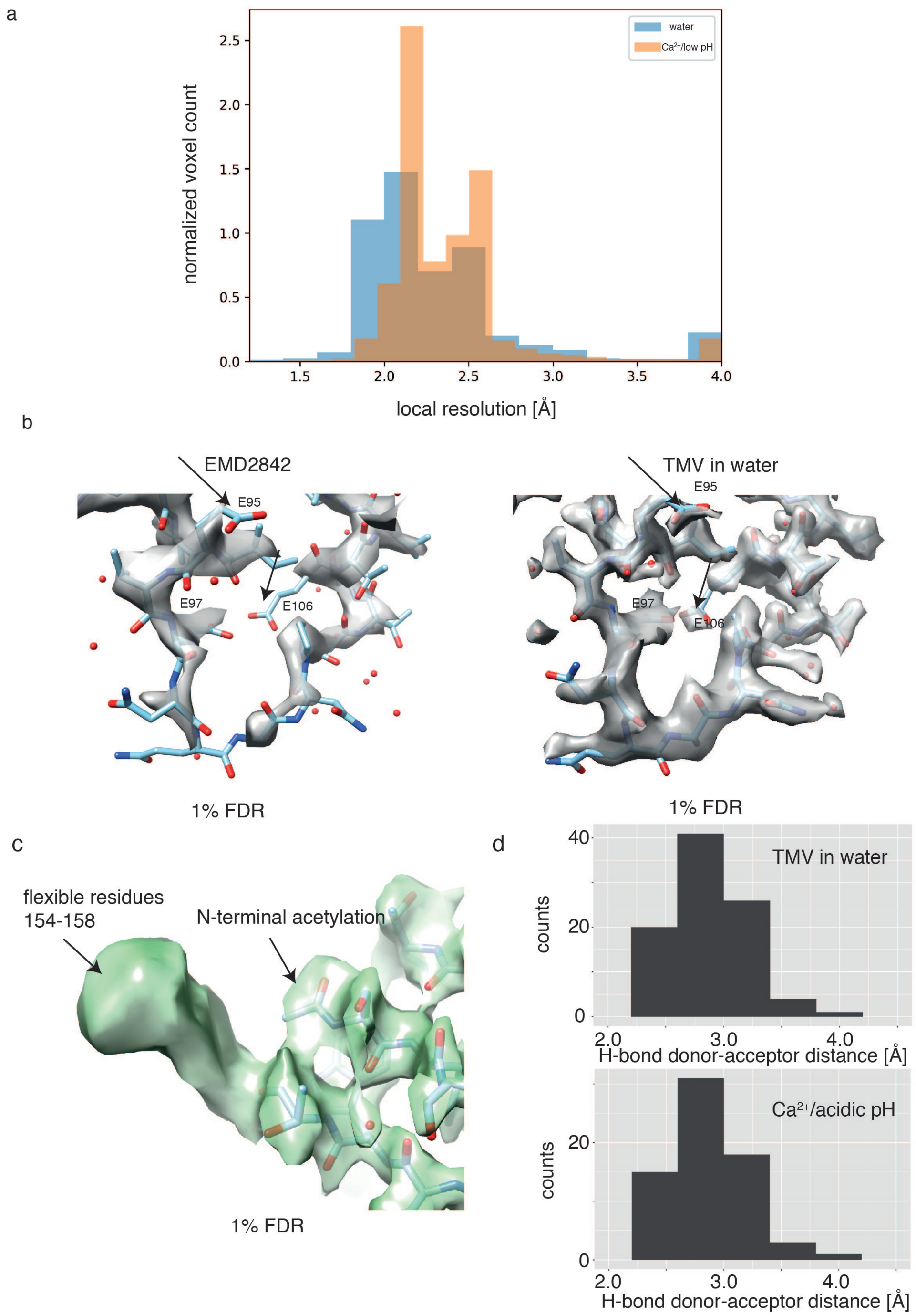


Figure B.2: Local resolution assessment of Ca^{2+}/acidic pH and water structure. (a) Overlay of local resolution histograms computed with BlocRes in the Ca^{2+}/acidic pH (orange) and water condition (blue). Resolution of the water map is slightly higher. (b) Map comparison from previous study (left) [16] with this study in water (right) including overlaid current atomic model thresholded at a FDR of 1%. The here determined structure shows additional significant and defined map features for the lower radius region. This figure is taken from [156] and was prepared by myself.

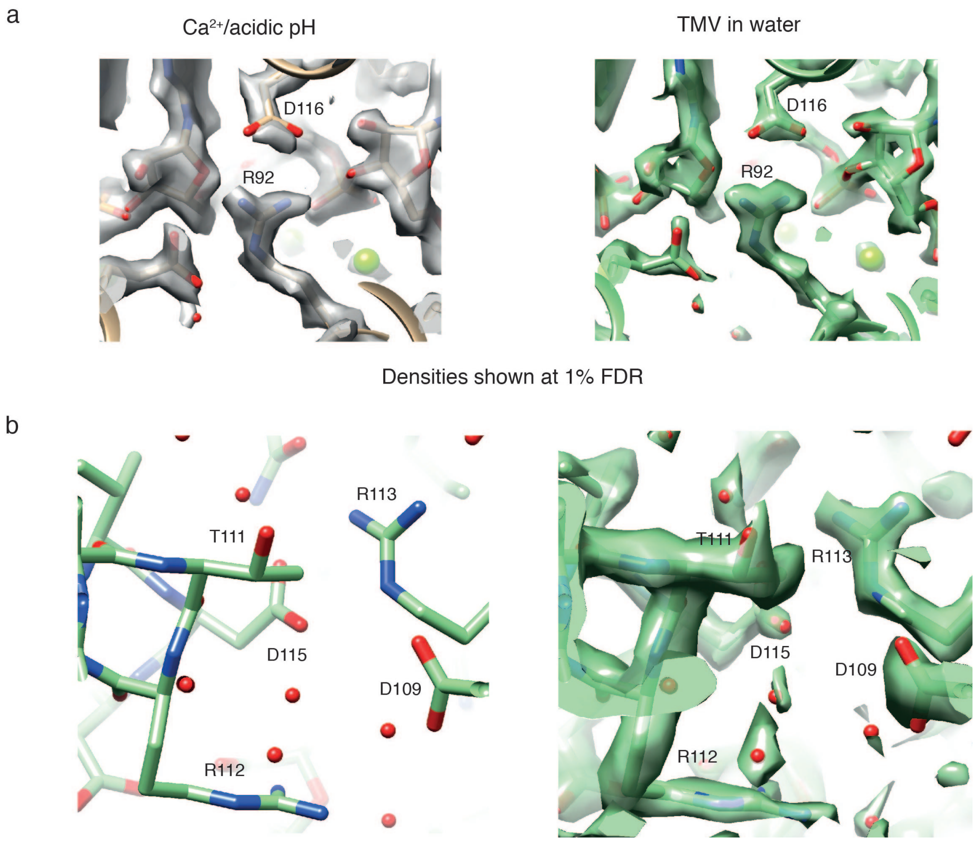


Figure B.3: Structural details of Ca^{2+}/acidic pH and water states. (a) Atomic models shown with the respective maps of TMV in Ca^{2+}/acidic pH (left) and water condition (right) at the previously proposed location of a second Ca^{2+} site in proximity to the RNA. No compatible Ca^{2+} ion map features could be detected. (b) Residue Asp109 and its environment: no obvious interaction with other carboxylates is evident in our structures (left). The same view is shown with corresponding map at 1% FDR (right). This figure is taken from [156] and was prepared by myself.

Table B.1: Model validation statistics. Model statistics for the Ca^{2+}/acidic pH atomic model and the 3 different models from TMV in water.

Model quantity	Ca^{2+}/acidic pH	Model 1	Model 2	Model 3
Ramachandran outliers	0.00%	0.00%	0.00%	0.00%
Ramachandran favored	97.35%	97.35%	96.69%	96.69%
Rotamer outliers	0.00%	0.00%	0.00%	0.00%
Clashscore	1.6	3.6	4.4	4.4
RMS(bonds)	0.0037	0.0051	0.0056	0.0052
RMS(angles)	0.63	0.76	0.82	0.82
C-beta deviations	0	0	0	1
MolProbity score	1.03	1.28	1.43	1.43
DipCheck Chi-score	-0.59	-0.61	-0.99	-0.67
RSCC(mask)	0.86	0.86	0.86	0.86

List of Tables

List of Figures