

From Tomograms to molecular Structure

IMAGE PROCESSING IN CRYO-ELECTRON TOMOGRAPHY

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Physik

der Goethe-Universität Frankfurt

in Frankfurt am Main

von

Zhou Yu

aus Beijing, China

Frankfurt am Main 2014

D30

vom Fachbereich Physik der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Joachim Stroth

Gutachter: Prof. Dr. Achilleas Frangakis

Prof. Dr. Jochen Triesch

Datum der Disputation: 25. März 2015

Zusammenfassung

Kryoelektronentomographie (engl. Cryo-Electron Tomography, Abk. CET) ist die einzige Technik, die es erlaubt biologische Objekte in ihrem quasi natürlichen Zustand, atomar aufzulösen. Die biologischen Objekte werden vor dem CET-Prozess mittels Flüssigstickstoff fixiert, wodurch das enthaltene Wasser in einen glasartigen (vitrifizierten) Zustand überführt wird. Der gesamte CET-Prozess findet unterhalb der Devitrifikationstemperatur statt. Dadurch können die Proben ohne Anwendung von chemischen Fixiermitteln und ohne Dehydration unter den Hochvakuumbedingungen des Mikroskops analysiert werden. In kryofixierten biologischen Proben entstehen weniger Fixationsartefakte und der natürliche Zustand der Probe wird für die Aufnahme besser erhalten.

Die Elektronen werden auf 200~300 keV beschleunigt und durchstrahlen die Probe. Die Interaktionen der Elektronen mit der Probe verursachen Elektronenstreuung. Die gestreuten Elektronen werden durch einen Detektor aufgefangen und in ein Bild umgewandelt, welches eine Projektion der Probe darstellt. Bei Tomographie wird die Probe physikalisch um eine definierte Achse gekippt und eine sog. Kippserie aufgenommen, die aus den Projektionen der verschiedenen Winkel besteht. Die Kippserie wird anschließend mit einer Rekonstruktionsmethode, wie z.B. gewichteter Rückprojektion (engl. Weighted Back Projection, Abk. WBP), algebraischer Rekonstruktionstechnik (Abk. ART) sowie ihrer Variante SART/SIRT, zu einem Tomogramm rekonstruiert. Aufgrund der Geometrie des Halters ist der mögliche Kippwinkel im Mikroskop beschränkt. Daraus folgt ein nicht abgetasteter Wedge-förmiger Bereich im Fourierraum des Tomogramms. Dieser Effekt wird als "Missing-Wedge" bezeichnet.

Aufgrund der beschränkten Stabilität kryofixierter biologischer Proben gegenüber hochbeschleunigten Elektronen muss die kumulative Dosis der Elektronen während des gesamten Tomographievorgangs unterhalb einer vorgegebenen proben-abhängigen Grenze liegen. Die Bildgebung mit niedriger

Zusammenfassung

Dosis hat Projektionen mit einem schlechten Signal-Rausch-Verhältnis zur Folge, was die erreichbare Auflösung auf ca. 5 Nanometer beschränkt. Um eine höhere Auflösung zu erreichen, werden mehrfach vorkommende biologische Strukturen in Tomogrammen gemittelt. Dafür kommen automatische Berechnungsmethoden wie z.B. „Template Matching“, „Sub-Tomogram Averaging“ und Klassifizierungen zum Einsatz.

Template Matching ist eine auf Mustererkennung basierende Technik zur Identifizierung von Makromolekülen in Tomogrammen. Diese Technik wurde zuerst durch [1] in die Strukturbiologie eingebracht. Sind die 3D Strukturen von Interesse bereits durch Röntgenstrukturanalytik oder Einzelpartikelanalytik bekannt, können sie als Referenzstruktur für eine systematische Suche im Tomogramm benutzt werden. Zuerst wird eine Referenz aus der Struktur erstellt, welche eine ähnliche Auflösung wie die Struktur in den Tomogrammen hat. Der Grauwert eines Voxels ergibt sich aus der Anzahl aller Atome, die sich innerhalb dieses Voxels befinden. Diese Referenz wird dann durch die Kontrastübertragungsfunktion (engl. Contrast Transfer Function, Abk. CTF) des Elektronenmikroskops gefiltert, um den Vorgang der Tomographie zu simulieren. Eine anschließende Tiefpassfilterung der Referenz ergibt eine drei-dimensionale Struktur, welche dem zu suchenden Makromolekül im Tomogramm sehr ähnlich ist. Danach wird diese erstellte Referenz in jede mögliche Richtung rotiert und mit dem Tomogramm kreuz-korreliert, wobei nur die Bereiche mit Informationen in Fourierraum berücksichtigt werden, um den Missing-Wedge Effekt zu kompensieren [2]. Aus dieser Suche ergibt sich eine Kreuz-Korrelationsfunktion, worin jedes echte Maximum einen mutmaßlichen Treffer repräsentiert. Der gesamte Vorgang erfolgt vollautomatisch, wobei kein Eingreifen vom Benutzer erforderlich ist. Die Ergebnisse sind daher komplett reproduzierbar.

Das Template Matching kann zur Identifizierung von bekannten Makromolekülen in Tomogrammen und zum Bestimmen von ihren relativen Orientierungen sowie ihrer räumlichen Interaktion eingesetzt werden. Wenn das Template Matching aufgrund unbekannter Struktur der Makromoleküle, oder eines nicht ausreichenden Signal-Rausch-Verhältnis nicht eingesetzt werden

kann, müssen die Makromoleküle aus den Tomogrammen manuell selektiert werden. Um ihre relativen Orientierungen zu bestimmen und anschließend mitteln zu können, wird die Methode des sog. Sub-Tomogram Averaging eingesetzt. Das Sub-Tomogram Averaging ist eine Referenz-basierte iterative Berechnungsmethode. Zuerst wird hier eine Referenz als Startpunkt erstellt. Dafür kann sowohl die Kristallstruktur als auch eine Kugel ohne jegliche Merkmale verwendet werden. Letztere Strategie wird „Referenzfrei“ genannt und hat den Vorteil dass kein bzw. sehr wenig Referenzbias eingeführt wird [3, 4]. Die Referenz wird dann rotiert und mit jedem Sub-Tomogramm kreuz-korreliert, um seine Orientierungen und Verschiebungen im Vergleich zur Referenz zu bestimmen, wobei der Missing-Wedge Effekt berücksichtigt wird. Die Summe aller ausgerichteten Sub-Tomogramme wird als Referenz für die nächste Iteration angewendet. Der gesamte Vorgang läuft iterativ weiter bis zur Konvergenz. Durch Sub-Tomogram Averaging wird das Signal-Rausch-Verhältnis erhöht, womit die Auflösung verbessert wird. Die Mittelungen erreichen auch isotrope Auflösung, da der Missing-Wedge Effekt durch Mittelung von zahlreichen Sub-Tomogrammen unterschiedlicher Orientierungen überwunden werden kann.

Mit der Weiterentwicklung der CET werden auch Template Matching Techniken immer häufiger eingesetzt. Dadurch fällt das Augenmerk zunehmend auf den Referenzbias, welcher durch referenzbasierte Techniken eingebracht wird, und welcher unsichere und falsche Detektion von biologischen Makromolekülen verursacht. Der Grund dafür liegt in der Zuweisung von hohen Kreuz-Korrelationswerten an die nicht gesuchten Objekte. Aufgrund niedrigerer Verschiedenartigkeit von Daten hat die Sub-Tomogram Averaging Technik deutlich weniger Probleme mit Referenzbias. Aber dennoch spielt der Bias-Effekt eine wichtige Rolle bei Sub-Tomogram Averaging.

Studien zeigen die Existenz von Biaseffekten in der Einzelpartikelanalytik, wo es möglich ist, dass hoch aufgelöste Strukturen sogar aus Phantomen gewonnen werden können [5]. In [6] wurde der Referenzbias durch Elimination von Daten bestimmter Frequenzen anhand der unterschiedlichen Verhältnisse der Fourier Shell Korrelation (FSC) analysiert. Um das Problem von Bias zu überwinden

Zusammenfassung

wurden zahlreiche Methoden wie z.B. die „Gold Standard“ [6, 7] und die Substitution von Rauschen hoher Frequenzen [8] zur Minimierung oder Kompensierung des Effekts entwickelt und eingesetzt.

Im Vergleich zur Einzelpartikelanalytik könnte der Referenzbias aufgrund höherer Freiheitsgrade beim Suchen und niedrigerer Verschiedenartigkeit von Daten in der CET noch prominenter sein. Insbesondere beim Template Matching, wo der Referenzbias schon in niedrigen Frequenzen eingebracht werden kann, kann die Technik leider nicht immer bedenkenlos eingesetzt werden. Deswegen ist die Einführung eines Qualitätsmerkmals für einen erfolgreichen Einsatz der Technik Voraussetzung. Dieser muss sowohl die Qualität der Berechnung als auch die Menge der Fehldetektionen erfassen. Die bisher entwickelten Methoden in Fourierraum sind schwierig für den Einsatz in Tomographie anzupassen.

Zu diesem Zweck wurde eine neuartige Methode mit dem Namen „M-free“ zur Einschätzung des Bias entwickelt. Diese Methode basiert auf einer Analyse im Realraum und kann für Template Matching und Sub-Tomogram Averaging Techniken eingesetzt werden. Der Kern der Methode ist das Unterteilen der benutzten Maske im Realraum in einen Arbeits- und einen Testbereich (engl. *Working area* und *Testing area*). Es werden ausschließlich die Informationen in der *Working area* für die Ausrichtungsvorgänge verwendet. Die *Testing area* bleibt während des gesamten Vorgangs unberührt. Nach dem Vorgang werden Informationen aus beiden Bereichen für die Ermittlung der „M-free“ Werte genutzt.

Die Theorie des „M-free“ basiert auf der Kohärenz von echten Signalen im Realraum [9]. D.h. die Werte verschiedener Voxel sind bei echten Signalen gebunden. Der Referenzbias dagegen, welcher durch Rauschen verursacht wird, hat keine Kohärenz. Das hat zur Folge, dass der Signalanteil in der *Working area* automatisch den Anteil in der *Testing area* ausrichtet. Die Ausrichtung des Rauschens in der *Working area* hat im Gegensatz dazu keinerlei Einfluss auf die *Testing area*. Die Messung des „M-free“ repräsentiert den Unterschied zwischen den gemessenen Signalen in beiden Bereichen. In der vorliegenden Arbeit wurde

mathematisch und experimentell bewiesen, dass der „M-free“ eine präzise Einschätzung von Referenzbias erlaubt.

Das sehr starke Rauschen innerhalb der Tomogramme verursacht eine hohe Verschiedenartigkeit der Daten und dadurch Referenzbias im Template Matching und Sub-Tomogram Averaging Prozess. Die praktische Einsetzbarkeit dieser Techniken ist hierdurch nicht immer gewährleistet. Um die Verschiedenartigkeit der Daten zu reduzieren kommen Klassifikationsalgorithmen zum Einsatz. Die Klassifikation von tomographischen Daten ist aufgrund des niedrigen Signal-Rausch-Verhältnis und des Missing-Wedge Effekts [2] eine herausfordernde Aufgabe.

Selbst-organisierende neuronale Netze (engl. Self-Organizing Map, Abk. SOM) sind eine Art von neuronalen Netzwerken, welche hoch-dimensionale Daten auf ein zwei-dimensionales Gitter abbilden. Jeder Knoten des Gitters wird ein „Codevector“ genannt [10]. Die Abbildung erfolgt nach dem Prinzip, dass die Daten auf benachbarten Codevektoren ähnlicher sind, als die auf entfernten Codevektoren. Die Codevektoren werden durch die durchfließenden Daten aktualisiert.

SOMs sind Klassifikationsmethoden, die keine Vorkenntnisse von Daten erfordern. Sie fassen die originalen, verrauschten Daten in guten „Repräsentanten“ zusammen, dadurch wird die Aufgabe der Klassifikation erleichtert, allerdings ohne den Verlust der Datenstruktur [11]. Die originale Kohonen'sche SOM wurde 1994 von Carazo für zwei-dimensionale Daten in der Strukturbiologie angewendet [12]. Eine erweiterte Variante bekannt als KerDenSOM basierend auf der Schätzung von Wahrscheinlichkeitsdichten wurde für Klassifikation zwei- und drei-dimensionaler Daten eingeführt [11, 13]. Ein wesentlicher Vorteil dieser Variante liegt in einer sicheren Konvergenz auf das globale Maximum, womit dieser Algorithmus weniger abhängig von den jeweiligen Startbedingungen ist. In [14] wurde der KerDenSOM durch eine Strategie der Kompensation des Missing-Wedge Effekts für drei-dimensionale Daten erweitert. Diese fortgeschrittene Variante wird KerDenSOM3D genannt und zeigte eine hervorragende Leistung in Experimenten mit synthetischen und

Zusammenfassung

realen Daten. In einem umfassenden Vergleich mit anderen Algorithmen sticht sie durch besondere Qualität hervor.

Abstract

Cryo-electron tomography (CET) is a unique technique to visualize biological objects under near-to-native conditions at near-atomic resolution. CET provides three-dimensional (3D) snapshots of the cellular proteome, in which the spatial relations between macromolecular complexes in their near native cellular context can be explored. Due to the limitation of the electron dose applicable on biological samples, the achievable resolution of a tomogram is restricted to a few nanometers, higher resolution can be achieved by averaging of structures occurring in multiples. For this purpose, computational techniques such as template matching, sub-tomogram averaging and classification are essential for a meaningful processing of CET data.

This thesis introduces the techniques of template matching and sub-tomogram averaging and their applications on real biological data sets. Subsequently, the problem of reference bias, which restricts the applicability of those techniques, is addressed. Two methods that estimate the reference bias in Fourier and real space are demonstrated. The real space method, which we have named the “M-free” score, provides a reliable estimation of the reference bias, which gives access to the reliability of the template matching or sub-tomogram averaging process. Thus, the “M-free” score makes those approaches more applicable to structural biology. Furthermore, a classification algorithm based on Neural Networks (NN) called “KerDenSOM3D” is introduced, which is implemented in 3D and compensates for the missing-wedge. This approach helps extracting different structural states of macromolecular complexes or increasing the class purity of data sets by eliminating outliers. A comprehensive comparison with other classification methods shows superior performance of KerDenSOM3D.

List of publications

- I. **Yu Z.**, Frangakis AS. "Classification of electron sub-tomograms with neural networks and its application to template matching", *J Struct Biol.* 2011 Jun;174(3):494-504
- II. **Yu Z.**, Frangakis AS. "M-free: Scoring the reference bias in sub-tomogram averaging and template matching", *J Struct Biol.* 2014; Jul;187(1):10-19.

Supporting publications

- II. Perkovic M., Kunz. M., Endesfelder U., Bunse S., Wigge C., **Yu Z.**, Hodoranau VV., Scheffer M., Seybert A., Malkusch S., Schuman EM., Heilemann M., Frangakis AS. "Correlative Light- and Electron Microscopy with Chemical Tags", *J Struct Biol.* 2014 May;186(2):205-213.
- IV. Eltsov M., Dubé N., **Yu Z.**, Haselmann-Weiss U., Brunner D., Frangakis AS. "Large-volume electron tomography reveals the mechanism of epithelial tissue sealing", *Submitted manuscript*

Important note

The results of the experiments and the description of the algorithms were published by the author of this thesis in the following articles:

(2011) Z. Yu & A. S. Frangakis, Classification of electron sub-tomograms with neural networks and its application to template-matching.

(2014) Z. Yu & A. S. Frangakis, M-free: Scoring the reference bias in sub-tomogram averaging and template matching.

(2014) M. Perkovic et. al., Correlative Light- and Electron Microscopy with chemical tags.

The dissertation is a more detailed version of the publications, containing the same results. Thus, throughout this work, original figures, formulations and excerpts from the publications are adopted word-by-word. For reasons of clarity, and in order to enhance the flow of reading the used passages are not put in quotation marks and sentences are rearranged. In physics this is recognized as practice of quotation and was explicitly approved by the board of examiners (Promotionsausschuss) for this thesis. The publisher of the above cited articles (Journal of Structural Biology) allows the usage of the publications or excerpts of the publications in the author's dissertation.

Table of contents

| | |
|--|----|
| Zusammenfassung..... | 1 |
| Abstract..... | 7 |
| List of publications..... | 9 |
| Supporting publications..... | 9 |
| Important note..... | 10 |
| Table of contents..... | 11 |
| List of abbreviations..... | 13 |
| 1. Introduction..... | 15 |
| 1.1. Cryo-electron tomography..... | 15 |
| 1.2. Data acquisition..... | 18 |
| 1.3. Reconstruction..... | 21 |
| 2. Methods..... | 25 |
| 2.1. Similarity measurement with cross-correlation..... | 25 |
| 2.2. Resolution estimation with Fourier shell correlation..... | 28 |
| 2.3. Template matching..... | 29 |
| 2.4. Sub-tomogram averaging..... | 31 |
| 3. Classification of tomographic data sets..... | 35 |
| 3.1. Classification with self-organizing maps..... | 36 |
| 3.2. Theory of KerDenSOM3D..... | 37 |
| 3.3. Choice of the smoothness parameters..... | 41 |
| 3.4. Data acquisition conditions..... | 42 |
| 3.5. Results..... | 43 |
| 3.5.1. Simulated data sets..... | 43 |
| 3.5.2. Comprehensive performance comparison to other algorithms..... | 47 |

Table of contents

| | |
|---|-----|
| 3.5.3. <i>In vitro</i> GroEL and GroEL/ES data sets..... | 51 |
| 4. Reference bias in reference based techniques | 59 |
| 4.1. Estimation of reference bias in Fourier space | 60 |
| 4.2. Estimation of reference bias in real space: M-free score | 61 |
| 4.3. Detailed derivation of M-free score | 64 |
| 4.4. Results | 68 |
| 4.4.1. Simulated data sets | 69 |
| 4.4.2. <i>In vitro</i> GroEL data sets | 73 |
| 4.4.3. Sub-tomogram averaging of SIV Spikes | 79 |
| 4.4.4. Template matching of whole Mycoplasma cells..... | 82 |
| 5. Conclusion | 87 |
| 6. References | 89 |
| A. Appendix..... | 101 |
| A.1. PAPER I..... | 101 |
| A.2. PAPER II | 113 |
| A.3. PAPER III..... | 124 |
| A.4. PAPER IV | 134 |
| Acknowledgements..... | 135 |
| Curriculum vitae..... | 137 |

List of abbreviations

| | |
|-----------|--|
| 1D | one-dimensional |
| 2D | two-dimensional |
| 3D | three-dimensional |
| ART | Algebraic reconstruction technique |
| CCC | Constraint cross-correlation |
| CCD | Charge-coupled device |
| CCF | Cross-correlation function |
| CET | Cryo-electron tomography |
| CTF | Contrast transfer function |
| DDD | Direct detection device |
| DOF | Degree of freedom |
| DQE | Detection quantum efficiency |
| FEG | Field emission gun |
| FRC | Fourier ring correlation |
| FSC | Fourier shell correlation |
| INFR | Iterative non-uniform fast Fourier transform based reconstruction method |
| KerDenSOM | Kernel density estimator self-organizing map |

List of abbreviations

| | |
|-------------|---|
| KerDenSOM3D | Kernel density estimator self-organizing map with 3D missing-wedge compensation |
| KSOM | Kohonen self-organizing map |
| MCC | Maximum cross-correlation |
| MLTOMO | Maximum likelihood for tomography |
| MTF | Modulation transfer function |
| NN | Neural network |
| NUFFT | Non-uniform fast Fourier transform |
| PCA | Principal component analysis |
| PDB | Protein data bank |
| PDF | Probability density function |
| PSF | Point-spread function |
| SART | Simultaneous algebraic reconstruction technique |
| SIRT | Simultaneous iterative reconstruction technique |
| SNR | Signal-to-noise ratio |
| SOM | Self-organizing map |
| SSNR | Spatial signal-to-noise ratio |
| TEM | Transmission electron microscope |
| WBP | Weighted back-projection |

1. Introduction

This thesis introduces computational methods for 3D structure determination of macromolecular complexes from cryo-electron tomographic data sets. The presented algorithms offer an automated pipeline for tomogram processing, including KerDenSOM3D for classification, template matching for identification, sub-tomogram averaging for alignment and M-free score for cross-validation of the results. The material is presented in five chapters (1-5) and four papers (I-IV). Chapter 1 gives a brief introduction to cryo-electron tomography. Chapter 2 describes the methods. Chapter 3 and paper I give the theory and results from the experiments of KerDenSOM3D. Chapter 4 and paper II give a detailed description of the problem of reference bias in reference-based approaches and the ways to estimate it. Results from experiments of M-free score are presented. Chapter 5 concludes the thesis. In appendix the papers are summarized. The research papers III and IV provide additional examples of real data in which the developed algorithms have been applied to the study of biological structures.

1.1. Cryo-electron tomography

Tomography is a technique that images the interior of an object by computationally slicing or sectioning through the 3D data volume. During cryo-electron tomography, the biological object is first fixed at cryogenic temperature and subsequently two-dimensional (2D) projections are recorded as tilt series during which the sample is physically rotated within the transmission electron microscope (TEM) with a defined angular increment. These projection images are subsequently used to reconstruct the whole biological object in its entirety

1. Introduction

[15]. CET is the unique technique to provide high-resolution 3D visualization of cells and organelles at their near-to-native state.

The specimen preparation is the most important step as it determines the meaningfulness of the complete CET process with the goal to optimally preserve the ultrastructure of the biological samples. To preserve high-resolution information, cryo-fixation is the method of choice. The technique avoids the use of chemical fixatives and dehydration of the samples. At the same time, cryo-fixation makes the sample resistant to the high vacuum conditions of TEM imaging and protects it from decay or deformation during the microscopy process.

Cryo-fixation preserves the native structure of biological samples by vitrification of its water content. Biological objects such as small cells or organelles can be prepared for CET by plunge freezing [16-19], which is carried out by rapid propulsion of the sample into liquid ethane cooled to liquid nitrogen temperature [20]. In this process a thin layer of amorphous, vitreous ice is generated that is devoid of crystallization, which otherwise would destroy the structure of the specimen, scatter the electron beam, and thus would lead to loss of high-frequency information.

Even when using a high voltage TEM (e.g. 300kV), the application of plunge freezing is restricted to thin samples typically <500nm. During sample tilting in the TEM, the electron path through the specimen is increased by the factor of $1/\cos(\alpha)$, where α is the tilt angle. A too long electron path however leads to multiple electron scattering events that result in blurred images that not truly represent projection images of the specimen [15]. Moreover, thicker samples prevent vitrification of their water content [15, 21] and lead to formation of damaging ice crystals. Therefore, for thicker samples such as large cells or organelles, high-pressure freezing and subsequent cryo-ultramicrotomy is the method of choice for sample preparation [16, 21, 22]. High-pressure freezing is based on the purely physical principle, that under normal atmospheric pressure the homogeneous nucleation (supercooling) of water starts already at -40 °C, whereas this point can be lowered under high pressure conditions to -90 °C

according to the phase diagram of water. High pressure of >2000 bar is applied to the sample within a couple of milliseconds followed by the instantaneous freezing of the sample with very high cooling rates that lead to vitrification of sample layers in the range of 200 μm . Subsequent ultramicrotomy sectioning of the frozen-hydrated samples extends the application of CET to larger biological objects such as bulk cells and tissues in their near-to-native states. The cryo-ultramicrotomy technique can produce sections with thicknesses from 70 nm to 350 nm [15]. Thinner sections result in higher resolution with the same amount of electron dose. Therefore, a trade-off decision between higher tomographic resolution and the 3D field of view of the specimen needs to be made by the user.

During cryo-fixation techniques, the biological specimens are prepared and maintained below the devitrification temperature of water [16] during the whole CET process to prevent crystallization of their water content. CET images biological objects closest to their native state, since in contrast to classical methods that employ chemical fixation, dehydration, and post-staining with heavy metals, specimens are imaged directly in the vitreous state and hence preparation artifacts are minimized.

Specimens prepared by cryo-fixation techniques are highly radiation-sensitive. If the specimen receives an electron dose higher than its endurable amount, the specimen itself as well as the ice layer will be damaged due to energy transferred from the electron beam to the sample. Therefore, to prevent irradiation damage, imaging conditions need to be optimized such that the signal-to-noise ratio (SNR) is maximized within the limit of the acceptable cumulative electron dose. For tomography, where projection images from multiple directions are recorded, this total tolerable dose needs to be further divided by the total amount of projections [23, 24]. At the same time, sufficient electrons are required for every individual image to give sufficient high SNR on the detector [25]. For those reasons, the tolerated electron dose is, in principle, the limiting factor for CET [15].

CET provides 3D snapshots of biological specimens at high-resolution (typically ~ 5 nm). In combination with automated computational techniques such as

1. Introduction

template matching, sub-tomogram averaging, and classification, which is introduced in the Chapter 2 of this thesis, structures with a resolution between 1 and 2 nm could be achieved by averaging of their multiple occurring copies [26-29]. Occasionally it even breaks the 1 nm barrier [30]. Beside high-resolution structures, CET can also provide a complete 3D visualization of the entire proteome of a cell (“Visual proteomics”) and hence depict the location, orientation, and interactions of large macromolecular complexes, which underlie cellular functions [31]. Successful applications of CET with appropriate automated image processing techniques are listed in the reviews [32] and [33].

1.2. Data acquisition

A typical TEM for cryo experiments is equipped with a field emission gun (FEG), and a specimen holder that is cooled by liquid nitrogen. The biological specimen is prepared below the devitrification temperature of water as described in the earlier paragraph. The highly coherent electron beam produced by the FEG is accelerated by an anode typically set to 200-300 kV, focused by electromagnetic lenses, and transmitted through the specimen. The interactions with the specimen cause electron scattering, which is the process that makes transmission microscopy feasible [34]. To prevent electron irradiation damage of the samples, a low cumulative electron dose must be used for exposure (typically ~ 70 electrons/ \AA^2) [16, 35, 36]. Low-dose imaging results in poor amplitude contrast due to the prevalence of low mass elements like carbon in biological specimens, which makes it hard to distinguish biological relevant information from the crowded background. Therefore, to increase the overall contrast, images are collected in underfocus to enhance phase contrast. The underfocus has the effect that different spatial frequencies are transferred unequally as described by the contrast transfer function (CTF) [37-41].

The CTF depends on several parameters such as resolution, wavelength, defocus and spherical aberration [37, 40, 42]:

$$CTF(f) = A(\sin(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s))) + B \cos(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s)),$$

(Eq. 1.2.1)

where f is the spatial frequency (in \AA^{-1}), Δz the defocus (in \AA), λ the electron wavelength (in \AA), which depends on the acceleration voltage, and C_s is the spherical aberration of the objective lens of the microscope (in \AA). A is the envelope function that depends on defocus and B is the fraction of the amplitude contrast [30].

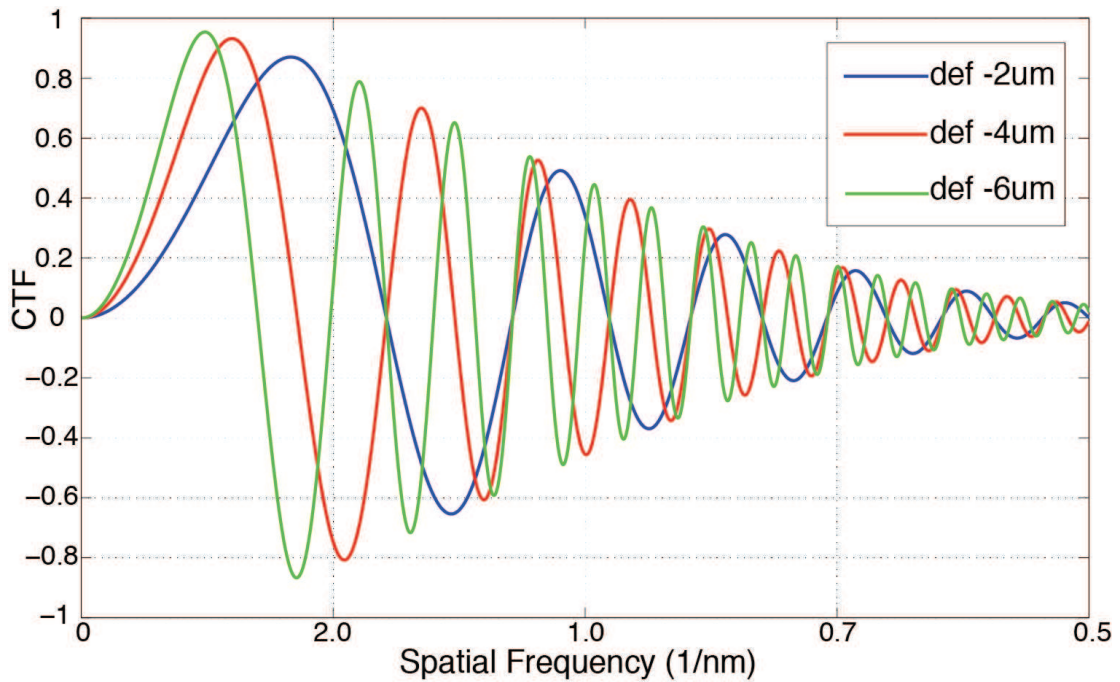


Figure 1.1: Three representative CTF curves with different defocus level. The parameters are: pixel size 0.5 nm; acceleration voltage 300kV; spherical aberration of the objective lens 2.26 mm; fraction of the amplitude contrast 0.1. The green, red and blue curves are generated with defocus level of -6 um, -4 um and -2 um sequentially.

Furthermore, due to the imperfections of the microscope system, the CTF is attenuated by an envelope function that depends on factors such as the coherence of the beam and specimen drift [37, 39, 40, 42]. Computer simulated representative CTF curves for different defocus on a 300kV FEG microscope are

1. Introduction

illustrated in Figure 1.1. The CTF oscillates between positive and negative values throughout the spatial frequency range, making features of the object appear with frequencies differently enhanced or suppressed in the image. Furthermore, the frequencies at the zero crossings of the CTF are not transferred, which results in loss of information. Without compensation or correction of the CTF effect, the achievable resolution of CET is normally limited to the first zero crossing of the CTF curve.

To achieve higher resolution, it is suggested to combine data from tomograms that span a range of defocus levels [42-44]. The CTF can be corrected by applying phase flipping or a Wiener filter [38, 45]. In CET, the defocus value varies on projection images due to the tilting of the sample. Therefore, CTF correction needs to proceed by performing phase flipping or Wiener filter to individual stripes parallel to the tilt axis, in which the defocus is approximately constant.

Traditionally, the images are recorded on photographic film or on charge-coupled device (CCD) cameras, with the CCD substantially speeding up data collection and giving improved SNR ratio at low- to medium-resolution [46]. Current revolutionary developments of direct detector devices (DDD), which directly detect the electrons on the image sensor and eliminate the use of scintillators with fiber optics or lens coupling, offer an improved detective quantum efficiency (DQE) especially at high-frequencies [47]. The much higher internal frame rate of the DDDs (e.g. 400 frames/s for a K2 summit) makes it possible to record videos of the sample during the total exposure time. Processing these videos can compensate for beam-induced movement or charging of the specimen during imaging [48-50]. Therefore, employing DDDs in CET can improve the structure determination of macromolecules at atomic precision [47, 49, 51].

1.3. Reconstruction

In CET, 2D projection images of the biological specimen are recorded from different viewing angles. Those projection images are subsequently synthesized into a 3D mass density map, which is called a tomogram. This process is referred to as reconstruction.

Due to mechanical imperfections of the specimen holder and the specimen itself, the region of recording could move when the specimen holder is tilted [15]. Therefore, to compensate for these movements, the projection images need to be mutually aligned prior to reconstruction. There exist mainly two approaches to align tilt series. The most commonly applied approach is based on using electron-dense fiducials, typically colloidal gold (10-15 nm in diameter) [27, 52-55]. The other approach is fiducial-less alignment that tracks structural elements within the images throughout the tilt series and subsequently employs them as surrogates for fiducial markers [56-62]. The fiducial-less approaches have the advantage that no additional preparation step needs to be done and no artifacts are produced from the electron-dense fiducials in the reconstruction volume. On the down side, the fiducial-less alignment method is not as reliable as the fiducial-based approach yet.

The existing reconstruction algorithms can be categorized into mainly three groups: direct Fourier reconstruction methods [63-70], weighted back-projection (WBP) [71-75] and algebraic reconstruction techniques (ART) [76-82].

Direct Fourier reconstruction methods are based on Fourier transformation. The projection images are transformed into Fourier space and a Fourier synthesis is then carried out, which sums up the Fourier components contributed by all the images for each point of the 3D Fourier map [63]. The subsequent inversed Fourier transformation results in the 3D reconstruction in real space. These Fourier based methods have been shown to result in accurate reconstructions

1. Introduction

[64]. In recent research from [68], a new direct Fourier method called Iterative Nonuniform fast Fourier transform (NUFFT) based Reconstruction method (INFR) [83] has been shown to result in excellent reconstruction of CET data. However, due to the high computational effort of the direct Fourier methods, they are not as widely applied to CET as the other two methods, especially as the WBP.

The WBP is based on the Fourier slice theorem: “*the 2D Fourier transform of a projection of the object is identical to a central section of the object’s 3D Fourier transform*” [15]. A WBP is a simple back-projection, where the projections are smeared out and subsequently summarized in real space. The 3D back-projected volume is subsequently deconvoluted with the point-spread function (PSF) by dividing its Fourier transform by its weighting function. The weighting function depends on the geometrical arrangement of all the projections.

The ART was first introduced by Gordon in 1970 [76]. ART can be considered as a solver of a linear equation system, where each linear equation is considered as a ray coming from an individual pixel of a particular projection image, which intersects with the reconstruction volume. The reconstruction volume is updated by sequentially applying the error correction terms computed for each ray in a given projection. The ART algorithm provides a fast convergence, however it produces very noisy images because of the inconsistencies in the set of linear equations [78]. For noise reduction, the Simultaneous Iterative Reconstruction Technique (SIRT) was introduced in 1972 [77], where all error correction terms in all the projections are applied simultaneously. However, while the noise level is successfully suppressed in comparison to ART, the SIRT convergences slowly and several iterations are needed. To combine the rapid convergence of ART and the smooth reconstructions produced by SIRT, the Simultaneous ART (SART) algorithm was developed by Andersen and Kak in 1984 [78]. By SART, all error correction terms for rays in a particular projection are applied simultaneously. No relaxation factor is employed here and one single iteration results in the best reconstruction quality and also short processing time.

Algebraic reconstruction methods provide better contrast than WBP. However, WBP is preferable as it could achieve better resolution in the subsequent averaging of multiple occurring structures and is significantly faster than the algebraic approaches. Current development of SART improves its reconstruction quality by e.g. super-sampling [84]. The quality of the reconstruction volume is comparable or even better than WBP. The implementation on multiple GPUs accelerates the reconstruction and improves the applicability of the algorithms [84-88].

In CET, the sample is fastened to the specimen holder and physically tilted around the tilt axis. The slab geometry of the specimen holder restricts the tilt range as the holder blocks the electron path at high tilts. Consequently, the angular tilt range in CET was usually limited to $\pm 65^\circ$. In addition, the increased path within the specimen at high tilt angles prevents the penetration of electrons.

This limited tilt range means, a fraction of the Fourier space is left unsampled during the recording of a tomogram, which leads to incomplete sampling. This effect is referred to as the "missing-wedge", as the unsampled region in Fourier space has the shape of a wedge. The missing-wedge results in an anisotropic resolution, thus the image features are distorted in the direction of the beam [2].

There are ways and means developed to minimize the missing-wedge effect, such as dual-axis electron tomography [89, 90]. During a double-axis tomography process, two tilt series are consecutively recorded at the same position of the specimen, where the second tilt series is recorded after a physical rotation of the specimen by 90° . The resulting missing information in Fourier space has only the shape of a pyramid and is hence less of the conventional single-axis tomography wedge. For classical room temperature electron tomography, double-axis electron tomography has been widely applied. However, for CET where the electron dose and the resulting SNR of the individual projection images are the limiting factors, the burden of splitting the dose on two tilt series instead one is usually avoided [91].

1. Introduction

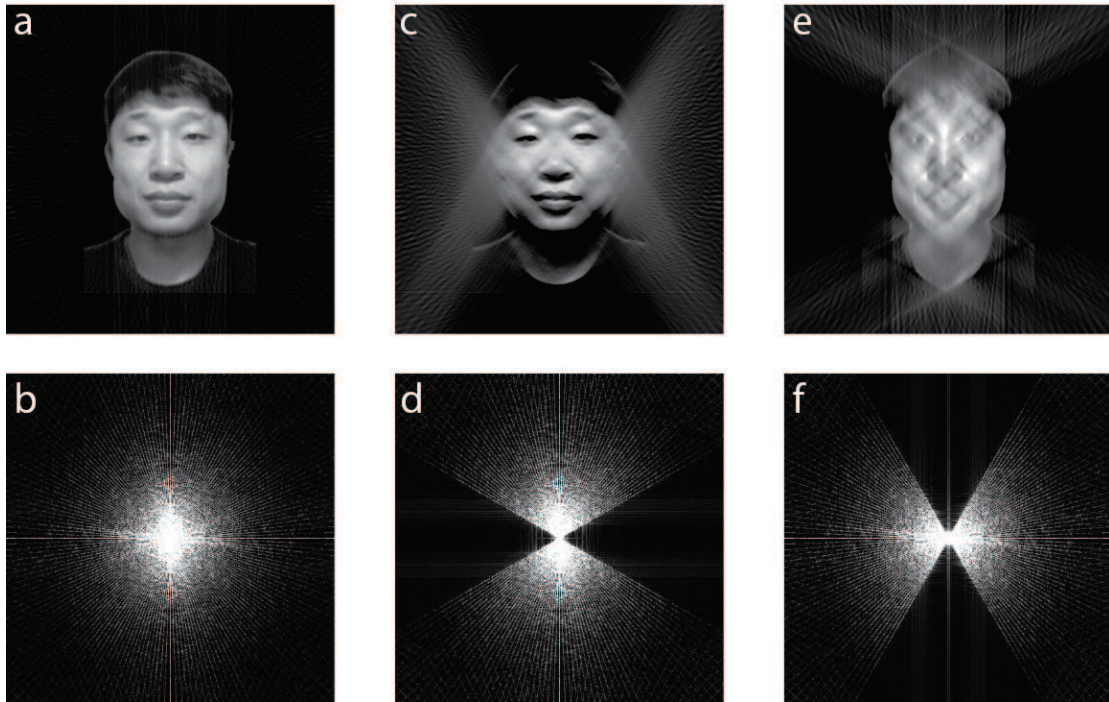


Figure 1.2: A graphical representation of the missing-wedge effect. (a) The reconstructed image from all the projection images. (b) The power spectrum of (a), shows isotropic resolution in all directions and no missing information. (c) The reconstructed image from a subset of the projection images, where the projections from 60° to 120° are absent. The reconstruction was blurry in the horizontal direction. (d) The power spectrum of (c), shows an empty area, where the information is missing. (e) Same as (c), but reconstructed from a subset of the projection images, where the projections from 0° to 15° from 75° to 105° and from 165° to 180° are absent. The reconstruction was blurry in the vertical direction. (f) The power spectrum of (e), shows a different empty area as in (d), where the information is missing.

2. Methods

2.1. Similarity measurement with cross-correlation

A prerequisite of the tomographic image processing approaches such as template matching, sub-tomogram averaging, and classification is a figure of merit that measures the similarity between two (sub-) volumes. The cross-correlation, also known as the sliding dot product, is commonly used in image processing for this purpose.

The cross-correlation of two real space discrete signals f and g is defined in one-dimensional (1D) space as:

$$CC(f, g)(d) \stackrel{\text{def}}{=} \sum_n f(n)g(n - d), \quad (\text{Eq. 2.1.1})$$

where d is the relative shift between the signals.

This equation (2.1.1) for cross-correlation can be easily extended into 3D cases. For image processing applications, in which the contrast may vary from image to image, the intensity of the images needs to be normalized. The normalization of the image is carried out by subtraction of the mean with subsequent division by the standard deviation:

$$\hat{f} = \frac{f - \bar{f}}{\sqrt{\sum(f - \bar{f})^2}} \text{ and } \hat{g} = \frac{g - \bar{g}}{\sqrt{\sum(g - \bar{g})^2}}. \quad (\text{Eq. 2.1.2})$$

The normalized cross-correlation is then defined as:

$$NCC(f, g)(d) \stackrel{\text{def}}{=} CC(\hat{f}, \hat{g})(d) = \frac{\sum_n (f(n) - \bar{f})(g(n-d) - \bar{g})}{\sqrt{\sum_n (f(n) - \bar{f})^2} \sqrt{\sum_n (g(n-d) - \bar{g})^2}}. \quad (\text{Eq. 2.1.3})$$

The normalized cross-correlation lies in the range of [-1 1], where 1 indicates identical signal up to scaling and 0 indicates uncorrelated signals (Figure 2.1).

2. Methods

Minus normalized cross-correlation values represent inversed contrast. In the following thesis, the normalized cross-correlation is exclusively referred without further indication.

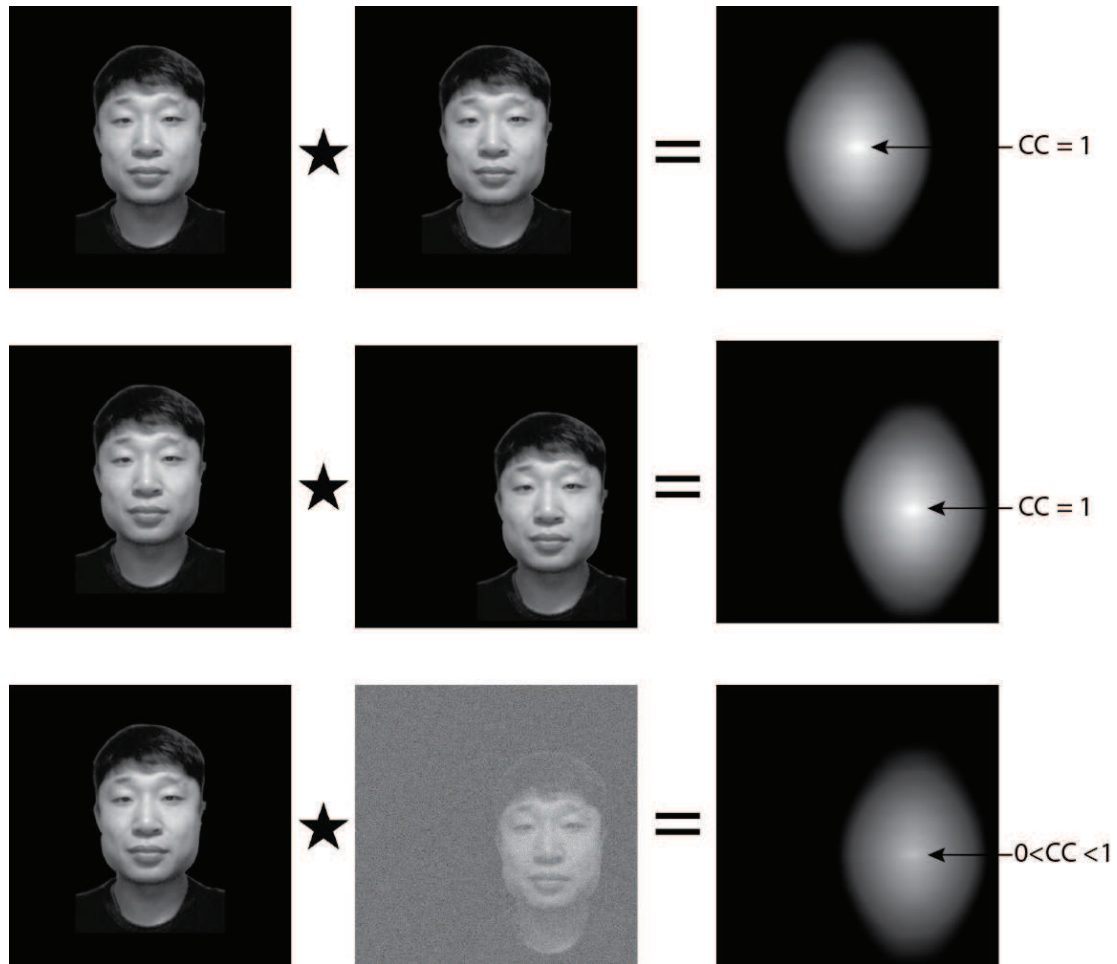


Figure 2.1: An illustration of the properties of the normalized cross-correlation function. (a) An image cross-correlated with itself results in a cross-correlation function, which has the value 1 in the origin. (b) An image cross-correlated with its shifted version results in a cross-correlation function, which has the value 1 in the coordinate of the shift. (c) An image cross-correlated with its shifted version plus additive white Gaussian noise results in a cross-correlation function, which has a value between 0 and 1 according to the noise level in the coordinate of the shift.

The calculation of the cross-correlation function (CCF), which gives the cross-correlation coefficient for all the possible relative shifts d between f and g , can be accelerated in Fourier space by the help from the correlation theorem:

$$\mathcal{F}(CC(f, g)) = \mathcal{F}(f)^* \cdot \mathcal{F}(g), \quad (\text{Eq. 2.1.4})$$

where \mathcal{F} denotes Fourier transformation and $*$ denotes complex conjugation.

Therefore, the similarity between two volumes and their relative translational shift can easily be calculated in Fourier space with modest computational effort [1].

As already discussed in an earlier paragraph the missing-wedge results in an anisotropic resolution, which affects the following processing. When similarity is measured between two volumes from which at least one is missing-wedge affected, the cross-correlation is not accurate. The presumption of cross-correlation between two identical signals does not hold true any longer due to the missing information in the wedge. Together with the extremely low SNR of cryo-electron tomograms, the analysis can be severely hampered. For example, the classification algorithms tend to group the volumes rather according to their orientation of the missing-wedge than to the underlying signal [2, 92]. To overcome this problem, the similarity between missing-wedge affected signals should be calculated exclusively by using the information from the common region of the pair of volumes in Fourier space [1, 2].

This constraint cross-correlation (CCC) is defined as:

$$CCC(f, g) = NCC(f^{\Omega_{fg}}, g^{\Omega_{fg}}) \text{ with}$$

$$f^{\Omega_{fg}} = \mathcal{F}^{-1}(\mathcal{F}(f) \cdot \Omega_{fg}) \text{ and } g^{\Omega_{fg}} = \mathcal{F}^{-1}(\mathcal{F}(g) \cdot \Omega_{fg}) \quad (\text{Eq. 2.1.5})$$

The \mathcal{F}^{-1} denotes the inverse Fourier transformation and Ω_{fg} denotes the intersection of the volumes in Fourier space. More detailed description of CCC can be found in [2]. The CCC gives more precise and reliable similarity measurement for missing-wedge affected volumes and is therefore widely utilized in image processing of electron tomographic data sets [14, 93-96].

2. Methods

2.2. Resolution estimation with Fourier shell correlation

The 3D Fourier shell correlation (FSC) measures the normalized cross-correlations between two volumes over spatial shells in Fourier space. It can be considered as the 3D version of the Fourier ring correlation (FRC). The FSC is frequently employed both in single particle analysis and electron tomography to estimate the resolution of 3D average volumes [7, 97].

To estimate the resolution of an average volume, the whole data set is randomly splitted into two halves. From each half of the data set, an average is independently generated. The FSC between these two average volumes F and G is defined as:

$$FSC(R) = \frac{\sum_R FG^*}{\sqrt{\sum_R |F|^2 \sum_R |G|^2}}, \quad (\text{Eq. 2.2.1})$$

where R is a spatial shell in Fourier space.

The FSC gives a measurement of the spectral signal-to-noise Ratio (SSNR) [98, 99] in the reconstruction. For a large number of averaged volumes, it exists:

$$FSC = \frac{SSNR}{SSNR+1} \text{ and } SSNR = \frac{FSC}{1-FSC} \quad (\text{Eq. 2.2.2})$$

[7, 100-102].

For the definition of useful information, the energy of the signal should overcome the energy of the noise. Therefore, $SSNR = 1$ is conventionally chosen as the lower limit of reliable information [7]. This leads to the criterion that the outermost spatial shell with its $SSNR \geq 1$ indicates the resolution. This is according to (Eq. 2.2.2) equivalently to $FSC \geq 0.5$ [15, 44]. Although a fix-valued threshold for FSC was argued to be based on incorrect statistical assumptions [103], the “0.5-Criterion” is most widely used in cryo-electron microscopy for estimation of the resolution.

2.3. Template matching

Template matching is a molecular identification approach based on pattern recognition to detect some previously defined patterns within the given tomographic volumes. Provided the high-resolution 3D density maps of the macromolecules of interest obtained from single-particle analysis or x-ray crystallography are available, they can be used as a reference for a systematic search of tomograms [32, 104]. The search is exhaustive and performed in a fully automatic manner, which requires no user-interference [1]. The outcome is, therefore, fully reproducible.

The technique was first introduced in [1]. A schematic flowchart representing the process is given in Figure 2.2. A brief introduction is given here.

Firstly, the higher-resolution structure of the macromolecules of interest is used to generate a reference. Hereby, the voxel size of the reference is adjusted to the voxel size of the tomogram, on which the search will be performed. The value of a voxel from the reference volume is calculated as the summation of the number of atoms positioned in this particular voxel. This density map is then filtered in Fourier space with the CTF, which is simulated from the recording condition of the tomogram. A subsequent low-pass filter at approximately the resolution of the tomogram produces a reference, which adequately resembles the copies of the macromolecules of interest in the tomogram. Afterwards, an exhaustive search based on constraint cross-correlation is performed. For better computational performance, the reference volume is pasted into an empty box of the size of the tomogram and the CCF is calculated with Eq. 2.1.4 in Fourier space. While the tomogram is missing-wedge affected, the reference is essentially missing-wedge free. Taking the missing-wedge into account, the constraint CCF is calculated exclusively in the region of Fourier space where the tomogram contains information. With the Eq. 2.1.4, the CCC is derived for a single orientation of the reference. For an exhaustive search, the calculation of the CCF

2. Methods

is performed for every possible angular orientation of the reference separately. The maximum similarity for every position can then be derived.

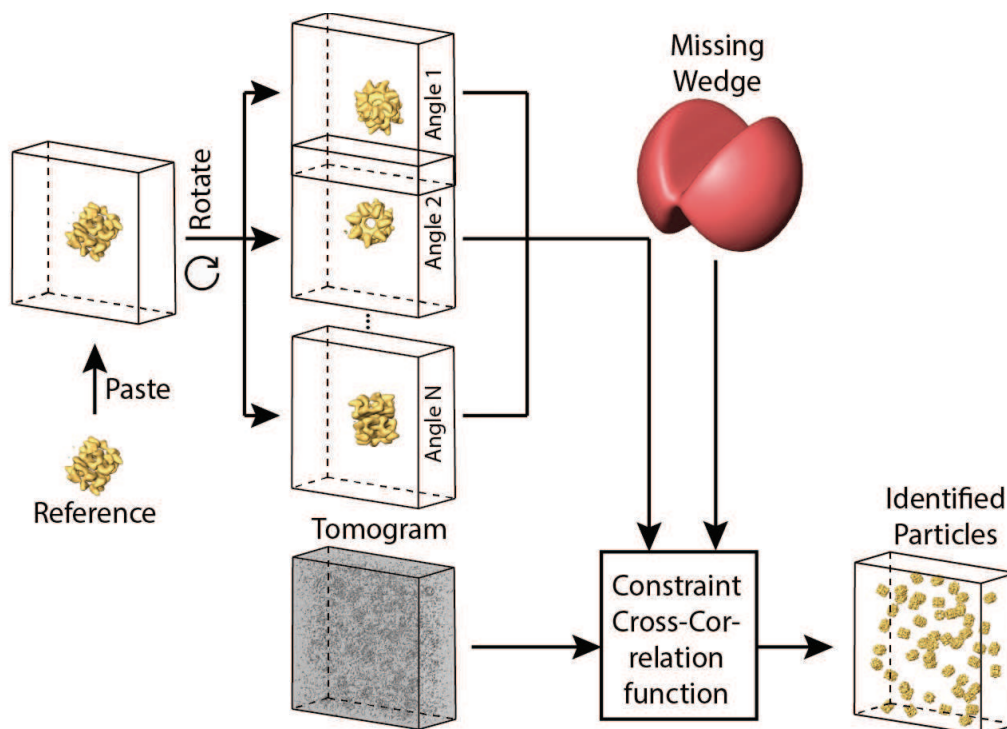


Figure 2.2: A schematic flow diagram showing the template matching procedure. The reference is pasted into a blank box and then rotated to every possible orientation. The rotated reference is then cross-correlated with the tomogram, where the missing-wedge is taken into account. The position and orientation of the sought particles worked out from peak detection.

The outcome of the described template matching procedure is a CCF containing the maximum CCC values and the corresponding rotation angles for each individual position. The true maxima of the CCF can be extracted in descending order. To prevent multiple extractions of a macromolecule, the CCF within a certain radius is set to zero after determination of a true maximum. Each extracted position with its corresponding angles represents one putative hit and is subsequently subject to further analysis.

The Fourier space approach of calculating the CCF allows optimal parallelization with excellent speed-up characteristics [1]. The tomogram can be divided into smaller sub-volumes. The processing of each sub-volume can be carried out with one computing unit (e.g. one CPU core). The subdivision enables a speed-up factor of n , where n is the number of available computing units. In the meanwhile, it also dramatically reduces the amount of memory required for processing of large tomograms, which would be problematic otherwise [105].

The template matching is an excellent approach for exploring the positions and orientations of macromolecules as well as for elucidating the network of macromolecular interactions (resulting in a cellular proteome atlas). At current resolution of ~ 5 nm, the application of template matching is limited to large macromolecular complexes with a molecular weight of 200-2000 kDa [1, 32, 106]. The method is successfully applied on mapping ribosomes in tomograms of the thin bacterium *Spiroplasma melliferum* [105]. In [107], polyribosomes from intact human glioma cells are successfully identified by this technique. Multiple template matching was applied to extract proteins such as ribosomes, GroEL, GroEL/ES and RNA polymerase II from intact *Leptospira interrogans* [106]. Because of the low SNR nature of the cryo-electron tomograms, sometimes even the positions corresponding to the highest cross-correlation values do not represent the searched macromolecules [4]. Validation of template matching, which should provide a Boolean judgment (true/false) for each putative hit, becomes an important task [32]. The applicability of template matching would be improved with a reliable validation technique.

2.4. Sub-tomogram averaging

Sub-tomogram averaging is a technique that reveals the high-resolution information by averaging of multiple occurrences of the macromolecular structure in the tomograms. The core of the sub-tomogram averaging approach is the rotational and translational alignment of the individual sub-tomograms.

2. Methods

The alignment, which is conceptually similar to single particle analysis, is an iterative procedure based on CCC. At each iteration, the relative orientation and shift of the sub-tomograms is assigned with help from a reference structure, and then used to update the reference for the next iteration [32]. This iterative procedure is carried on until convergence.

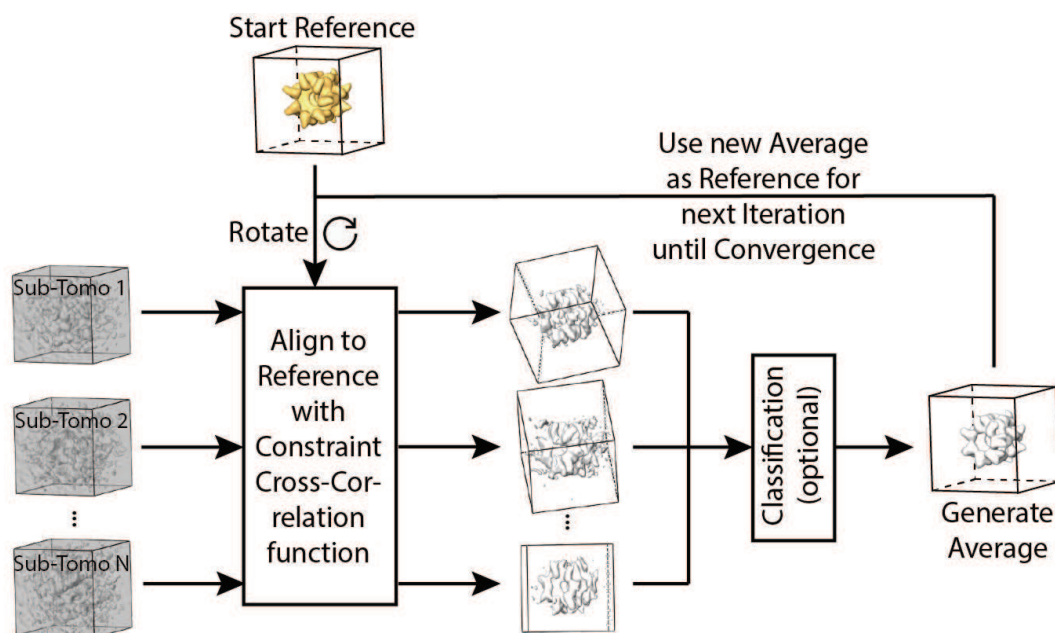


Figure 2.3: A schematic flow diagram showing the sub-tomogram averaging procedure. Sub-tomograms are rotationally and translationally aligned against a start reference. The aligned sub-tomograms are then averaged to generate a new reference after an optional classification step. The new reference is then used for alignment of the sub-tomograms in the next iteration. This procedure is iterated until convergence.

A schematic flowchart representing the process is given in Figure 2.3. Prior to the sub-tomogram averaging process, a starting reference needs to be generated, which is deployed for the 1st iteration. For generating the starting reference, three different methods could be applied. First, it could be generated from the high-resolution structure of the macromolecule to study, if this is available from single particle analysis or x-ray crystallography. Second, a so-called “reference-

free” strategy could be used. A featureless Gaussian sphere or the raw average of the not-aligned sub-tomograms is used as starting template. This approach does not require any *a-priori* information of the data and is much less prone to overfitting and bias introduced from an initial reference [3, 4, 108]. Lastly, the starting average could be generated from the sub-tomograms and their available pre-orientation and pre-localization information. This information can be obtained by the known spatial relationship between the macromolecules and reference objects with high contrast. For instance, in the case of membrane proteins, some are orientated perpendicular to the membrane, which has normally very high contrast in tomograms [109-112].

In each iteration, it resolves the problem:

$$\arg \max_{(a,R)}(CCC(r, \tau_a \Lambda_R p_i), i \in (1,2, \dots, n), \quad (\text{Eq. 2.4.1})$$

where r is the reference volume and p_i is the i -th sub-tomogram, τ_a is a translational operation with a 3D vector a and Λ_R is a rotational operation with a set of Euler angles R .

In a similar fashion to template matching, a CCF between the reference volume and each individual sub-tomogram is calculated for every possible angular orientation. To compensate for the missing-wedge, the CCF is calculated within exclusively the information from the region where the particular sub-tomogram was sampled. The maxima of the CCFs determine the angles and the shifts, which are then used to rotate and translate the sub-tomograms to a common orientation and position. They are subsequently averaged to generate the reference for the next iteration. An additional step of classification can improve the class purity of the sub-tomogram by eliminating the outliers or misaligned sub-tomograms. In addition it can also discern between different conformational states if they exist. This optional classification step can be executed at each iteration before generating the new average. In the averaging step, the correct weighting is performed in Fourier space according to the information density [112]. The procedure is then iteratively repeated with the updated reference, by which the orientation and position of the individual sub-tomograms can be

2. Methods

further improved. The whole procedure is iterated until no further improvement takes place.

The sub-tomogram averaging techniques improve the SNR and therefore the resolution of the macromolecular structure by aligning and averaging its multiple copies in the sub-tomograms. The final average has also isotropic resolution in all directions since the sub-tomograms are ideally randomly orientated and have therefore different orientations of the missing-wedge. In comparison to the conceptually similar single particle analysis, sub-tomogram averaging methods can provide high-resolution 3D structure of macromolecular complexes *in situ* without prior purification [32, 113]. The results of sub-tomogram averaging provide not only the 3D structure of the macromolecular complexes *in situ*, but also their spatial distribution and relationship, which can be obtained by projecting the resolved structure and spatial information back into the tomograms [113].

The sub-tomogram averaging is a relatively young method but already applied on revealing the structure and arrangement of a wide range of macromolecular complexes such as polysomes [107, 114], virus surface glyco-proteins [115-118], internal virus structural proteins [119-125], microtubule binding proteins [27], nuclear pore complexes [126-128] and many others as listed in [113].

Similar to the template matching technique, the application of sub-tomogram averaging is currently limited to large complexes (>~750kDa) in cellular tomograms or small complexes located on the surface of viruses or vesicles (>~300kDa) [113]. The problem of reference bias and overfitting could potentially lead to false results. Therefore, the sub-tomogram averaging process should be carried out with extreme caution and cross-validation of the resolved structure obtained by sub-tomogram averaging is mandatory.

3. Classification of tomographic data sets

Classification methods are essential techniques to improve the class purity of electron sub-tomogram or to discern between different conformational states or interaction partners of the macromolecules [2]. Due to the low SNR and the missing-wedge effect, the classification of electron sub-tomograms is a challenging task. Classification algorithms tend to classify data according to their orientation of the missing-wedge, rather than to the underlying signal [14].

Classification procedures are well established in single-particle analysis as an integral part of the methodology. However, their application in tomography is still developing slowly. The reasons for this are multiple (a) The unknown nominal defocus (b) the defocus variation within the sample (c) the alignment errors of the 3D-reconstruction, (d) the specimen deterioration due to irradiation damage, (e) the missing-wedge effect and (f) the significant increase in computing time. In addition to those, an additional effect arising from the crowded nature of the cytoplasm is experienced, in particular through the CTF influence of neighboring macromolecules [2].

Classification in single-particle electron microscopy is very often performed with multivariate statistical analysis [15]. Here the data set is reduced to a few statistically significant features. In this analysis the covariance matrix is involved, which includes the mutual cross-correlation of every two sub-tomograms. Foerster et al. incorporated in the calculation of the covariance matrix the constrained cross-correlation, which measures the similarity of two sub-tomograms only in those areas where signal exists in both sub-tomograms [2], also known as constrained principle component analysis (PCA). In this way, it is guaranteed that two identical sub-tomograms will have a cross-correlation value of approximately one, independent of the orientation of the missing-wedge. Otherwise, due to the erroneous normalization, the cross-correlation value can

3. Classification of tomographic data sets

vary significantly, meaning that identical images with different orientations would not be identified as such, and hence would not be classified into the same class. In a similar type of analysis, Bartesaghi et al. included classification in the alignment process, which was also accelerated by spherical harmonics [96]. Winkler et al. incorporated classification further in the averaging procedure, by also including a weighted average [129]. In addition to these methods, maximum likelihood approaches have recently been proposed, which perform the alignment and the sub-tomogram averaging together [94, 130]. Maximum likelihood methods are shown to be less prone to reference bias compared to other classification methods [131].

3.1. Classification with self-organizing maps

Self-organizing maps (SOMs) are a type of artificial NN, which map a set of high-dimensional data vectors onto a typically 2D set of nodes. These nodes are referred to as the “code-vectors” [10]. The mapping process is executed in such a way that data projected onto neighboring code-vectors are more similar than those projected onto distant code-vectors [12]. The 2D distances on the output space represent the dissimilarity measurement in the high-dimensional input space. Each input vector is compared simultaneously with every code-vector, where the dissimilarity is calculated with a distance function. The input vector is then used to update the code-vector with the smallest distance as well as its neighborhood code-vectors. The magnitude of the modification of individual code-vectors is given by a neighborhood function, which is inverse proportional to the distance function.

The initial code-vectors are randomly generated and will tend to approximate the probability density function (PDF) of the input data as the mapping and update process are iterated [10, 12]. The procedure results in a set of code-vectors and the assignment of each input data to a certain code-vector, which gives the classification.

SOMs are unsupervised classification methods that require no *a priori* knowledge of the data set such as the number of classes and the underlying genuine signal, which are mostly unknown factors in classification tasks of cryo electron tomographic data sets. The SOMs summarize the original data, which are mostly very noisy, to a set of high-quality “representatives”, namely the code-vectors, while the basic pattern of the data is rigorously preserved [11]. In comparison to the raw input data, the code-vectors have increased SNR and homogenous information distribution in Fourier space. Therefore, SOMs are excellent tools for classification of electron tomographic data sets. The original Kohonen self-organizing map (KSOM) was first introduced in the field of electron microscopy for classification of 2D images by Marabini and Carazo in 1994 [12]. A modified algorithm named Kernel Probability Density Estimator Self-organizing Map (KerDenSOM) that is based on estimation of the PDF of the input data is proposed for classification of both 2D and 3D data sets [11, 13]. In comparison to the KSOM, the KerDenSOM is less sensitive to initial conditions since an additional deterministic annealing strategy applied to the smoothness factor of the output grid actuates the convergence toward the global maximum [11]. In paper I, the KerDenSOM was extended and implemented in 3D having compensated for the missing-wedge. It shows superior performance in a comprehensive comparison with previously used algorithms.

3.2. Theory of KerDenSOM3D

The classification should distinguish between different types of macromolecules contained in the sub-tomograms and correctly assign each individual sub-tomogram to the correct type of particles. We always assume that the sub-tomograms are translationally and rotationally aligned to a common origin by e.g. the sub-tomogram averaging procedure as described in chapter 2.4. Misaligned particles, in the case of sub-tomogram averaging, as well as false positive hits in the case of template matching should be recognized by the classification procedure and classified into classes with low occupancy.

3. Classification of tomographic data sets

The data set for classification includes n sub-tomograms $P_i(x, y, z)$ containing putative macromolecules of interest that were extracted from larger tomograms. The positions from which the extractions are taken place can be either manually depicted or automatically detected with e.g. template matching as described in chapter 2.3. The (x, y, z) denotes the spatial Euclidean coordinates of the voxels and i denotes the index of the sub-tomogram. In a similar fashion, the code-vectors are denoted as V_j with j denotes the index. Base on the CCC described in chapter 2.1, the similarity between a sub-tomogram P_i and a code-vector V_j can be calculated by a similarity metric that takes both the missing-wedge and the contrast difference between the sub-tomogram and the code-vector into account. The code-vectors, which are calculated from information present in a number of sub-tomograms, have better SNR compared to the individual sub-tomograms. In addition, since the sub-tomograms have generally arbitrarily oriented missing-wedges, they are, in general, missing-wedge free. Hence, it is only required to apply the constraint of every individual sub-tomogram to compensate for the missing-wedge. The sub-tomograms are compared to the representative code-vectors rather than the other individual sub-tomograms. For the sake of missing-wedge compensation, no extra portion of signal has to be sacrificed when calculating the similarity within only the overlapping region in Fourier space not affected by missing-wedge.

The individual sub-tomograms P_i may have different contrast due to their spatial position in tomograms. In order to eliminate the difference in the calculation of similarity, they are normalized to zero mean and unit variance prior to calculation. The normalization is carried out by subtraction of their mean value \bar{P}_i followed by dividing by their standard deviation. The normalized sub-tomograms $P_i'(x, y, z)$ are defined as:

$$P_i'(x, y, z) = \frac{P_i(x, y, z) - \bar{P}_i}{\sqrt{\sum_{x, y, z} (P_i(x, y, z) - \bar{P}_i)^2}} \quad (\text{Eq. 3.2.1})$$

Presumed that the sub-tomograms P_i are aligned to a common origin prior to classification. This involves rotating each of them by a different set of Euler angles, which can be obtained by performing sub-tomogram averaging or

template matching. The region in Fourier space that is not missing-wedge affected in each rotated individual sub-tomogram corresponds to the rotation of the originally sampled region by these angles. This region is denoted by Ω_i for the i -th sub-tomogram.

The normalization of the sub-tomogram P_i and the code-vector V_j are restricted to the overlapping region Ω_i , when we constrain the correlation of P_i and V_j to Ω_i in Fourier space, as described in chapter 2.1. Since the code-vectors are considered as missing-wedge free, Ω_i is the region in Fourier space that is not affected by the missing-wedge in the i -th sub-tomogram. We denote by $V_j^{\Omega_i}$ the artificial missing-wedge affected code-vector that is obtained by removing the Fourier components outside of the region Ω_i from V_j , that is:

$$V_j^{\Omega_i} = \mathcal{F}^{-1}(\mathcal{F}(V_j) \cdot \Omega_i). \quad (\text{Eq. 3.2.2})$$

Ω_i in the Eq. 3.2.2 can be interpreted as a binary mask in Fourier space, where the value in the missing-wedge affected region is 0, and is 1 elsewhere.

It is recommended to mask the sub-tomograms and code-vectors with a suitable real space mask M to focus and limit the classification on certain features and reduce the effect of noise located outside of the area containing signal. Applying an appropriate real space mask M could significantly increase the precision of classification. When using a mask M in real space, the sub-tomograms and code-vectors can be then normalized as:

$$P_i'(x, y, z) = \frac{M(x, y, z)[P_i'(x, y, z) - \bar{P}_i]}{\sqrt{\sum_{x', y', z'} [M(x', y', z')(P_i'(x', y', z') - \bar{P}_i)]^2}}, \text{ and} \quad (\text{Eq. 3.2.3})$$

$$V_{j,i}'(x, y, z) = \frac{M(x, y, z)[V_i^{\Omega_i}(x, y, z) - \bar{V}_{j,i}]}{\sqrt{\sum_{x', y', z'} [M(x', y', z')(V_i^{\Omega_i}(x', y', z') - \bar{V}_{j,i})]^2}}, \quad (\text{Eq. 3.2.4})$$

where the code-vectors need to be constraint both in Fourier space with Ω_i and in real space with the mask M .

The mean values \bar{P}_i and $\bar{V}_{j,i}$ have also to be constrained to the mask M or to Ω_i and M respectively:

3. Classification of tomographic data sets

$$\bar{P}_i = \frac{\sum_{x,y,z} P_i(x,y,z)}{\sum_{x',y',z'} M(x',y',z')}. \quad (\text{Eq. 3.2.5})$$

$$\bar{V}_{j,i} = \frac{\sum_{x,y,z} V_i^{\Omega_i}(x,y,z)}{\sum_{x',y',z'} M(x',y',z')}. \quad (\text{Eq. 3.2.6})$$

The KerDenSOM3D algorithm is based on the approach described in [11]. The algorithm is expanded in 3D with the missing-wedge compensation as described by Eq. 3.2.3 and Eq. 3.2.4.

The expressions that involve measuring the similarity of the sub-tomograms P_i to the code-vectors V_j need to be expanded with missing-wedge compensation, for instance:

$$\sum_{i=1}^n \sum_{j=1}^c \|P_i - V_j\|^2 u_{j,i}, \quad (\text{Eq. 3.2.7})$$

where n is the number of sub-tomograms, c is the number of code-vectors, and $u_{j,i}$ is a matrix giving the weight of sub-tomogram P_i in the computation for the update of code-vector V_j . Evaluating each term of $\|P_i - V_j\|^2$ is required for the calculation and update of the weight matrix $u_{j,i}$. The missing-wedge effect has to be taken into account at this place and compensated accordingly.

The L_2 -norm is applied for measuring the similarity between a sub-tomogram and a code-vector, this similarity is formulated as the following:

$$\|P_i - V_j\|^2 = \sum_{x,y,z} [P_i(x,y,z) - V_j(x,y,z)]^2, \quad (\text{Eq. 3.2.8})$$

which in the new KerDenSOM3D approach is substituted by the L_2 -norm of the normalized, missing-wedge compensated sub-volumes $\|P_i' - V_{j,i}'\|^2$. Note that this norm in L_2 , can be substituted by other norm in L_p , defining different self-organized maps for each definition of distance.

Further steps of the algorithm remain unmodified from the original as described in [11]. The KerDenSOM3D algorithm aims to converge to a set of code-vectors that fulfills two requirements simultaneously: (a) their similarity to the sub-tomograms according to the similarity metric is maximized, and (b) the smoothness of their distribution among the network is maximized. The

smoothness of the network corresponds to the similarity to all the other neighboring code-vectors and is described as part of the cost function in the equation 18 in [11]. Both requirements are implemented in a cost function based on the equation 14 in [11] expanded with missing-wedge compensation. Note that the balance between the both requirements of similarity and of smoothness is regulated by a smoothness parameter ϑ , which decreases with each step of the process, from an initial value ϑ_1 to a final value ϑ_0 . This is referred in the numerical algorithm to as the annealing strategy. The value of ϑ in a certain annealing step can be calculated using the following formula:

$$\vartheta = \exp (\ln(\vartheta_1) - (\ln(\vartheta_1) - \ln(\vartheta_0)) * \text{iter}/\text{MaxIter}). \quad (\text{Eq. 3.2.9})$$

For each annealing step, the code-vectors are updated with the current value of ϑ by iterating on t the relation:

$$V_j^{(t+1)} = \frac{\sum_{i=1}^n u_{j,i}^{(t)} P_i + \vartheta \overline{V_j^{(t)}}}{\sum_{i=1}^n u_{j,i}^{(t)} + \vartheta}; \quad u_{j,i}^{(t+1)} = u_{j,i}^{(t)} \left(\left\{ \left\| P_{i'} - V_{j'}^{(t)} \right\| \right\}_{i',j'}, \{V_{j'}^{(t)}\}, \vartheta \right), \quad (\text{Eq. 3.2.10})$$

until convergence is achieved. In the formula of updating the code-vectors, the term $\overline{V_j^{(t)}}$ is the numerical average of the neighboring code-vectors of $V_j^{(t)}$ on the network in the t -th iteration, not including $V_j^{(t)}$ itself.

3.3. Choice of the smoothness parameters

The grid size of the network and the smoothness parameter ϑ for the annealing strategy are relevant for the performance of a classification process. However, a full scan of these parameters may often be the most burdensome and time-consuming part of the classification procedure. To achieve good classification performance despite avoiding the exhaustive search of the parameters, good empirical values of the smoothness parameter depending on the number of particles n and the size of the grid c need to be utilized.

3. Classification of tomographic data sets

The part of the formula where the smoothness parameters ϑ is involved in Eq. 3.2.10 can be derived by assuming that the contribution of all the sub-tomograms to all code-vectors is the same. That is, all the entries of the $u_{j,i}$ have the same value $u_{j,i} = \frac{1}{c}, \forall j, i$. Thus, the code-vectors are updated in the following way: $V_j^{(t+1)} = \frac{\frac{n}{c}\bar{x} + \vartheta \overline{V_j^{(t)}}}{\frac{n}{c} + \vartheta}$, where $V_j^{(t+1)}$ is the j -th code-vector in $(t+1)$ -th step. With this assumption, all code-vectors are supposed to be initialized identically. For data sets containing sub-tomograms with similar image size, recorded at similar conditions and have similar SNR, the ratio $\frac{n/c}{\vartheta}$ has to be set to the fixed good empirical values in order to assure good classification accuracy. Thus, an exhaustive search of parameters can be avoided, which is of great advantage to any user.

From our numerous experiments, the following empirical values: $\vartheta_1 = 30 \cdot \frac{n}{c}$ and $\vartheta_0 = 3 \cdot \frac{n}{c}$ are suggested for the classification of cryo-electron tomographic data sets. The number of annealing steps can be set to 10 or 50 and a grid size of 5x5, 7x7 or larger should be employed.

3.4. Data acquisition conditions

The tilt-series were recorded on a 300 kV, Tecnai G2 Polara FEG and a CM300 transmission electron microscope (FEI, Eindhoven, The Netherlands) equipped with a Gatan post-column GIF 2002 energy filter (Pleasanton, CA). Tomographic single tilt-series were acquired using either the TOM Toolbox or the UCSF software package with zero-loss filtering (slit-width of 30 eV) [55, 132]. The individual projection images recorded with the microscope were interactively aligned with respect to a common origin using 10nm colloidal gold particles distributed in the samples as fiducial markers. Reconstructions were performed using WBP. Visualization for the figures was prepared with Matlab® and the Amira package for tomography [133].

3.5. Results

3.5.1. Simulated data sets

The procedure to produce simulated data sets in [2] was also used here in order to allow a direct comparison of the performance of the different algorithms. The procedure faithfully simulates the physical recording process of electron tomography and produces realistic sub-tomograms, as already proved in [2]. For each sub-tomogram a tilt-series is simulated: the macromolecular structure for simulation was rotated around a single tilt axis and projected onto a projection plane simulating the detector. One fraction of Gaussian noise is then added to the projections. The CTF is then applied on the resulting 2D projection image, which is composed of two terms as described in chapter 1.2: (a) the calculated CTF under ideal conditions assuming pure phase contrast at a particular defined defocus and (b) a modulation-transfer function (MTF) or envelope function that describes the imaging properties of the detector. The envelope function was approximated with a Gaussian function that drops to 20% at half Nyquist frequency. On the resulting image, another fraction of Gaussian noise was added simulating background noise due to the readout noise of the detector. The resulting projections were then used to reconstruct the sub-tomogram using WBP, where the projections are weighted in the Fourier space using analytical weighting. The data sets were simulated with different half-opening angle W of the missing-wedge (e.g. a half-opening angle W of 20 degrees corresponds to a simulated tilt-series from -70 to +70 degrees) and SNR.

We applied KerDenSOM3D on the simulated data sets. The simulated object for testing the classification algorithm was the archaeal chaperonin thermosome in its open and closed conformations. The protein data bank (PDB) coordinates 1A6D from the atomic structure obtained by x-ray crystallography were used for the closed conformation [134]. For its open conformation, an atomic structure

3. Classification of tomographic data sets

derived from a molecular model fitted to a density map obtained from single particle-analysis was used [135]. The constrained PCA was tested on the same data sets. This provides the possibility for a direct comparison of the two techniques.

We performed the classification with KerDenSOM3D for data sets with various SNR and half opening angles of the missing-wedge. A typical 5x5 code-vector network was used for the classification. The *a priori* knowledge of the number of populations present in the data set is not required. In contrast to the conventional classification algorithms, the number of code-vectors does not correspond specifically to the number of anticipated classes. Instead, an excess of code-vectors results in the vast majority not having any sub-tomograms assigned to them, and can be easily discarded for further analysis (Figure 3.1).

The relevant annealing parameter ϑ has been set to 8000 and is decreased to 100 in 20 annealing steps. The classification shows that for all the simulated missing-wedge angles W (0, 10, 20, 30 and 40 degrees) and SNRs (from 1, 10^{-1} , 10^{-2} , 5×10^{-3} , 10^{-3}), KerDenSOM3D performs a perfect classification, where two classes are clearly distinguishable from each other (Figure 3.2 for W at 20 and Figure 3.3 for W at 30 degrees). At the SNR of 5×10^{-4} , the classification algorithm shows its limitation, such that the clusters representing the two classes can no longer be unambiguously distinguished any more. However, the proposed KerDenSOM3D with missing-wedge compensation still shows superior performance, specifically, the classifying clusters are significantly more compact than clusters obtained with the original version of the algorithm that is not missing-wedge compensated. This matches the observation as in [2], which shows overall superior distinguishing capability and classification performance of the missing-wedge corrected algorithm, especially at low SNRs.

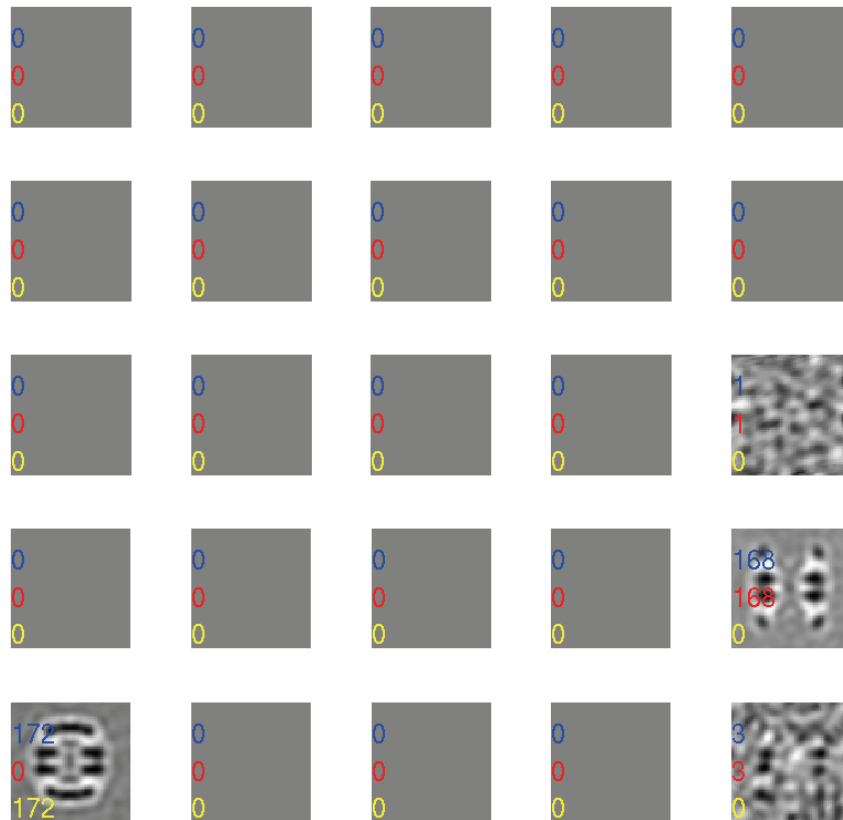


Figure 3.1 An example of a KerDenSOM3D network with 5x5 nodes. The simulated data set has the SNR of 10^{-2} and W of 30 degree. The numbers in blue, red and yellow are the number of sub-tomograms classified to the node, from class open and from class closed respectively. The nodes not having any sub-tomograms assigned to them are rendered grey.

3. Classification of tomographic data sets

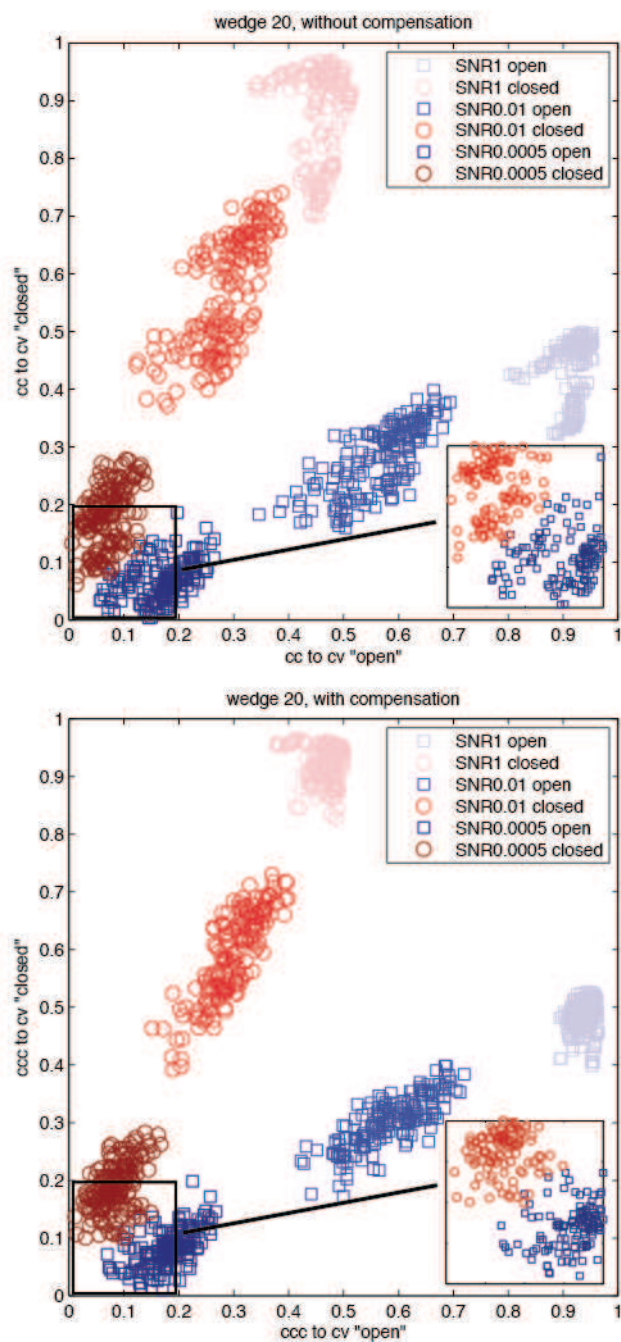


Figure 3.2 Comparison of the classification results of simulated open thermosomes (squares) and closed thermosomes (circles) at different SNR (1, 0.01 and 0.0005) and missing-wedge half opening angle of 20 degree, which correspond to a tilt-series of ± 70 degrees, with the original KerDenSOM algorithm without missing-wedge compensation in the upper row and the novel KerDenSOM3D in the lower row. The color-coding is given in the upper right corner of each graph. The similarity of a certain sub-tomogram to the classifying code-vector is measured. At high SNR, the clusters can be clearly distinguished

from one other. The clusters are much more compact with missing-wedge compensation. With decreasing SNR the clusters come closer together, harder to distinguish from each other and merge together at a SNR of 0.0005. Classification with missing-wedge compensation still produces almost a perfect classification even at a SNR of 0.0005, but here the algorithm begins to reach its limit. On the lower right corner of each graph, an inset shows a zoomed-in view of the clusters at SNR of 0.0005.

3.5.2. Comprehensive performance comparison to other algorithms

In order to perform a comprehensive comparison of the performance of the purposed KerDenSOM3D to other existing approaches, we performed classifications of the same simulated artificial thermosome data sets with missing-wedge 20 and 40, SNR from 10^{-4} to 10^{-2} using different techniques. For the constrained PCA, the approach and implementation described in [2] is used. For the original SOM and maximum likelihood (MLTOMO) we use the Xmipp package [136]. For the correlation approach as used for comparison in [94], we use maximum cross-correlation (MCC) and hierarchical clustering.

3. Classification of tomographic data sets

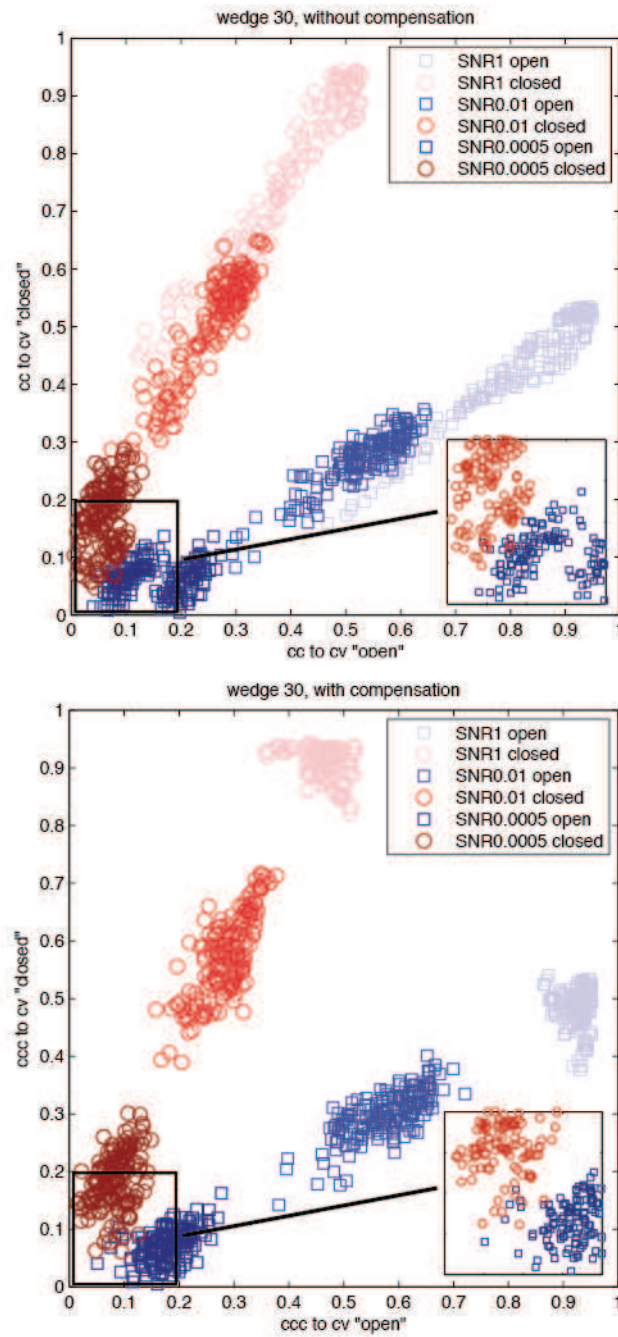


Figure 3.3 Same as the figure 3.2, of data sets with missing-wedge half opening angle of 30 degrees, which correspond to a tilt-series of ± 60 degrees.

The performance criterion used for this comparison is the class purity; the same as the one used in [130]. Class purity is calculated as the ratio of correctly classified sub-tomograms to the total number of sub-tomograms. We evaluate the performance of every individual algorithm and make a comprehensive comparison of the different approaches (Figure 3.4, Figure 3.5, Table 3.1, Table 3.2). For every approach, we tuned the parameter setting to achieve best possible classification results. In a direct comparison to the constrained PCA approach, which failed in obtaining an error-free classification with the simulated data set at $W = 40$ and $SNR = 10^{-3}$, KerDenSOM3D classifies the same data set perfectly. In comparison to the other listed approaches, KerDenSOM3D always shows a superior performance. The class-purity at lower SNR and larger missing-wedge angle is also always better in KerDenSOM3D compared to MLTOMO, MCC and SOM. The performance difference can be explained as the following: During KerDenSOM3D, each sub-tomogram is compared to a set of code-vectors that have higher SNR and are missing-wedge free. On the other hand, during constrained PCA and MCC, the covariance matrix is calculated by mutual comparison of the individual sub-tomograms. The procedures are thus much more sensitive to noise level and to missing-wedge.

Due to the extra computational effort required from the missing-wedge compensation strategy, which involves one Fourier transform and one inversed Fourier transform for every single calculation of similarity, the KerDenSOM3D is computationally more expensive than the constrained PCA, MCC and the simple SOM approaches. But one classification process could still be done on a single core of an Intel XEON E5620 processor at 2.4GHz within a few hours. Due to the nature of KerDenSOM3D, it could be easily parallelized onto multiple CPUs or GPUs, which allows a significant acceleration of the algorithm procedure. With missing-wedge compensation, the KerDenSOM3D approach outperforms all the previous algorithms. With our results, we also address that KerDenSOM3D as well as constrained PCA perform significantly better than the MCC approach followed by hierarchical clustering.

3. Classification of tomographic data sets

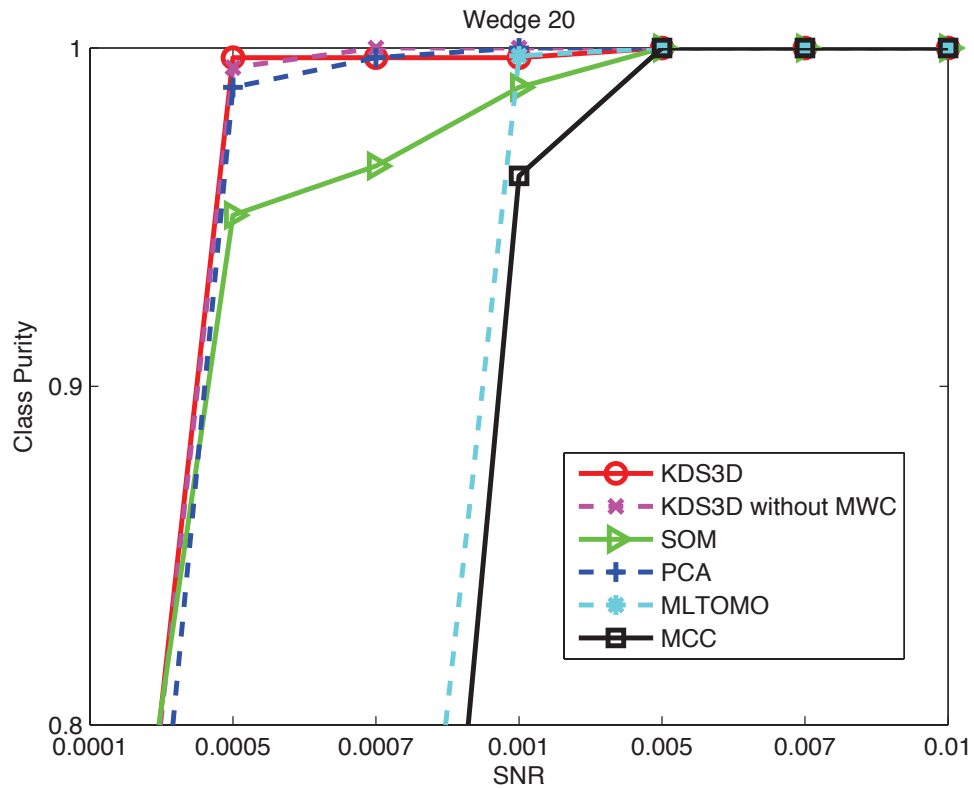


Figure 3.4 Comparison of several techniques based on the class purity at a missing-wedge half opening angle of 20 degrees. The abbreviations used in the graph are: KDS3D for KerdenSOM3D (red circle), KDS3D without MWC for KerdenSOM3D without missing-wedge compensation (purple x), SOM for Self-organizing map according to the simplest edition (green triangle), PCA for the constrained PCA (blue cross), MLTOMO for the maximum likelihood approach (light blue dot), and MCC for maximum cross-correlation (black square). KerdenSOM3D performs better than the previous techniques and the non-missing-wedge compensated algorithm.

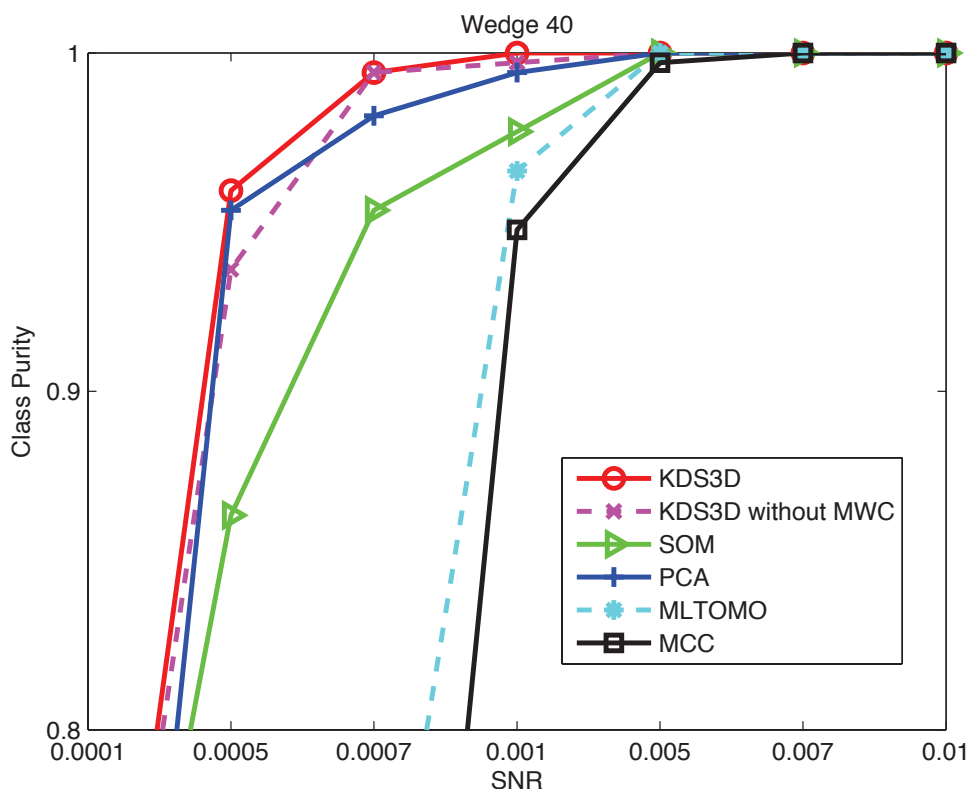


Figure 3.5 Same as figure 3.4 but with a missing-wedge half opening angle of 40 degrees. The improved performance of the algorithm is better appreciated for a higher missing-wedge angle of 40 degrees, where KerDenSOM3D outperforms any previous approach.

3.5.3. *In vitro* GroEL and GroEL/ES data sets

In [2] the classification performance of the missing-wedge corrected constraint PCA on the example of the *in vitro* GroEL and GroEL/ES tomograms was explored. The data sets have the advantage that we know from which tomograms the different species arise. In the GroEL/ES data sets, the majority of the GroEL molecules form a GroEL/ES complex. But as evidenced at a later stage, not all molecules have a GroES cap. This knowledge from the data sets allows us to perform the most accurate analysis under the most realistic and controlled conditions. For comparison purposes, the same sub-tomograms were used as in

3. Classification of tomographic data sets

Foerster et al. and Scheres et al.. The sample preparation: 1 μM GroEL14 was incubated with and without 5 μM GroES7 in a buffer containing 12.5 mM Hepes (pH 7.5), 5 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 5 mM ADP for 15 min at 30 °C. Protein solutions (3.5 μl) were mixed with 0.5 μl of a 10 nm BSA-colloidal gold suspension, applied to 300 mesh grids coated with Lacey carbon films (Plano, Wetzlar, Germany) and vitrified by plunge-freezing. Single-axis tilt-series were collected covering an angular range from -65° to $+65^\circ$ with 2° angular increment. Projection images were recorded on a 2kx2k pixel CCD camera at a nominal defocus level of $-4 \mu\text{m}$ to mimic the real experimental conditions of cellular tomography. The pixel size at the specimen level was 0.6 nm.

We then assess the classification performance of KerDenSOM3D on sub-tomograms extracted from these GroEL and GroEL/ES tomograms. A network of 5x5 code-vectors was used for the classification. The data set consists of 786 sub-tomograms, 214 of them extracted from the GroEL tomograms and 572 from the GroEL/GroES tomograms. The positions of the macromolecules are depicted manually. The extracted sub-tomograms are mutually aligned using the sub-tomogram averaging technique introduced in chapter 2.4. The majority of the sub-tomograms is well aligned. However, due to the low SNR and unfavorably orientated missing-wedge, completely misaligned sub-tomograms still exist.

The KerDenSOM3D classification algorithm separates the sub-tomograms into 25 classes and their class-averages are visualized in Figure 3.6. The complete network could be visually grouped into four major classes. On the left side of the network, class averages resembling a GroEL structure are present and indicated with “Class1”. The individual classes inside “Class1” differ slightly in the height of the GroEL density in x-direction, which may arise from the inaccurate sub-tomogram averaging process. The selection of specific classes according to the visual shape of the classes remains user-dependent. In the summarized “Class1”, 203 out of 214 sub-tomograms extracted from the GroEL tomograms are classified, providing a class-purity of at least 95%. In Figure 3.6, on the upper-right corner of the network, three classes resembling a GroEL/ES structure are indicated with “Class2”. In comparison to the “Class1”, the average density maps have an extra GroES density on the top of the GroEL density. Here 177 out of 181

sub-tomograms classify in this class are extracted from the tomograms containing GroEL/ES particles, results in a class purity of 98%. Beside these two expected populations from the data sets, two other class averages can be found in the network, they are namely “Class3” on the lower middle panel and “Class4” in the lower right corner of the network. These two classes together contain in total 181 sub-tomograms. They depict structures that are significantly smaller than a GroEL molecule. They are rather in the size of GroES and likely consist of (a) half particles (b) an outcome of radiation damage to the sample (c) misaligned particles. Analyzing the origin of the 181 sub-tomograms classified to these classes shows that the 96% of these sub-tomograms are extracted from the GroEL/GroES tomograms. This is the indication that a number of these sub-tomograms include GroES particles might not bind to GroEL. Not surprisingly, similar ring-structures were also obtained by [2] (Figure 3.8). The classification method introduced in [130] aligns and classifies the sub-tomograms iteratively, thus the structure for this class looks different (Figure 3.9).

3. Classification of tomographic data sets

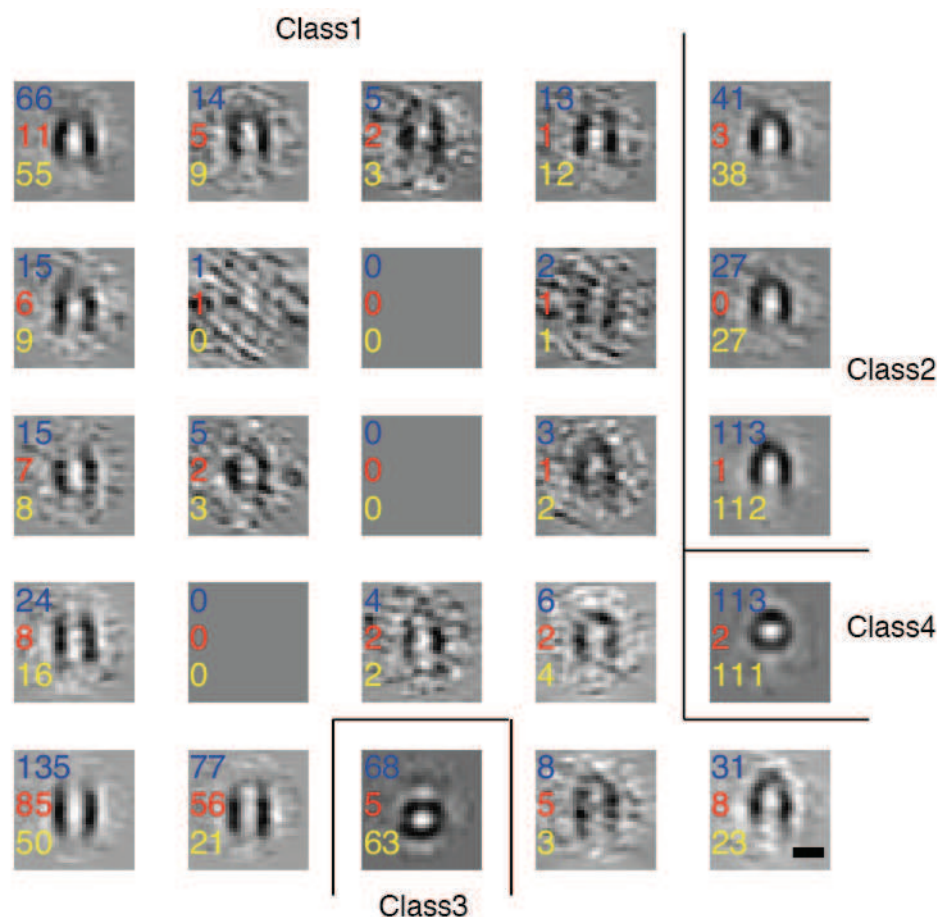


Figure 3.6 KerDenSOM3D network of classification of the *in vitro* GroEL and GroEL/GroES data sets. The central slice of the class averages in a 5x5 grid is visualized. In the left side of each class average, the number of sub-tomograms classifying to each code-vector is indicated. In blue, the total number of sub-tomograms in the class; in red, sub-tomograms originating from the GroEL dataset; in yellow, sub-tomograms originating from the GroEL/GroES dataset. Four classes can be distinguished, namely, Class 1 which represents GroEL (left part of the map); Class 2 which represents GroEL/GroES complexes (upper right corner of the map); and two classes with other particles (Classes 3 and 4). The vast majority of the half particles originate from the GroEL/GroES dataset. The scale bar in the lower right corner applies to all and is 10 nm in length.

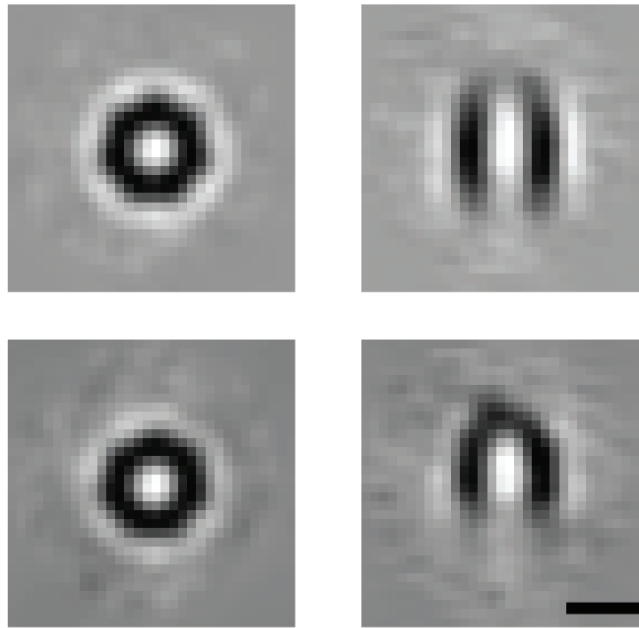
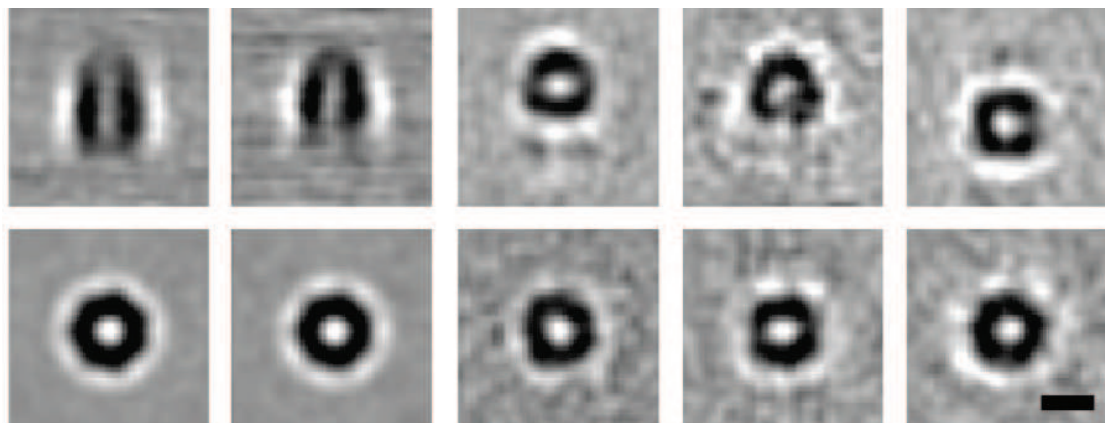


Figure 3.7 Upper row shows xy section and yz section of the averages of the particles classifying to a GroEL-similar class. A clearly 7-fold symmetry can be observed. Lower row shows the xy section and yz section of the average of the particles classifying to a GroEL/GroES-similar class. The scale bar is 10 nm.



| | | |
|-----|-----|--------------------|
| 323 | 269 | 194 (95% GroEL/ES) |
|-----|-----|--------------------|

Figure 3.8 Classification results of the same *in vitro* GroEL and GroEL/GroES data sets from Foerster et. al.. The scale bar is 10 nm.

3. Classification of tomographic data sets

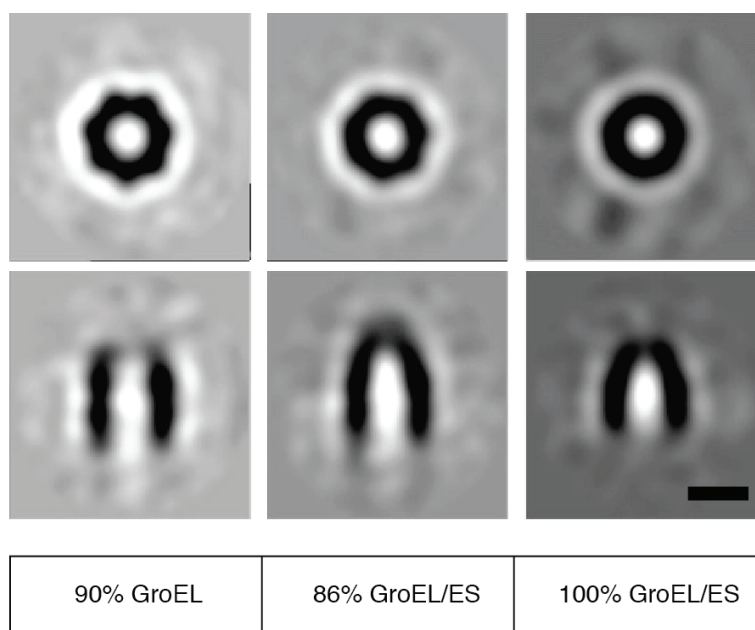


Figure 3.9 Classification results of the same *in vitro* GroEL and GroEL/GroES data sets from Scheres et. al.. The classes look different from the ones in this thesis and in Foerster et. al., as the sub-tomograms are iteratively aligned and classified. The alignment of the sub-tomograms is improved during the iterative procedure. The scale bar is 10 nm.

The *in vitro* GroEL and GroEL/ES data sets allow us to preform a direct comparison of the performance of KerDenSOM3D with constrained PCA and MLTOMO on real data sets. We analyze the performance of the algorithms based on the complete data sets of 786 sub-tomograms, and the results present in [2] and [130]. For both the GroEL and the GroEL/ES classes KerDenSOM3D delivers the highest class purity among all. The class purity obtained for the GroEL tomograms is 90 % for MLTOMO and 90.5% for constrained PCA, compared to 95% for KerDenSOM3D. The class purity for the GroEL/GroES tomograms is 86% for MLTOMO and 95.5% for constrained PCA, compared to 98% for KerDenSOM3D. Besides the higher class purity obtained by KerDenSOM3D, KerDenSOM3D also selects for every class the largest number of particles compared to the other two approaches, where a greater number of sub-tomograms was classified into classes with low occupancy and treated as debris.

This shows both qualitative and quantitative superiority (Table 3.1 and Table 3.2). Therefore, KerDenSOM3D delivers classes with higher class purities and greater numbers of sub-tomograms, which allow a better achievable resolution from the same data sets in the subsequent processing.

| Class Purity | KerDenSOM3D | ML_TOMO | PCA |
|--------------|-------------|---------|-----|
| GroEL | 95% | 90% | 90% |
| GroEL/ES | 98% | 86% | 95% |

Table 3.1: Comparison of the different methods on the basis of the class-purity. The class-purity in KerDenSOM3D is higher for both classes.

| | KerDenSOM3D | | | MLTOMO | | | PCA | | |
|----------------|-------------|--------|-----------|--------|--------|-----------|--------|--------|-----------|
| | Tot al | Gro EL | Gro EL/ES | Tot al | Gro EL | Gro EL/ES | Tot al | Gro EL | Gro EL/ES |
| GroEL Class | 424 | 203 | 221 | ? | 193 | ? | 377 | 182 | 195 |
| GroEL/ES Class | 181 | 4 | 177 | 150 | 21 | 129 | 222 | 10 | 212 |

Table 3.2: Comparison of the different methods on the basis of the total number of sub-tomograms associated to each class. From the complete set of 786 sub-tomograms KerDenSOM3D associated 424 sub-tomograms to GroEL and 181 to GroEL/ES. The remaining particles are associated to other classes. In contrast PCA associated 377 sub-tomograms to GroEL and 222 to GroEL/ES. The absolute number of particles associated to the classes in KerDenSOM3D is larger. The absolute number of particles classified in the class GroEL/ES with MLTOMO is unknown and marked with question mark.

4. Reference bias in reference based techniques

There is a growing awareness that cross-correlation based template matching techniques can introduce a significant amount of reference bias that results in uncertain detections or false-positive hits. This is because high cross-correlation values may be assigned to positions that do not represent the sought macromolecular structure. Alternatively, the search may succeed in finding the correlating macromolecule but in a wrong orientation due to the high noise level of the tomograms [4]. While sub-tomogram averaging is less prone to reference bias than template matching, because the heterogeneity of the data is lower, the bias still plays a significant role.

Previous studies have shown that reference bias is present in single particle analysis approaches as well. In a recent evaluation [5], the 6 Å structures of HIV's surface proteins gp120 and gp41, which are published in [137], are suspected to be indeed structures of phantom molecules. The samples, from which the data were recorded, were doubted to contain any HIV particles. Reference based methods are subject to reference bias. Given enough degree of freedom, the reference can be recovered from even pure noise. It has been shown and cited by many publications that 1000 computer-generated images of white noise can be aligned to create a picture of Einstein sticking out his tongue.

In single particle analysis there are numerous benefits such as higher SNR, lower degree of freedom (DOF) of the alignment and lower heterogeneity that may act favorably towards a reduced reference bias in comparison to CET [4]. Thus, in combination with well-developed classification techniques such as RELION [138, 139] and FREALIGN [140], the number of false-positives and misaligned particles can be successfully reduced. Therefore, the amount of reference bias can be effectively limited. In tomography however, especially in the case of template matching, the bias is mostly already introduced at low frequencies.

4. Reference bias in reference based techniques

Therefore, the confidence that even the highest cross-correlation values truly represent properly aligned macromolecular complexes of interest is generally low.

4.1. Estimation of reference bias in Fourier space

The reference bias in single particle analysis is addressed in [6] by analyzing the behavior of the FSC curves. The FSC curves in which the whole Fourier data is used for the alignment are compared to the FSC curves in which certain Fourier shells are excised. When real data with sufficient high SNR are aligned, the differences between the curves are minimal, indicating marginal amount of reference bias. However, when pure noise are aligned, the FSC within the excised shells drops to nearly zero, indicating that the bias introduced by the reference is high. Methods such as the “gold standard” are developed to minimize the reference bias [7, 108]. The core of the “gold standard” approach is to split the data set into two halves and then proceed the alignment of each half of the images using independently generated initial reference structures. The FSC between the 3D models derived from the two halves measures the self-consistency of the data set and gives an estimation of the achieved resolution, but does not assess the amount of reference bias. In [8], data above a certain frequency was substituted by noise in order to assess the amount of overfitting during the refinement process. The FSC curve derived from the approach gives a more accurate estimation of the resolution. In a similar fashion, high frequency data was excised during flexible fitting to assess the model bias introduced by the initial model [141].

In tomography, analogous to the experiments in [6], high-frequency data is excised during the sub-tomogram averaging process of artificial GroEL data sets. The different behavior between the FSC where the whole Fourier data is included versus the FSC where a shell is excluded is investigated. As illustrated in Figure 4.1, the FSC curves cannot be discerned from each other when sub-

tomograms containing real information with sufficient high SNR are aligned, indicating no reference bias (Figure 4.1a). When the SNR decreases, the FSC where high-frequency data is excluded is slightly lower than the FSC where all data is included, showing minimal amount of reference bias (Figure 4.1b). Hence, the sub-tomogram averaging process can be still considered as reliable. However, when data containing only noise is aligned, the FSC where high-frequency data is excluded drops to nearly zero (Figure 4.1c). Not surprisingly, almost no difference could be observed from the three averages, when all of them are low-pass filtered to the excised frequency (Figure 4.1d-f). However, while the averages in Figure 4.1d and 4.1e are averaged out of real data of GroEL, the average in Figure 4.1f emerges exclusively from random noise, through the effect of reference bias. The behavior of the FSC curves in this experiment matches the observations in [6]. Therefore, the method introduced in [6] as well as the argument and conclusion derived in the paper is also applicable to tomography.

4.2. Estimation of reference bias in real space: M-free score

From the above experiments, we can conclude that the effect of reference bias exists in single particles analysis as well as in tomography. The methods of exclusion or substitution of data in Fourier space, which aim to study the influence imposed by the initial reference model, are difficult to transfer to sub-tomogram averaging [4, 113], and even more to template matching. The primary prerequisite for improving the applicability of template-based approaches is a figure of merit that assesses the quality of the search, limits the reference bias, and reliably estimates the amount of false positives. For the template matching approach, the real space method of omitting an area from the search was proposed in [14]. The method was shown to increase the specificity of the template matching of *in vitro* ribosome tomograms [142]. However, the amount of false-positives, and moreover the reference bias, could not be estimated.

4. Reference bias in reference based techniques

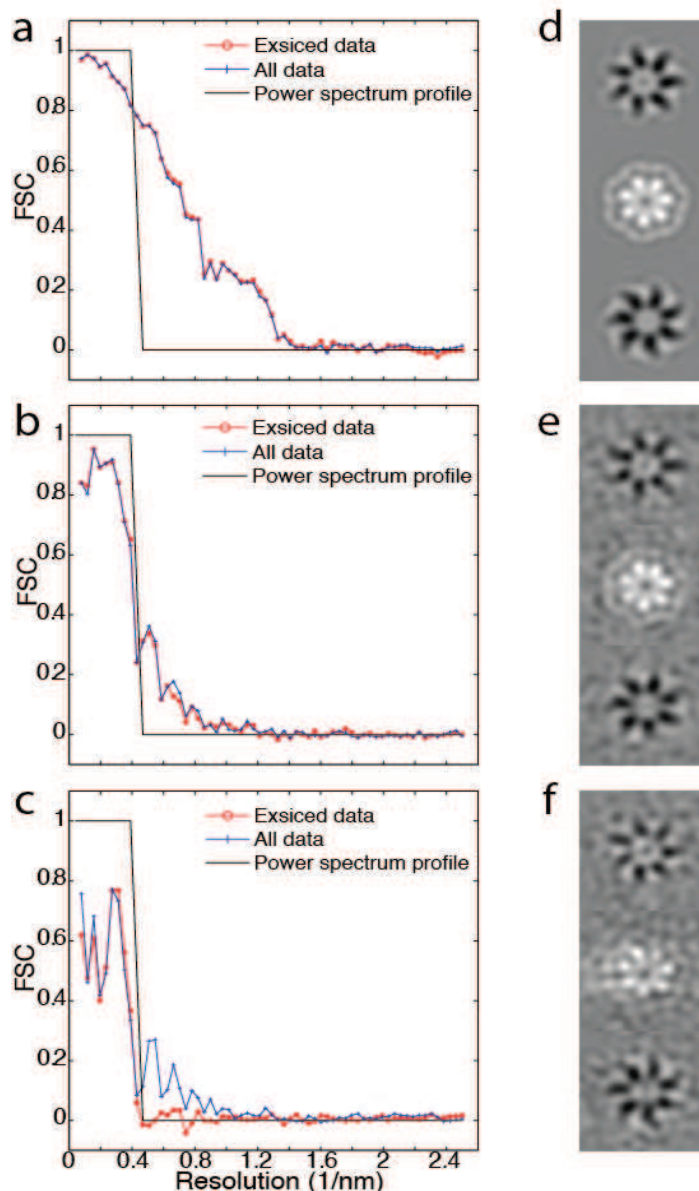


Figure 4.1: Fourier shell correlation curves and the average slices for real data with different SNRs and for pure noise. FSC curves were generated from a single round of alignment of 100 sub-tomograms either of real data (a) and (b) or noise (c). (a) Alignments of real data carried out using all Fourier data from the reference in blue and the Fourier data above 2.5 nm^{-1} excised in red. The curves hardly differ from each other (b) Same as (a), but of real data with lower SNR. The red curve drops slightly from the blue one at excised frequencies. (c) Same as (a) and (b), but of random noise. The red curve holds on with the blue one but exclusively at the included frequencies. At excised frequencies above 2.5 nm^{-1} the red curve drops to almost 0.

In this thesis, a heuristic named “M-free” that is conceptually based on the method introduced in [14] and [142], which operates in real space, was developed. The M-free score estimates the bias imposed by the reference in both template matching and sub-tomogram averaging procedures. The real space mask, which outlines the area of the macromolecules in real space, has been splitted into a *working area* and a *testing area*. While the signal in the *working area* is used for the alignment process, the signal from both areas is explored to access the reference bias by calculating the M-free score. The theory of the M-free score is based on the homogeneity and coherency of the underlying genuine signal [9]. The value of any voxel in an image containing underlying signal is correlated with the value of any other voxel in the image. The reference bias on the other hand, which is imposed by random noise, has no coherence. This means that alignment of the underlying genuine signal in the *working area* will also automatically align the signal in the *testing area*. The aligned underlying genuine signal will result in homogenous average signals in the *testing* and *working areas*. The average of the noise imposed by reference bias lacks the coherency and resembles the reference density exclusively within the *working area*, where the noise was aligned. The M-free score reflects the different behavior of the average density map in the *working* and *testing areas* and can therefore be used to estimate the amount of reference bias. A schematic flowchart representing the process is given in Figure 4.2.

4. Reference bias in reference based techniques

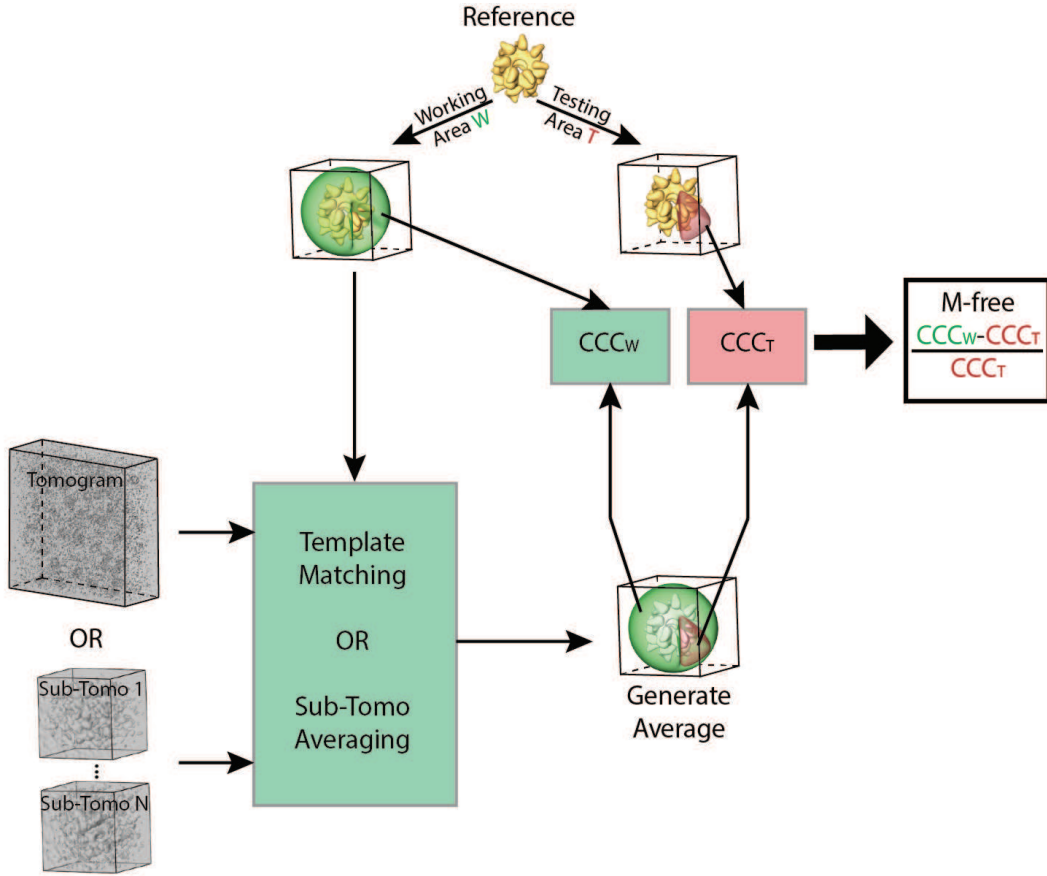


Figure 4.2: Schematic flow diagram of the M-free heuristic. The start reference is first splitted into a *working area* W in green and a complementary *testing area* T in red. The information exclusively in the *working area* is then used to perform template matching or sub-tomogram averaging. The generated average volume is then compared to the start reference in both *working* and *testing areas* by calculating the cross-correlation, from which the M-free score is determined.

4.3. Detailed derivation of M-free score

A brief derivation of the M-free score is given in the paper II. The more detailed version of the mathematics can be found in the following paragraph.

The core of the M-free heuristic is the separation of the mask applied on the reference into a *working area* (W) and a *testing area* (T), in a similar fashion to [143]. W and T are disjoint and their conjunction is the full mask area, that is:

$$W \cap T = \emptyset \text{ and } W \cup T = ALL . \quad (\text{Eq. 4.2.1})$$

We consider the average of the sub-tomograms $p = s + n$, where $n = n_{bias} + n_{ran}$.

We dissect the signal into three types: The macromolecular signature s , the noise which is considered to be composed of random noise n_{ran} , and biased noise n_{bias} . Since the signal in the *testing area* never "sees" the reference, the amount of biased noise in the *testing area* $n_{bias,T} = 0$, while $n_{bias,W}$ in the *working area* is what we aim to measure. The reference is considered to be uncorrelated with the random noise, and correlated to the biased noise, in a similar fashion to [8].

The normalized cross-correlation coefficient between the reference r and the average of the sub-tomograms p , with r and p being mean-value free within the area of the mask, within the *working area* can be expressed as:

$$\begin{aligned}
 CCC_W(r, p) &= \frac{\sum_W rp}{\sqrt{\sum_W r^2 \sum_W p^2}} \\
 &= \frac{\sum_W(rs + rn_{bias,W} + rn_{ran})}{\sqrt{\sum_W r^2 \sum_W (s^2 + n_{bias,W}^2 + n_{ran}^2 + 2sn_{ran} + 2n_{bias,W}n_{ran} + 2sn_{bias,W})}} \\
 &= \frac{\sum_W(rs + rn_{bias,W})}{\sqrt{\sum_W r^2 \sum_W (s^2 + n_{bias,W}^2 + n_{ran}^2 + 2sn_{bias,W})}} \quad (\text{Eq. 4.2.2})
 \end{aligned}$$

Although the term $2sn_{bias,W}$ under the square root in the denominator is unequal zero, it is considerably small in comparison to the other terms. This term could be therefore neglected, in a similar fashion to the neglected terms of SN_{bias} in Eq. (1) in [8].

The Eq. (4.2.2) is simplified:

$$CCC_W(r, p) = \frac{\sum_W(rs + rn_{bias,W})}{\sqrt{\sum_W r^2 \sum_W (s^2 + n_{bias,W}^2 + n_{ran}^2)}} = \frac{\sum_W(rs + rn_{bias,W})}{\sqrt{\sum_W r^2 \sum_W (s^2 + n^2)}} \quad (\text{Eq. 4.2.3})$$

where $\sum_W n^2$ is the energy of the total noise in the *working area*.

Within the *testing area* the information has not been used for the alignment thus $n_{bias,T} = 0$, and the normalized cross-correlation coefficient can be expressed as:

4. Reference bias in reference based techniques

$$\begin{aligned}
CCC_T(r, p) &= \frac{\Sigma_T r p}{\sqrt{\Sigma_T r^2 \Sigma_T p^2}} \\
&= \frac{\Sigma_T (rs + rn_{bias,T} + rn_{ran})}{\sqrt{\Sigma_T r^2 \Sigma_T (s^2 + n_{bias,T}^2 + n_{ran}^2 + 2sn_{ran} + 2n_{bias,T}n_{ran} + 2sn_{bias,T})}} \\
&= \frac{\Sigma_T rs}{\sqrt{\Sigma_T r^2 \Sigma_T (s^2 + n_{ran}^2)}} = \frac{\Sigma_T rs}{\sqrt{\Sigma_T r^2 \Sigma_T (s^2 + n^2)}}, \tag{Eq. 4.2.4}
\end{aligned}$$

where $\Sigma_T n^2$ is the total energy of the total noise in the *testing area*.

The ratio of Eq. (4.2.3) and (4.2.4) can be written as:

$$\begin{aligned}
\frac{CCC_W(r, p)}{CCC_T(r, p)} &= \frac{\frac{\Sigma_W (rs + rn_{bias,W})}{\sqrt{\Sigma_W r^2 \Sigma_W (s^2 + n^2)}}}{\frac{\Sigma_T rs}{\sqrt{\Sigma_T r^2 \Sigma_T (s^2 + n^2)}}} \\
&= \frac{\Sigma_W (rs + rn_{bias,W})}{\Sigma_T rs} \cdot \frac{\sqrt{\Sigma_T r^2 \Sigma_T (s^2 + n^2)}}{\sqrt{\Sigma_W r^2 \Sigma_W (s^2 + n^2)}} \tag{Eq. 4.2.5}
\end{aligned}$$

We choose W and T such that the variance of the reference within each region is approximately the same. We also assume that the motif signal s has approximately the same variance in both areas. The variance of the total noise n is approximately equal in both areas.

The energies can be written in the following way:

$$\begin{aligned}
\Sigma_W r^2 &= k \cdot \Sigma_T r^2 \\
\Sigma_W s^2 &= k \cdot \Sigma_T s^2 \\
\Sigma_W n^2 &= k \cdot \Sigma_T n^2, \tag{Eq. 4.2.6}
\end{aligned}$$

where k is a constant describing the quotient of the size of the *working area* to the *testing area*.

Thus equation (4.2.5) can be reformulated as:

$$\frac{CCC_W(r, p)}{CCC_T(r, p)} = \frac{\Sigma_W (rs + rn_{bias,W})}{\Sigma_T rs} \cdot \sqrt{\frac{\Sigma_T r^2}{\Sigma_W r^2}} \cdot \sqrt{\frac{\Sigma_T s^2 + \Sigma_T n^2}{\Sigma_W s^2 + \Sigma_W n^2}} \tag{Eq. 4.2.7}$$

and further simplified as:

$\frac{CCC_W(r,p)}{CCC_T(r,p)} = \frac{\sum_W(rs+rn_{bias,W})}{\sum_T rs} \cdot \sqrt{\frac{1}{k}} \cdot \sqrt{\frac{1}{k}} = \frac{\sum_W(rs+rn_{bias,W})}{k \cdot \sum_T rs}$, and we can write:

$$\frac{CCC_W(r,p)-CCC_T(r,p)}{CCC_T(r,p)} = \frac{\sum_W rn_{bias,W}}{\sum_W rs}, \quad (\text{Eq. 4.2.7})$$

with $\sum_W rs = k \cdot \sum_T rs$, which following our hypothesis that an ideal signal s results in the same normalized cross-correlation coefficient between s and r in the *testing* and the *working areas*.

Using the vector form of the cross-correlation function between the signals in the *working area*, Eq. (4.2.7) could be written as the ratio of the scalar products:

$$\frac{CCC_W(r,p)-CCC_T(r,p)}{CCC_T(r,p)} = \frac{R \cdot N_{bias,W}}{R \cdot S} = \frac{\|R\| \|N_{bias,W}\| \cos \theta_{R,N_{bias,W}}}{\|R\| \|S\| \cos \theta_{R,S}} \quad (\text{Eq. 4.2.8})$$

If the reference \mathbf{R} is similar to the signal \mathbf{S} , which is valid for most of the real cases, the magnitude of the angle between the vector \mathbf{R} and \mathbf{S} is similar to the magnitude of the angle between \mathbf{R} and $\mathbf{N}_{bias,W}$, which by definition has to be similar. Thus in our effort to express the M-free score independently from the reference r we assume the cosine of the angles $\theta_{\mathbf{R},\mathbf{N}_{bias,W}}$ and $\theta_{\mathbf{R},\mathbf{S}}$ to be equal. The right hand side of Eq. (4.2.8) can then be written as:

$$\frac{CCC_W(r,p)-CCC_T(r,p)}{CCC_T(r,p)} = \frac{\|\mathbf{N}_{bias,W}\|}{\|\mathbf{S}\|}. \quad (\text{Eq. 4.2.9})$$

Thus the term $\frac{CCC_W(r,p)-CCC_T(r,p)}{CCC_T(r,p)}$ that we name M-free gives an estimate of the reference bias $\mathbf{N}_{bias,W}$ on a specific signal \mathbf{S} . The CCC_T in the denominator of Eq. (4.2.9) can be negative, showing an anti-correlation between the reference and average in this area, which the magnitude is irrelevant. Thus we consider negative values to be equal to zero resulting in a M-free score range $[0 +\infty)$, with 0 indicating no template bias, and $+\infty$ indicating non-existence of a macromolecular signature. The assumptions, which were necessary in order to allow the description of the reference bias by measurable cross-correlation values, affect the correctness of the M-free score only at very high bias.

4. Reference bias in reference based techniques

4.4. Results

In the process of sub-tomogram averaging and template matching, a real space mask is applied around the region of interest on the reference to concentrate the alignment on the feature of interest and to reduce the influence of noise. This mask also defines the area in which the variance for the denominator of the cross-correlation function is calculated. In theory, the real space masks can be arbitrarily shaped. During the alignment, it is rotated with the reference and always applied on the same region of the reference. For the sake of measuring the M-free score, ~10% of the mask area is reserved as the *testing area*. The information within the *testing area* is not included in the calculation of the cross-correlation (Figure 4.3). The *testing area* is selected in such a way that the variance of the signal is equal to the signal variance in the *working area*, as required for the derivation of M-free in chapter 4.3. In similar fashion to the complete mask, the *testing area* can also have arbitrary shapes.

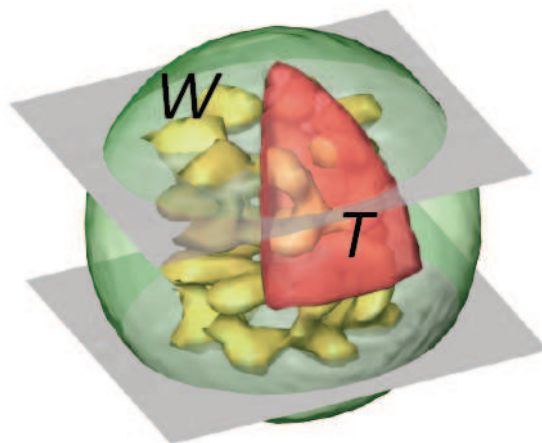


Figure 4.3 An illustration of the *working area* W in green and the *testing area* T in red encircling the isosurface of the GroEL in yellow.

4.4.1. Simulated data sets

We first explored the M-free score in sub-tomogram averaging of simulated GroEL data sets. The same simulation procedure is used to create data sets in chapter 3.5.1 for evaluation of the classification algorithms. The data sets have different SNR, random orientations and different half opening angle of missing-wedge. We performed sub-tomogram averaging using the GroEL structure from PDB as a starting reference. For data sets with SNR above 0.003, the sub-tomogram averaging produces averages that have fully 7-fold symmetry. The averages all show well-defined densities in both the *working area* and the *testing area* (Figure 4.4a-d). In comparison, the sub-tomogram averaging of the data sets with SNR below 0.003 do not produce average densities in the *testing area* (Figure 4.4e-f). We calculate the M-free score from the averages for all SNR and analyze the results. For the averages from data sets with SNR above 0.003 the M-free score is close to zero. At SNR lower than 0.003, the M-free score raises exponentially (Figure 4.5). This shows a fundamentally different characteristic to the cross-correlation value between the reference and the sub-tomogram average, which slowly reduces in a linear fashion depending on the SNR. The cross-correlation value is also dependent on the number of sub-tomograms, with data sets containing more sub-tomograms the curve would drop slower. Examining the reason for the alignment failure shows that for a number of particles the proper rotation and translation parameters were not found correctly.

4. Reference bias in reference based techniques

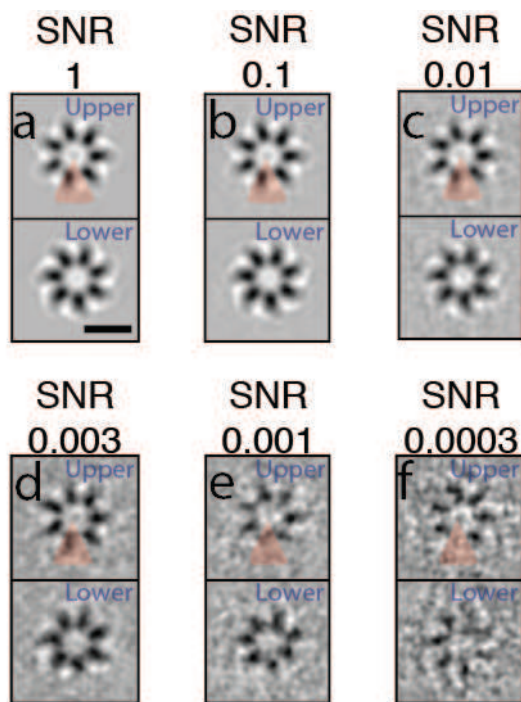


Figure 4.4 Two computational sections through the apical domains intersecting with the *testing* and *working area*, respectively. The slices of the averages from data sets with SNR (a) 1, (b) 0.1, (c) 0.01, (d) 0.003, (e) 0.001, and (f) 0.0003. In the upper slices, the *testing area* is indicated with transparent red. The M-free scores measured are 0, 0, 0, 0.05, 0.4, and 12 respectively. The scale bar is 10 nm.

As next experiment, we exam the validity of the M-free score on simulated data sets with contamination. The contaminated data sets have a SNR of 0.003, which is on the lower limit of SNR necessary for success of sub-tomogram averaging. The contamination that has been added to the data sets of the simulated GroEL is thermosome particles with the same half opening angle of missing-wedge and SNR at 0.003. The idea is that, at this critical SNR, the GroEL sub-tomograms could be correctly aligned using a GroEL as starting reference, but not the thermosome sub-tomograms. Therefore, the amount of misaligned sub-tomograms is equal to the amount of contamination at this SNR. The M-free score shows a linear dependency on the percentage of contamination in the data sets, that is, on the amount of misaligned sub-tomograms (Figure 4.6).

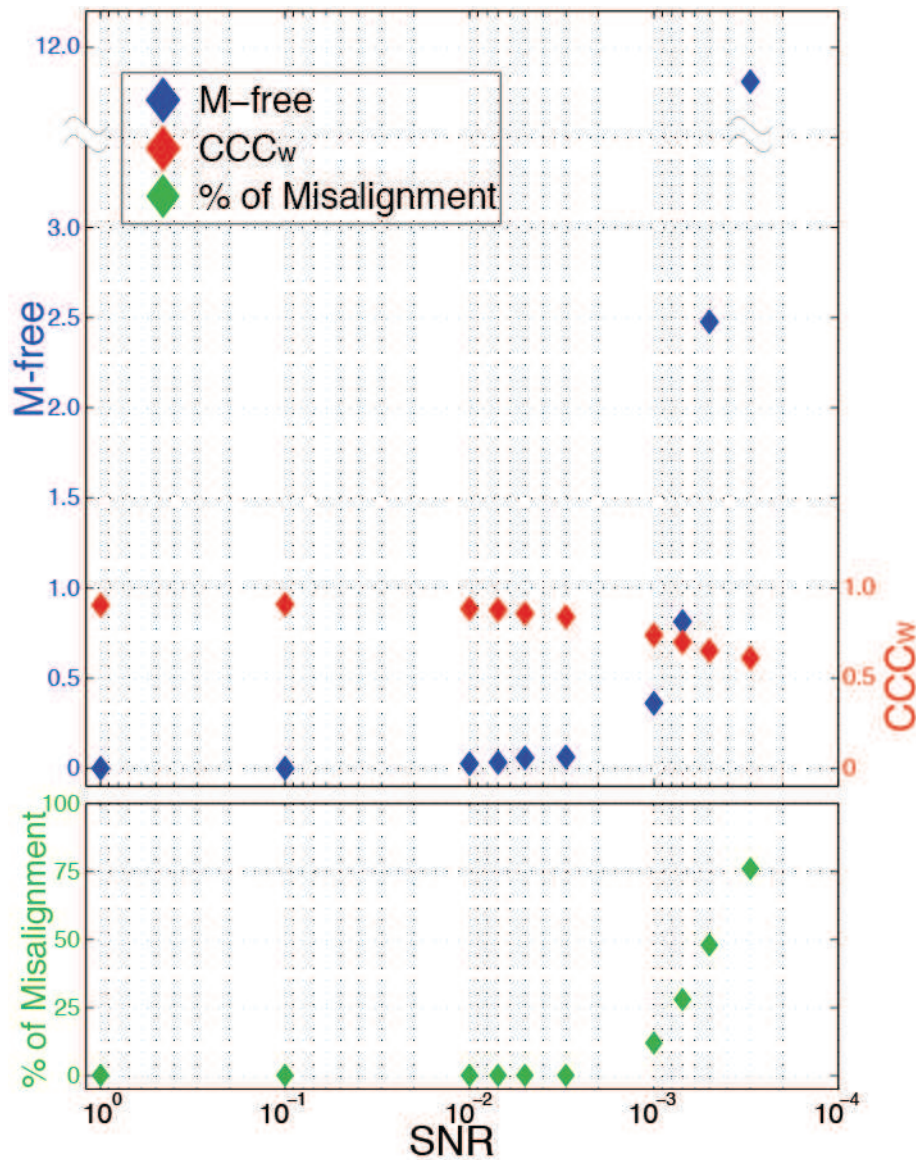


Figure 4.5 Upper graph shows M-free scores measured for data sets with different SNRs aligned using GroEL as a starting reference. M-free scores remain at ~ 0 until the SNR reaches 0.003. When the SNR decreases further, the M-free score rises exponentially. In comparison, the cross-correlation value decreases slowly. More importantly, the cross-correlation value depends on the number of particles, while the M-free score is independent of this criterion. Lower graph shows the percentile of misaligned particles as a function of the SNR. The M-free score appropriately reflects the amount of particles for which the orientation and/or translation was not found properly due to the SNR.

4. Reference bias in reference based techniques

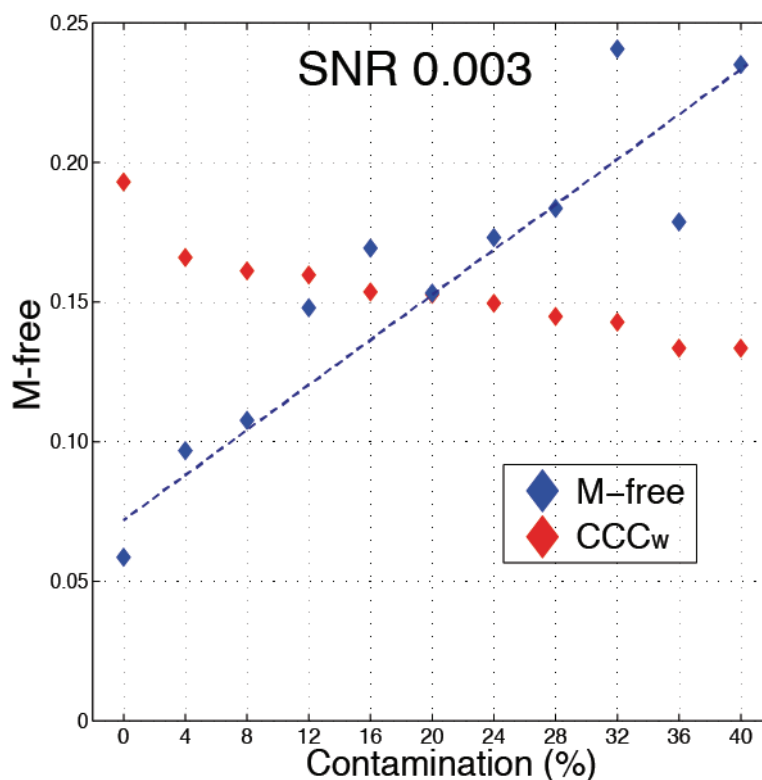


Figure 4.6 Results of sub-tomogram averaging experiments with contaminated artificial GroEL data sets. The M-free scores and cross-correlation values measured from sub-tomogram averaged data sets with an SNR equal to 0.003 and a contamination of 0% to 40% in 4% steps.

We then performed template matching on the artificial tomograms containing GroEL particles contaminated with thermosome. We performed an exhaustive search with six DOF: three for translation and three for rotation. With the experiment of template matching, the M-free score is even more informative. When the GroEL template is used to search in an artificial tomogram containing exclusively GroEL, the M-free score is measured at 0.16. When the template matching is performed with the same GroEL template on the tomogram with 40% contamination, a much higher M-free score at close to 0.7 is measured. In case no GroEL is present in the tomogram, the template matching procedure delivers an average from which the M-free score is measured at 17. In contrast to the M-free score, the cross-correlation value between the average and the GroEL reference in the above experiments are 0.81, 0.80 and 0.72 successively and

shows almost no difference (Figure 4.7). Concluding, rather than the cross-correlation value, the M-free score gives a meaningful value that reflects the bias. The M-free score is characteristic for the amount of contamination and reflects the quality of the alignment.

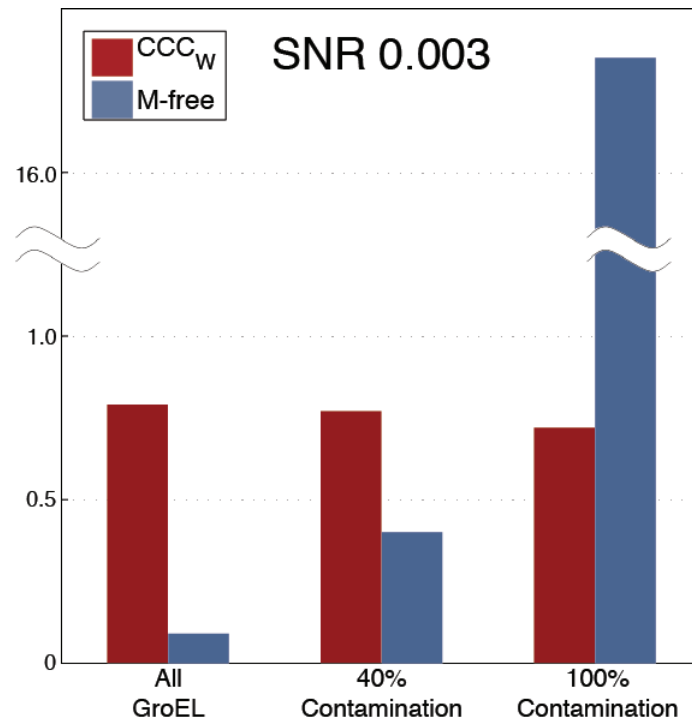


Figure 4.7 Results of template matching experiments with contaminated artificial GroEL data sets. The bar chart shows M-free scores and cross-correlation values. While the cross-correlation value stays approximately constant the M-free score rises significantly.

4.4.2. *In vitro* GroEL data sets

To validate the M-free score at realistic experimental conditions, we use one tomogram from the GroEL data sets introduced in chapter 3.5.2 (Figure 4.8). With this tomogram, we performed both sub-tomogram averaging and template matching experiments.

4. Reference bias in reference based techniques



Figure 4.8 An 18nm thick projection slice of the tomogram. The scale bar is 100 nm. In the inset, an illustration of the *testing area* is shown in red in relation to the isosurface of the GroEL derived from the crystal structure (PDB 2C7E) in yellow.

First, we manually depicted ~ 100 positions corresponding to putative GroEL particles and performed sub-tomogram averaging either using the GroEL crystal structure (Figure 4.9 upper left) [144] as the starting reference or a Gaussian blob (Figure 4.9 lower left). We used a spherical mask, from which the *testing area* was a wedge representing 10% of the total volume, in the same fashion as in the experiments with artificial data sets (Figure 4.8 inset). The variance within the *testing* and *working area* is the same. The sub-tomogram averaging was carried out for 5 iterations, with the M-free score calculated after every iteration. For the final structures, the M-free score for the GroEL crystal structure approach was 0.74 while the M-free score for the Gaussian sphere (Figure 4.9) approach was estimated at 0.08. The cross-correlation value is 0.82 for both approaches and shows no difference (Figure 4.11). The visual comparison between the two averages in Figure 4.9 shows that in the case of the PDB reference the density does not recover, and in the *testing area* an electron lucent area remains.

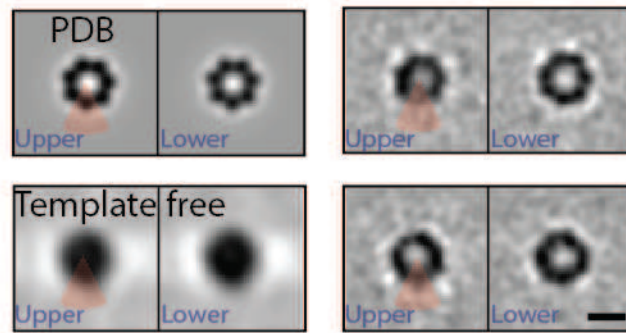


Figure 4.9 Upper left: Two slices through the upper and lower apical domains of the density map derived from the GroEL crystal structure. The *testing area* is marked with transparent red throughout all images. Upper right: The same slices through the sub-tomogram average after 5 iterations using the GroEL crystal structure as a starting reference. The M-free score is 0.74. Lower left: The slices of the Gaussian blob used for “template-free” sub-tomogram averaging. Lower right: The two slices of the result of “template-free” sub-tomogram averaging after 5 iterations. The M-free score is 0.08. Scale bar is 10 nm.

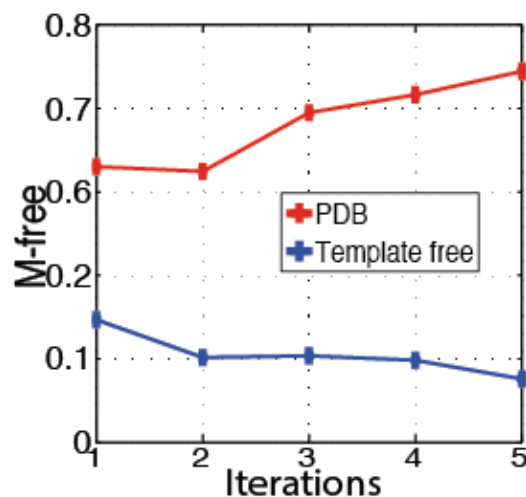


Figure 4.10 The development of the M-free score depending on the number of iterations. The red curve shows the M-free scores with GroEL crystal structure as a starting reference, while the blue curve shows the “template-free” approach.

4. Reference bias in reference based techniques

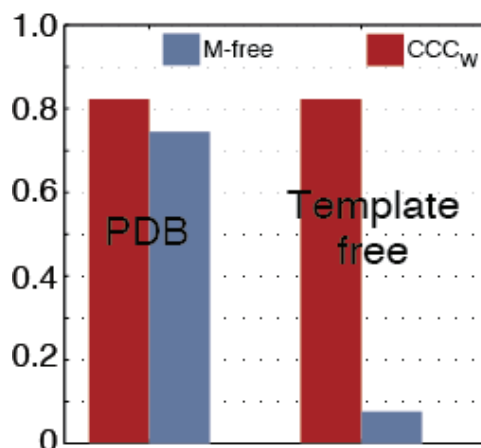


Figure 4.11 A bar chart where M-free scores are shown in blue and the cross-correlation coefficients calculated in the *working area* (CCC_w) are shown in red, measured from the averages in Figure 4.9. The CCC_w do not differ from each other in the different experiments, however the M-free scores are 10-fold higher in experiments using a crystal-structure-derived template rather than a Gaussian sphere.

After the experiments of sub-tomogram averaging, we also tested the M-free score with template matching on the same tomogram. In this experiment, the complete tomogram is scanned for putative particles. We scanned the complete tomogram exhaustively in two independent experiments.

We first searched with the average obtained from the sub-tomogram averaging using the template-free strategy. First, there exists no hard threshold of the cross-correlation value from which the putative positions of the sought macromolecule should be selected. Second, the number of putative particles is unknown. Therefore, we selected all the 716 maximum peaks of the resulting cross-correlation map. For this average the M-free score is calculated at 1.5. For the first 200 peaks corresponding to the highest cross-correlation values the M-free score is 0.59. When the next 200 peaks are added, the M-free score increases to 0.60, indicating similar amount of bias. When the 401st to the 600th peaks are selected, the M-free score rises to almost 20. Summarizing the results, while the highest 400 peaks have similar heterogeneity, the next 200 contain barely particles. The cross-correlation value between the template and the average

increases slowly depending linearly on the number of selected sub-tomograms (Figure 4.12). Interestingly the FSC also shows increasing resolution as more sub-tomograms are selected. Calculating the M-free score of the fraction of the peaks corresponding to the manually selected sub-tomograms results in a score of 0.09 (Figure 4.13), which is similar to the score in the sub-tomogram averaging experiment indicating a marginal amount of bias.

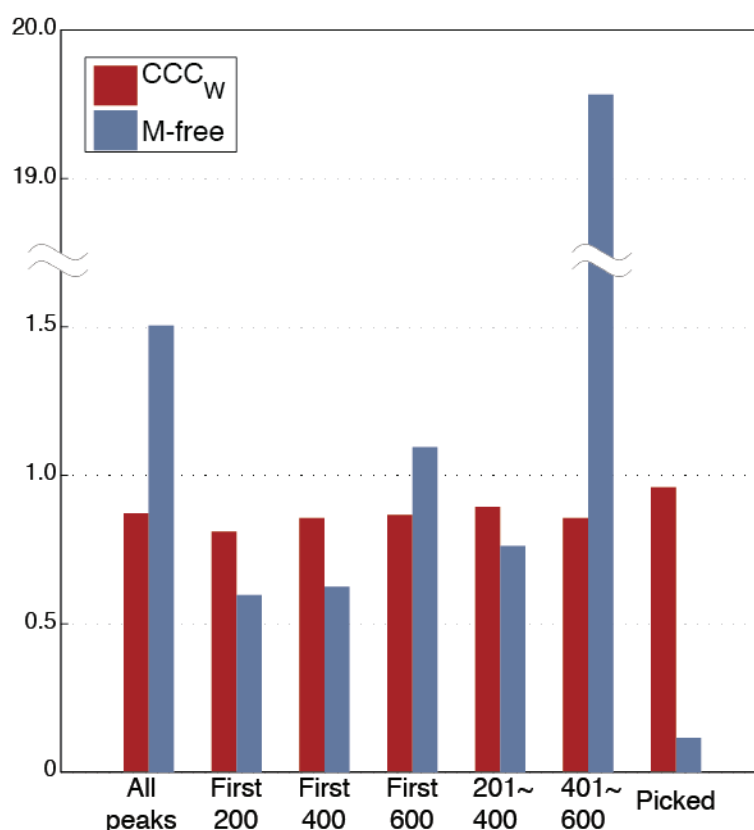


Figure 4.12 A bar diagram of the M-free scores in blue and the cross-correlation coefficients in red, measured from averages generated from batches of sub-tomograms selected according to the peaks detected from the template matching with a GroEL template. The M-free scores are similar for the averages generated from the 200 highest peaks, the 201st – 400th peaks and all the 400 highest peaks together. The average created from the 401st – 600th peaks has a much higher M-free score. An average generated from the manually selected positions has the lowest M-free score of 0.08, which indicates that when the amount of contamination is low, the M-free score is very low. In contrast, the cross-correlation coefficients in W are similarly high for all averages.

4. Reference bias in reference based techniques

When we performed essentially the same experiment but using the density map derived from the crystal structure as a template, the M-free score is almost 0.8 for the manually selected peaks (Figure 4.13). The cross-correlation again shows almost no difference (Figure 4.14).

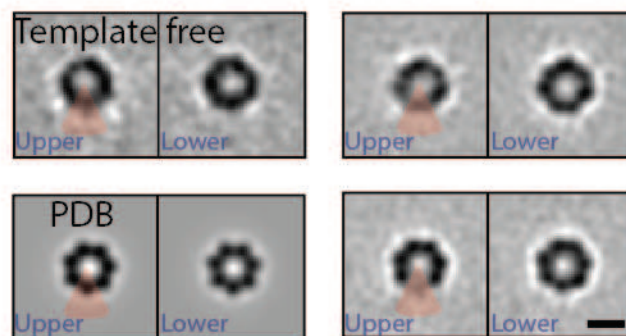


Figure 4.13 Upper left: The result of the sub-tomogram averaging used as a template for the template matching. Upper right: The two xy slices of the result from template matching using the average on the left as a reference. The M-free score was calculated to be 0.08. Lower left: Two slices through the upper and lower apical domains of the density map derived from the GroEL crystal structure. Lower right: The two xy slices of the result from template matching using the reference from the crystal structure. The M-free score is 0.80. The density in the *testing area* is recovered much better than the upper one. Scale bar is 10 nm.

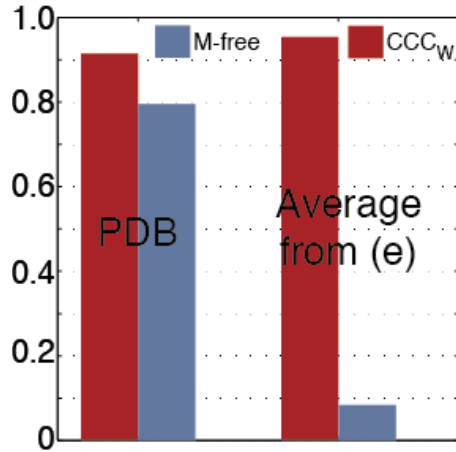


Figure 4.14 A bar chart for M-free and CCC_w for template matching. As in figure 4.11.

When we scan the same tomogram containing only GroEL with a thermosome template (pixel size and CTF adapted to the recording conditions of the tomogram with the same mask used for GroEL), the distribution of the cross-correlation values is very similar compared to the ones of the GroEL template. However, the M-free score is infinite, indicates high reference bias (Figure 4.15) and non-existing of underlying signal from thermosome. Thus, while the cross-correlation value does not discern between real signal and reference bias, the M-free score again shows major differences.

4.4.3. Sub-tomogram averaging of SIV Spikes

We then exam the sub-tomogram averaging process of SIV spikes in [111]. We performed the procedure exactly as described in [111] and aimed to reproduce their result. As the real space mask, we used an elliptical mask from which a wedge of 10% the volume was carved out as *testing area* (Figure 4.15a-c). This mask can be varied significantly without changing the outcome of the processing. The same processing procedure was carried out: The positions were manually depicted as described in the paper. The sub-tomograms are then pre-aligned

4. Reference bias in reference based techniques

according to their spatial position on the surface of the virus, in regard to the center of the virus. As result from the pre-aligned sub-tomogram, we obtain a globular structure on top of a membrane patch (Figure 4.15b). With progressing iterations a structure with a resolution of 2.8 nanometer was reached (Figure 4.15c-d), which is judged by the 0.5 criterion of the FSC curve (Figure 4.16). Our average resembles the structure in the paper of [111]. The imprint of carved out region is rarely visible in the average structure. During the sub-tomogram averaging process, the M-free score decreases from ~ 0.17 at the 1st iteration to around 0.01 after 16 iterations indicating almost no reference bias (Figure 4.17).

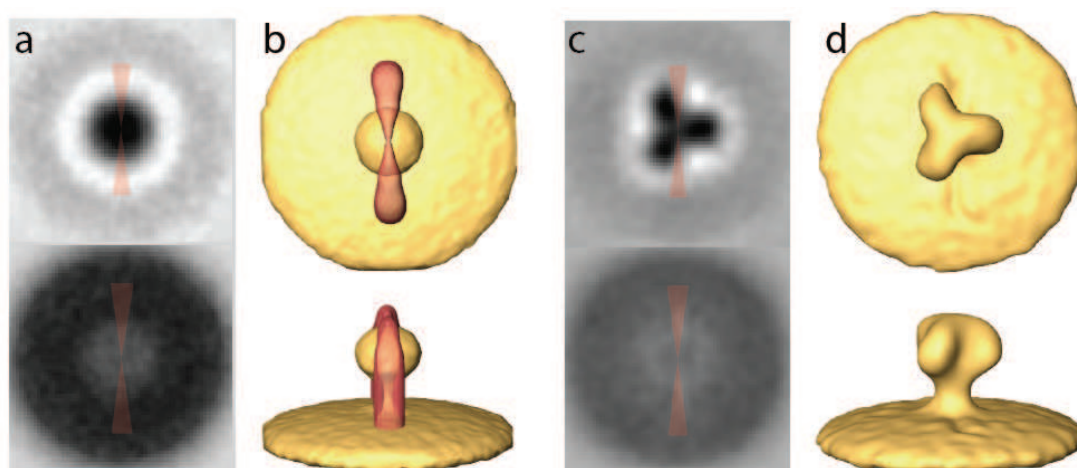


Figure 4.15 Results of the sub-tomogram averaging with SIV-spike data sets [111] (a) Two xy-slices of the starting template generated using the information from the pre-alignment. The upper slice cuts through the spike and the lower slice cuts through the membrane. The *testing area* is marked in red. (b) The same xy-slices as in (a) of the average after 16 iterations. The scale bar is 10 nm. (c) The top view and side view of the isosurface from (a). (d) The top view and side view of the isosurface from the result in (b). The average clearly has 3-fold symmetry and corresponds to the result as shown in Figure 2 in [111].

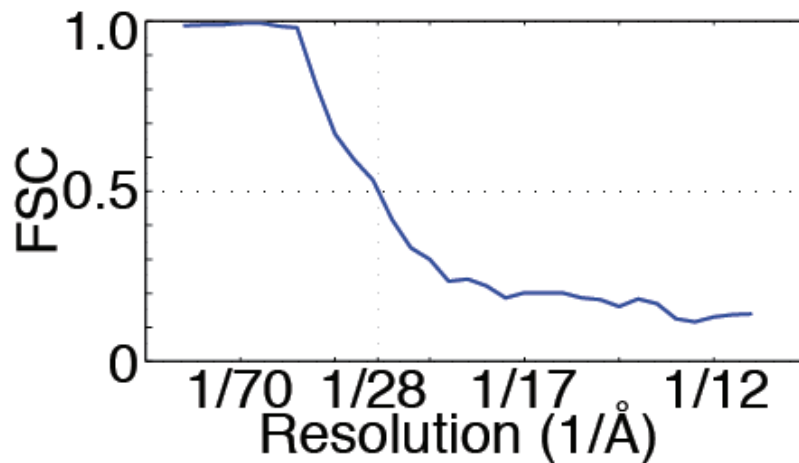


Figure 4.16 The Fourier Shell Correlation (FSC) shows that the average in (4.15c) and (4.15d) has a resolution of 28\AA judged using the 0.5 criterion, which matches the result in [111].

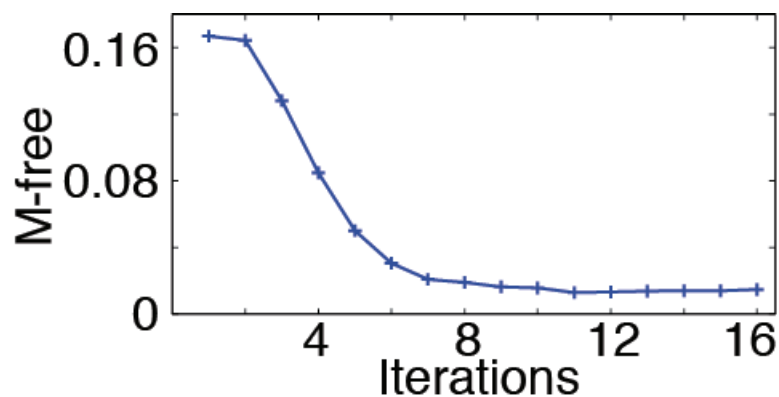


Figure 4.17 The development of the M-free score depending on the number of iterations. After the first iteration the M-free score is relatively high with a value of 0.17, but it decreases linearly with the quality improvement of the average. The M-free score converges to 0.01 after 16 iterations.

To prove that the M-free score and the final average we obtained is independent of the shape and position of the *testing area*, we repeated the same process with a variety of masks using different regions of the starting reference of the SIV spikes (Figure 4.18). The resulting sub-tomogram average exhibited a similar structure with similar resolution. The M-free score is almost equal (Figure 4.18).

4. Reference bias in reference based techniques

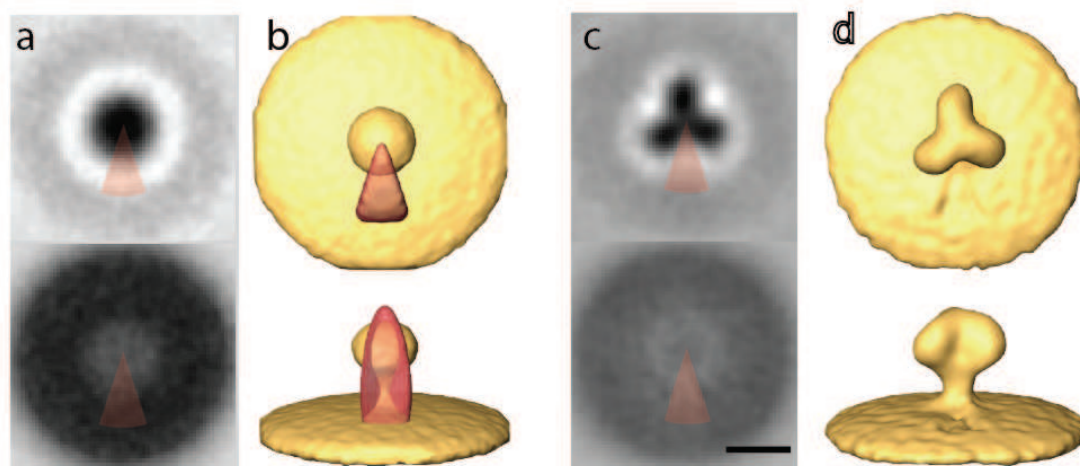


Figure 4.18 Same as figure 4.15, with a different *testing area* that is shown in (a), (b) and (c). The result is extremely similar to the one above. The scale bar in (c) is 10 nm.

4.4.4. Template matching of whole Mycoplasma cells.

We next tested the validity of the M-free score on cellular tomograms. For this purpose we performed template matching on two cryo-electron tomograms, in which Mycoplasma cells were scanned with a ribosome template derived from the crystal structure [145]. As mask we used a sphere, which is slightly bigger than the ribosome structure. From this spherical mask, we reserve a wedge representing 10% of the total voxels as *testing area*. In order to valid the independency of the M-free score to the *testing area*, we performed the experiment several times with the *testing area* placed at random orientations (Figure 4.19, 4.20). For the imaging of the Mycoplasma cells, they are embedded into ice films of similar thickness. The tomograms are recorded with a similar cumulative electron dose ($70 e^-/\text{\AA}^2$). The major difference between the two tomograms is the detector used for recording: the first cell was recorded on a Gatan 2002 energy filter with a 2k x 2k pixel Gatan Megascan 795 CCD camera (Short Megascan), while the second was recorded on a Quantum energy filter with a 4k x 4k pixel K2 Summit direct detector (Short K2). From both tomograms, the template matching delivers numerous positions corresponding to putative

particles. The indicated positions are localized all over the tomogram, most of which are inside the cells. But a small portion of the positions are on the cell membranes and rarely even outside the cells (Figure 4.19, 4.20). From each experiment, the positions corresponding to the highest 150 cross-correlation peaks were selected, which include the majority of the true positive hits. The M-free score for the average from the first cell recorded on the Megascan is 0.41 (Figure 4.19, 4.20), while on the K2 a M-free score of 0.09 and 0.10 were measured for different *testing areas* (Figure 4.19, 4.20). From the visual impression of the averages of the putative hits, we can conclude that the average created on the K2 looks significantly better, with almost as much contrast in the *testing area* as elsewhere. In contrast, the average on the Megascan has much lower intensity in the *testing area* than in the *working area*. Using the values of the M-free score, we obtained from the artificial data sets as a reference, we can summarize that the quality of ribosome detection using template matching on the tomogram recorded with the K2 is very good, while on the one recorded with the Megascan the result is not as reliable.

4. Reference bias in reference based techniques

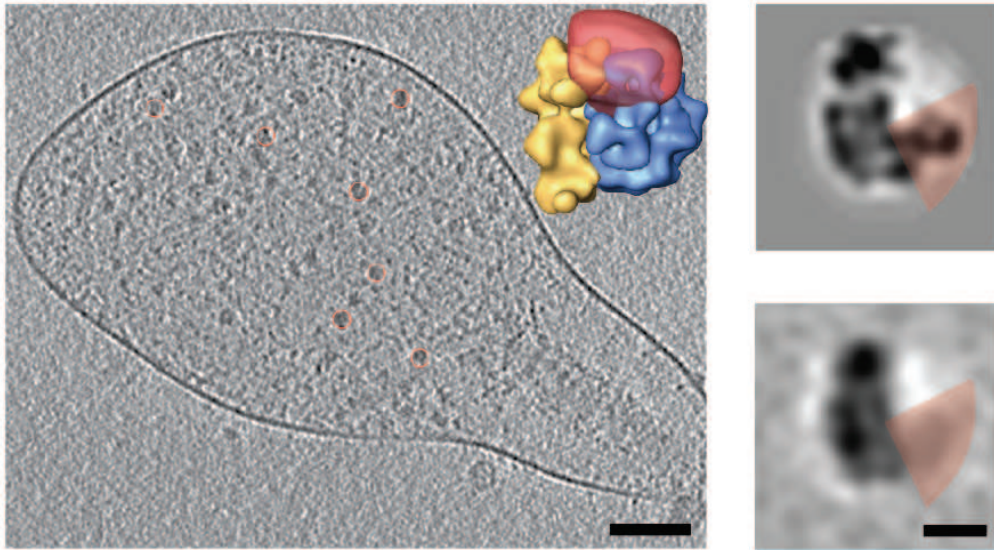


Figure 4.19 Results of the template matching with whole cell tomograms of Mycoplasma cells. Left: A 7nm xy-slice of the tomogram of a ~250nm thick Mycoplasma cell recorded on the Megascan camera. Inset in the upper right corner shows an isosurface of the ribosome template derived from the crystal structure (2AW7 in yellow and 2AWB in blue) with the *testing area* in transparent red. The top 150 detected positions that contain most of the ribosomes on this slice are shown in red circles. The scale bar is 100 nm. Upper right: The central slice from the template derived from the crystal structure with the *testing area* shown in red. Lower right: The central slice from the average of the top 150 detected positions with the *testing area* shown in red. M-free score is 0.41. The scale bar is 10 nm.

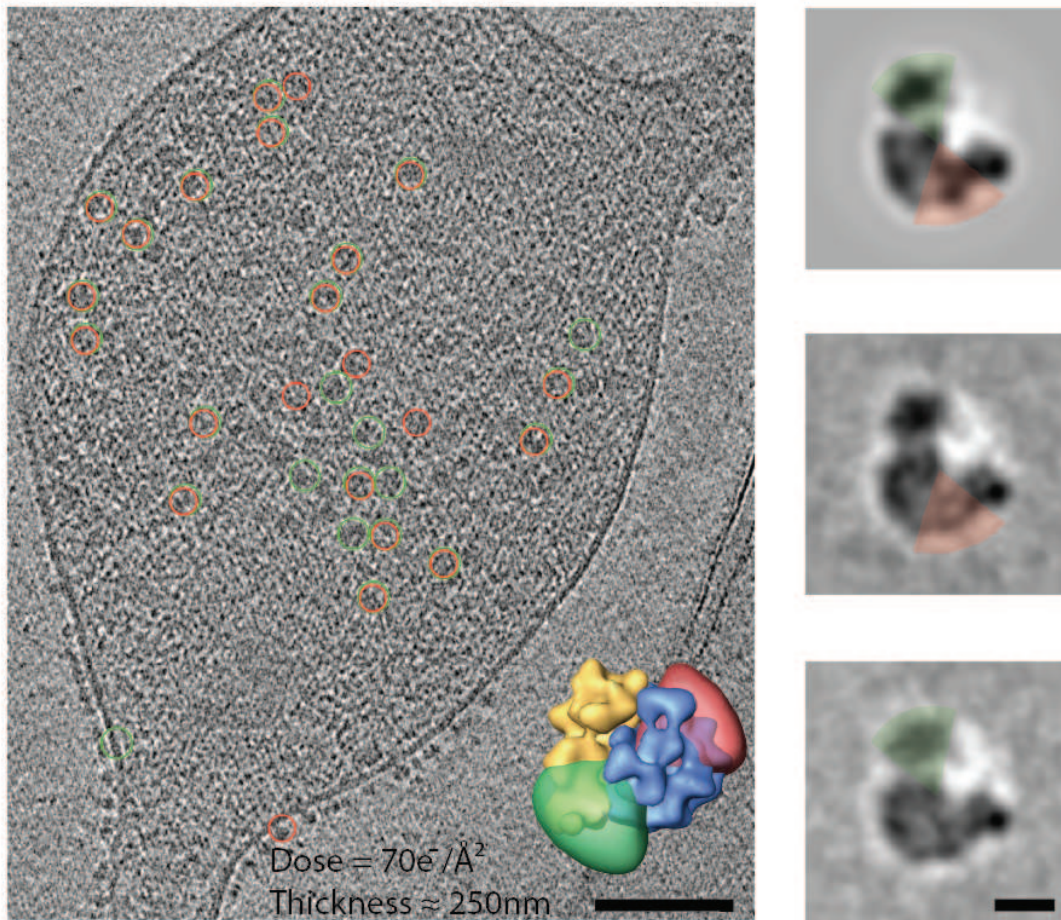


Figure 4.20 Results of the template matching with whole cell tomograms of Mycoplasma cells. Left: A 7nm xy-slice of a ~250nm thick Mycoplasma cell recorded on the K2 direct detector camera. The x-ray derived template is shown in the upper right corner. Two different *testing areas* are marked in red and green. The top 150 detected positions of putative ribosomes on this slice are shown in red or green circles, corresponding to the red or green *testing areas*. The scale bar is 100 nm. Upper right: The central slice from the template derived from the x-ray structure and the *testing areas* in red or green. Middle right: The central slice from the average of the detected positions with the red *testing area*. M-free score is 0.09. Lower right: Same as above, with the green *testing area*. The average shows minimal difference in comparison to the one above. The amount of template bias is similar. The M-free score is 0.10. The scale bar is 10 nm.

5. Conclusion

Cryo-electron tomography provides snapshots of the cellular proteome. With the development of the recording techniques, the demand for automated image processing techniques is rapidly increasing. With cross-correlation based techniques such as template matching and sub-tomogram averaging, the spatial positions and orientations of various macromolecular complexes within their native cellular context can be explored. However, the applicability of the cross-correlation based techniques is restricted due to the problem of reference bias. With our M-free heuristic, in a similar fashion to the R-free value in crystallography, a *testing area* is excluded from the mask in real space used in template matching and/or sub-tomogram averaging. The information within this area can be utilized to examine the amount of bias imposed by the reference after processing. It is proved that sacrificing 10% of the mask as *testing area* can provide reliable estimation of the amount of reference bias. The sacrifice of the signal might introduce deterioration of the alignment. However, we show that the averages of data sets having a strong signal are not severely affected by the 10% missing information in the *testing area*.

The M-free score works in real space instead of in Fourier space as in the previous single-particle approaches. Excising an area in real space still keeps the information from the whole frequency domain that is essential for the averaging purposes. For the mathematical derivation of the M-free score, two assumptions are necessary, both of which are valid when a well-recovered macromolecular signature and a reasonably small amount of bias exist. The M-free score should be zero in perfectly aligned data sets. It rises with increasing misalignment error and gives a reliable estimate of the reference bias. It is shown experimentally that the M-free score leads to sensible results, and can be used for estimating the amount of reference bias present in all the experiments.

5. Conclusion

Classification algorithms are a necessary supplement to the sub-tomogram averaging and template matching techniques for extracting different structural states of macromolecular complexes or reducing heterogeneity of data sets. The increased class purity can lead to improved resolution. Here we use a novel neural network approach. To improve the performance of the algorithm, we have compensated for the missing-wedge by measuring the similarity of the code-vectors to the sub-tomograms with constrained similarity metric. The algorithm performs slightly better compared to the constrained cross-correlation followed by PCA. In contrast to the constraint PCA no additional signal is sacrificed for the sake of missing-wedge compensation. This becomes obvious in the superior performance of the algorithm compared to previous approaches on the GroEL and GroEL/ES tomograms. Even though a perfect classification is not expected for noisy data such as cryo-electron sub-tomograms, it is plausible that the classification is hampered beyond this constitutive level by systematic errors.

Template matching and sub-tomogram averaging are automated approaches to identify the macromolecular complexes, explore their spatial relationship and determinate their structural information within their cellular context. The M-free score provides a precise estimation of the reference bias and gives a reliability measurement of the processes. It makes those approaches therefore more applicable to structural biology. In combination with the introduced high-performance classification algorithm KerDenSOM3D, which eliminates outliers and improves class purity of the data sets, high-resolution structure of macromolecular complexes can be resolved, which can be great advantageous to solve complex biological problems.

6. References

- [1] A. S. Frangakis, J. Bohm, F. Forster, S. Nickell, D. Nicastro, D. Typke, *et al.*, "Identification of macromolecular complexes in cryoelectron tomograms of phantom cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 14153-8, Oct 29 2002.
- [2] F. Forster, S. Pruggnaller, A. Seybert, and A. S. Frangakis, "Classification of cryo-electron sub-tomograms using constrained correlation," *J Struct Biol*, vol. 161, pp. 276-86, Mar 2008.
- [3] S. Subramaniam, "The SIV surface spike imaged by electron tomography: one leg or three?," *PLoS Pathog*, vol. 2, p. e91, Aug 2006.
- [4] Z. Yu and A. S. Frangakis, "M-free: Scoring the reference bias in sub-tomogram averaging and template matching," *J Struct Biol*, vol. 187, pp. 10-9, Jul 2014.
- [5] J. Cohen, "Structural biology. Is high-tech view of HIV too good to be true?," *Science*, vol. 341, pp. 443-4, Aug 2 2013.
- [6] T. R. Shaikh, R. Hegerl, and J. Frank, "An approach to examining model dependence in EM reconstructions using cross-validation," *J Struct Biol*, vol. 142, pp. 301-10, May 2003.
- [7] P. A. Penczek, "Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms," *J Struct Biol*, vol. 138, pp. 34-46, Apr-May 2002.
- [8] S. Chen, G. McMullan, A. R. Faruqi, G. N. Murshudov, J. M. Short, S. H. Scheres, *et al.*, "High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy," *Ultramicroscopy*, vol. 135, pp. 24-35, Dec 2013.
- [9] A. Stewart and N. Grigorieff, "Noise bias in the refinement of structures derived from single particles," *Ultramicroscopy*, vol. 102, pp. 67-84, Dec 2004.
- [10] T. Kohonen, "The Self-Organizing Map," *Proceedings of the IEEE*, vol. 78, pp. 1464-1480, Sep 1990.
- [11] A. Pascual-Montano, L. E. Donate, M. Valle, M. Barcena, R. D. Pascual-Marqui, and J. M. Carazo, "A novel neural network technique for analysis

- and classification of EM single-particle images," *J Struct Biol*, vol. 133, pp. 233-45, Feb-Mar 2001.
- [12] R. Marabini and J. M. Carazo, "Pattern-Recognition and Classification of Images of Biological Macromolecules Using Artificial Neural Networks," *Biophysical Journal*, vol. 66, pp. 1804-1814, Jun 1994.
- [13] A. Pascual-Montano, K. A. Taylor, H. Winkler, R. D. Pascual-Marqui, and J. M. Carazo, "Quantitative self-organizing maps for clustering electron tomograms," *J Struct Biol*, vol. 138, pp. 114-22, Apr-May 2002.
- [14] Z. Yu and A. S. Frangakis, "Classification of electron sub-tomograms with neural networks and its application to template-matching," *J Struct Biol*, vol. 174, pp. 494-504, Jun 2011.
- [15] J. Frank, *Electron tomography : methods for three-dimensional visualization of structures in the cell*, 2nd ed. New York ; London: Springer, 2006.
- [16] J. Dubochet, M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowell, *et al.*, "Cryo-electron microscopy of vitrified specimens," *Q Rev Biophys*, vol. 21, pp. 129-228, May 1988.
- [17] M. Adrian, B. ten Heggeler-Bordier, W. Wahli, A. Z. Stasiak, A. Stasiak, and J. Dubochet, "Direct visualization of supercoiled DNA molecules in solution," *EMBO J*, vol. 9, pp. 4551-4, Dec 1990.
- [18] M. Cyrklaff, M. Adrian, and J. Dubochet, "Evaporation during preparation of unsupported thin vitrified aqueous layers for cryo-electron microscopy," *J Electron Microscop Tech*, vol. 16, pp. 351-5, Dec 1990.
- [19] M. J. Dobro, L. A. Melanson, G. J. Jensen, and A. W. McDowell, "Plunge freezing for electron cryomicroscopy," *Methods Enzymol*, vol. 481, pp. 63-82, 2010.
- [20] H. Sosa, D. Popp, G. Ouyang, and H. E. Huxley, "Ultrastructure of skeletal muscle fibers studied by a plunge quick freezing method: myofilament lengths," *Biophys J*, vol. 67, pp. 283-92, Jul 1994.
- [21] J. Dubochet and N. Sartori Blanc, "The cell in absence of aggregation artifacts," *Micron*, vol. 32, pp. 91-9, Jan 2001.
- [22] H. Sitte, "Advanced instrumentation and methodology related to cryoultramicrotomy: a review," *Scanning Microscop Suppl*, vol. 10, pp. 387-463; discussion 463-6, 1996.
- [23] R. Hegerl and W. Hoppe, "Influence of Electron Noise on Three-dimensional image Reconstruction," *Zeitschrift Naturforschung Teil A*, vol. 31, p. 1717, 1976.

- [24] B. F. McEwen, K. H. Downing, and R. M. Glaeser, "The relevance of dose-fractionation in tomography of radiation-sensitive specimens," *Ultramicroscopy*, vol. 60, pp. 357-73, Oct 1995.
- [25] R. Grimm, H. Singh, R. Rachel, D. Typke, W. Zillig, and W. Baumeister, "Electron tomography of ice-embedded prokaryotic cells," *Biophys J*, vol. 74, pp. 1031-42, Feb 1998.
- [26] D. Nicastro, C. Schwartz, J. Pierson, R. Gaudette, M. E. Porter, and J. R. McIntosh, "The molecular architecture of axonemes revealed by cryoelectron tomography," *Science*, vol. 313, pp. 944-8, Aug 18 2006.
- [27] J. Cope, S. Gilbert, I. Rayment, D. Mastronarde, and A. Hoenger, "Cryo-electron tomography of microtubule-kinesin motor complexes," *J Struct Biol*, vol. 170, pp. 257-65, May 2010.
- [28] C. L. Schwartz, J. M. Heumann, S. C. Dawson, and A. Hoenger, "A detailed, hierarchical study of *Giardia lamblia*'s ventral disc reveals novel microtubule-associated protein complexes," *PLoS One*, vol. 7, p. e43783, 2012.
- [29] O. Kuybeda, G. A. Frank, A. Bartesaghi, M. Borgnia, S. Subramaniam, and G. Sapiro, "A collaborative framework for 3D alignment and classification of heterogeneous subvolumes in cryo-electron tomography," *J Struct Biol*, vol. 181, pp. 116-27, Feb 2013.
- [30] F. K. Schur, W. J. Hagen, A. de Marco, and J. A. Briggs, "Determination of protein structure at 8.5Å resolution using cryo-electron tomography and sub-tomogram averaging," *J Struct Biol*, vol. 184, pp. 394-400, Dec 2013.
- [31] S. Nickell, C. Kofler, A. P. Leis, and W. Baumeister, "A visual approach to proteomics," *Nat Rev Mol Cell Biol*, vol. 7, pp. 225-30, Mar 2006.
- [32] V. Lucic, A. Rigort, and W. Baumeister, "Cryo-electron tomography: the challenge of doing structural biology in situ," *J Cell Biol*, vol. 202, pp. 407-19, Aug 5 2013.
- [33] A. Hoenger, "High-resolution cryo-electron microscopy on macromolecular complexes and cell organelles," *Protoplasma*, vol. 251, pp. 417-27, Mar 2014.
- [34] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, 2nd ed.: Springer, 2009.
- [35] M. Adrian, J. Dubochet, J. Lepault, and A. W. McDowell, "Cryo-electron microscopy of viruses," *Nature*, vol. 308, pp. 32-6, Mar 1-7 1984.
- [36] R. A. Steinbrecht and K. Zierold, *Cryotechniques in biological electron microscopy*. Berlin ; New York: Springer-Verlag, 1987.

- [37] H. P. Erickson and A. Klug, "Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs," *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, vol. 261, pp. 105-&, 1971.
- [38] J. Frank, A. Verschoor, Y. Li, J. Zhu, R. K. Lata, M. Radermacher, *et al.*, "A model of the translational apparatus based on a three-dimensional reconstruction of the Escherichia coli ribosome," *Biochem Cell Biol*, vol. 73, pp. 757-65, Nov-Dec 1995.
- [39] R. H. Wade and J. Frank, "Electron-Microscope Transfer-Functions for Partially Coherent Axial Illumination and Chromatic Defocus Spread," *Optik*, vol. 49, pp. 81-92, 1977.
- [40] R. H. Wade, "A Brief Look at Imaging and Contrast Transfer," *Ultramicroscopy*, vol. 46, pp. 145-156, Oct 1992.
- [41] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, *et al.*, "Single-particle electron cryo-microscopy: towards atomic resolution," *Quarterly Reviews of Biophysics*, vol. 33, pp. 307-369, Nov 2000.
- [42] T. S. Baker and R. Henderson, "Electron cryomicroscopy of biological macromolecules," *International Tables for Crystallography*, vol. F, pp. 593-614, 2012.
- [43] P. N. Unwin and R. Henderson, "Molecular structure determination by electron microscopy of unstained crystalline specimens," *J Mol Biol*, vol. 94, pp. 425-40, May 25 1975.
- [44] B. Bottcher, S. A. Wynne, and R. A. Crowther, "Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy," *Nature*, vol. 386, pp. 88-91, Mar 6 1997.
- [45] Y. Chen, S. Pfeffer, T. Hrabe, J. M. Schuller, and F. Forster, "Fast and accurate reference-free alignment of subtomograms," *J Struct Biol*, vol. 182, pp. 235-45, Jun 2013.
- [46] B. Sander, M. M. Golas, and H. Stark, "Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy," *J Struct Biol*, vol. 151, pp. 92-105, Jul 2005.
- [47] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, *et al.*, "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM," *Nat Methods*, vol. 10, pp. 584-90, Jun 2013.
- [48] R. Henderson, "Image contrast in high-resolution electron microscopy of biological macromolecules: TMV in ice," *Ultramicroscopy*, vol. 46, pp. 1-18, Oct 1992.

- [49] X. C. Bai, I. S. Fernandez, G. McMullan, and S. H. Scheres, "Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles," *Elife*, vol. 2, p. e00461, 2013.
- [50] N. Grigorieff, "Direct detection pays off for electron cryo-microscopy," *Elife*, vol. 2, p. e00573, 2013.
- [51] X. Li, S. Q. Zheng, K. Egami, D. A. Agard, and Y. Cheng, "Influence of electron dose rate on electron counting images recorded with the K2 camera," *J Struct Biol*, vol. 184, pp. 251-60, Nov 2013.
- [52] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies*. San Diego: Academic Press, 1996.
- [53] J. C. Fung, W. Liu, W. J. de Ruijter, H. Chen, C. K. Abbey, J. W. Sedat, *et al.*, "Toward fully automated high-resolution electron tomography," *J Struct Biol*, vol. 116, pp. 181-9, Jan-Feb 1996.
- [54] J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, "Computer visualization of three-dimensional image data using IMOD," *Journal of Structural Biology*, vol. 116, pp. 71-76, Jan-Feb 1996.
- [55] S. Nickell, F. Forster, A. Linaroudis, W. D. Net, F. Beck, R. Hegerl, *et al.*, "TOM software toolbox: acquisition and analysis for electron tomography," *J Struct Biol*, vol. 149, pp. 227-34, Mar 2005.
- [56] S. Brandt, J. Heikkonen, and P. Engelhardt, "Automatic alignment of transmission electron microscope tilt series without fiducial markers," *J Struct Biol*, vol. 136, pp. 201-13, Dec 2001.
- [57] Y. Liu, P. A. Penczek, B. F. McEwen, and J. Frank, "A marker-free alignment method for electron tomography," *Ultramicroscopy*, vol. 58, pp. 393-402, Jun 1995.
- [58] C. H. Owen and W. J. Landis, "Alignment of electron tomographic series by correlation without the use of gold particles," *Ultramicroscopy*, vol. 63, pp. 27-38, Apr 1996.
- [59] D. Castano-Diez, M. Scheffer, A. Al-Amoudi, and A. S. Frangakis, "Alignator: a GPU powered software package for robust fiducial-less alignment of cryo tilt-series," *J Struct Biol*, vol. 170, pp. 117-26, Apr 2010.
- [60] D. Castano-Diez, A. Al-Amoudi, A. M. Glynn, A. Seybert, and A. S. Frangakis, "Fiducial-less alignment of cryo-sections," *J Struct Biol*, vol. 159, pp. 413-23, Sep 2007.
- [61] D. Castano-Diez, A. Al-Amoudi, A. M. Glynn, A. Seybert, and A. S. Frangakis, "Reprint of "Fiducial-less alignment of cryo-sections" [*J. Struct. Biol.* 159 (2007) 413-423]," *J Struct Biol*, vol. 161, pp. 249-59, Mar 2008.

- [62] C. O. Sorzano, C. Messaoudi, M. Eibauer, J. R. Bilbao-Castro, R. Hegerl, S. Nickell, *et al.*, "Marker-free image registration of electron tomography tilt-series," *BMC Bioinformatics*, vol. 10, p. 124, 2009.
- [63] D. J. De Rosier and A. Klug, "Reconstruction of three dimensional structures from electron micrographs," *Nature*, vol. 217, pp. 130-4, Jan 13 1968.
- [64] P. A. Penczek, R. Renka, and H. Schomberg, "Gridding-based direct Fourier inversion of the three-dimensional ray transform," *J Opt Soc Am A Opt Image Sci Vis*, vol. 21, pp. 499-509, Apr 2004.
- [65] K. Sandberg, D. N. Mastronarde, and G. Beylkin, "A fast reconstruction algorithm for electron microscope tomography," *J Struct Biol*, vol. 144, pp. 61-72, Oct-Nov 2003.
- [66] N. Grigorieff, "Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 2.2 Å in ice," *J Mol Biol*, vol. 277, pp. 1033-46, Apr 17 1998.
- [67] W. Zhang, M. Kimmel, C. M. Spahn, and P. A. Penczek, "Heterogeneity of large macromolecular complexes revealed by 3D cryo-EM variance analysis," *Structure*, vol. 16, pp. 1770-6, Dec 10 2008.
- [68] Y. X. Chen and F. Forster, "Iterative reconstruction of cryo-electron tomograms using nonuniform fast Fourier transforms," *Journal of Structural Biology*, vol. 185, pp. 309-316, Mar 2014.
- [69] S. Matej, J. A. Fessler, and I. G. Kazantsev, "Iterative tomographic image reconstruction using Fourier-based forward and back-projectors," *Ieee Transactions on Medical Imaging*, vol. 23, pp. 401-412, Apr 2004.
- [70] T. Knopp, S. Kunis, and D. Potts, "A note on the iterative MRI reconstruction from nonuniform k-space data," *Int J Biomed Imaging*, vol. 2007, p. 24727, 2007.
- [71] G. Harauz and M. Vanheul, "Exact Filters for General Geometry 3-Dimensional Reconstruction," *Optik*, vol. 73, pp. 146-156, Jul 1986.
- [72] M. Radermacher, T. Wagenknecht, R. Grassucci, J. Frank, M. Inui, C. Chadwick, *et al.*, "Cryo-EM of the native structure of the calcium release channel/ryanodine receptor from sarcoplasmic reticulum," *Biophys J*, vol. 61, pp. 936-40, Apr 1992.
- [73] M. Radermacher, T. Wagenknecht, A. Verschoor, and J. Frank, "A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of E. coli," *J Microsc*, vol. 141, pp. RP1-2, Jan 1986.
- [74] A. D. Poularikas, *Handbook of Formulas and Tables for Signal Processing*, 1st ed.: CRC Press, 1998.

- [75] G. N. Ramachandran and A. V. Lakshminarayanan, "Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms," *Proc Natl Acad Sci U S A*, vol. 68, pp. 2236-40, Sep 1971.
- [76] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *J Theor Biol*, vol. 29, pp. 471-81, Dec 1970.
- [77] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J Theor Biol*, vol. 36, pp. 105-17, Jul 1972.
- [78] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm," *Ultrason Imaging*, vol. 6, pp. 81-94, Jan 1984.
- [79] P. Penczek, M. Radermacher, and J. Frank, "Three-dimensional reconstruction of single particles embedded in ice," *Ultramicroscopy*, vol. 40, pp. 33-53, Jan 1992.
- [80] X. H. Wan, F. Zhang, Q. Chu, K. Zhang, F. Sun, B. Yuan, *et al.*, "Three-dimensional reconstruction using an adaptive simultaneous algebraic reconstruction technique in electron tomography," *Journal of Structural Biology*, vol. 175, pp. 277-287, Sep 2011.
- [81] R. Marabini, G. T. Herman, and J. M. Carazo, "3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs)," *Ultramicroscopy*, vol. 72, pp. 53-65, Apr 1998.
- [82] J. G. Colsher, "Iterative 3-Dimensional Image-Reconstruction from Tomographic Projections," *Computer Graphics and Image Processing*, vol. 6, pp. 513-537, 1977.
- [83] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *Ieee Transactions on Signal Processing*, vol. 51, pp. 560-574, Feb 2003.
- [84] M. Kunz and A. S. Frangakis, "Super-sampling SART with ordered subsets," *J Struct Biol*, vol. 188, pp. 107-15, Nov 2014.
- [85] K. Mueller, R. Yagel, and J. J. Wheller, "Anti-aliased three-dimensional cone-beam reconstruction of low-contrast objects with algebraic methods," *Ieee Transactions on Medical Imaging*, vol. 18, pp. 519-537, Jun 1999.
- [86] J. Kim, H. Guan, D. Gersten, and T. Zhang, "Evaluation of algebraic iterative image reconstruction methods for tetrahedron beam computed tomography systems," *Int J Biomed Imaging*, vol. 2013, p. 609704, 2013.
- [87] W. Xu, F. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard, *et al.*, "High-performance iterative electron tomography reconstruction with long-

- object compensation using graphics processing units (GPUs)," *Journal of Structural Biology*, vol. 171, pp. 142-153, Aug 2010.
- [88] W. M. Pang, J. Qin, Y. Q. Lu, Y. M. Xie, C. K. Chui, and P. A. Heng, "Accelerating simultaneous algebraic reconstruction technique with motion compensation using CUDA-enabled GPU," *International Journal of Computer Assisted Radiology and Surgery*, vol. 6, pp. 187-199, Mar 2011.
- [89] D. N. Mastronarde, "Dual-axis tomography: An approach with alignment methods that preserve resolution," *Journal of Structural Biology*, vol. 120, pp. 343-352, Dec 1997.
- [90] P. Penczek, M. Marko, K. Buttle, and J. Frank, "Double-tilt electron tomography," *Ultramicroscopy*, vol. 60, pp. 393-410, Oct 1995.
- [91] S. Nickell, R. Hegerl, W. Baumeister, and R. Rachel, "Pyrodictium cannulae enter the periplasmic space but do not enter the cytoplasm, as revealed by cryo-electron tomography," *Journal of Structural Biology*, vol. 141, pp. 34-42, Jan 2003.
- [92] J. Walz, D. Typke, M. Nitsch, A. J. Koster, R. Hegerl, and W. Baumeister, "Electron Tomography of Single Ice-Embedded Macromolecules: Three-Dimensional Alignment and Classification," *J Struct Biol*, vol. 120, pp. 387-95, Dec 1997.
- [93] M. Xu, M. Beck, and F. Alber, "High-throughput subtomogram alignment and classification by Fourier space constrained fast volumetric matching," *J Struct Biol*, vol. 178, pp. 152-64, May 2012.
- [94] M. Stolken, F. Beck, T. Haller, R. Hegerl, I. Gutsche, J. M. Carazo, *et al.*, "Maximum likelihood based classification of electron tomographic data," *J Struct Biol*, vol. 173, pp. 77-85, Jan 2011.
- [95] J. M. Heumann, A. Hoenger, and D. N. Mastronarde, "Clustering and variance maps for cryo-electron tomography using wedge-masked differences," *Journal of Structural Biology*, vol. 175, pp. 288-299, Sep 2011.
- [96] A. Bartesaghi, P. Sprechmann, J. Liu, G. Randall, G. Sapiro, and S. Subramaniam, "Classification and 3D averaging with missing wedge correction in biological electron tomography," *J Struct Biol*, vol. 162, pp. 436-50, Jun 2008.
- [97] W. O. Saxton and W. Baumeister, "The correlation averaging of a regularly arranged bacterial cell envelope protein," *J Microsc*, vol. 127, pp. 127-38, Aug 1982.
- [98] M. Unser, C. O. Sorzano, P. Thevenaz, S. Jonic, C. El-Bez, S. De Carlo, *et al.*, "Spectral signal-to-noise ratio and resolution assessment of 3D reconstructions," *J Struct Biol*, vol. 149, pp. 243-55, Mar 2005.

- [99] M. Unser, B. L. Trus, and A. C. Steven, "A new resolution criterion based on spectral signal-to-noise ratios," *Ultramicroscopy*, vol. 23, pp. 39-51, 1987.
- [100] N. J. Bershad and A. J. Rockmore, "On estimating signal-to-noise ratio using the sample correlation coefficient," *IEEE Trans. Inf. Theory*, vol. 20, pp. 112-113, 1974.
- [101] J. Frank and L. Alali, "Signal-to-Noise Ratio of Electron Micrographs Obtained by Cross-Correlation," *Nature*, vol. 256, pp. 376-379, 1975.
- [102] W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy*. the University of Virginia: Academic Press, 1978.
- [103] M. van Heel and M. Schatz, "Fourier shell correlation threshold criteria," *Journal of Structural Biology*, vol. 151, pp. 250-262, Sep 2005.
- [104] W. Baumeister, "From proteomic inventory to architecture," *FEBS Lett*, vol. 579, pp. 933-7, Feb 7 2005.
- [105] J. O. Ortiz, F. Forster, J. Kurner, A. A. Linaroudis, and W. Baumeister, "Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition," *J Struct Biol*, vol. 156, pp. 334-41, Nov 2006.
- [106] M. Beck, J. A. Malmstrom, V. Lange, A. Schmidt, E. W. Deutsch, and R. Aebersold, "Visual proteomics of the human pathogen *Leptospira interrogans*," *Nat Methods*, vol. 6, pp. 817-23, Nov 2009.
- [107] F. Brandt, L. A. Carlson, F. U. Hartl, W. Baumeister, and K. Grunewald, "The three-dimensional organization of polyribosomes in intact human cells," *Mol Cell*, vol. 39, pp. 560-9, Aug 27 2010.
- [108] S. H. Scheres and S. Chen, "Prevention of overfitting in cryo-EM structure determination," *Nat Methods*, vol. 9, pp. 853-4, Sep 2012.
- [109] A. Al-Amoudi, D. Castano-Diez, D. P. Devos, R. B. Russell, G. T. Johnson, and A. S. Frangakis, "The three-dimensional molecular structure of the desmosomal plaque," *Proc Natl Acad Sci U S A*, vol. 108, pp. 6480-5, Apr 19 2011.
- [110] A. Al-Amoudi, D. C. Diez, M. J. Betts, and A. S. Frangakis, "The molecular architecture of cadherins in native epidermal desmosomes," *Nature*, vol. 450, pp. 832-7, Dec 6 2007.
- [111] G. Zanetti, J. A. Briggs, K. Grunewald, Q. J. Sattentau, and S. D. Fuller, "Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ," *PLoS Pathog*, vol. 2, p. e83, Aug 2006.
- [112] F. Forster, O. Medalia, N. Zauberman, W. Baumeister, and D. Fass, "Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography," *Proc Natl Acad Sci U S A*, vol. 102, pp. 4729-34, Mar 29 2005.

- [113] J. A. Briggs, "Structural biology in situ--the potential of subtomogram averaging," *Curr Opin Struct Biol*, vol. 23, pp. 261-7, Apr 2013.
- [114] F. Brandt, S. A. Etchells, J. O. Ortiz, A. H. Elcock, F. U. Hartl, and W. Baumeister, "The native 3D organization of bacterial polysomes," *Cell*, vol. 136, pp. 261-71, Jan 23 2009.
- [115] J. T. Huiskonen, J. Hepojoki, P. Laurinmaki, A. Vaheri, H. Lankinen, S. J. Butcher, *et al.*, "Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses," *J Virol*, vol. 84, pp. 4889-97, May 2010.
- [116] J. Liu, A. Bartesaghi, M. J. Borgnia, G. Sapiro, and S. Subramaniam, "Molecular architecture of native HIV-1 gp120 trimers," *Nature*, vol. 455, pp. 109-13, Sep 4 2008.
- [117] E. E. Tran, M. J. Borgnia, O. Kuybeda, D. M. Schauder, A. Bartesaghi, G. A. Frank, *et al.*, "Structural mechanism of trimeric HIV-1 envelope glycoprotein activation," *PLoS Pathog*, vol. 8, p. e1002797, 2012.
- [118] T. A. White, A. Bartesaghi, M. J. Borgnia, J. R. Meyerson, M. J. de la Cruz, J. W. Bess, *et al.*, "Molecular architectures of trimeric SIV and HIV-1 envelope glycoproteins on intact viruses: strain-dependent variation in quaternary structure," *PLoS Pathog*, vol. 6, p. e1001249, 2010.
- [119] T. A. M. Bharat, J. D. Riches, L. Kolesnikova, S. Welsch, V. Kraehling, N. Davey, *et al.*, "Cryo-Electron Tomography of Marburg Virus Particles and Their Morphogenesis within Infected Cells," *Plos Biology*, vol. 9, Nov 2011.
- [120] J. A. Briggs, J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H. G. Krausslich, "Structure and assembly of immature HIV," *Proc Natl Acad Sci U S A*, vol. 106, pp. 11090-5, Jul 7 2009.
- [121] L. A. Carlson, A. de Marco, H. Oberwinkler, A. Habermann, J. A. Briggs, H. G. Krausslich, *et al.*, "Cryo electron tomography of native HIV-1 budding sites," *PLoS Pathog*, vol. 6, p. e1001173, 2010.
- [122] A. de Marco, B. Muller, B. Glass, J. D. Riches, H. G. Krausslich, and J. A. G. Briggs, "Structural Analysis of HIV-1 Maturation Using Cryo-Electron Tomography," *Plos Pathogens*, vol. 6, Nov 2010.
- [123] L. Liljeroos, J. T. Huiskonen, A. Ora, P. Susi, and S. J. Butcher, "Electron cryotomography of measles virus reveals how matrix protein coats the ribonucleocapsid within intact virions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 18085-18090, Nov 1 2011.
- [124] E. R. Wright, J. B. Schooler, H. J. Ding, C. Kieffer, C. Fillmore, W. I. Sundquist, *et al.*, "Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells," *EMBO J*, vol. 26, pp. 2218-26, Apr 18 2007.

- [125] P. W. Keller, C. S. Adamson, J. B. Heymann, E. O. Freed, and A. C. Steven, "HIV-1 maturation inhibitor bevirimat stabilizes the immature Gag lattice," *J Virol*, vol. 85, pp. 1420-8, Feb 2011.
- [126] M. Beck, V. Lucic, F. Forster, W. Baumeister, and O. Medalia, "Snapshots of nuclear pore complexes in action captured by cryo-electron tomography," *Nature*, vol. 449, pp. 611-5, Oct 4 2007.
- [127] D. Frenkiel-Krispin, B. Maco, U. Aebi, and O. Medalia, "Structural analysis of a metazoan nuclear pore complex reveals a fused concentric ring architecture," *J Mol Biol*, vol. 395, pp. 578-86, Jan 22 2010.
- [128] T. Maimon, N. Elad, I. Dahan, and O. Medalia, "The Human Nuclear Pore Complex as Revealed by Cryo-Electron Tomography," *Structure*, vol. 20, pp. 998-1006, Jun 6 2012.
- [129] H. Winkler, P. Zhu, J. Liu, F. Ye, K. H. Roux, and K. A. Taylor, "Tomographic subvolume alignment and subvolume classification applied to myosin V and SIV envelope spikes," *J Struct Biol*, vol. 165, pp. 64-77, Feb 2009.
- [130] S. H. Scheres, R. Melero, M. Valle, and J. M. Carazo, "Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization," *Structure*, vol. 17, pp. 1563-72, Dec 9 2009.
- [131] F. J. Sigworth, "A maximum-likelihood approach to single-particle image refinement," *J Struct Biol*, vol. 122, pp. 328-39, 1998.
- [132] Q. S. Zheng, M. B. Braunfeld, J. W. Sedat, and D. A. Agard, "An improved strategy for automated electron microscopic tomography," *J Struct Biol*, vol. 147, pp. 91-101, Aug 2004.
- [133] S. Pruggnaller, M. Mayr, and A. S. Frangakis, "A visualization and segmentation toolbox for electron microscopy," *J Struct Biol*, vol. 164, pp. 161-5, Oct 2008.
- [134] L. Ditzel, J. Lowe, D. Stock, K. O. Stetter, H. Huber, R. Huber, *et al.*, "Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT," *Cell*, vol. 93, pp. 125-38, Apr 3 1998.
- [135] M. Nitsch, J. Walz, D. Typke, M. Klumpp, L. O. Essen, and W. Baumeister, "Group II chaperonin in an open conformation examined by electron tomography," *Nat Struct Biol*, vol. 5, pp. 855-7, Oct 1998.
- [136] C. O. Sorzano, R. Marabini, J. Velazquez-Muriel, J. R. Bilbao-Castro, S. H. Scheres, J. M. Carazo, *et al.*, "XMIPP: a new generation of an open-source image processing package for electron microscopy," *J Struct Biol*, vol. 148, pp. 194-204, Nov 2004.
- [137] Y. Mao, L. Wang, C. Gu, A. Herschhorn, A. Desormeaux, A. Finzi, *et al.*, "Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer," *Proc Natl Acad Sci U S A*, vol. 110, pp. 12438-43, Jul 23 2013.

- [138] S. H. Scheres, "A Bayesian view on cryo-EM structure determination," *J Mol Biol*, vol. 415, pp. 406-18, Jan 13 2012.
- [139] S. H. Scheres, "RELION: implementation of a Bayesian approach to cryo-EM structure determination," *J Struct Biol*, vol. 180, pp. 519-30, Dec 2012.
- [140] D. Lyumkis, A. F. Brilot, D. L. Theobald, and N. Grigorieff, "Likelihood-based classification of cryo-EM images using FREALIGN," *J Struct Biol*, vol. 183, pp. 377-88, Sep 2013.
- [141] B. Falkner and G. F. Schroder, "Cross-validation in cryo-EM-based structural modeling," *Proc Natl Acad Sci U S A*, vol. 110, pp. 8930-5, May 28 2013.
- [142] T. Hrabe, Y. Chen, S. Pfeffer, L. K. Cuellar, A. V. Mangold, and F. Forster, "PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis," *J Struct Biol*, vol. 178, pp. 177-88, May 2012.
- [143] A. T. Brunger, "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures," *Nature*, vol. 355, pp. 472-5, Jan 30 1992.
- [144] N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, *et al.*, "ATP-bound states of GroEL captured by cryo-electron microscopy," *Cell*, vol. 107, pp. 869-79, Dec 28 2001.
- [145] B. S. Schuwirth, M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton, *et al.*, "Structures of the bacterial ribosome at 3.5 angstrom resolution," *Science*, vol. 310, pp. 827-834, Nov 4 2005.

A. Appendix

A.1. PAPER I

Yu Z., Frangakis AS. “Classification of electron sub-tomograms with neural networks and its application to template matching”, *J Struct Biol.* 2011 Jun;174(3):494-504

Main conclusion: Classification of electron sub-tomograms is a challenging task, due to the missing-wedge and the low signal-to-noise ratio of the data. Classification algorithms tend to classify data according to their orientation to the missing-wedge, rather than to the underlying signal. Here we use a neural network approach, called the Kernel Density Estimator Self-Organizing Map (KerDenSOM3D), which we have implemented in three-dimensions (3D), also having compensated for the missing-wedge, and we comprehensively compare it to other classification methods. For this purpose, we use various simulated macromolecules, as well as tomographically reconstructed *in vitro* GroEL and GroEL/GroES molecules. We show that the performance of this classification method is superior to previously used algorithms. Furthermore, we show how this algorithm can be used to provide an initial cross-validation of template matching approaches. For the example of sub-tomogram classification extracted from cellular tomograms of *Mycoplasma pneumonia* and *Spiroplasma melliferum* cells, we show the bias of template matching, and by using differing search and classification areas, we demonstrate how the bias can be significantly reduced.

My contributions: Design of research (together with ASF), method development (together with ASF), and analysis of results, data processing and figure preparation, writing of the manuscript (together with ASF).



Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Classification of electron sub-tomograms with neural networks and its application to template-matching

Zhou Yu, Achilleas S. Frangakis *

Frankfurt Institute for Molecular Life Sciences and Institute of Biophysics, Goethe University Frankfurt, Max-von-Laue Str.1, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:

Received 16 December 2010
 Received in revised form 25 February 2011
 Accepted 28 February 2011
 Available online 5 March 2011

Keywords:

Sub-tomogram classification
 Image processing
 Cryo-electron tomography

ABSTRACT

Classification of electron sub-tomograms is a challenging task, due to the missing-wedge and the low signal-to-noise ratio of the data. Classification algorithms tend to classify data according to their orientation to the missing-wedge, rather than to the underlying signal. Here we use a neural network approach, called the Kernel Density Estimator Self-Organizing Map (KerDenSOM3D), which we have implemented in three-dimensions (3D), also having compensated for the missing-wedge, and we comprehensively compare it to other classification methods. For this purpose, we use various simulated macromolecules, as well as tomographically reconstructed *in vitro* GroEL and GroEL/GroES molecules. We show that the performance of this classification method is superior to previously used algorithms. Furthermore, we show how this algorithm can be used to provide an initial cross-validation of template-matching approaches. For the example of sub-tomogram classification extracted from cellular tomograms of *Mycoplasma pneumonia* and *Spiroplasma melliferum* cells, we show the bias of template-matching, and by using differing search and classification areas, we demonstrate how the bias can be significantly reduced.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Cryo-electron tomography (CET) provides unique three-dimensional (3D) images of cells and organelles at molecular resolution. The application range varies from tissue and whole cells to organelles and large pleiomorphic *in vitro* samples like viruses (Leis et al., 2009; Li and Jensen, 2009; Milne and Subramaniam, 2009). Ultimately, CET aims to visualize the spatial organization and interaction of various macromolecules in the cellular context. Currently, most of the applications are based on multiple occurring structures, which are averaged in order to improve the resolution (Al-Amoudi et al., 2007; Beck et al., 2004; Briggs et al., 2009). Thus, sub-tomogram averaging and classification techniques play an essential role in data interpretation.

A number of classification methods have been presented over the last years, all of which were inspired by their counterparts used in single-particle electron microscopy. For the purpose of tomography, where the signal already exists in 3D, the classification aims to identify populations of macromolecules with different conformations and various interaction partners. In addition, an emerging application is classification of the sub-tomograms and separation of the selected sub-tomograms into true and false-positives hits after the application of template-matching algorithms. True-positive hits are positions selected from the template-matching that

correspond to the sought macromolecule, and false-positive hits are those that do not correspond to the sought macromolecule, but rather represent other features such as membranes or debris. Template-matching algorithms use a template which is translated and rotated in 3D space and its similarity is measured locally against each position in a tomogram (Frangakis et al., 2002). Finally, at every pixel of the tomogram, a similarity value to the specific template is provided, usually in the form of a normalized cross-correlation score. In any case, it is not known if the sought macromolecule has truly been localized, since an absolute measurement cannot be provided. Independent of the underlying signal, the template is always recovered, which is an effect, known as “template bias”, and is inevitable.

A number of studies have been published recently, which analyze the location, orientation, and identity of macromolecular complexes, based on the outcome of template-matching (Brandt et al., 2009; Ortiz et al., 2010). However, in most cases, selection of the positions at which the macromolecular complexes are placed, i.e. detection of the ‘true’ positive hits, is carried out by assigning an arbitrary cutoff value for the similarity measurement provided by the template-matching. Thus, no computational cross-validation is performed. Classification methods based on the same similarity measurements as the template-matching, cannot discern between true and false-positive hits – both have almost identical similarity values, since this is the criterion according to which the hits are selected. Thus, discernment of true-positive hits from false-positive-hits, i.e. debris arising from pattern-recognition algorithms, becomes a more ad-

* Corresponding author.

E-mail address: achilleas.frangakis@biophysik.org (A.S. Frangakis).

vanced task since the similarity measurement was already reserved for the template-matching. Thus, additional criteria, independent of the similarity criterion used for the template-matching are needed in order to provide a cross-validation for the detection results.

In this work, we introduce KerDenSOM3D, which is the extended 3D variant of KerDenSOM self-organized map (SOM) (Pascual-Montano et al., 2001), (Pascual-Montano et al., 2002), as a new classification method for sub-tomograms which takes the missing-wedge into account. SOMs, which is a category of Neural networks were introduced in electron microscopy for the classification of 2D images by Marabini and Carazo in 1994 (Marabini and Carazo, 1994). The SOMs take the original large set of data and produces a reduced set of good-quality representatives, which are called “code-vectors”. SOMs are excellent tools for classification and can be applied to both cases addressed here: (i) For the classification of individual and distinct populations, such as required when classifying two different macromolecular complexes, and (ii) when distinct populations do not exist, as in the case of sub-tomogram extraction from a tomogram after a template-matching approach.

Classification procedures are well established in single-particle analysis as an integral part of the methodology; however their application in tomography, even though already available, is still developing slowly. The reasons for this are multiple: (i) the defocus variation within the sample, (ii) the alignment errors of the 3D-reconstruction, (iii) specimen deterioration due to irradiation damage, (iv) the missing-wedge, due to the slab geometry of the specimen holder, (v) the significant increase in computing time, (vi) the crowded nature of the cytoplasm, and (vii) the sometimes-unknown nature of the complexes. Of those, the most important and ubiquitous effect arising from the crowded nature of the cytoplasm is that through the influence of the contrast transfer function (CTF) of neighboring macromolecules, the densities of the sought complex are modified.

Foerster et al. incorporated constrained correlation in the calculation of the covariance matrix of the Principal Components Analysis (constrained PCA). Constrained correlation measures the similarity of two sub-tomograms only in those areas where signal exists in both sub-tomograms (Foerster et al., 2008). In this way, it is guaranteed that two identical images within sub-tomograms will have a cross-correlation value of approximately one, independent of the orientation of the missing-wedge. Otherwise, due to the erroneous normalization, the cross-correlation value can vary significantly, meaning that identical images with different orientations would not be identified as such, and hence would not be classified into the same class. In a similar type of analysis, Bartesaghi et al. included classification in the alignment process, which was also accelerated by spherical harmonics (Bartesaghi et al., 2008). Winkler et al. further incorporated classification in the averaging procedure, by also including a weighted average (Winkler et al., 2009). In addition to these methods, maximum likelihood approaches have recently been proposed, which perform the alignment and the sub-tomogram averaging together (Scheres et al., 2009a) and (Stolken et al., 2010). Maximum likelihood methods are shown to be less prone to template bias compared to other classification methods (Sigworth, 1998).

Firstly, we describe the extension of the algorithm in 3D and incorporate the missing-wedge. We then validate the applicability of KerDenSOM3D on simulated data and comprehensively compare it to a number of the above mentioned classification methods. We show that KerDenSOM3D is superior to other classification methods, and that it is better able to distinguish between different macromolecules. We further present the classification of various sub-tomograms extracted from cellular tomograms and use the output of this algorithm as a cross-validation criterion, based on the classification of additional densities deliberately excluded from the template, which can indicate the correct detection of the underlying macromolecular complexes.

2. Theory

For the description of the algorithms, we assume that the sub-tomograms are translationally and rotationally aligned to a common origin. Misaligned particles are treated as debris, and together with debris should be recognized by the classification procedure and classified into classes with low occupancy.

The learning procedure of the neural network is an iterative process. The code-vectors are initialized with similar values to the sub-tomogram average. The code-vectors are related to each other by a two-dimensional grid, and progressively differentiate to represent the different populations in the data.

The similarity between a sub-tomogram (P_i) and a code-vector (V_j) is measured by a function that takes into account both the missing-wedge and the contrast of the sub-tomogram, in a similar fashion as explained in (Foerster et al., 2008). The sub-tomograms are compared to representative code-vectors. These are calculated according to the information present in several sub-tomograms, and thus have a better signal-to-noise ratio (SNR) compared to individual sub-tomograms and they are, in general, missing-wedge free (Fig. 1). Thus, no signal has to be sacrificed when measuring the similarity only at the overlapping region not affected by the missing-wedge, for the sake of an identical similarity value for identical particles with differently oriented missing-wedges.

We assume that we have n sub-tomograms $P_i(x,y,z)$ containing putative macromolecules that were extracted from a tomogram, where (x,y,z) denotes the spatial coordinates of each voxel and i denotes the sub-tomogram index.

The individual sub-tomograms P_i often differ in contrast. In order to adjust them to similar values, a subtraction of their mean value \bar{P}_i and normalization to their standard deviation is performed. Thus, the normalized sub-tomograms $P'_i(x,y,z)$ are defined as:

$$P'_i(x,y,z) = \frac{P_i(x,y,z) - \bar{P}_i}{\sqrt{\sum_{x,y,z} (P_i(x,y,z) - \bar{P}_i)^2}} \quad (1)$$

Alignment of the sub-tomograms P_i to the same orientation requires rotating each by a different set of Euler angles. The region of Fourier space sampled in this rotated sub-tomogram corresponds to the rotation of the originally sampled region by this set of angles, and is denoted by Ω_i .

To constrain the correlation of the sub-tomogram (P_i) and the code-vector (V_j) to their intersection sub-tomogram Ω_i in Fourier

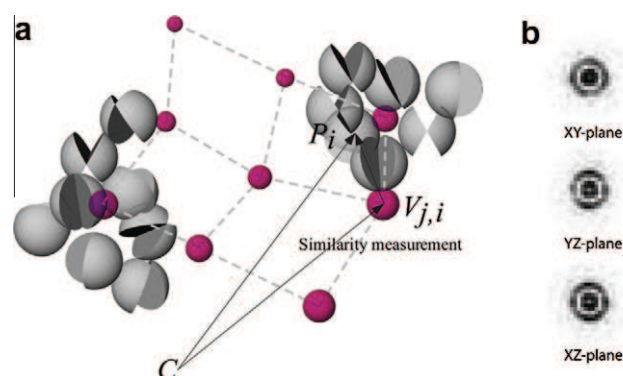


Fig. 1. An illustration of the self-organized map placed in the context of the particles influenced by the missing-wedge. (a) The virtual positions of the code-vectors are indicated by purple spheres. The neighboring code-vectors are related to each other by a two-dimensional network. The code-vectors are typically comprised of more than one sub-tomogram which, if arbitrarily oriented, render the code-vectors without a missing-wedge. (b) Three center slices in the x - y , x - z and y - z planes of the power spectrum (z is the direction of the beam) of a code-vector arising from classification of the GroEL structure presented in Fig. 4, which does not show effects of the missing-wedge.

space, we need to restrict their normalization to the overlapping region Ω_i , that is, the region not affected by the missing-wedge in the sub-tomogram (as the code-vector is considered missing-wedge free). We will denote by $V_j^{\Omega_i}$, the code-vector that results from filtering the Fourier components not included in Ω_i from V_j , that is:

$$V_j^{\Omega_i} = FT^{-1}(FT(V_j) \times \Omega_i), \quad (2)$$

where FT denotes the Fourier transformation and FT^{-1} the inverse Fourier transformation.

In addition, the sub-tomograms are masked with a suitable mask M to focus the classification on certain features or reduce the effect of noise located away from the area containing signal. The sub-tomograms are normalized using the mask M in real space:

$$P'_i(x, y, z) = \frac{M(x, y, z) \times [P_i(x, y, z) - \bar{P}_i]}{\sqrt{\sum_{x', y', z'} [M(x', y', z') \times (P_i(x', y', z') - \bar{P}_i)]^2}} \quad (3)$$

where the mean value \bar{P}_i is constrained to M :

$$\bar{P}_i = \frac{\sum_{x, y, z} P_i(x, y, z)}{\sum_{x', y', z'} M(x', y', z')}. \quad (4)$$

Using the constraint in Fourier space Ω_i and the mask M in real space, the code-vectors are also normalized as:

$$V'_{j,i}(x, y, z) = \frac{M(x, y, z) \times (V_j^{\Omega_i}(x, y, z) - \bar{V}_{j,i})}{\sqrt{\sum_{x', y', z'} [M(x', y', z') \times (V_j^{\Omega_i}(x', y', z') - \bar{V}_{j,i})]^2}} \quad (5)$$

The mean value of the code-vector $\bar{V}_{j,i}$ also has to be constrained to Ω_i and M :

$$\bar{V}_{j,i} = \frac{\sum_{x, y, z} V_j^{\Omega_i}(x, y, z)}{\sum_{x', y', z'} M(x', y', z')}. \quad (6)$$

These definitions originally allowed a *normalized constrained cross-correlation* between sub-tomograms to be defined;

$$CCC(P_i, P_j) = \sum_{x, y, z} P'_{j,i}(x, y, z) \times P'_{i,i}(x, y, z) \quad (7)$$

which was then used as the basis for classification by constrained PCA, that takes the missing-wedge into account (Foerster et al., 2008). In the present work, we use the same approach as described in (Pascual-Montano et al., 2001), but we expand it in 3D and introduce a missing-wedge compensation.

The expression for measuring the similarity of the sub-tomograms P_i to the code-vectors V_j is denoted by:

$$\sum_{i=1}^n \sum_{j=1}^c \|P_i - V_j\|^2 u_{j,i}, \quad (8)$$

where n is the number of sub-tomograms, c is the number of code-vectors, and each $u_{j,i}$ measures the weight of sub-tomogram P_i in the computation for the update of code-vector V_j (see equation (18) in Pascual-Montano). Note also that computation of the $u_{j,i}$ requires evaluating each term $\|P_i - V_j\|$, where the influence of the missing-wedge has to be taken into account.

The similarity of a sub-tomogram and a code-vector is thus computed as the Euclidean norm of the difference,

$$\|P_i - V_j\|^2 = \sum_{x, y, z} (P_i(x, y, z) - V_j(x, y, z))^2, \quad (9)$$

which in our approach is simply substituted by the Euclidean norm of the difference of the normalized, constrained sub-tomograms $\|P'_i - V'_{j,i}\|^2$.

The KerDenSOM3D algorithm converges to a set of code-vectors that simultaneously maximizes two requirements: (a) their similarity to a set of sub-tomograms and (b) their smoothness on a set grid, which corresponds to a similarity to other neighboring code-vectors. Both requirements are implemented in a missing-wedge compensated cost-function, similarly to the one mentioned in Eq. (14) in Pascual-Montano (Pascual-Montano et al., 2001). Note that the balance between similarity and smoothness is controlled by a smoothness parameter ϑ , which appears in the numerical algorithm in an annealing strategy: it decreases slowly with each step of the annealing process, from an initial value ϑ_1 to a final value ϑ_0 following:

$$\vartheta = \exp(\ln(\vartheta_1) - (\ln(\vartheta_1) - \ln(\vartheta_0)) * \text{iter}/\text{MaxIter}). \quad (10)$$

For each value of ϑ the code-vectors are computed by iterating on t the relation

$$V_j^{(t+1)} = \frac{\sum_{i=1}^n u_{j,i}^{(t)} P_i + \vartheta \bar{V}_j^{(t)}}{\sum_{i=1}^n u_{j,i}^{(t)} + \vartheta}; u_{j,i}^{(t)} = u_{j,i}^{(t)} \left(\left\{ \|P_i - V_j^{(t)}\| \right\}_{i,j}, \left\{ V_j^{(t)} \right\}_j, \vartheta \right), \quad (11)$$

until convergence is achieved ($\bar{V}_j^{(t)}$ is the average of the code-vectors that are neighbors of $V_j^{(t)}$ on the grid, excluding $V_j^{(t)}$ itself from the average).

The smoothness parameter ϑ and the grid size are practically the only parameters relevant for a classification. The following empirical values are suggested for the classification of datasets: $\vartheta_1 = 30 \frac{n}{c}$, $\vartheta_0 = 3 \frac{n}{c}$, the number of annealing steps set to 50 and a grid size of 5×5 or larger. The smoothness parameter depends on the number of sub-tomograms n and the size of the grid c .

The formula for the smoothness parameters ϑ can be derived by assuming that the contribution of all the sub-tomograms to all code-vectors is the same, thus: $u_{j,i} = \frac{1}{c}$, $\forall j, i$ and $V_j^{(t+1)} = \frac{\vartheta \bar{X}_j + \vartheta \bar{V}_j^{(t)}}{\vartheta + \vartheta}$, where $V_j^{(t+1)}$ is the j th code-vector in $(t+1)$ th step, n is the number of the sub-tomograms, and all code-vectors are initially identical. For datasets with same image size, recorded at similar conditions and similar SNR, the ratio $\frac{n}{c}$ has to be set to the empirical values suggested above in order to obtain a good performance. Thus, an exhaustive search of parameters can be avoided, which is of great advantage to any user, since a scan of these parameters may often be the most burdensome and time-consuming part of the classification procedure.

3. Materials and methods

The algorithm was implemented in MATLAB (The MathWorks, Inc.) and C/C++ for massive parallel processing. The algorithms are available upon request, and the source code can be downloaded at <http://www.biophys.uni-frankfurt.de/frangakis/>.

3.1. Acquisition conditions

The tilt-series were recorded on a 300 kV, Tecnai G2 Polara FEG and a CM300 transmission electron microscope (FEI, Eindhoven, The Netherlands) equipped with a Gatan post-column GIF 2002 energy filter (Pleasanton, CA). Tomographic single-tilt-series were acquired using either the TOM Toolbox or the UCSF software package with zero-loss filtering (slit-width of 30 eV) (Nickell et al., 2005; Zheng et al., 2004). The individual projection images recorded with the microscope were interactively aligned with respect to a common origin using 10 nm colloidal gold particles distributed in the samples as fiducial markers. Reconstructions were performed

using weighted backprojection. Visualization was performed with the Amira package (Pruggnaller et al., 2008).

3.2. Simulated data

The simulated data used in Foerster et al. was also used here in order to allow for a direct comparison of the performance of the algorithms. A detailed description for generating the data is given in Foerster et al. chapter 3.2 (Foerster et al., 2008). The goal of the simulation was to produce realistic tomograms. In brief: The density of the macromolecule was rotated around a single axis, projected onto a plane and superimposed with Gaussian noise. On the resulting 2D-image, the CTF was applied. The CTF was composed of two terms: (i) the idealized CTF assuming pure phase contrast at a particular defocus and (ii) a modulation-transfer function (MTF) that describes the imaging properties of the CCD-camera. On the resulting image, noise was added representing background noise due to non-elastic scattering events and the readout noise of the CCD-camera. The resulting projections were then merged into the sub-tomogram using weighted backprojection. The main parameters varying in the simulation are the half-opening angle W of the missing-wedge (as the half angle of the opening of the missing-wedge, i.e. a half-opening angle W of 20 degrees would correspond to a tilt-series from -70 to $+70$ degrees) and the SNR.

3.3. Acquisition of the *in vitro* GroEL and GroEL/GroES tomograms

For comparison purposes, the same sub-tomograms were used as in Foerster et al. and Scheres et al. In brief: 1 μM GroEL14 was incubated with and without 5 μM GroES7 in a buffer containing 12.5 mM Hepes (pH 7.5), 5 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 5 mM ADP for 15 min at 30 °C. Protein solutions (3.5 μl) were mixed with 0.5 μl of a 10 nm BSA-colloidal gold suspension, applied to 300 mesh grids coated with Lacey carbon films (Plano, Wetzlar, Germany) and vitrified by plunge-freezing. Single-axis tilt-series were collected covering an angular range from -65° to $+65^\circ$ with 2° or 2.5° angular increment. Images were recorded on a $2\text{ k} \times 2\text{ k}$ pixel CCD camera at a defocus level of $-4\ \mu\text{m}$ to mimic the real experimental conditions of cellular tomography. The pixel size at the specimen level was 0.6 nm.

3.4. Cryo-electron tomography of *Mycoplasma pneumonia* and *Spiroplasma melliferum*

For cellular tomography, we analyzed tomograms used in previous studies (Kuerner et al., 2005; Seybert et al., 2006). Detailed description of the cultivation of the strains can be found in the references. The tilt-series were acquired between -66° and $+66^\circ$ for *M. pneumonia* and -60° to $+60^\circ$ for *S. melliferum* with a tilt increment of 1.5° . The images were recorded at defocus between -5 and $-10\ \mu\text{m}$ and a pixel size on the CCD camera of 0.6 nm or 0.68 nm at the specimen plane.

4. Results

4.1. Simulated data: "Open" and "Closed" thermosomes

In order to assess the quality of the classification with KerDenSOM3D, we applied it to the same simulated tomographic data as in Foerster et al. (Foerster et al., 2008), where the constrained correlation was tested. This provides the possibility for direct comparison of the two techniques. The simulated object used was the archaeal chaperonin thermosome in its open and closed conformations. For the closed conformation, the pdb coordinates from the

X-ray structure were used (1A6D)(Ditzel et al., 1998), and for the open conformation, a structure derived from a molecular model fitted to a single-particle EM density map was used (Nitsch et al., 1998).

We have calculated the classification with KerDenSOM3D for various SNR and opening angles of the missing-wedge. The parameter ϑ has been set to 8000 and is decreased to 100 in 20 steps. The classification shows that for all simulated missing-wedge half-opening angles W (from 0° to 40° with 10° steps) and SNRs (from 1 to 10^{-3} with a geometric progression of 10), a perfect classification is possible (Fig. 2, only the results for W at 20° and 30° shown). Decreasing the SNR to 5×10^{-4} shows the limits of the classification procedure, where the two classes can no longer be unambiguously distinguished. KerDenSOM3D shows clusters which are significantly more compact than clusters obtained with a version of the algorithm which is not missing-wedge compensated. This is the same observation as in constrained-correlation, and has the effect that when the SNR decreases and the clusters are more compact, the distinguishing capability of the missing-wedge corrected algorithm is retained. Typically 5×5 code-vectors are used for the classification. The requirement for *a priori* knowledge of the number of classes present in the data is less stringent than in other techniques, as an excess of code-vectors does not have to correspond specifically to the number of anticipated classes. Instead, an excess of code-vectors results in the vast majority not having any sub-tomograms assigned to them, and can be easily discarded for further analysis.

In order to compare the performance of KerDenSOM3D to other approaches, the same artificial thermosome datasets (missing-wedge with a half-opening angle of 20° and 40° , as well as an SNR variation from 10^{-4} to 10^{-2}) have been classified using different techniques. For the constrained PCA, the approach described in Foerster et al. is used (Foerster et al., 2008). For the self-organized maps (SOM) and maximum likelihood (MLTOMO) we use the Xmipp package (Sorzano et al., 2004). For the correlation approach as used in (Stolken et al., 2010) we use maximum cross-correlation (MCC) and hierarchical clustering (Foerster et al., 2008).

We use the same "class purity" measurement as used in (Scheres et al., 2009a), which is defined as the ratio of correctly classified sub-tomograms to the total number of sub-tomograms, in order to evaluate the performance of the algorithm and make a comprehensive comparison of the different approaches (Fig. 3) and (Table 1). In comparison to the other approaches, KerDenSOM3D shows a superior performance. The class-purity at lower SNR and larger missing-wedge angle is always better in KerDenSOM3D compared to constrained PCA, MLTOMO, MCC and SOM. In direct comparison to the constrained PCA approach, which could not offer a perfect classification with the simulated datasets at the stage $W = 40$, SNR = 10^{-3} , KerDenSOM3D classifies the data perfectly. The performance difference can be explained by the fact that during KerDenSOM3D, each sub-tomogram is compared to code-vectors that have a high signal-to-noise ratio, while during constrained PCA the covariance matrix is calculated by mutual comparison of the individual sub-tomograms, and is thus more prone to be affected by noise and the missing-wedge. The missing-wedge compensation improves the performance. Even more importantly, the KerDenSOM3D approach performs significantly better than previous algorithms. KerDenSOM3D is computationally more expensive than the constrained PCA and SOM approaches, but one classification process could still be done on a computer with Intel XEON E5620 at 2.4 GHz within a few hours. We also find that KerDenSOM3D performs better than MCC, while the constrained PCA is still better than MCC. For the comparison with MLTOMO we used the XMIPP package after we tuned the parameters to achieve the best possible classification (Fig. 3).

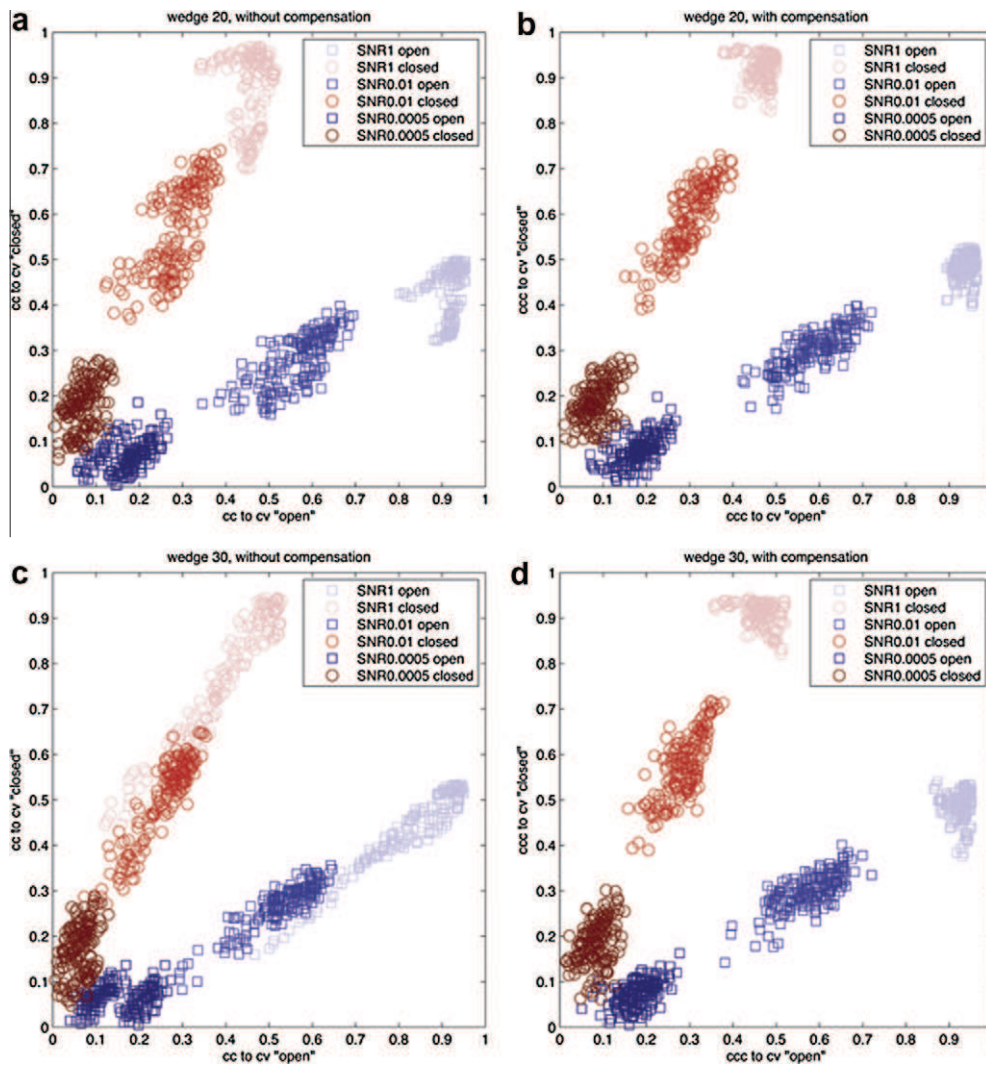


Fig. 2. Comparison of the classification results of simulated open thermosomes (squares) and closed thermosomes (circles) at different SNR (1, 0.01 and 0.0005) and different missing-wedge opening angles, with KerDenSOM3D without missing-wedge compensation in the left-hand column and KerDenSOM3D in the right-hand column. The upper row corresponds to a missing-wedge half-opening angle of 20° and the lower row to 30° (This would correspond to a tilt-series of $\pm 70^\circ$ and $\pm 60^\circ$, respectively). Every color associates to different SNRs, where lighter colors correspond to higher SNR. The similarity of a certain particle to the classifying code-vector is measured. (a) Missing-wedge half-opening angle of twenty degrees. At high SNR, the clusters (which occur when the similarity of a sub-tomogram average with the two highest populated code-vectors is plotted) can be distinguished from one other. (b) The algorithm's performance with missing-wedge compensation. Note that the clusters are much more compact. (c) Missing-wedge half-opening angle of 30 degrees. With decreasing SNR the clusters come closer together, and merge together at a SNR of 0.0005. (d) Classification with missing-wedge compensation still produces an almost a perfect classification even at a SNR of 0.0005, but here the algorithm begins to reach its limits. The clusters remain compact at high SNR. The better performance of the algorithm is justified by the more compact clusters produced by the missing-wedge compensation incorporated into the algorithm (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

5. *In vitro* GroEL and GroEL/GroES tomograms

We further assess the classification performance of KerDenSOM3D on the example of *in vitro* GroEL and GroEL/GroES tomograms (Foerster et al., 2008) and (Scheres et al., 2009b). A network of 5×5 code-vectors was used for the classification. In this case, the dataset consisted of 786 sub-tomograms, 214 of them extracted from the GroEL tomograms and 572 from the GroEL/GroES tomograms. The outcome of the classification is a separation of the dataset into 25 classes and their class-averages are visualized in Fig. 4. On the left-hand side of the grid, classes resembling a GroEL structure are visualized and are indicated with "Class1" (The selection of specific classes depends on the number of sub-tomograms classifying to a specific class, but remains user-dependent). 203 out of 214 sub-tomograms extracted from the GroEL tomograms are classified in this class, providing a class-purity of at least 95% (individual GroELs from the GroEL/GroES dataset

might also be included). On the upper-right corner of the grid, several classes resembling a GroEL/GroES structure are apparent and are indicated with "Class2". Here 177 from 181 sub-tomograms classify in this class, with a class purity of 98%. Two other class-averages are visible, namely "Class3" and "Class4", which are comprised of 181 sub-tomograms. These classes depict structures that are significantly smaller than a GroEL molecule and potentially contain either half particles – an outcome of radiation damage to the sample – or erroneously aligned side-views. Ninety-six percents of these sub-tomograms are extracted from the GroEL/GroES tomograms. Similar ring-structures were also obtained by other classification methods (Foerster et al., 2008; Scheres et al., 2009a) (Supplementary Fig. 1).

We directly compare the performance of KerDenSOM3D with constrained PCA and MLTOMO, based on the complete data set of 786 sub-tomograms, and the results obtained in Foerster et al. and Scheres et al. The class purity obtained for the GroEL tomo-

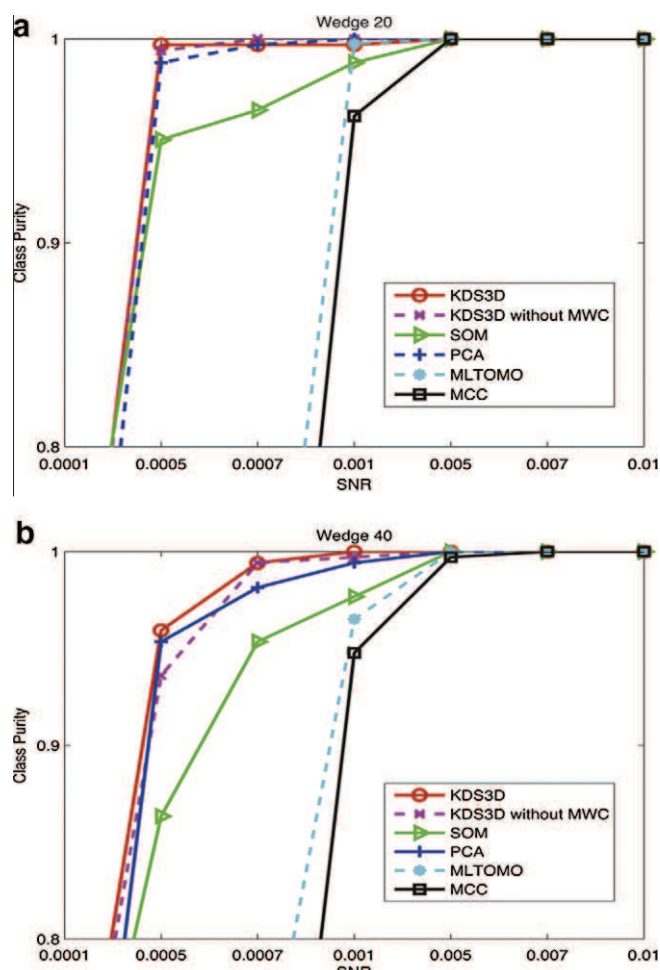


Fig. 3. Comparison of several techniques based on the class-purity criterion. The abbreviations used in the graph are: KDS3D for KerdenSOM3D (red circle), KDS3D without MWC for KerdenSOM3D without missing-wedge compensation (purple x), SOM for Self-organized map according to the simplest edition (green triangle), PCA for the constrained principal components analysis (blue cross), MLTOMO for the maximum likelihood approach (light blue dot), and MCC for maximum constrained cross-correlation (black square). (a) At a missing-wedge half-opening angle of 20° (corresponding to a tilt-series from $\pm 70^\circ$), KerdenSOM3D performs better than previous techniques and the non-missing-wedge compensated algorithm. (b) The improved performance of the algorithm is better appreciated for a missing-wedge half-opening angle of 40° (corresponding to a tilt-series from $\pm 50^\circ$), where KerDenSOM3D outperforms any previous approach (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Table 1

Comparison of the different methods on the basis of class-purity. The class-purity in KerDenSOM3D is better, but also the absolute number of particles associated to the classes is larger (Table 2).

| Class Purity | KerDenSOM3D | MLTOMO | PCA |
|--------------|-------------|--------|-----|
| GroEL | 95% | 90% | 90% |
| GroEL/ES | 98% | 86% | 95% |

grams is 90% for MLTOMO and 90.5% for constrained PCA, compared to 95% for KerDenSOM3D. The class purity for the GroEL/GroES tomograms is 86% for MLTOMO and 95.5% for constrained PCA, compared to 98% for KerDenSOM3D. Also, the absolute number of particles selected from KerDenSOM3D is much larger compared to the previous approaches, where a number of sub-tomograms was discarded as debris. This shows both a qualitative and a quantitative superiority (Table 1 and Table 2).

6. Classification and cross-validation of template-matching results

When tomograms are scanned with a particular template, a number of hits (corresponding to the positions and orientations of the putative macromolecule) are obtained. However, how many of these hits actually correspond to the sought macromolecule, i.e. how many false-positives are present in a particular population, is unclear. Hence, in order to increase the confidence in the “particle authenticity” of the final selected volumes after classification, we developed the following approach and applied it to the case of the ribosome. Instead of using the whole ribosome as a template, we used only the large ribosomal subunit (50S) for scanning, and after classification with KerDenSOM3D we analysed whether the smaller ribosomal subunit (30S) was recovered. In principle, this approach is also applicable when a macromolecule is separated into arbitrary pieces which are not related to any particular structural component. As the tomogram is correlated with the 50S subunit, the resulting probability map only contains the probabilities that the large ribosomal subunit is located at particular positions in the electron tomogram. Our approach is based on the fact that in a functional ribosome, the 50S subunits will have the 30S subunit attached. Hence, when classifying the hits obtained for the 50S subunit, we expect that at least some of the classes will show the ribosome as a complete entity (Fig. 5a). Correspondingly, we expect that the sub-tomograms composing those classes will have a significantly reduced false-positive rate, since the evidence of the recovery of an additional density not included in the original template increases the level of reliability. Thus we use this approach as a cross-validation criterion for the hits.

We used two whole-cell tomograms, the first from *M. pneumoniae* (Henderson and Jensen, 2006; Seybert et al., 2006) and the second from *S. melliferum* (Kuerner et al., 2005) (Fig. 6). By means of normalized cross-correlation-based techniques, 919 sub-tomograms containing putative ribosomes were extracted from the *Mycoplasma* cell and 876 sub-tomograms from the *Spiroplasma* cell (Frangakis et al., 2002). The cells were individually processed from this point but using exactly the same approach for both. The number of sub-tomograms extracted from a tomogram corresponds to the number of all local maxima in the cross-correlation map. Thus, no user-interaction or arbitrarily set thresholds are involved in this step. Here we assume that the template approach works sufficiently well to produce a local maximum at the position of a particle, independent of the intensity of the local maximum in the cross-correlation map, which depends largely on the normalization (i.e. the local cross-correlation value). The local cross-correlation value is the result of the interplay of a number of parameters such as the CTF, molecular crowding, orientation of the particle, and SNR. The sub-tomogram average created from the selected positions is almost identical to the 50S subunit template. Template bias is an inherent consequence of the template-matching procedure, which results in the average of the positions recovered to appear identical to the template used for the search. No other sub-tomogram average can be obtained by this process.

Without a doubt, the selected positions, i.e. the hits, do not always represent the sought macromolecule (Fig. 6). In order to separate the map into true-positive and false-positive hits, we masked the region of the location of the 30S subunit with respect to the template in each one of the extracted sub-tomograms and classified them using KerDenSOM3D. The classification should be able to distinguish the extracted sub-tomograms with putative ribosomes into classes where the full ribosome appears and classes of the remaining positions, which we will treat as false-positive hits. For the KerDenSOM3D classification, a network of 5×5 code-vectors was used. Each image of the grid corresponds to an individual class-average in which the central section is presented.

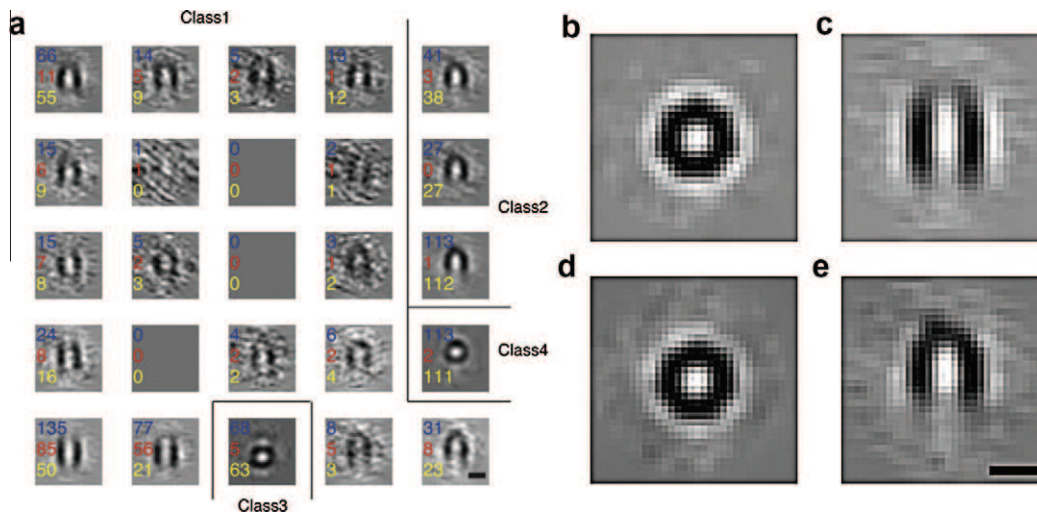


Fig. 4. Classification of the GroEL and GroEL/GroES *in vitro* datasets. (a) Visualization of the central slice of the class-averages in a 5×5 self-organized map. In the inset of each class-average, the number of sub-tomograms classifying to each code-vector is indicated. In blue, the total number of sub-tomograms in the class; in red, sub-tomograms originating from the GroEL dataset; in yellow, sub-tomograms originating from the GroEL/GroES dataset. Four classes can be distinguished, namely, Class 1 which represents GroEL (left part of the map); Class 2 which represents GroEL/GroES complexes (upper right corner of the map); and two classes with half particles (Classes 3 and 4). The vast majority of the half particles originate from the GroEL/GroES dataset. The upper-right class contains GroEL/GroES-similar particles and the lower-right class contains GroEL-similar particles. The scale bar in the lower-right corner applies to all images and is 10 nm in length. (b) x–y section and (c) y–z sections of the averages of the particles classifying to a GroEL-similar class. 7-fold symmetry can be observed. 203 out of 214 sub-tomograms extracted from the GroEL tomograms contribute to this average. (d) x–y section and (e) y–z section of the average of the particles classifying to a GroEL/GroES-similar class. 177 out of 181 particles classified to this class originate from the GroEL/GroES tomograms, presenting a class purity of 98%. The scale bar is 10 nm (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Table 2

Comparison of the different methods on the basis of the total number of sub-tomograms associated to each class. From the complete set of 786 sub-tomograms, KerDenSOM3D associated 424 sub-tomograms to GroEL and 181 to GroEL/ES. The remaining particles are associated to other classes. In contrast, PCA associated 377 sub-tomograms to GroEL and 222 to GroEL/ES. The class-purity in KerDenSOM3D is better, but also the absolute number of particles associated to the classes is larger. The absolute number of particles classified to the class GroEL/ES with MLTOMO is unknown (Scheres et al., 2009a) and is marked with question mark.

| | KerDenSOM3D | | | MLTOMO | | | PCA | | |
|----------------|-------------|-------|----------|--------|-------|----------|-------|-------|----------|
| | Total | GroEL | GroEL/ES | Total | GroEL | GroEL/ES | Total | GroEL | GroEL/ES |
| GroEL Class | 424 | 203 | 221 | ? | 193 | ? | 377 | 182 | 195 |
| GroEL/ES Class | 181 | 4 | 177 | 150 | 21 | 129 | 222 | 10 | 212 |

In the upper-right part of the map (second row, last column), a density next to the ribosome is shown, which resembles the density of the 30S subunit well (Fig. 5a and b).

Unfortunately, classifying density within a mask at a specific position also introduces a bias. As a thought-experiment, we note that the sub-tomograms which are used for the averaging are classified according to the high intensity values within the mask. Consequently, the resulting class-averages look similar to the total average of all the sub-tomograms (the 50S subunit), with an increased density at the location of the mask which was used for the classification. In order to increase the confidence that we recover the 30S subunit density in its true position, and that no possible bias is introduced by either the mask or the shape of the mask, we performed an additional cross-validation experiment. We exhaustively scanned the surrounding of the 50S subunit with various transformed 30S subunit masks, and verified that the properties are different, when the mask is at the true 30S subunit position compared to any other position.

Let us call M the mask that tightly covers the true position of the 30S subunit. We define a set of affine transformations of M that produce a set of masks that cover the neighborhood of the 50S subunit. Each transformed mask, $M_d = R_d(M)$, is called a *decoy* (Fig. 5c). Here, each R_d represents the action of a rotation followed by a translation, and d ranges from 1 to 104, where decoy 104 is reserved for the true position of the 30S subunit. The set of transfor-

mations was chosen arbitrarily, but it was ascertained that all voxels in the immediate surroundings of the 50S subunit were located within some decoys (Fig. 5c).

For each decoy M_d , we classify the entire set of sub-tomograms $\{P_i\}_{i \in I}$ using only the voxels included in the decoy. The results of each classification can be represented by a partition of the set I of sub-tomogram indexes into C subsets:

$$I = \bigcup_{c=1}^C I^{d,c}. \quad (12)$$

Here, $I^{d,c}$ represents the indices of the sub-tomograms that were assigned to a class c by the classifying algorithm, where c is the number of classes generated. The sub-tomogram average of each class is denoted by $X^{d,c}$ with the definition:

$$P^{d,c}(x, y, z) = \frac{\sum_{i \in I^{d,c}} P_i(x, y, z)}{\#I^{d,c}}, \quad (13)$$

where $\#$ expresses the number of elements in a set. Now, we compute in each classification the class number of the average which best resembles the template of the 30S subunit (transformed as the corresponding decoy). This resemblance, denoted by s_d is measured by cross-correlation, only taking into consideration the voxels included in M_d . The resulting class number is denoted by c^d .

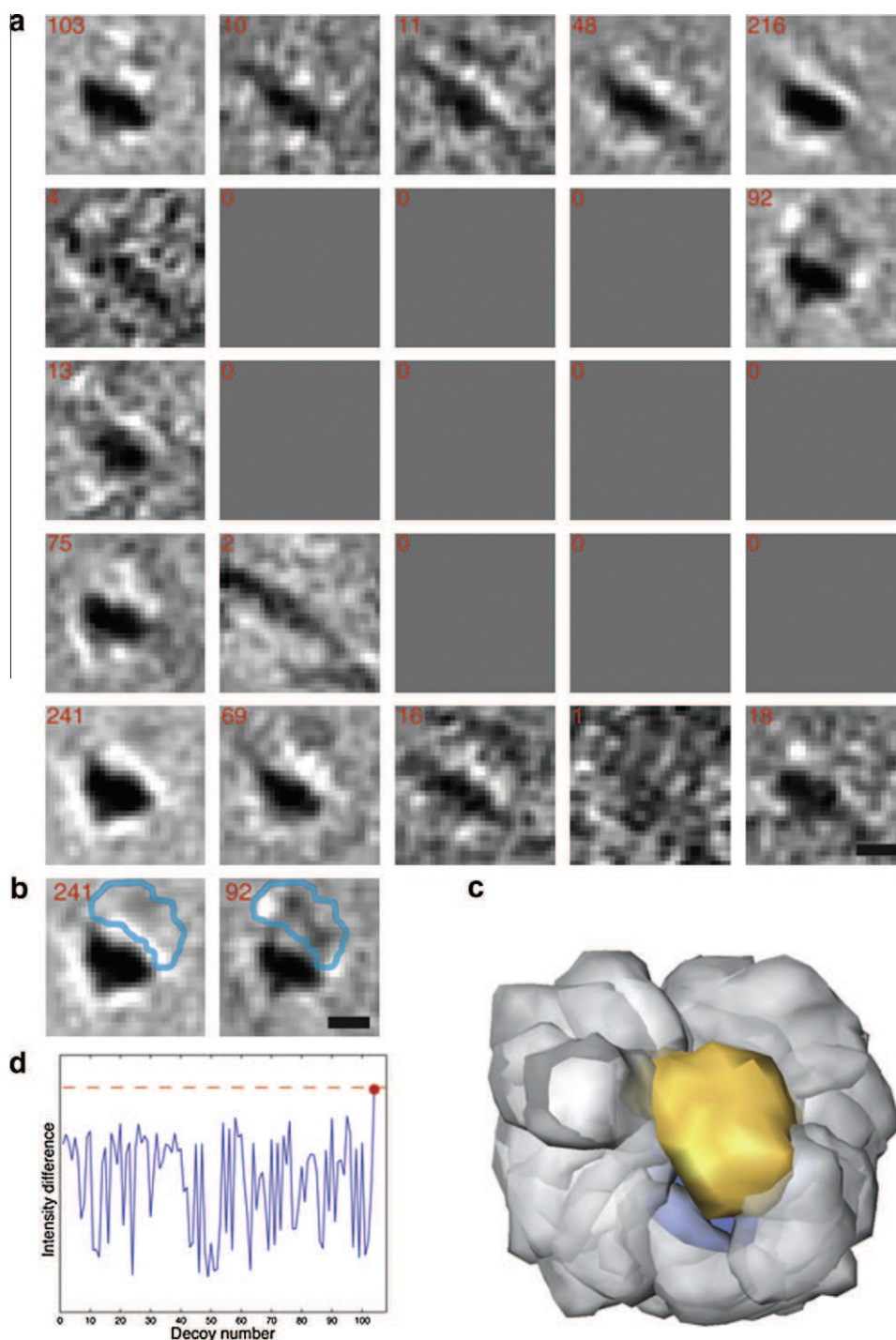


Fig. 5. Application of KerDenSOM3D on the selected sub-tomograms of a *Mycoplasma pneumoniae* cell after template-matching with the large ribosomal subunit (50S). (a) A 5×5 self-organized map depicting the central slices of the class-averages after classifying all sub-tomograms. Various class-averages can be observed, all of which appear similar to the 50S subunit, which was used as the scanning template. In the upper-right corner, a ribosome-similar structure with significant density in the 30S subunit area is observed. The values in the red inset indicate the number of sub-tomograms contributing to that specific class-average. The scale bar is 10 nm. (b) Two selected class-averages: the left slice shows a class with a 50S subunit with no density in area of the 30S subunit mask and the right slice shows a class with density similar to the 30S subunit. (c) Visualization of the positions of the various 30S-shaped decoys for the confidence measurement in the context of the density next to the 50S subunit. One hundred and three different decoys visualized with transparent gray were used to verify the statistical significance of the recovered density in the 30S subunit region. Decoy 104 was reserved for the true position of the 30S subunit and is visualized in yellow. (d) The cross-validation method for the decoys. A plot of the normalized decoy similarity shows that in the true region of the 30S subunit a statistically significant density is present (decoy 104) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Within the decoys we establish a validation using the overall intensities measured in the neighborhood of the 50S subunit. The properties of the decoy $d = 104$ are different compared to all other decoys, suggesting that this is the true position of the 30S subunit

(Fig. 5c). If the 50S subunit is attached to the 30S subunit, the measurement of the similarity in the area of the decoy ($d = 104$) covering the 30S subunit, should attain a clear maximum. We measure this value for each possible pair of decoys d, d' by the quantity

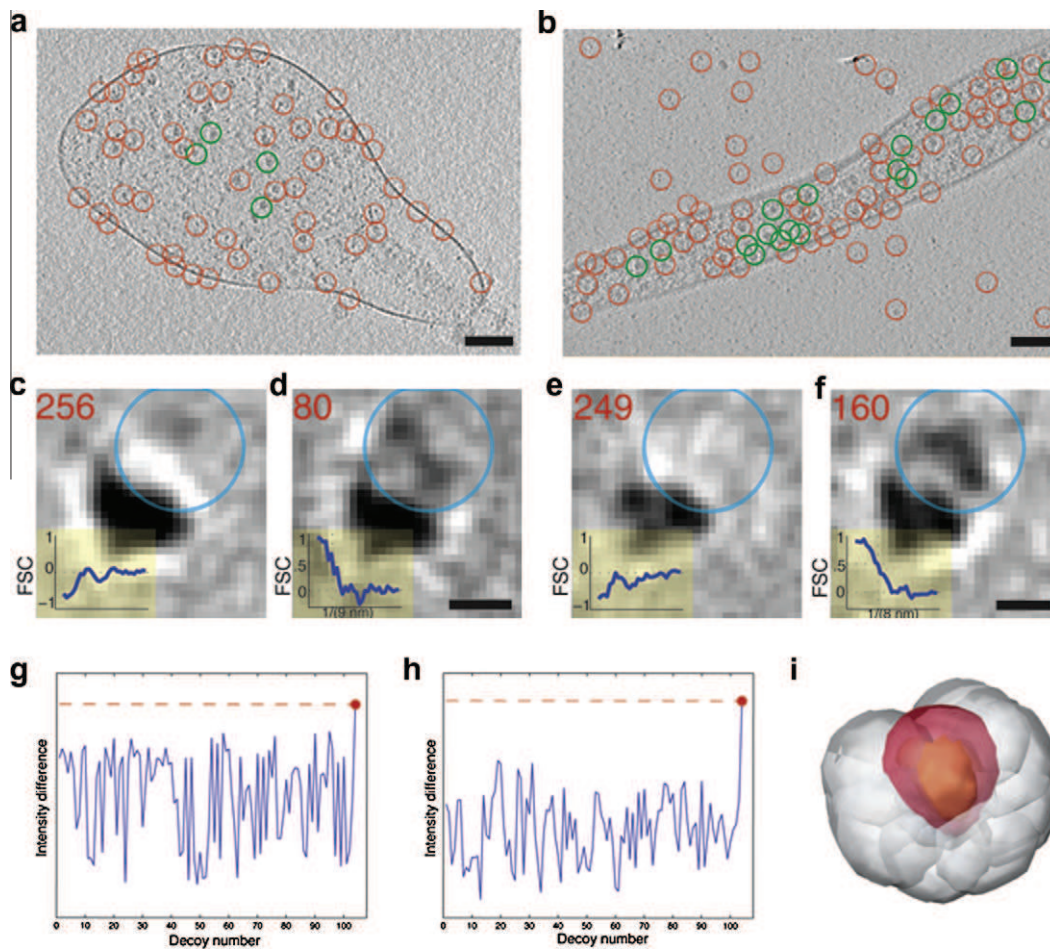


Fig. 6. Application of the classification method on the example of *Mycoplasma* and *Spiroplasma* cells. Cross-validation of the experiment shown in Fig. 5 using a spherical mask to rule out bias resulting from the mask shape. (a) A 12 nm thick slice through the tomogram of a *Mycoplasma* cell with circles indicating the positions of the hits from the template-matching with the 50S subunit template. Red circles indicate the hits which contribute to a class-average that does not contain a density at the true area of the 30S subunit, and are considered as false-positives. The green circles indicate hits which have a prominent density in the area of the 30S subunit, and thus are considered as true-positive hits. The scale bar is 50 nm. (b) A 13.6 nm thick slice through the tomogram of a *Spiroplasma* cell, with the same convention as used in (a). (c and d) Two class-averages selected from the classification result from KerDenSOM3D for the *Mycoplasma* cell. At top-left corner in red, the number of sub-tomograms contributing to each average is indicated. In the bottom-left corner, the Fourier shell correlation is shown in blue. The FSC is calculated exclusively within the area of the light-blue contour representing the spherical mask used, which covers a slightly larger area than the 30S subunit. While the FSC is negative for most of the low frequencies in (c) the FSC in (d) shows a resolution of 9 nm at 0.5 cut-off. The scale bar is 10 nm. (e and f) are the equivalent representation of (c and d) for the *Spiroplasma* cell. The resolution is estimated to be ~8 nm. The scale bar is 10 nm. (g and h) Plot of the confidence measurements of the various decoys. In the case of *Mycoplasma* (g) the difference to the background is lower, compared to the higher confidence measurement of the decoys in *Spiroplasma* (h). This indicates that the tomogram of *Spiroplasma* cell has a better SNR than that of the *Mycoplasma* cell. (i) Visualization of the positions of the various spherical decoys in the context of the ribosome. 103 different decoys visualized with transparent gray were used to verify the statistical significance of the recovered density. Decoy 104 was reserved for the true position of the 30S subunit and is visualized in red (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

$$D[d, c^d; d'] = \frac{\sum_{x,y,z \in M_d} P^{d,c^d}(x,y,z)}{\#M_d}, \quad (14)$$

which contains the mean intensities of the sub-tomogram average of the best class attained using mask M_d as a decoy, computed only in the voxels in M_d .

The argumentation behind this definition is the following: one would expect that, if the 30S subunit is actually present in some sub-tomogram P_i , the classification generated using a mask situated at the real position of the 30S subunit will generate some classes that correspond to an isolated 30S subunit. Classifications using arbitrary decoys will also find combinations of sub-tomograms that accumulate intensities in the respective decoys, but these intensities will not be consistent with the size or the shape of the 30S subunit. They might be created by other possible high intensity objects neighboring each individual copy of the 50S sub-

unit. As these objects are not necessarily restrained in the area of the respective decoy (which we have defined in a random manner) they might extend to areas outside of the decoy, or fail to fill the decoy consistently.

We observe that the decoy placed at the true position of the 30S subunit shows different properties. We measure these properties by the intensity difference,

$$D[d, c^d; d] - \max_{1 \leq d' \leq 104, d' \neq d} D[d, c^d; d'], \quad (15)$$

which shows a clear peak for $d = 104$. This indicates that the classification based on the true decoy detects intensities and shapes concentrated on the decoy itself, while classification based on other decoys detects intensities that are less concentrated on the decoy in question and have a different shape. This is evidence that we are effectively recovering the true 30S subunit when using the above mentioned criterion.

In order to also verify that the density which appears next to the 50S subunit is not dependant on the shape of the mask, we rerun the process described above with featureless spherical decoys. For sub-tomograms extracted from both cells, a density resembling the shape of the 30S subunit appears again at the proper position with respect to the 50S subunit. Using the Fourier shell correlation (FSC), the resolution of these sub-tomogram averages was estimated to be 9 and 8 nm for the *Mycoplasma* and *Spiroplasma* datasets respectively (Fig. 6d and f). The FSC is calculated only within the recovered density, and not for the complete structure, in order to avoid any improvement originating from template bias. When the intensity difference values within the decoys are measured, the decoy at the true position shows a significant higher value than the surrounding (Fig. 6g and h).

The above procedure was applied on both cellular tomograms, and the results are similar, with the *Spiroplasma* tomogram showing a slightly higher SNR and a higher confidence in the detection of the 30S subunit. In order to verify the improved selection of sub-tomograms, we perform sub-tomogram averaging with the selected sub-tomograms from both cells after classification. In contrast to previous studies we do not use the ribosome as a template for the sub-tomogram averaging. Instead we use a sphere, since even low-pass filtered ribosome templates generate a high amount of template bias. We further randomize all translations and reset all rotations of the sub-tomograms. Interestingly, our results show that sub-tomogram averaging of putative ribosomes extracted from cellular tomograms does not converge to a ribosome structure without the use of a ribosome as a starting template. This is however the case when the dataset has been pre-classified with KerDenSOM3D (Fig. 7).

7. Discussion

Classification algorithms are necessary for extracting different structural states of macromolecular complexes. Here we use KerDenSOM3D for this purpose, but also as a cross-validation of template-matching algorithms applied on electron tomograms. We have performed a thorough comparison with already existing algorithms, and find that KerDenSOM3D performs better than previously suggested approaches, as can be seen in the thermosome and the GroEL/ES datasets. The missing-wedge compensation of the algorithm slightly improves the performance. The reason for this is twofold: (i) The sub-tomograms are compared to code-vectors with no missing-wedge, and (ii) The objects used in our comparison have almost evenly distributed information. The better performance compared to constrained PCA or MCC can be explained by the fact that no additional signal is sacrificed for the sake of a better measurement. Our performance measurements do not match those presented in (Stolken et al., 2010), where it is shown that the MCC performs significantly better compared to the constrained PCA. However, it is implausible that constrained PCA performs worse than MCC, since exactly the same measurement is used for the creation of the covariance matrix in both cases (Foerster et al., 2008). We have performed a thorough comparison between the various methods, and have tuned the parameters, in order to get the best possible classification results from all the approaches.

Even though a perfect classification is not expected for noisy data such as cryo-electron sub-tomograms, it is plausible that the classification is hampered beyond the optimal level by systematic errors. Arguably, the major source of misclassification of sub-tomograms arises from misalignment of sub-tomograms with respect to another. The classification algorithm assumes that the particles are aligned with respect to each other, since the initial alignment from the template matching is performed with a tem-

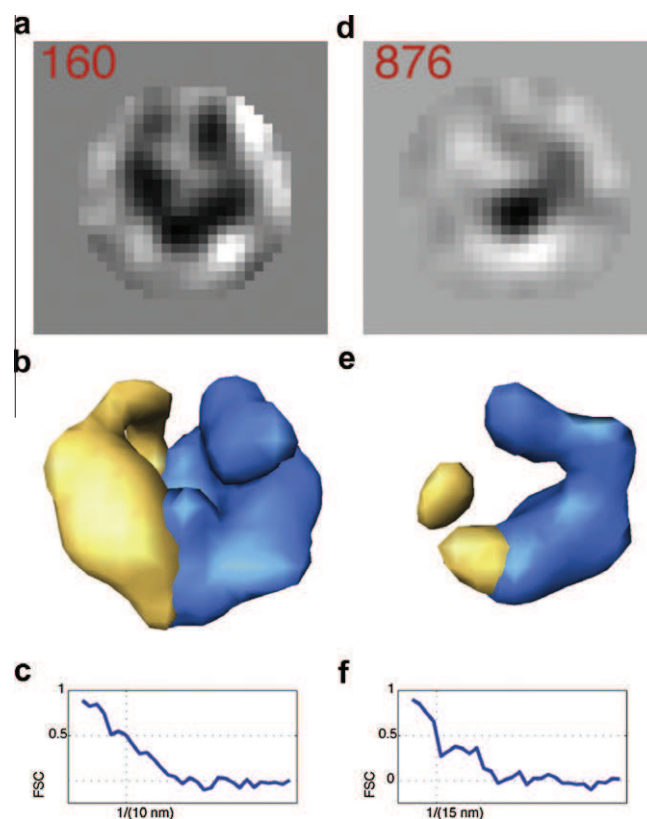


Fig. 7. Comparison of the sub-tomogram averaging results of the sub-tomograms extracted from the *Spiroplasma* cell after randomizing their orientations, using a Gaussian sphere as a starting reference. (a, b and c) The 160 sub-tomograms, which were classified by the class-average containing an additional density at the position of the 30S subunit, converge to a structure similar to the ribosome. (a) Central slice through the sub-tomogram average depicting the general shape of the ribosome. (b) Isosurface representation of the ribosome, indicating the two sub-units in different colors. (c) FSC against the entire ribosome obtained by X-ray crystallography shows a resolution of ~ 10 nm using the 0.5 cut-off criterion. (d, e, and f) Sub-tomogram averaging of all selected sub-tomograms converges to a meaningless feature, with a resolution of ~ 15 nm. Interestingly, if the angles are not randomized, these sub-tomograms converge to the 50S subunit, which was used as the search template, and do not show the additional 30S subunit. (d) Central slice through the sub-tomogram average. (e) Isosurface of the resulting structure aligned with the ribosome template, and colored according to the positions of the sub units. (f) FSC correlation against the ribosome template (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

plate that has higher resolution than one that can be achieved with sub-tomogram averaging. In the case of template-matching, the situation becomes worse due to molecular crowding effects, which in combination with the CTF, affects the density of the search pattern. Thus, cross-validation is necessary to quantify the outcome of the template-matching.

We developed a procedure for cross-validating the hits of template-matching and found that KerDenSOM3D can perform well for the classification with additional masks. This method of decoy search can be also be utilized with other classification methods. At the core of the approach, is the validation of a structure based on additional density appearing in the area of classification. We find that cross-correlation based methods cannot be used for template-matching and for classification in parallel. If template-matching selects objects according to their similarity, the same similarity measurement cannot also be used for classification. Thus, additional criteria are necessary. Despite the different measurements that were implemented for cross-validating the template-matching results, the methods for recovering the complete ribosome are still coupled. The template bias arising from

template-matching remains contained within each sub-tomogram. The resulting ribosome structure has a prominent white stripe between the 50S subunit and 30S subunit (Fig. 7), which indicates that some false-positives still exist despite the fact that the classification successfully removed a significant number. This is not unusual, and the effect will likely never be completely eliminated, and unfortunately the number of false positives cannot be estimated or even eliminated. Even in single-particle analysis where the signal is stronger and the search area more restricted, a template bias is still present (Shaikh et al., 2003). However, when some of the signal is recovered, it is an indication that the algorithm performs well. This is the first approach for improving the confidence of the detection results, and we combined three criteria: (i) classifying density within a mask that covers an area of the sought macromolecule which was not used for template matching; (ii) the introduction of decoy masks around the true area of density in order to verify the statistical significance of the finding, and (iii) the FSC between the recovered density and the true missing part of the sought macromolecule. Nevertheless, continued algorithmic development is required in order to increase the confidence. Increased computational power will allow for statistical improvement of the results and the class purity.

Acknowledgments

Daniel Castano and Sabine Pruggnaller for the first implementation of the algorithm, Wolfgang Baumeister for the *Spiroplasma melliferum* data, and the European Research Council for the ERC starting grant to AF.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2011.02.009.

References

- Al-Amoudi, A., Diez, D.C., Betts, M.J., Frangakis, A.S., 2007. The molecular architecture of cadherins in native epidermal desmosomes. *Nature* 450, 832–837.
- Bartesaghi, A., Sprechmann, P., Liu, J., Randall, G., Sapiro, G., Subramaniam, S., 2008. Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* 162, 436–450.
- Beck, M., Forster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W., Medalia, O., 2004. Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* 306, 1387–1390.
- Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., Baumeister, W., 2009. The native 3D organization of bacterial polysomes. *Cell* 136, 261–271.
- Briggs, J.A., Riches, J.D., Glass, B., Bartonova, V., Zanetti, G., Krausslich, H.G., 2009. Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. USA* 106, 11090–11095.
- Ditzel, L., Lowe, J., Stock, D., Stetter, K.O., Huber, H., Huber, R., Steinbacher, S., 1998. Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* 93, 125–138.
- Foerster, F., Pruggnaller, S., Seybert, A., Frangakis, A.S., 2008. Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* 161, 276–286.
- Frangakis, A.S., Bohm, J., Forster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R., Baumeister, W., 2002. Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc. Natl. Acad. Sci. USA* 99, 14153–14158.
- Henderson, G.P., Jensen, G.J., 2006. Three-dimensional structure of *Mycoplasma pneumoniae*'s attachment organelle and a model for its role in gliding motility. *Mol. Microbiol.* 60, 376–385.
- Kuerner, J., Frangakis, A.S., Baumeister, W., 2005. Cryo-Electron Tomography Reveals the Cytoskeletal Structure of *Spiroplasma melliferum*. *Science* 307, 436–438.
- Leis, A., Rockel, B., Andrees, L., Baumeister, W., 2009. Visualizing cells at the nanoscale. *Trends Biochem. Sci.* 34, 60–70.
- Li, Z., Jensen, G.J., 2009. Electron cryotomography: a new view into microbial ultrastructure. *Curr. Opin. Microbiol.* 12, 333–340.
- Marabini, R., Carazo, J.M., 1994. Pattern recognition and classification of images of biological macromolecules using artificial neural networks. *Biophys. J.* 66, 1804–1814.
- Milne, J.L., Subramaniam, S., 2009. Cryo-electron tomography of bacteria: progress, challenges and future prospects. *Nat. Rev. Microbiol.* 7, 666–675.
- Nickell, S., Forster, F., Linaroudis, A., Net, W.D., Beck, F., Hegerl, R., Baumeister, W., Plitzko, J.M., 2005. TOM software toolbox: acquisition and analysis for electron tomography. *J. Struct. Biol.* 149, 227–234.
- Nitsch, M., Walz, J., Typke, D., Klumpp, M., Essen, L.O., Baumeister, W., 1998. Group II chaperonin in an open conformation examined by electron tomography. *Nat. Struct. Biol.* 5, 855–857.
- Ortiz, J.O., Brandt, F., Matias, V.R., Sennels, L., Rappsilber, J., Scheres, S.H., Eibauer, M., Hartl, F.U., Baumeister, W., 2010. Structure of hibernating ribosomes studied by cryoelectron tomography in vitro and in situ. *J. Cell Biol.* 190, 613–621.
- Pascual-Montano, A., Donate, L.E., Valle, M., Barcena, M., Pascual-Marqui, R.D., Carazo, J.M., 2001. A novel neural network technique for analysis and classification of EM s-particle images. *J. Struct. Biol.* 133, 233–245.
- Pascual-Montano, A., Taylor, K.A., Winkler, H., Pascual-Marqui, R.D., Carazo, J.M., 2002. Quantitative self-organizing maps for clustering electron tomograms. *J. Struct. Biol.* 138, 114–122.
- Pruggnaller, S., Mayr, M., Frangakis, A.S., 2008. A visualization and segmentation toolbox for electron microscopy. *J. Struct. Biol.* 164, 161–165.
- Scheres, S.H., Melero, R., Valle, M., Carazo, J.M., 2009a. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Structure* 17, 1563–1572.
- Scheres, S.H., Valle, M., Grob, P., Nogales, E., Carazo, J.M., 2009b. Maximum likelihood refinement of electron microscopy data with normalization errors. *J. Struct. Biol.* 166, 234–240.
- Seybert, A., Herrmann, R., Frangakis, A.S., 2006. Structural analysis of *Mycoplasma pneumoniae* by cryo-electron tomography. *J. Struct. Biol.* 156, 42–54.
- Shaikh, T.R., Hegerl, R., Frank, J., 2003. An approach to examining model dependence in EM reconstructions using cross-validation. *J. Struct. Biol.* 142, 301–310.
- Sigworth, F.J., 1998. A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* 122, 328–339.
- Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., Pascual-Montano, A., 2004. XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* 148, 194–204.
- Stolken, M., Beck, F., Haller, T., Hegerl, R., Gutsche, I., Carazo, J.M., Baumeister, W., Scheres, S.H., Nickell, S., 2010. Maximum likelihood based classification of electron tomographic data. *J. Struct. Biol.* 173, 77–85.
- Winkler, H., Zhu, P., Liu, J., Ye, F., Roux, K.H., Taylor, K.A., 2009. Tomographic subvolume alignment and subvolume classification applied to myosin V and SIV envelope spikes. *J. Struct. Biol.* 165, 64–77.
- Zheng, Q.S., Braunfeld, M.B., Sedat, J.W., Agard, D.A., 2004. An improved strategy for automated electron microscopic tomography. *J. Struct. Biol.* 147, 91–101.

A.2. PAPER II

Yu Z., Frangakis AS. “M-free: Scoring the reference bias in sub-tomogram averaging and template matching”, *J Struct Biol.* 2014; Jul;187(1):10-19.

Main conclusion: Cryo-electron tomography provides a snapshot of the cellular proteome. With template matching, the spatial positions of various macromolecular complexes within their native cellular context can be detected. However, the growing awareness of the reference bias introduced by the cross-correlation based approaches, and more importantly the lack of a reliable confidence measurement in the selection of these macromolecular complexes, has restricted the use of these applications. Here we propose a heuristic, in which the reference bias is measured in real space in an analogous way to the R-free value in x-ray crystallography. We measure the reference bias within the mask used to outline the area of the template, and do not modify the template itself. The heuristic works by splitting the mask into a *working* and a *testing area* in a volume ratio of 9:1. While the *working area* is used during the calculation of the cross-correlation function, the information from both areas is explored to calculate the M-free score. We show using artificial data, that the M-free score gives a reliable measure for the reference bias. The heuristic can be applied in template matching and in sub-tomogram averaging. We further test the applicability of the heuristic in tomograms of purified macromolecules, and tomograms of whole *Mycoplasma* cells.

My contributions: Design of research (together with ASF), method development (together with ASF), and analysis of results, data processing and figure preparation, writing of the manuscript (together with ASF).



M-free: Scoring the reference bias in sub-tomogram averaging and template matching



Zhou Yu, Achilleas S. Frangakis*

Goethe University Frankfurt, Buchmann Institute for Molecular Life Sciences and Institute for Biophysics, Max-von-Laue Str.15, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:

Received 19 March 2014

Received in revised form 23 April 2014

Accepted 15 May 2014

Available online 22 May 2014

Keywords:

Template matching

Sub-tomogram averaging

Assessment of reference bias

Reference bias

Template bias

ABSTRACT

Cryo-electron tomography provides a snapshot of the cellular proteome. With template matching, the spatial positions of various macromolecular complexes within their native cellular context can be detected. However, the growing awareness of the reference bias introduced by the cross-correlation based approaches, and more importantly the lack of a reliable confidence measurement in the selection of these macromolecular complexes, has restricted the use of these applications. Here we propose a heuristic, in which the reference bias is measured in real space in an analogous way to the R-free value in X-ray crystallography. We measure the reference bias within the mask used to outline the area of the template, and do not modify the template itself. The heuristic works by splitting the mask into a *working* and a *testing area* in a volume ratio of 9:1. While the *working area* is used during the calculation of the cross-correlation function, the information from both areas is explored to calculate the M-free score. We show using artificial data, that the M-free score gives a reliable measure for the reference bias. The heuristic can be applied in template matching and in sub-tomogram averaging. We further test the applicability of the heuristic in tomograms of purified macromolecules, and tomograms of whole Mycoplasma cells.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Cryo-electron tomography is currently the only imaging technique that can visualize large macromolecular assemblies in an unperturbed cellular environment. The tomograms are three-dimensional (3D) images with a resolution in the range of a few nanometers. In contrast to fluorescence-based microscopy that displays only a few fluorescently labeled macromolecules together, electron tomograms visualize the entire proteome of a cell and contain the information of spatial relationships of the involved macromolecules (Nickell et al., 2006). The most promising way of extracting this information from the tomograms is by utilizing template-based approaches, which use known macromolecular structures and scan the tomogram for their spatial positions (Baumeister, 2005; Frangakis et al., 2002; Lucic et al., 2013). In particular, with the current availability of direct-detector cameras and exponentially growing computational resources, template-based approaches can become more applicable. The primary prerequisite for applying template-based approaches is a figure of merit that assesses the quality of the search, limits the reference bias, and reliably estimates the amount of false positives.

There is a growing awareness that cross-correlation based approaches can introduce a significant amount of reference bias that leads to uncertain or wrong interpretations. This is because high cross-correlation values may be assigned to features that are not the sought macromolecular structure or that the search has delivered the correct macromolecule but in the incorrect orientation. Thus, confidence that the highest cross-correlation values truly represent the properly aligned macromolecular complex is low. When the signal is sufficiently strong, correlation-based sub-tomogram averaging of individual molecules has provided a multitude of high-resolution structures, sometimes with a resolution of ~ 1 nm (Schur et al., 2013). While sub-tomogram averaging is less prone to reference bias than template matching, the bias still plays a significant role. Approaches such as the gold-standard method conceptually address the bias introduced by overfitting in high frequencies (Penczek, 2002; Scheres and Chen, 2012). However, in the case of low-resolution structures such as those in electron tomography, if the bias is already introduced at low frequencies, the outcome is already compromised.

Previous studies have shown that reference bias is also present in single particle approaches, and they have addressed ways and means to minimize it (Scheres and Chen, 2012; Shaikh et al., 2003). Fortunately, in single particle analysis there are numerous benefits that may act favorably towards a reduced reference bias

* Corresponding author.

E-mail address: achilleas.frangakis@biophysik.org (A.S. Frangakis).

in comparison to tomography. Firstly, the higher SNR and the higher resolution achieved by this method reduce the possibility of errors. Secondly, the reduced number of degrees of freedom (DOF) during the alignment search is advantageous. Thirdly, since the particles analyzed are generally purified, the heterogeneity is less compared to sub-tomogram averaging and template matching, which are usually applied to entire cells. Thus, in combination with well-developed classification techniques such as RELION (Scheres, 2012a,b) and FREALIGN (Lyumkis et al., 2013), the number of false positives and wrongly oriented particles is much less or can be reduced during the alignment process.

Reference bias was previously addressed in single particle analysis by utilizing the Fourier Shell Correlation (FSC) (Shaikh et al., 2003). However, the FSC and subsequent approaches such as the gold standard method (Penczek, 2002; Scheres and Chen, 2012), measure the self-consistency of two halves of the data set and do not assess the amount of reference bias. While data can be removed or substituted by noise in order to assess the amount of bias during the refinement process and during flexible fitting (Chen et al., 2013; Falkner and Schroder, 2013), these methods operate in Fourier space and are difficult to transfer to sub-tomogram averaging (Briggs, 2013), and even more so to template matching. For template matching and sub-tomogram averaging, approaches that operate in real space have been suggested that investigate the signal recovery in an area excluded from the search and the classification process (Yu and Frangakis, 2011). These approaches were shown to increase the specificity of the template matching when the signal is strong (Hrabe et al., 2012).

The reason for reference bias is complex. Firstly, template matching will, by definition, indicate positions marking putative objects of interest such as macromolecular complexes. Whether this recognition truly corresponds to a macromolecular complex with the correct translation and rotation depends primarily on the signal-to-noise ratio (SNR), which cannot be measured. Furthermore, it depends on normalization of the cross-correlation function, which is affected by the choice of the mask and the molecular crowding. Finally, the missing wedge additionally plays an important role, as characteristic features of a macromolecule can be missing, making the identification in one particular direction especially complicated. Thus, in summary, reference bias can be introduced by (i) noise alone, (ii) misaligned particles of the macromolecule of interest, (iii) falsely identified features in the tomograms.

Here we develop a heuristic in real space, named the “M-free score”, which allows an estimation of the amount of reference bias, without the knowledge of the genuine underlying signal. The name “M-free” originates from “Mask-free”, indicating that a particular region of the mask is excluded during the template matching or sub-tomogram averaging process, and is only used to estimate the amount of reference bias. In contrast to previous approaches, which are based on measuring the recovery of frequencies in Fourier space, we investigate how the signal behaves and recovers in a defined area of the mask in real space, which contains information from all the frequencies in Fourier space. In this regard, the mask is split into a *working area* and a *testing area*. Comparable to the R-free value in X-ray crystallography, the *testing area* is not “seen” during the search or alignment process correspondingly (Brunger, 1992). In our heuristic, the behavior of the signal in the *testing area* is compared to the signal in *working area* after the alignment. The main argument for using a real space heuristic is that the medium and high frequencies, which are typically used for the Fourier based approaches (Chen et al., 2013; Falkner and Schroder, 2013; Shaikh et al., 2003), are almost absent in the raw data used for template matching and sub-tomogram averaging. Further, the contrast of the signal in the *testing* and *working areas* provides a good visual impression that helps users to validate the results.

We will first derive the mathematics on which the M-free heuristic is based. Next we will apply the heuristic on artificial data sets, and compare the results with the cross-correlation coefficient. We will show that the M-free score estimates the amount of reference bias, independently of the underlying signal, the SNR and the search range. To provide more realistic recording conditions, we use a tomographic data set of purified GroEL particles. We will show how the heuristic behaves when the tomogram is searched with an incorrect template. Lastly, we will show using the example of the ribosome in cellular data sets, that when the signal is improved, the M-free score is smaller allowing an estimation of the reference bias.

2. Materials and methods

The template matching and sub-tomogram averaging algorithms are implemented in C/C++ for massive parallel processing. The script for M-free score calculation implemented in Matlab (The MathWorks, Inc.) is available upon request. Visualization was performed with the Amira package (Pruggnaller et al., 2008).

2.1. Simulated data set

For the simulated data we used the same procedures and data sets first used in Foerster et al. (2008), which have subsequently also been used in several other studies. The simulated images include all artefacts present in electron tomography, and provide tomograms with a very realistic impression.

2.2. Biological data sets

We used previously published publicly available data sets (cited at each particular experiment), recorded on various microscopes and under various conditions. We use in particular an *in vitro* GroEL data set, which has already been used in many other publications. We used the results of sub-tomogram averaging from the SIV-spike data sets in (Zanetti et al., 2006) for validating the M-free score. Further, we used the Mycoplasma data set from (Seybert et al., 2006) to compare with cellular data sets recorded on modern direct detector cameras.

Mycoplasma cells were grown as described in (Hayflick, 1965). Prior to vitrification by plunge freezing in liquid ethane, 10 nm colloidal gold particles were applied.

Single-axis tilt-series were collected covering an angular range from -66° to $+66^\circ$ with 1.5° angular increment on a Titan Krios microscope operated at 300 kV (FEI, Eindhoven, The Netherlands) cooled to liquid nitrogen temperature and equipped with a Quantum energy filter (Gatan, Pleasanton, CA, USA). Data acquisition was carried out under low-dose conditions using the Digital Micrograph Software (Gatan, Pleasanton, CA, USA). Images were recorded on a 4×4 k pixel K2 Summit direct detector (Gatan, Pleasanton, CA, USA) at a nominal defocus of $-5 \mu\text{m}$. The pixel size at the specimen level is 0.34 nm.

3. Theory

In the core of our heuristic is the separation of the mask applied on the reference into a *working area* (W) and a *testing area* (T), in a similar fashion to (Brunger, 1992). W and T are disjoint and their conjunction is the full mask area, that is:

$$W \cap T = \emptyset \text{ and } W \cup T = \text{ALL}. \quad (1)$$

We consider the average of the sub-tomograms $p = s + n$, with $n = n_{\text{bias}} + n_{\text{ran}}$.

We dissect the signal into three types: the macromolecular signal s , the noise n which is considered to be composed of random noise n_{ran} , and biased noise n_{bias} . Since the signal in the *testing area* never “sees” the reference, the amount of biased noise in the *testing area* $n_{\text{bias},T} = 0$, while $n_{\text{bias},W}$ in the *working area* is what we aim to measure. The reference is considered to be uncorrelated with the random noise, and correlated to the biased noise, in a similar fashion to Chen et al. (2013).

The normalized cross-correlation coefficient between the reference r and the average of the sub-tomograms p within the *working area* can be expressed as:

$$CCC_W(r, p) = \frac{\sum_W(rp)}{\sqrt{\sum_W r^2 \sum_W p^2}} = \frac{\sum_W(rs + m_{\text{bias},W})}{\sqrt{\sum_W r^2 \sum_W p^2}}, \quad (2)$$

with r and p being mean-value free within the area of the mask.

Within the *testing area* the information has not been used for the alignment thus $n_{\text{bias},T} = 0$, and the normalized cross-correlation coefficient can be expressed as:

$$CCC_T(r, p) = \frac{\sum_T rs}{\sqrt{\sum_T r^2 \sum_T p^2}} \quad (3)$$

The ratio of Eqs. (2) and (3) can be written as:

$$\frac{CCC_W(r, p)}{CCC_T(r, p)} = \frac{\sum_W(rs + m_{\text{bias},W})}{\sqrt{\sum_W r^2 \sum_W p^2}} \cdot \frac{\sqrt{\sum_T r^2 \sum_T p^2}}{\sum_T rs} \quad (4)$$

We choose W and T such that the variance of the reference within each region is approximately the same. We also assume that the macromolecular signal s has approximately the same variance in both areas. The variance can be written in the following way:

$\sum_W p^2 = k \sum_T p^2$ and $\sum_W r^2 = k \sum_T r^2$, where k is a constant describing the quotient of the size of the *working area* to the *testing area*.

Thus Eq. (4) can be simplified as: $\frac{CCC_W(r, p)}{CCC_T(r, p)} = \frac{\sum_W(rs + m_{\text{bias},W})}{k \sum_T rs}$, and we can write:

$$\frac{CCC_W(r, p) - CCC_T(r, p)}{CCC_T(r, p)} = \frac{\sum_W m_{\text{bias},W}}{\sum_W rs} \quad (5)$$

with $\sum_W rs = k \sum_T rs$ assuming that a signal s results in the same cross-correlation coefficient between s and r in the *testing* and the *working area*.

Using the vector form of the cross-correlation function between the signals in the *working area*, Eq. (5) could be written as the ratio of the scalar products:

$$\begin{aligned} \frac{CCC_W(r, p) - CCC_T(r, p)}{CCC_T(r, p)} &= \frac{\mathbf{R} \cdot \mathbf{N}_{\text{bias},W}}{\mathbf{R} \cdot \mathbf{S}} \\ &= \frac{\|\mathbf{R}\| \|\mathbf{N}_{\text{bias},W}\| \cos \theta_{\mathbf{R}, \mathbf{N}_{\text{bias},W}}}{\|\mathbf{R}\| \|\mathbf{S}\| \cos \theta_{\mathbf{R}, \mathbf{S}}} \end{aligned} \quad (6)$$

If the reference \mathbf{R} is similar to the signal \mathbf{S} , which is valid for most of the real cases, the magnitude of the angle between the vector \mathbf{R} and \mathbf{S} is similar to the magnitude of the angle between \mathbf{R} and $\mathbf{N}_{\text{bias},W}$, which by definition has to be similar. Thus in our effort to express the M-free score independently from the reference we assume the cosine of the angles $\theta_{\mathbf{R}, \mathbf{N}_{\text{bias},W}}$ and $\theta_{\mathbf{R}, \mathbf{S}}$ to be equal. The right hand side of Eq. (6) can then be written as:

$$\frac{CCC_W(r, p) - CCC_T(r, p)}{CCC_T(r, p)} = \frac{\|\mathbf{N}_{\text{bias},W}\|}{\|\mathbf{S}\|} \quad (7)$$

Thus the term $\frac{CCC_W(r, p) - CCC_T(r, p)}{CCC_T(r, p)}$ that we name **M-free** gives an estimate of the reference bias $\mathbf{N}_{\text{bias},W}$ on a signal \mathbf{S} . The CCC_T in the denominator of Eq. (7) can be negative, showing an anti-correlation between the reference and the average in this area, in which case the magnitude is irrelevant since it only reveals that there is no density in the testing area. Thus we set all negative

values to be equal to zero resulting in an M-free score range of $(0 +\infty)$, with 0 indicating no template bias, and $+\infty$ indicating non-existence of a macromolecular signal. Three assumptions were necessary in order to allow a description of the reference bias independent of the reference: (i) the signal s and the reference r in the *testing* and the *working area* relate in the following way: $\sum_W rs = k \cdot \sum_T rs$; (ii) the variance of the noise in the *working* and in the *testing area* relate in this way: $\sum_W n^2 = k \cdot \sum_T n^2$, assuming that after the alignment the total noise variance remains the same in the *working* and *testing areas*; and (iii) the magnitude of the angle between the reference and the biased noise is similar to the magnitude of the angle between the reference and the signal $\theta_{\mathbf{R}, \mathbf{N}_{\text{bias},W}} = \theta_{\mathbf{R}, \mathbf{S}}$. These assumptions only affect the correctness of the M-free score when the reference bias is very high. In the following section, we will show experimentally that the M-free score gives sensible results, and can be used for estimating the amount of reference bias present in the experiment.

4. Results

In sub-tomogram averaging and template matching, the area around the reference is masked. This is to delineate the boundaries of the template from the surrounding, as well as to define the area in which the variance for the denominator of the cross-correlation function is calculated. The masks can have arbitrary shapes, and are rotated together with the template. Here, we reserve $\sim 10\%$ of the mask as a *testing area* that does not contribute to the cross-correlation function (Fig. 1a). The *testing area* is selected such that the signal variance within the *testing area* is identical to the signal variance in the *working area* and can have an arbitrary shape.

4.1. Artificial data

We first measured the M-free score in sub-tomogram averaging of simulated GroELs with different SNR, random orientations and missing wedge. After sub-tomogram averaging using the original GroEL structure as a reference, the particles with a SNR above 0.003 produce an average that is fully symmetric with a well-defined density in the *testing area* (Fig. 1b–e). The average structures from the particles with a SNR below 0.003 do not have a density in the *testing area* (Fig. 1f, g). When the M-free score is measured for an SNR above 0.003, the M-free score is close to zero, while for lower SNR the M-free value rises quickly (Fig. 1h). This is a fundamentally different behavior to the cross-correlation coefficient between the reference and the sub-tomogram average, which slowly decreases in a fashion that is dependent on the SNR, and the number of sub-tomograms (with more sub-tomograms the decrease would be less). Examining the reason for the increased in the M-free score shows that a number of particles were misaligned in the final average (Fig. 1i).

We next explored the validity of the M-free score against contamination. For this we chose the particles with a SNR of 0.003 which is at the lower limit of SNR necessary for success of the alignment. We contaminated the simulated GroELs with Thermosomes having the same missing wedge and SNR as the GroELs. The M-free score rises linearly depending on the amount of contamination present in the data set, which corresponds to the amount of misaligned sub-tomograms (Fig. 2a).

When template matching is simulated on an artificial data set containing GroEL particles, and an exhaustive search with six DOF is performed, the M-free score is even more informative. While the cross-correlation value between the average and the GroEL reference stays approximately the same, the M-free score shows a completely different characteristic: When a GroEL reference is used to search for GroELs the M-free score is 0.16, but when

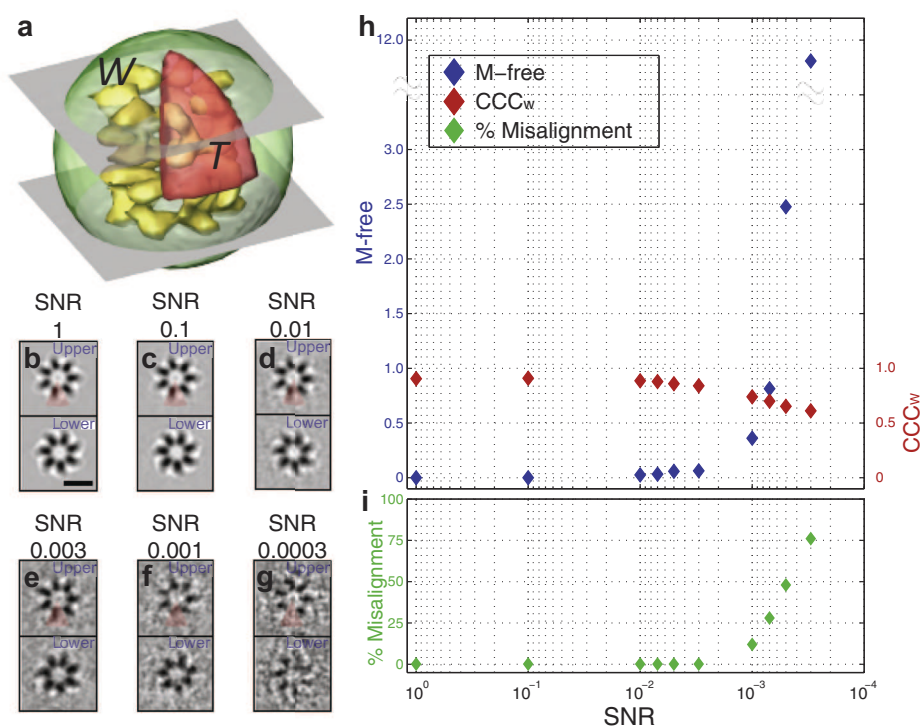


Fig. 1. Results of the sub-tomogram averaging experiments with artificial GroEL sub-tomograms. (a) An illustration of the *working area* *W* in green and the *testing area* *T* in red encircling the isosurface of the GroEL in yellow. The two computational sections through the apical domains intersecting with the *testing* and *working area* respectively are shown in (b)–(g). (b)–(g) The slices indicated in (a) of the averages from data sets with SNR (b) 1, (c) 0.1, (d) 0.01, (e) 0.003, (f) 0.001, and (g) 0.0003. In the upper slices, the *testing area* is indicated with transparent red. The M-free scores measured are 0, 0, 0, 0.05, 0.4, and 12 respectively. The scale bar is 10 nm. (h) M-free scores measured for data sets with different SNRs aligned using GroEL as a starting reference. M-free scores remain at ~0 until the SNR reaches 0.003. When the SNR decreases further, the M-free score rises exponentially. In comparison, the cross-correlation values decrease slowly. More importantly, the cross-correlation value depends on the number of particles, while the M-free score is independent of this criterion. (i) The percentile of misaligned particles as a function of the SNR is shown. The M-free score appropriately reflects the amount of particles for which the orientation and/or translation was not found properly due to the SNR.

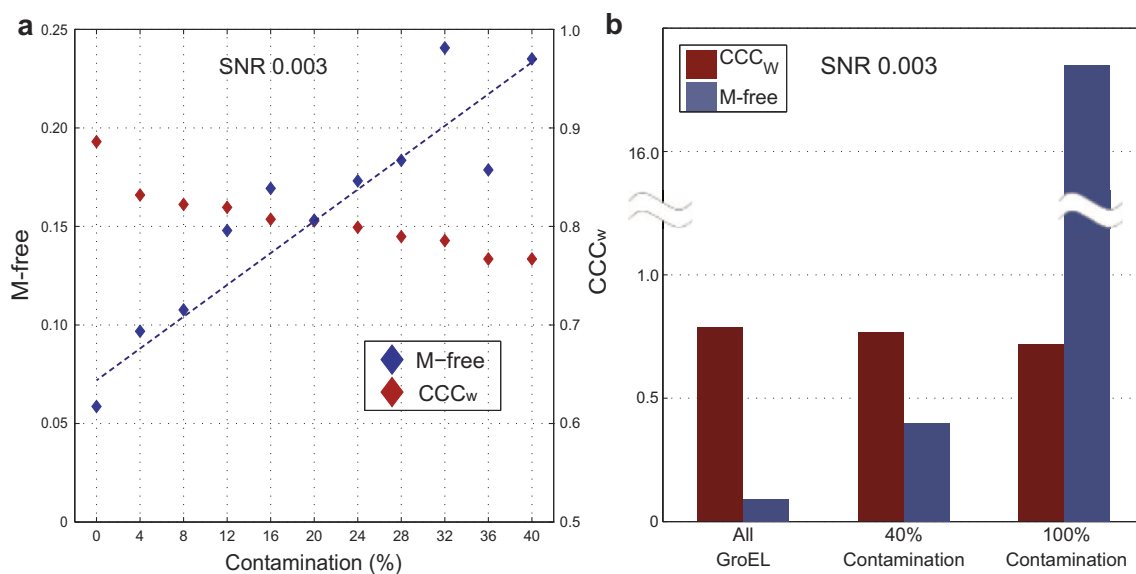


Fig. 2. Results of sub-tomogram averaging and template matching experiments with contaminated artificial GroEL data sets. (a) The M-free scores and cross-correlation values measured from sub-tomogram averaged data sets with an SNR equal to 0.003 and a contamination of 0–40% in 4% steps. (b) Template matching on an artificial data set containing GroELs contaminated with Thermosomes. The bar chart shows M-free scores and cross-correlation values. While the cross-correlation value stays approximately constant the M-free score rises significantly.

the contamination is at $\sim 40\%$ the M-free score is close to 0.7, and when only Thermosomes are present the M-free score is 17. The corresponding cross-correlation coefficients in the above experiments are 0.81, 0.81 and 0.72 (Fig. 2b). Concluding, in defined computational conditions the M-free score gives meaningful results that reflect the quality of the alignment and the amount of contamination.

4.2. In vitro data sets

To measure the information content of the M-free score at realistic experimental conditions we performed template matching on a tomogram of purified GroELs (Fig. 3a). We visually selected ~ 100 sub-tomograms depicting putative GroEL particles and performed sub-tomogram averaging. In a similar manner to the artificial data

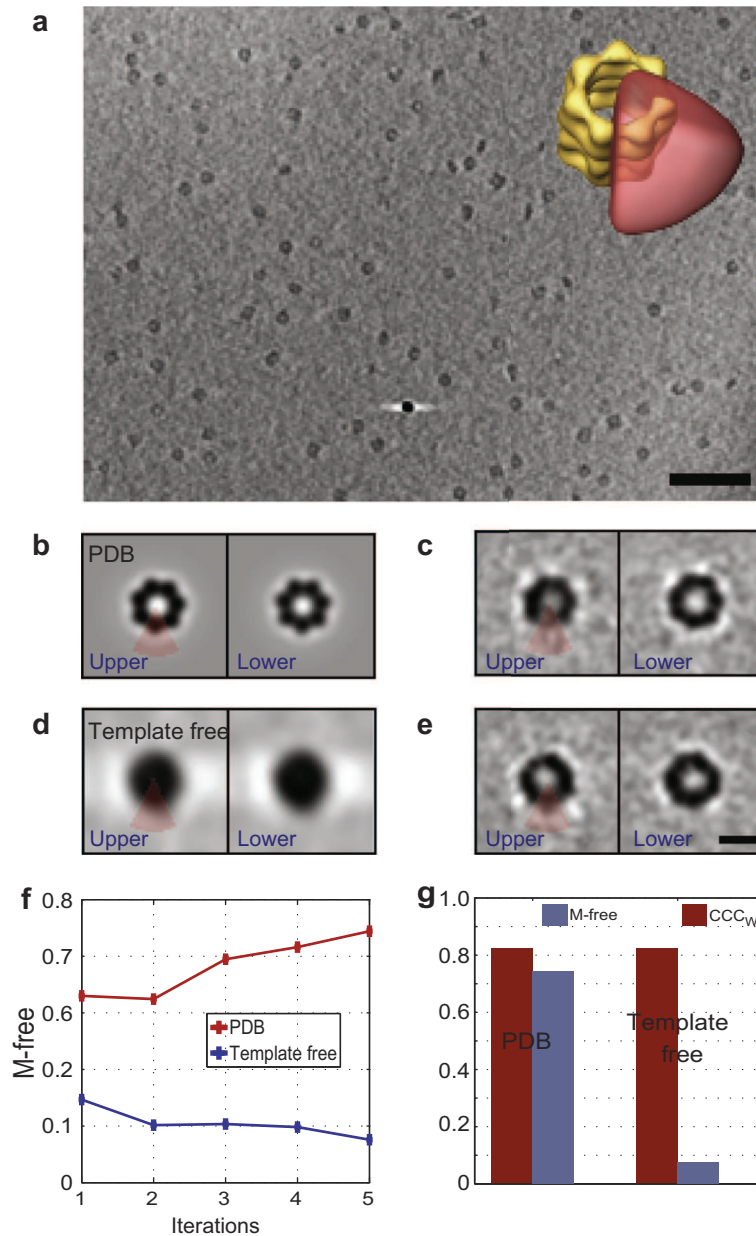


Fig. 3. Results of sub-tomogram averaging with a tomogram depicting purified GroELs. (a) An 18 nm thick projection slice of the tomogram. The scale bar is 100 nm. In the inset, an illustration of the *testing area* is shown in red in relation to the isosurface of the GroEL derived from the crystal structure (PDB: 2C7E) in yellow. (b) Two slices through the upper and lower apical domains of the density map derived from the GroEL crystal structure. The *testing area* is marked with transparent red throughout all images. (c) The same slices through the sub-tomogram average after 5 iterations using the GroEL crystal structure as a starting reference. The M-free score is 0.74. (d) The slices of the Gaussian blob used for “template-free” sub-tomogram averaging. (e) The two slices of the result of “template-free” sub-tomogram averaging after 5 iterations. The M-free score is 0.08. Scale bar is 10 nm. (f) The development of the M-free score depending on the number of iterations. The red curve shows the M-free scores with GroEL crystal structure as a starting reference, while the blue curve shows the “template-free” approach. (g) A bar chart where M-free scores are shown in blue and the cross-correlation coefficients calculated in the *working area* (CCC_w) are shown in red, measured from the averages in (c) and (e). The CCC_w do not differ from each other in the different experiments, however the M-free scores are 10-fold higher in experiments using a crystal-structure-derived template rather than a Gaussian sphere.

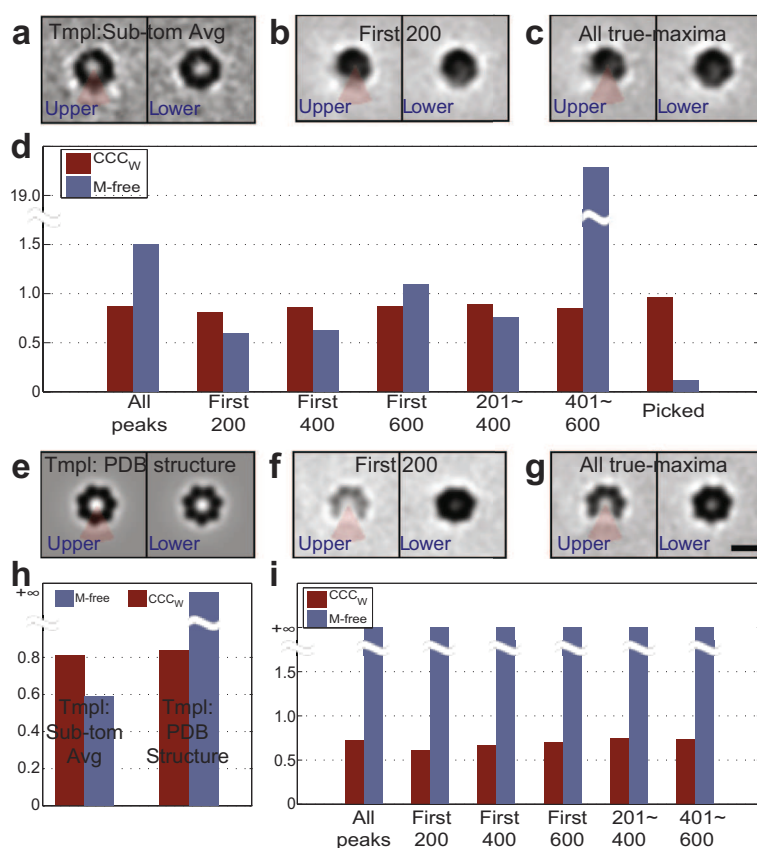


Fig. 4. Result of template matching with a tomogram of purified GroELs. (a) Two slices through the upper and lower apical domains of the sub-tomogram average, which was used as the reference for template matching. (b) The outcome after template matching and averaging the positions corresponding to the highest 200 true maxima. The M-free score was calculated to be 0.59. (c) The same two slices using all true maxima and an M-free score of 1.5. Scale bar is 10 nm. (d) A bar diagram of the M-free scores in blue and the cross-correlation coefficients in red, measured from averages generated from batches of sub-tomograms selected according to the true maxima detected from the template matching with a GroEL template. The M-free scores are similar for the averages generated from the 200 highest true maxima, the 201st–400th true maxima and all the 400 highest true maxima together. The average created from the 401st–600th true maxima has a much higher M-free score. An average generated from the manually selected positions has the lowest M-free score of 0.08, which indicates that when the amount of contamination is low the M-free score is very low. In contrast, the cross-correlation coefficients in W are similarly high for all averages. (e) Two slices through the upper and lower apical domains of filtered PDB structure, that was used as the reference for the second template matching experiment. (f) The density after sub-tomogram averaging of the sub-tomograms corresponding to the highest 200 true maxima. The M-free value is infinite. (g) The density after sub-tomogram averaging all putative particles. The M-free value is infinite again. Scale bar is 10 nm in all slices. In both cases the density in the *testing area* is much more poorly recovered than in (c). (h) The bar chart for M-free and CCC_w for template matching shows that the cross-correlation values are similar to sub-tomogram averaging, while the M-free score efficiently discriminates between the different amounts reference bias. (i) The GroEL tomogram was subjected to template matching using a Thermosome in its open conformation as a template, which does not occur in this tomogram. The cross-correlation values again show similar values for all the averages, and are only slightly lower compared to the experiment where GroEL is used as a search template. The M-free scores however, are all infinitely high indicating that no energy from the macromolecular signature is present at all, thereby unambiguously rejecting the Thermosome as a potential macromolecule in this tomogram.

sets, we used a spherical mask from which a wedge representing 10% of the total volume was removed to provide the *testing area* (Fig. 3a inset). The position of the *testing area* was chosen in such a way so that the variance within the *testing* and *working area* was approximately the same. The sub-tomogram averaging was performed for 5 iterations. Firstly, the GroEL crystal structure from the PDB was used as a reference (Fig. 3b) (Ranson et al., 2001), where the sub-tomogram averages converged to a density showing clear sevenfold symmetry, although with less density in the *testing area* than elsewhere (Fig. 3c). Using a Gaussian blob (Fig. 3d) as the starting reference, the sub-tomogram average converges to a slightly different structure, with less prominent sevenfold symmetry, but with a clear density in the *testing area* (Fig. 3e). For both cases the M-free score was calculated after every iteration (Fig. 3f). In the last iteration, the M-free score for the PDB reference was 0.74 while the M-free score for the Gaussian blob reference was 0.08. The cross-correlation value is 0.82 for both approaches and shows no difference (Fig. 3g). A visual comparison between Fig. 3c and e shows that in the case of the PDB reference the density does not recover in the *testing area*. Furthermore, the M-free

score reflects a common sense observation that the Gaussian blob reference does not introduce a bias, while the GroEL crystal structure does.

While for the sub-tomogram averaging only a selected number of particles are used, in template matching the complete tomogram is scanned for putative particles. For this experiment the complete tomogram was scanned exhaustively with two templates. We first searched with the density obtained from the sub-tomogram averaging using the Gaussian blob as a reference (Fig. 4a). Sorting the values of the local maxima of the cross-correlation function shows a steadily decreasing function for which no hard threshold for the putative positions of the sought macromolecule can be set. Additionally, the number of putative particles is not known. Therefore, we selected all 716 true maxima present in the cross-correlation map and calculated the average of the highest 200 first and of all 716 next (Fig. 4b, c), for which the M-free score was 0.59 and 1.5 respectively (Fig. 4d). When the 201st–400th true maxima are averaged, the M-free score increases slightly to 0.60 and for the average of the 401st–600th true maxima the score rises to 20 (Fig. 4d). The cross-correlation value between the template and

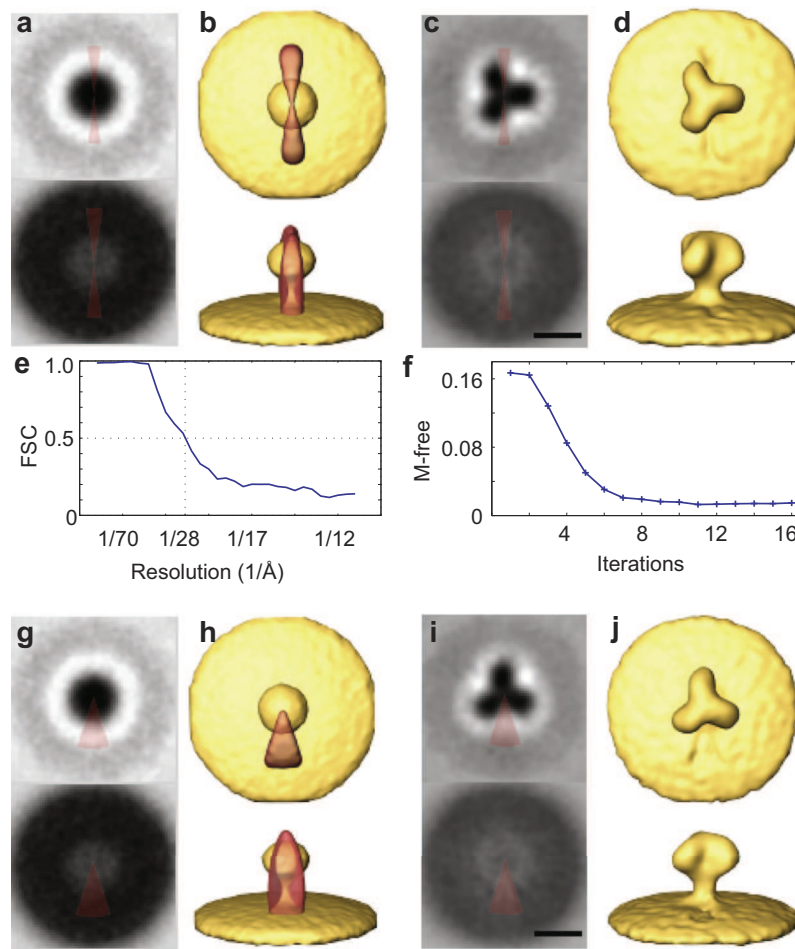


Fig. 5. Results of the sub-tomogram averaging with SIV-spike data sets (Zanetti et al., 2006). (a) 2 xy-slices of the starting template generated using the information from the pre-alignment. The upper slice cuts through the spike and the lower slice cuts through the membrane. The testing area is marked in red. (b) The top view and side view of the isosurface from (a). (c) The same xy-slices as in (a) of the average after 16 iterations. The scale bar is 10 nm. (d) The top view and side view of the isosurface from the result in (c). The average clearly has a threefold symmetry and corresponds to the result as shown in Fig. 2 in Zanetti et al. (2006). (e) The Fourier Shell Correlation (FSC) shows that the average in (c) and (d) has a resolution of 28 Å judged using the 0.5 criterion, which matches the result in Zanetti et al. (2006). (f) The development of the M-free score depending on the number of iterations. After the first iteration the M-free score is relatively high with a value of 0.17, but it decreases linearly with the quality improvement of the average. The M-free score converges to 0.01 after 16 iterations. (g)–(j), the same as (a)–(d), with a different testing area that is shown in (g)–(i). The result is extremely similar to the one above. The scale bar in (i) is 10 nm.

the average increases slowly as more sub-tomograms are selected (Fig. 4d). Interestingly, the Fourier Shell Correlation also shows increasing resolution as more sub-tomograms are added (not shown), despite the fact that more misaligned or non-true particles are included in the average. Calculating the M-free score of the fraction of the true maxima corresponding to the manually selected sub-tomograms results in a score of 0.09 (Fig. 4d), which is similar to the score in the sub-tomogram averaging experiment indicating a similar amount of bias, showing that for classified data the alignment is successful.

Repeating the above experiment using the crystal structure as a template (Fig. 4e) the M-free score is infinite for the first 200 highest true maxima (Fig. 4f) and stays infinite when all positions are selected (Fig. 4g). The cross-correlation again shows almost no difference between the performances of the two templates (Fig. 4h).

To perform the negative experiment, we used a Thermosome template (pixel size and CTF adapted to the recording conditions of the tomogram with the same mask used for GroEL) to scan the tomogram containing only GroELs. Even though the distribution of the cross-correlation values and the selected positions are very similar compared to the ones of the GroEL template, the M-free score is infinite (Fig. 4i). Thus, while the cross-correlation value

hardly discerns between GroEL and Thermosomes, the M-free score shows significant differences.

4.3. Sub-tomogram averaging of SIV spikes (Zanetti et al., 2006)

We next tested the information content on the sub-tomogram averaging of SIV spikes from Zanetti et al. (2006), especially with regard to misalignments, resolution short fallings and artifacts introduced by the shape of the working area. We used an elliptical mask from which a wedge of 10% the volume was carved out as the testing area (Fig. 5a, b). We used several masks and probed the outcome of the sub-tomogram averaging. Apart from the mask alterations, the processing was carried out as described in the original paper: the sub-tomograms were rotationally pre-aligned based on their position on the surface of the virus, which results in a globular structure on top of a membrane patch (Fig. 5b). With progressing iterations a structure resembling the one shown in the paper of Zanetti et al. (2006) emerges (Fig. 5c, d), with a resolution of 2.8 nm, as judged by the 0.5 criterion of the FSC curve (Fig. 5e). The shape of the testing area is not visible in the final structure at all. The M-free score decreases during the process and reaches a value of around 0.01 indicating no template bias (Fig. 5f).

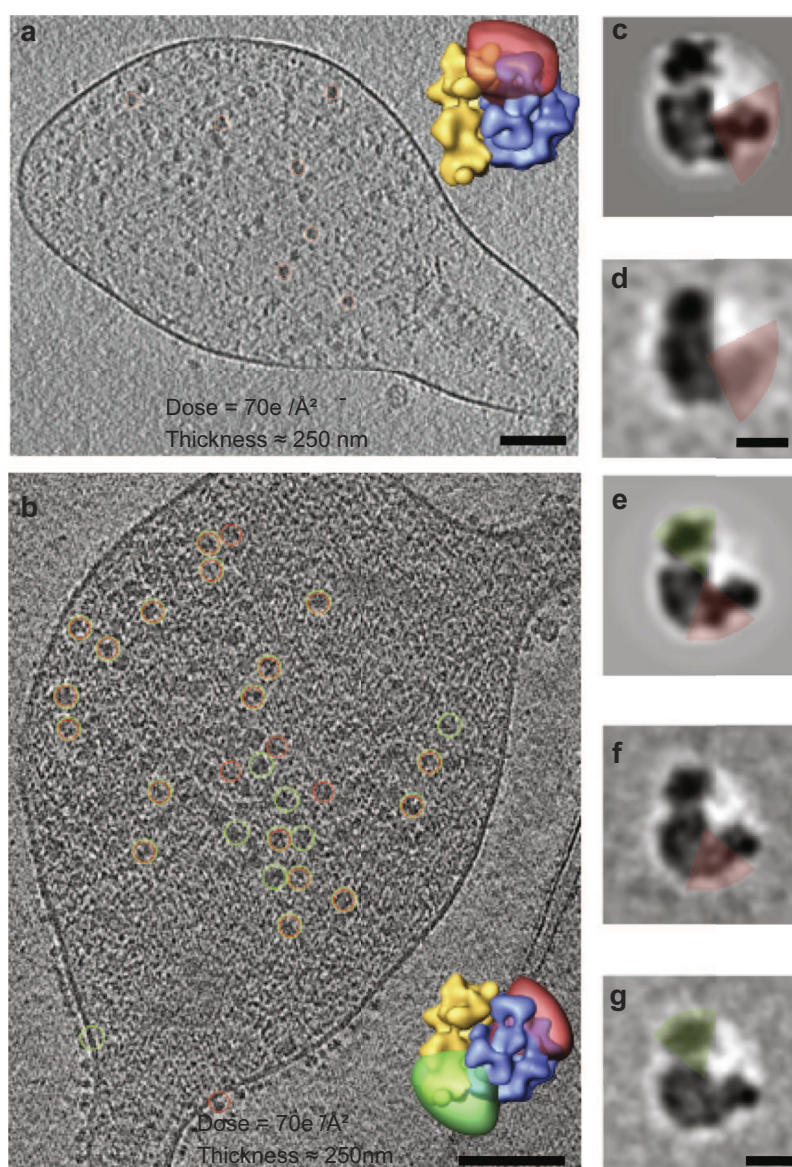


Fig. 6. Results of the template matching with whole cell tomograms of Mycoplasma cells. (a) A 7 nm xy-slice of the tomogram of a ~250 nm thick Mycoplasma cell recorded on the Megascan camera. Inset in the upper right corner shows an isosurface of the ribosome template derived from the crystal structure (2AW7 in yellow and 2AWB in blue) with the testing area in transparent red. The top 150 detected positions that contain most of the ribosomes on this slice are shown in red circles. The scale bar is 100 nm. (b) A 7 nm xy-slice of a ~250 nm thick Mycoplasma cell recorded on the K2 direct detector camera. The X-ray derived template is shown in the upper right corner. Two different testing areas are marked in red and green. The top 150 detected positions of putative ribosomes on this slice are shown in red or green circles, corresponding to the red or green testing areas. The scale bar is 100 nm. (c) The central slice from the template derived from the crystal structure with the testing area shown in red. (d) The central slice from the average of the top 150 detected positions with the testing area shown in red. M-free score is 0.41. The scale bar is 10 nm. (e) The central slice from the template derived from X-ray structure and the testing areas in red or green. (f) The center slice from the average of the detected positions with the red testing area. M-free score is 0.09. (g) Same as (f), with the green testing area. The average shows minimal difference in comparison to (f). The amount of template bias is similar. The M-free score is 0.10. The scale bar is 10 nm.

We repeated the same process with a variety of masks using different regions of the starting reference of the SIV spikes (Fig. 5g, h). As long as the variance within the testing and working area was approximately the same, the resulting sub-tomogram average exhibited a similar resolution and a similar M-free score (Fig. 5i, j).

4.4. Template matching – whole Mycoplasma cells

We then tested the information quality of the M-free score in template matching of cellular tomograms. We thereby scanned two cryo-electron tomograms of Mycoplasma cells with a ribosome template derived from the crystal structure (Schuwirth et al., 2005). We used a spherical mask, from which a wedge

representing 10% of the total volume was used as the testing area. In order to compare the M-free scores we performed the experiment multiple times with the testing area placed in random orientations. The cells were embedded in ice of similar thickness, and recorded at the same conditions with the same cumulative dose ($70 \text{ e}^-/\text{\AA}^2$). The only difference was that the first tomogram was recorded on a Gatan 2002 energy filter with a $2 \times 2 \text{ k}$ pixel Gatan Megascan 795 CCD camera (Short Megascan) (Fig. 6a), while the second was recorded on a Quantum energy filter with a $4 \times 4 \text{ k}$ pixel K2 Summit direct detector (Short K2) (Fig. 6b). The ribosome is by far the biggest complex within these cells, and occurs in high numbers. Indeed, template matching indicates numerous putative particles that are localized throughout the tomogram – sometimes

on the cell-boundaries and in some rare cases even outside the cell (Fig. 6a, b). The positions corresponding to the highest 150 true maxima after running the template matching with the filtered ribosome X-ray structure were selected for the Megascan (Fig. 6c). The M-free score was measured at 0.41 (Fig. 6d). In the slices of the sub-tomogram average the ribosome region under the testing area remains suspiciously electron lucent compared to the rest. For the tomogram recorded on the K2, M-free scores around 0.1 were measured for different *testing areas* (Fig. 6e). When visually examining the averages of the putative ribosome positions, the average created from data recorded on the K2 shows the similar electron density in the *testing area* as in the *working area* – which is not the case for the average created from data recorded on the Megascan (Fig. 6f, g). Comparison of these values with the values from the artificial data sets shows that the quality of ribosome detection on the tomogram recorded with the Megascan is very good, but the result is not as reliable for the tomogram recorded with the K2.

5. Discussion

With our heuristic, we suggest excluding a *testing area* from the mask used in template matching or sub-tomogram averaging in order to use the information in the area to estimate the amount of bias imposed by the reference. We show that sacrificing 10% of the mask area is sufficient to provide a reliable score for assessing the amount of template bias. Sacrificing signal from the sub-tomogram might worsen the alignment. Thus, the alignment could first be performed with a complete mask, and the result can be assessed later (for instance by a difference map) according to the same mask but excluding a *testing area*. Importantly, we show that for sub-tomogram averages that have a strong signal, the influence of the *testing area* is very small.

In our experiments we used arbitrarily oriented wedge shaped masks, which were automatically oriented such that the variance within the *testing* and *working area* is the same. For these cases we did not experience any major influence of the mask orientation or position in the magnitude of the M-free score. In the case of sub-tomogram averaging, the choice of the *testing area* was more complicated, because in some cases the average “moved-away” from the *testing area* and converged in such way, that the *testing area* was essentially empty, rendering it meaningless. For these cases the *testing area* had to be chosen in a way such that the variance equality in the *testing* and *working area* was fulfilled throughout the sub-tomogram averaging.

The M-free score focuses on real space rather than on Fourier space. In single particle approaches, the strong signal in medium–high frequencies allows analysis of the signal recovery in shells in Fourier space (Shaikh et al., 2003). In contrast, the strong signal in electron tomography for sub-tomogram averaging or template matching is only present in a few inner Fourier shells. At higher frequencies the signal can only be recovered after sub-tomogram averaging. Thus ignoring a few shells in the low frequency domains for testing purposes means rejecting a good portion of the information that is most necessary for the alignment. Excising an area in real space still keeps most of the information from the whole frequency domain that is essential for averaging purposes.

The smaller the M-free score, the smaller the amount of bias in the alignment process. In perfectly aligned artificial data sets without contamination, the M-free score is zero independent of the SNR; a property fundamentally different to the cross-correlation coefficient. When either the contamination increases or the SNR decreases below a certain threshold, the M-free score rises, thereby giving an estimate of the reference bias. Our experiments for real

data sets show that an average generated either from template matching or sub-tomogram averaging with an M-free score below 0.1 indicates an acceptably small reference bias. If the M-free score is in the range of 0.1–0.5, the process is still partially reliable, while any value higher than 0.5 makes the outcome questionable. There is always some reference bias in the alignment due to the existence of noise, and a score of zero could never be attained for real data, even for the best possible selected data sets. However, the scores were very close to zero.

Template matching and sub-tomogram averaging are essential approaches to explore the spatial information of macromolecular complexes within their cellular context. The M-free score that we introduce here provides a reliability measurement for those approaches, which makes them more applicable to structural biology.

Acknowledgments

We thank A. Seybert for the Mycoplasma tomogram on the Megascan CCD camera, L. Gonzales for the Mycoplasma tomogram on the K2 direct detector, and J. Briggs for the sub-tomograms of the SIV-spikes. We also thank M. Kunz, D. Castano-Diez, M. Scheffer and C. Wigge for critical reading of the manuscript. We would like to thank Reiner Hegerl for detailed advice on the theoretical section of the manuscript. This work was motivated by CRC 902 and was supported by an ERC starting grant to ASF.

References

- Baumeister, W., 2005. From proteomic inventory to architecture. *FEBS Lett.* 579, 933–937.
- Briggs, J.A., 2013. Structural biology in situ – the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* 23, 261–267.
- Brunger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
- Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H., Henderson, R., 2013. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135, 24–35.
- Falkner, B., Schroder, G.F., 2013. Cross-validation in cryo-EM-based structural modeling. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8930–8935.
- Foerster, F., Pruggnaller, S., Seybert, A., Frangakis, A.S., 2008. Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* 161, 276–286.
- Frangakis, A.S., Bohm, J., Forster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R., Baumeister, W., 2002. Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14153–14158.
- Hayflick, L., 1965. Tissue cultures and mycoplasmas. *Tex. Rep. Biol. Med.* 23 (Suppl. 1), 285+.
- Hrabe, T., Chen, Y., Pfeffer, S., Cuellar, L.K., Mangold, A.V., Forster, F., 2012. PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* 178, 177–188.
- Lucic, V., Rigot, A., Baumeister, W., 2013. Cryo-electron tomography: the challenge of doing structural biology in situ. *J. Cell Biol.* 202, 407–419.
- Lyumkis, D., Brilot, A.F., Theobald, D.L., Grigorieff, N., 2013. Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* 183, 377–388.
- Nickell, S., Kofler, C., Leis, A.P., Baumeister, W., 2006. A visual approach to proteomics. *Nat. Rev. Mol. Cell Biol.* 7, 225–230.
- Penczek, P.A., 2002. Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. *J. Struct. Biol.* 138, 34–46.
- Pruggnaller, S., Mayr, M., Frangakis, A.S., 2008. A visualization and segmentation toolbox for electron microscopy. *J. Struct. Biol.* 164, 161–165.
- Ranson, N.A., Farr, G.W., Roseman, A.M., Gowen, B., Fenton, W.A., Horwich, A.L., Saibil, H.R., 2001. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* 107, 869–879.
- Scheres, S.H., 2012a. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530.
- Scheres, S.H., 2012b. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* 415, 406–418.
- Scheres, S.H., Chen, S., 2012. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9, 853–854.
- Schur, F.K., Hagen, W.J., de Marco, A., Briggs, J.A., 2013. Determination of protein structure at 8.5 Å resolution using cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.* 184, 394–400.

- Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., Cate, J.H.D., 2005. Structures of the bacterial ribosome at 3.5 angstrom resolution. *Science* 310, 827–834.
- Seybert, A., Herrmann, R., Frangakis, A.S., 2006. Structural analysis of *Mycoplasma pneumoniae* by cryo-electron tomography. *J. Struct. Biol.* 156, 342–354.
- Shaikh, T.R., Hegerl, R., Frank, J., 2003. An approach to examining model dependence in EM reconstructions using cross-validation. *J. Struct. Biol.* 142, 301–310.
- Yu, Z., Frangakis, A.S., 2011. Classification of electron sub-tomograms with neural networks and its application to template-matching. *J. Struct. Biol.* 174, 494–504.
- Zanetti, G., Briggs, J.A., Grunewald, K., Sattentau, Q.J., Fuller, S.D., 2006. Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. *PLoS Pathog.* 2, e83.

A.3. PAPER III

Perkovic M., Kunz. M., Endesfelder U., Bunse S., Wigge C., **Yu Z.**, Hodirnau VV., Scheffer M., Seybert A., Malkusch S., Schuman EM., Heilemann M., Frangakis AS. "Correlative Light- and Electron Microscopy with Chemical Tags", *J Struct Biol.* 2014 May;186(2):205-213.

Main conclusion: Correlative microscopy incorporates the specificity of fluorescent protein labeling into high-resolution electron micrographs. Several approaches exist for correlative microscopy, most of which have used the green fluorescent protein (GFP) as the label for light microscopy. Here we use chemical tagging and synthetic fluorophores instead, in order to achieve protein-specific labeling, and to perform multicolor imaging. We show that synthetic fluorophores preserve their post-embedding fluorescence in the presence of uranyl acetate. Post-embedding fluorescence is of such quality that the specimen can be prepared with identical protocols for scanning electron microscopy (SEM) and transmission electron microscopy (TEM); this is particularly valuable when singular or otherwise difficult samples are examined. We show that synthetic fluorophores give bright, well-resolved signals in super-resolution light microscopy, enabling us to superimpose light microscopic images with a precision of up to 25nm in the x-y plane on electron micrographs. To exemplify the preservation quality of our new method, we visualize the molecular arrangement of cadherins in adherens junctions of mouse epithelial cells.

My contributions: Analysis of data (together with MP, MK, UE, CW, VVH, MS, AS), data processing and figure/video preparation (together with ASF, MP, MK, CW, VVH).



Correlative Light- and Electron Microscopy with chemical tags



Mario Perkovic^a, Michael Kunz^a, Ulrike Endesfelder^b, Stefanie Bunse^c, Christoph Wigge^a, Zhou Yu^a, Victor-Valentin Hodirna^a, Margot P. Scheffer^a, Anja Seybert^a, Sebastian Malkusch^b, Erin M. Schuman^c, Mike Heilemann^b, Achilleas S. Frangakis^{a,*}

^aGoethe University Frankfurt, Buchmann Institute for Molecular Life Sciences and Institute for Biophysics, Max-von-Laue Str. 15, 60438 Frankfurt am Main, Germany

^bGoethe University Frankfurt, Institute for Physical and Theoretical Chemistry, Max-von-Laue Str. 13, 60438 Frankfurt am Main, Germany

^cMax-Planck-Institute for Brain Research, Max-von-Laue Str. 4, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:

Received 19 February 2014

Received in revised form 20 March 2014

Accepted 24 March 2014

Available online 31 March 2014

Keywords:

Correlative electron and light microscopy

Electron tomography

Adherens junctions

ABSTRACT

Correlative microscopy incorporates the specificity of fluorescent protein labeling into high-resolution electron micrographs. Several approaches exist for correlative microscopy, most of which have used the green fluorescent protein (GFP) as the label for light microscopy. Here we use chemical tagging and synthetic fluorophores instead, in order to achieve protein-specific labeling, and to perform multi-color imaging. We show that synthetic fluorophores preserve their post-embedding fluorescence in the presence of uranyl acetate. Post-embedding fluorescence is of such quality that the specimen can be prepared with identical protocols for scanning electron microscopy (SEM) and transmission electron microscopy (TEM); this is particularly valuable when singular or otherwise difficult samples are examined. We show that synthetic fluorophores give bright, well-resolved signals in super-resolution light microscopy, enabling us to superimpose light microscopic images with a precision of up to 25 nm in the x - y plane on electron micrographs. To exemplify the preservation quality of our new method we visualize the molecular arrangement of cadherins in adherens junctions of mouse epithelial cells.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Correlative Light- and Electron Microscopy (CLEM) is an approach that complements the information from two microscopic contrasts. Imaging of fluorescently labeled proteins localized by light microscopy (LM) is combined with the visualization of subcellular structures by electron microscopy (EM). To date, super-resolution light microscopy techniques can achieve a protein localization in the range of ~ 20 nm, whereas biological EM of thin-sectioned, resin-embedded samples yields a resolution of ~ 2 – 6 nm. With CLEM, superimposed images allow the protein of interest to be localized in its cellular context within a few tens of nanometers, on specimen preserved at close-to-native conditions. If optimized CLEM can yield images with such an unprecedented richness of information that is expected to surpass traditional immuno-gold labeling techniques in terms of precision, sample preservation, versatility and interpretation capabilities.

CLEM can be applied as a pre- or post-embedding technique, depending on whether the LM is performed before or after the embedding for EM. Technically, when LM is performed prior to

fixation, there is a potentially small time delay (Verkade, 2008), in which the specimen can still change. Furthermore, samples might change during the EM embedding procedure, which might also obstruct CLEM. In general, the correlation of LM with EM data is challenging, since the z -resolution of the LM is much worse than the required thickness for TEM. Thus CLEM is preferably performed directly on sections on the TEM grid. For the imaging of thick samples without section preparation, block-face scanning electron microscopy can be performed (Denk and Horstmann, 2004; Heymann et al., 2006; Knott et al., 2008). This allows the generation of large volume data sets for 3D analysis of the sample in an automated fashion as well as CLEM (Murphy et al., 2011; Narayan et al., 2013).

Traditionally, protein-localization in TEM is performed by immuno-gold labeling. The gold could also be replaced by fluorophores for super-resolution microscopy. Both methods depend strongly on the quality of the available antibodies and accessibility of the epitopes on the surface (Nanguneri et al., 2012). Most closely related to the immuno-based techniques are quantum dots, which are both fluorescent and have a dense metal core that can be visualized in the EM. For increased specificity, protocols employing genetically-encoded tags applicable to photo-oxidation that utilize Tetracycline/ReAsH (Gaietta et al., 2002), GFP (Grabenbauer et al.,

* Corresponding author.

E-mail address: achilleas.frangakis@biophysik.org (A.S. Frangakis).

2005), and miniSOG/FMN (Shu et al., 2011) have been developed, these tags are suitable for photo-oxidation yielding an enrichment of electron-dense osmium tetroxide after illumination (reviewed in Ellisman et al. (2012)).

Electron microscopy of vitrified samples offers conditions closest to the native state (Sartori et al., 2007; Schwartz et al., 2007). However, vitrified samples can only be imaged at liquid nitrogen temperatures, which puts serious constraints on the light microscopy, prohibiting the use of super-resolution techniques as well as the use of high numerical aperture objectives, thus limiting the localization precision to a few hundred nanometers (van Driel et al., 2009). Notably, cryo-fixation techniques are generally considered to be the best methods for preserving biological samples in their native state and high-pressure freezing (HPF) is often implemented as a method in CLEM (Muller-Reichert et al., 2007). Most recently, HPF and freeze substitution (FS) EM-preparation protocols have managed to preserve the post-embedding fluorescence of GFP and other fluorescent proteins (Kukulski et al., 2011; Watanabe et al., 2011). Thereby, excellent protein localization has been achieved either by employing fiducials or by using super-resolution LM techniques. While these post-embedding fluorescence-retaining methods currently represent the state-of-the-art in CLEM, they have a particular disadvantage. These approaches require very low concentrations of heavy metals during sample preparation to avoid quenching of GFP fluorescence. However, heavy metals (uranyl acetate in particular) are important during the embedding procedure for structure preservation in EM. Thus, to date, a compromise between localization precision, fluorescence preservation, and EM quality has had to be made.

Here we use genetically-encoded chemical tags to label different cellular proteins with synthetic fluorophores, in order to circumvent the fluorescence quenching of GFP. We use SNAP (Keppler et al., 2003) and Halo (Los et al., 2008) tags expressed in living cells, that are coupled with the synthetic fluorophores Alexa Fluor 647, tetramethylrhodamine (TMR) and silicon-containing rhodamine derivative SiR-carboxyl (SiR). We show that these synthetic fluorophores conserve their post-embedding fluorescence independently of the uranyl acetate (UA) concentration, thus allowing CLEM for HPF/FS EM-preparation protocols. We demonstrate that our CLEM approach is suitable for simultaneously identifying the localization of up to two intracellular proteins using conventional fluorescence microscopy (FM) correlated with both TEM and SEM, thus opening up a new range of applications. Furthermore, it is easily possible to localize intra- and extracellular proteins of interest, with a few tens of nanometers precision, by combining super-resolution light microscopy and TEM.

2. Results

2.1. Preparation

CLEM allows the functional assignment of proteins prior to their EM analysis. Here we use a pre- and post-embedding CLEM approach. During HPF and FS we varied the concentrations of UA. We found that the post-embedding fluorescence of the synthetic fluorophores was well preserved at all UA concentrations (Fig. 1a–d). This is in contrast to GFP-based fluorescence retaining methods (Kukulski et al., 2011; Watanabe et al., 2011), which either avoid the use of UA, or keep it at an extremely low concentration in order to avoid fluorescence quenching. We find that 2% UA provides the best subcellular structural preservation, which is consistent with a previous report (Hawes et al., 2007). UA concentrations even higher than 2% might be advantageous for SEM imaging, since the higher heavy metal concentration increases the amount of detectable backscattered electrons. We quantified the

quenching of the fluorescence of embedded L-cells stably expressing SNAPf-NCadherin labeled with Alexa Fluor 647. We measured the intensity of several individual cell contours in live-cell confocal imaging and compared the fluorescence signal after embedding using exactly the same optical settings. The specific fluorescence in the embedded samples is decreased to ~25%, and is not decreased further at higher UA concentrations. We find that the auto-fluorescence increases in the green channel only in the area of the embedded cells by ~7-fold (Fig. 1e and Supplementary Fig. 2). Interestingly this was consistent on all investigated specimen, thus we further explored this for automated image superposition of post-embedding LM images and EM micrographs.

2.2. Computational image superposition

The precision of superposition of the FM images on the EM micrographs is important for accurately localizing the labeled proteins within the EM micrographs. Both for embedded single cells and for cell layers the cell boundaries were well resolved by their auto-fluorescence (Fig. 1c, d and Supplementary Fig. 2) providing sufficient signal for positional identification, when compared to the cell membranes visible in the EM micrographs (Supplementary Fig. 1). The affine transformations were necessary to compensate for distortions due to the different orientation of the sample in the optical path as well as sample shrinkage due to electron irradiation.

2.3. SEM experiments

We visualized the embedded L-cells stably expressing SNAPf-NCadherin labeled with Alexa Fluor 647 using confocal fluorescence microscopy (Fig. 2a); the sample was subsequently trimmed to a block face for ion-abrasion SEM. Then, we selected a region containing cells with several cell–cell junctions for ion abrasion SEM (Fig. 2a box). In the SEM, we used backscattered electrons to record an image stack of $33 \times 28 \times 17 \mu\text{m}^3$ with a pixel size of ~8 nm perpendicular to the beam direction, and ~20 nm along the beam direction. The computational slices from the image stack show a good structural preservation with sufficient contrast to recognize different cellular organelles (Fig. 2c). The N-cadherin fluorescence signal (red) overlapped with the positions of the junctions seen in the SEM, and in certain locations with the positions of filopodia on the cell surface, which should contain cadherins (Fig. 2c and d).

2.4. CLEM with various fluorophores

The wide range of genetically encoded chemical tags in combination with an even larger range of synthetic fluorophores offers various options for CLEM. Here we tested several tag-fluorophore combinations. In the first experiment we used HeLa cells stably expressing SNAPf-Histone 2B, which were labeled with TMR to show the feasibility of intra-nuclear labeling (Fig. 3a–d). In the overlay images of labeled Histone 2B (Fig. 3c), we observed that the fluorescence (orange) is excluded from the electron dense nucleolus that is separated from the electron lucent area of the nucleus showing areas of chromatin, in which the Histone 2B fluorescence signal is seen. To further verify the Histone 2B localization we recorded intranuclear tomograms (Fig. 3c box). In the tomographic slices the individual nucleosomes cannot be discerned, however the individual nuclear compartments can be clearly seen, and FM images superimpose nicely (Fig. 3d). In the second experiment we used HeLa cells stably expressing SNAPf-Cytochrome c oxidase subunit 8A labeled with TMR (Fig. 3e–h) to demonstrate cytoplasmic labeling. We find that the fluorescence of mitochondrial protein Cytochrome c oxidase (orange) is localized inside

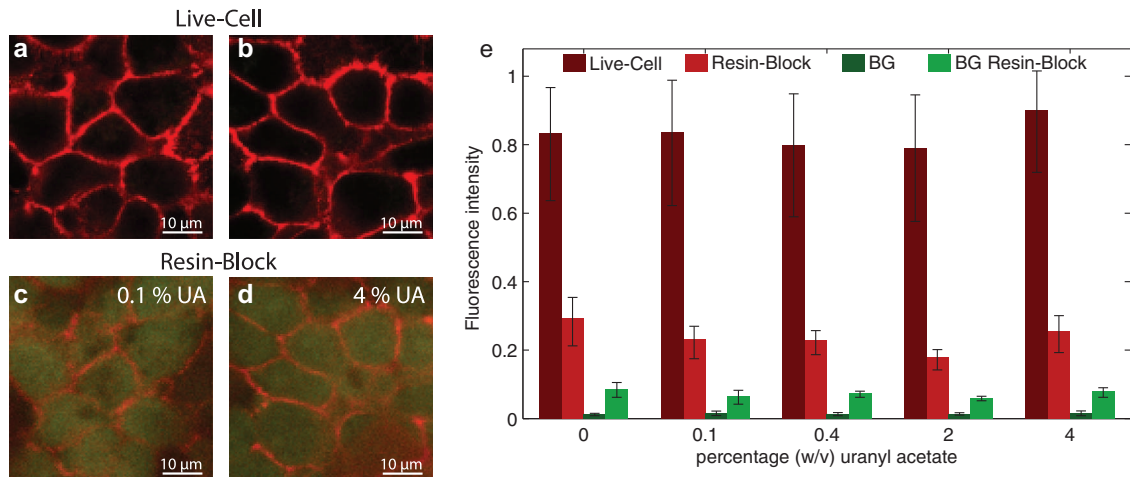


Fig. 1. Live-cell imaging and post-embedding fluorescence at different UA levels. (a, b) Live-cell imaging of L-cells stably expressing SNAPF-NCadherin labeled with Alexa Fluor 647. The cell–cell junctions are visualized in the red channel. The auto-fluorescence is shown in the green channel. (c, d) The corresponding fluorescence within the resin-block after high-pressure freezing and freeze-substitution with 0.1% and 4% UA. (e) Fluorescence intensity measured along the cell junctions in live-cell imaging (dark red), postembedding in the resin-block (light red), auto-fluorescence measured in the green channel in live-cell imaging (dark green), and the auto-fluorescence postembedding in the resin block (light green). At least three regions per uranyl acetate concentration were evaluated with 200 sample points each. Error bars represent the standard deviation of the intensity along the junction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

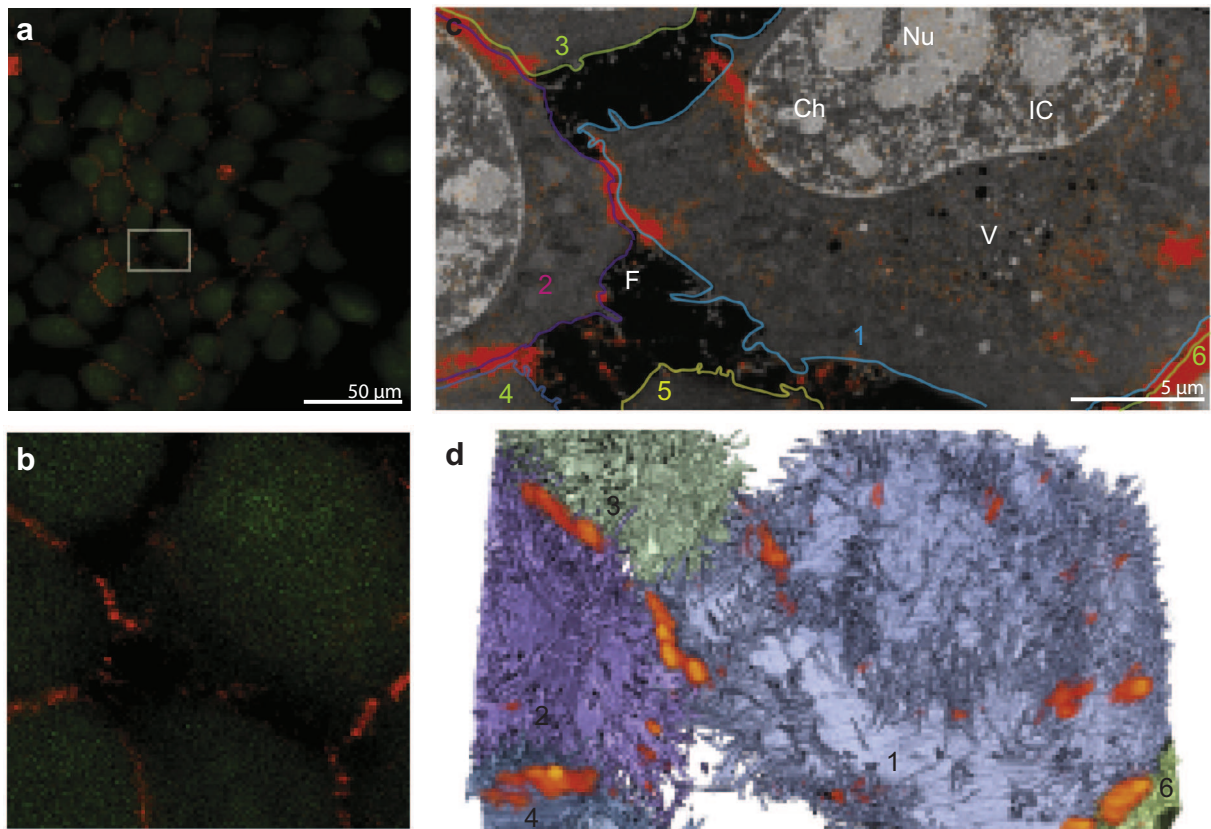


Fig. 2. ia-SEM CLEM of NCadherin in L-cells. (a) Fluorescence image of L-cells stably expressing SNAPF-NCadherin labeled with Alexa Fluor 647 after high-pressure freezing and freeze substitution with 2% UA. The box outlines the $32 \mu\text{m} \times 17 \mu\text{m}$ area enlarged in (b) and shows the interface of six neighboring cells. (c) Overlay of the confocal fluorescence image of SNAPF-NCadherin from (b) on a tomographic slice from the ia-SEM image stack showing the six neighboring cells outlined with different colors and numbered. The specific fluorescence of Alexa Fluor 647 shown in red is localized mainly at cell–cell contacts, and occasionally on the cell surface at densely clustered filopodia. The contrast in the SEM images allows the identification of sub-nuclear compartments: Nucleolus (Nu; medium grey), chromatin (Ch; light grey) and inter-chromatin compartment (IC; dark grey) as well as diverse cytoplasmic vesicles (V) and filopodia (F). (d) Volume representation of the same SEM dataset ($32 \mu\text{m} \times 17 \mu\text{m} \times 10 \mu\text{m}$), overlaid with the confocal fluorescence signal. The fluorescence is visualized by volume rendering in an orange heat color. The different cells are represented as transparent isosurfaces in corresponding colors to (c); blue (cell 1), purple (cell 2), grey-green (cell 3 and 6), and grey-blue (cell 4) with opaque nuclei shown in cells 1 and 2, corresponding to the contours in (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

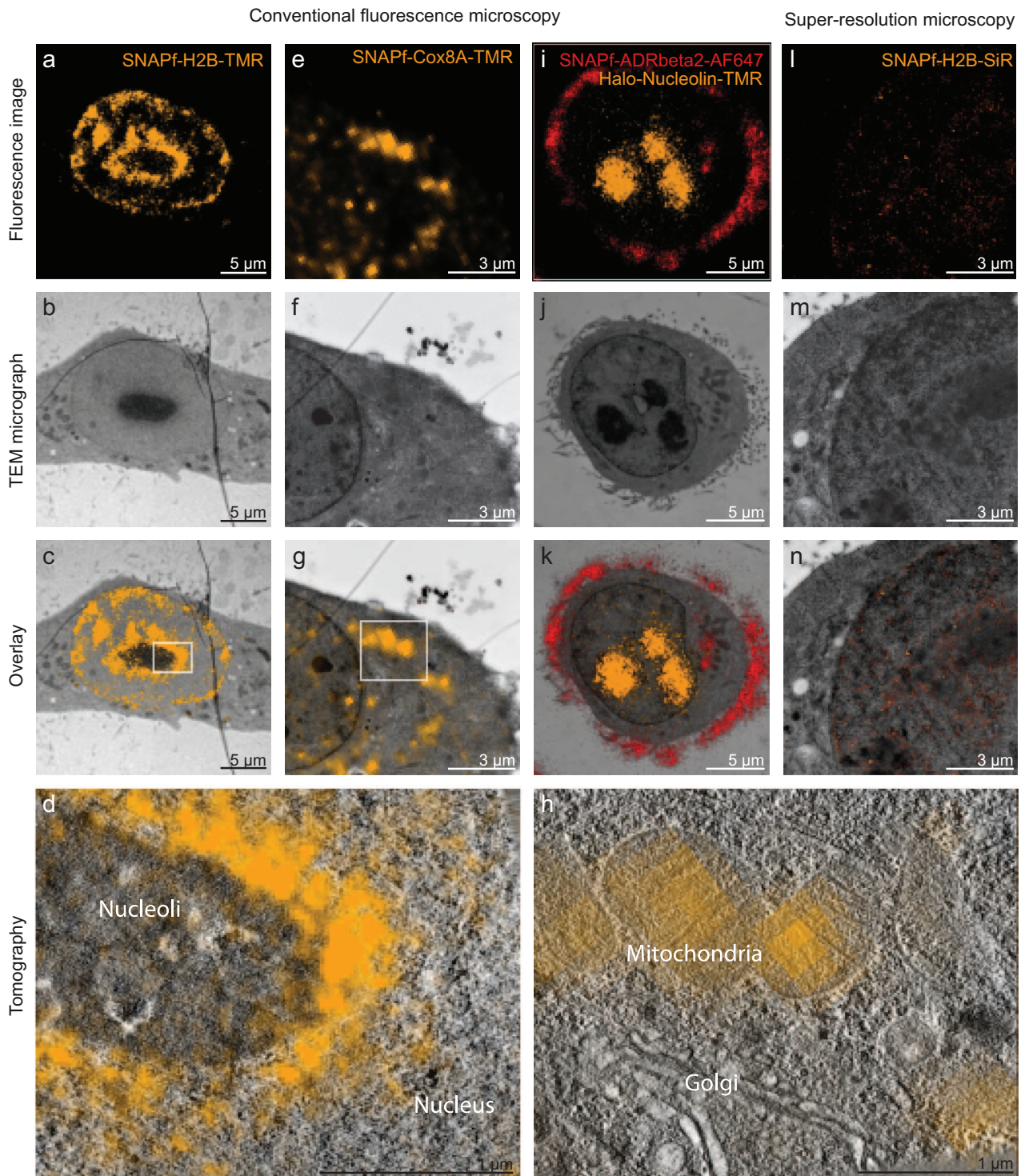


Fig. 3. CLEM of various intracellular and intranuclear proteins. (a) FM image of SNAPf-H2B-TMR. (e) FM image of SNAPf-Cox8A-TMR. (i) Dual color FM of SNAPf-ADRBeta2-AF647 (red) and Halo-Nucleolin-TMR (orange). (l) Super-resolution microscopy of SNAPf-H2B-SiR (orange). (b, f, j, m) TEM micrographs after FM imaging. (c, g, k, n) overlay of the FM images with the corresponding TEM micrographs. (c) FM signal overlaying with the nucleus sparing the darker stained nucleolus. (g) FM signal overlaying with intracellular mitochondria. (k) FM signal overlaying with cellular filipodia (red) as well as with nucleoli (orange). (n) Super-resolution signal overlaying with the nucleus. (d) tomogram acquired on the boxed area in (c). FM signal (orange) is sparing the dark nucleolus. (h) tomogram acquired on the boxed area in (g). FM signal (orange) is localized on the mitochondria. Mitochondrial cristae as well as cytoskeletal elements and parts of the Golgi apparatus are visible. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the mitochondria in the overlay image (Fig. 3g), and some spurious fluorescence puncta (4 out of 14) appearing within the nucleus, which were not created by the embedding process, and can be attributed to the excess of staining used to demonstrate the

process. The EM structural preservation is evidenced in the tomographic slices of this data. Overall the HeLa cells show high contrast membranes, individual cytoskeletal elements like microtubules can be seen as well as the Golgi apparatus (Fig. 3h). In

the third experiment we used HeLa cells stably expressing Halo-Nucleolin and SNAPf-Adrenergic receptor beta 2, which were labeled with TMR and Alexa Fluor 647 to demonstrate the possibilities of dual labeling (Fig. 3i–k). The dual labeling shows an accurate localization of Adrenergic receptor beta 2 (red) on the plasma membrane of the abundant filopodia and Nucleolin (orange) on the three electron dense intranuclear areas depicting the nucleoli (Fig. 3k). In the fourth experiment we used HeLa cells stably expressing SNAPf-Histone 2B labeled with SiR, to show the interchangeability of the different fluorophores as well as the possibility of super-resolution microscopy on the sections prepared for TEM (Fig. 3l–n). The super-resolution images show the specific overlay of Histone 2B (orange) with the nucleus (Fig. 3n). While SNAPf-Histone 2B could also be labeled with TMR enabling conventional LM, super-resolution imaging was best with SiR. We find that among the different synthetic fluorophores tested, SiR exhibited ideal photoswitching properties under the specific buffer, pH and laser intensities.

2.5. Precision of the protein localization in the EM tomogram

The precision of the protein localization in the EM images depends on the precision of the localization in the super-resolution images and the precision of the superposition of the LM on the EM images. Generally, conventional fluorescence microscopy indicates a continuous N-cadherin fluorescence signal across the length of intercellular boundaries, which implies that the cadherin arrangement forming cell-to-cell junctions is equally distributed across the intercellular space. FM imaging of the 300 nm TEM sections already revealed discrete clusters of cadherin molecules in the intercellular space (Fig. 4a). Those were further investigated by correlative super-resolution LM and electron tomography (ET) in order to estimate the precision with which the fluorescently labeled proteins can be localized on the EM images. Cell-cell junctions are well suited for this purpose since the cell membranes are clearly visible in EM. We estimated the precision of the protein localization of SNAPf-Ncadherin (~ 17 nm see Fig. 4a lower insert,

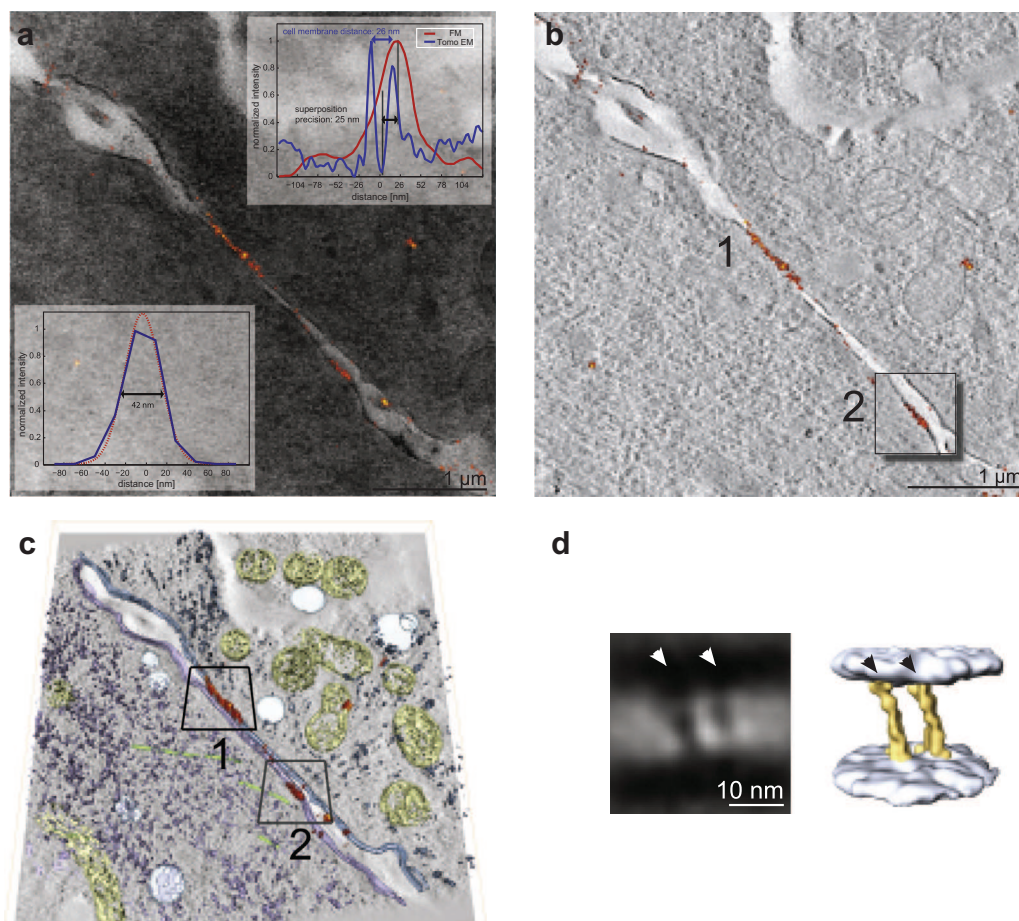


Fig. 4. Localization-based microscopy combined with ET, and sub-tomogram averaging of NCadherin expressing L-cells. (a) Superimposed images of the super-resolution signal of SNAPf-Ncadherin (orange) and the corresponding TEM micrograph. Lower insert is showing the precision of the protein localization with super-resolution microscopy. The blue line depicts the measured profile across the cellular junction in the super-resolution image, whereas the red line is the Gaussian fit to it. The full width at half maximum of the Gaussian is 42 nm, which is in good agreement with calculated precision of the localization precision of the super-resolution microscopy with sigma equal to 17 nm ($2.35 \times \text{sigma}$). Upper insert is showing the precision of proteins localization within the tomograms, on the example of the cadherin molecules. The blue plot shows the normalized intensity profile across the junction measured on the EM images, which shows a membrane spacing of ~ 26 nm (blue double arrow). The red plot shows the normalized intensity profile across the super-resolution signal. The shift between the red and the blue curve indicates the precision of protein localization within the EM tomogram. The shift (black double arrow) is measured between the peak of the super-resolution signal (red) and the center of the neighboring cell membranes (blue). The precision of the protein localization within the EM tomogram was estimated at ~ 25 nm. (b) Overlay of a 2 nm thick tomographic slice with the super-resolution signal (orange). The first fluorescent cluster is localized between two cell membranes with a spacing of ~ 26 nm. The second fluorescent cluster can be associated with the cell membrane of the left cell ~ 80 nm higher than the first and shows a gathering of cadherins, without a connection to the juxtaposed cell (Box 2). (c) Volume representation of intracellular organelles recognizable in the tomogram; mitochondria (yellow), vesicles (light blue), microtubules (green), cytoplasmic compounds (violet). (d) Sub-tomogram averaging of the first fluorescence cluster reveals two cell membranes at a distance of ~ 26 nm together with two densities spanning the distance between the membranes (white arrowheads). Volume representation of the membrane (light blue) and the obliquely oriented densities (yellow) with a thickness of 3 nm (black arrowheads). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

see Section 4: Fluorescence Microscopy), by measuring the average deviation of the fluorescence intensity peak from the center of the two neighboring cell membranes on a 300 nm TEM section. Based on the evaluation of the line-plot data from the super-resolution signal (red) and the cell membranes from TEM (blue), the precision of the superposition of the super-resolution image on the EM tomogram was estimated to be ~25 nm in the direction perpendicular to the cell membranes (Fig. 4a upper insert).

2.6. Preservation of adherens junctions

In order to evaluate the cellular EM preservation as well as the preservation of fine molecular assemblies, we selected a region on the section with two distinct N-Cadherin clusters (Fig. 4a), and performed ET (He et al., 2003). We see in the overlay with the tomogram that the first longer N-cadherin cluster (orange) (Fig. 4b1) corresponds to a region with a regular spacing of ~26 nm between the two juxtaposed membranes, which is consistent with the spacing of N-cadherin estimated from crystal structures (Harrison et al., 2011). The second N-cadherin cluster (Fig. 4b2) seen in the overlay image can be associated with a membrane region ~80 nm in z-height above the previous cluster with no apparent membrane on the juxtaposed position, indicating that cadherins might first cluster within one membrane (“cis-dimer”), and then find their juxtaposed neighbor. Cellular structural preservation is demonstrated by visualization of mitochondria (yellow) microtubules (green) cytoplasmic components (blue, purple) as well as vesicles (light blue) in the tomogram (Fig. 4c). For the analysis of the preservation of fine molecular assemblies we performed sub-tomogram averaging on the first N-cadherin cluster. We find two cell-membranes spaced ~26 nm apart, and obliquely oriented densities with a periodicity of 7 nm (Fig. 4d), which resemble the X-ray model of cadherins (Boggon et al., 2002; Harrison et al., 2011).

3. Discussion

The functional assignment of ultrastructural data with fluorescence microscopy has the capacity to transform our interpretations of biological structure and function. However, the functional assignment is challenging because the harsh EM preparation methods quench the fluorescence. During the EM preparation, the proteins are denatured due to the acidic environment of the embedding material (Watanabe et al., 2011), which has direct implications for the GFP based methods, such as the quenching of the GFP fluorescence, and an increase in the overall protein background fluorescent signal. Our experiments show that synthetic fluorophores do not degrade as severely, and thus the fluorescence signal withstands the increasing background signal for a broad range of EM preparation conditions. This higher stability allows the choice of the most appropriate EM preparation protocols, and, as demonstrated here, the heavy metal concentration can be increased to such an extent that the same sample can be used for TEM and SEM. Even more importantly, the higher stability increases the reproducibility of the experiments.

Synthetic fluorophores offer a much wider palette than fluorescent proteins and cover a broader range of the wavelength spectrum usable for FM. Here we have used two different dyes, which diffuse through the cell membrane without destroying it, and without compromising the EM quality. Other synthetic fluorophores that are not membrane permeable when conjugated with substrates for chemical tagging could be used intracellularly (ATTO655 as HaloTag substrate is permeable as opposed to the SNAP substrate) (Wilmes et al., 2012). For the deposition of specific synthetic fluorophores in the cytoplasm techniques like microinjection might be a valuable tool to further exploit the entire range

of synthetic fluorophores for CLEM. For the post-embedding super-resolution microscopy we tested a number of synthetic fluorophores. SiR and Alexa Fluor 647 have revealed high suitability for super-resolution imaging due to high photon yield per switching event and a large number of switching cycles.

For the thickness of the TEM section, the super-resolution images provide exclusively 2-dimensional information. Thus, in the 3D tomographic images, a z-localization without additional constraints – as in the case of the adherens junctions – is not possible. However, in any case, the 3D localization precision is better than in conventional immuno-gold labeling where epitopes must be available on the surface of the section in order to be detected. For an even better z-localization, correlative array tomography could be employed in which, a series of ultrathin sections (40–100 nm) is examined by CLEM (Wacker and Schroeder, 2013).

Generally, evaluation of the quality of preservation of the morphology in embedded samples is challenging, since, apart from specific caveats due to freezing damage, the quality cannot be assessed by objective criteria. Nevertheless, sample preservation remains the most crucial requirement in any CLEM approach, since it has direct impact on the biological interpretation. By exemplifying that the macromolecular organization of adherens junctions can be preserved to a large extent, it can be demonstrated that the molecular organization is retained. Fortunately, the broad range of preparation capabilities for EM using synthetic fluorophores allows for dedicated sample preparation. This is especially valuable for rare specimens, which cannot be processed multiple times and thus opens up new areas of biological exploration.

4. Materials and methods

4.1. Molecular cloning

The coding sequence of SNAP_f was amplified by PCR (sense primer: 5'-GCGGCGCCATGGACAAAGACTGCGAAATGAAGCG-3', antisense primer: 5'-GCGGCGGCCACCCAGCCAGGCTTGCCAG-3') using the pSNAP_f-vector (New England BioLabs) as a template and introducing *Ascl* sites at both ends. The amplified region was inserted via *Ascl* into the coding region of the extracellular domain of N-cadherin at a site (aa position 151 of the mature protein) previously reported to tolerate insertion of GFP (Kim et al., 2011).

4.2. Tissue culture, transfections and generation of stable cell lines

L cells (mouse fibroblasts, ATCC # CRL-2648) were grown in complete medium (DMEM GlutaMAX supplemented with 10% FCS and 1 mM sodium pyruvate (Life Technologies, Darmstadt, Germany)). Transfection was performed with the pNcad-SNAP_f using X-tremeGENE9 (Roche, Mannheim, Germany) and CombiMag beads (OZ Bioscience, Marseille, France). Cells were trypsinized 24 h post transfection, transferred to a 6 well plate and cultivated in selection medium (complete medium + 600 µg/ml Geneticin (Life Technologies)). For generation of monoclonal cell lines, cells were trypsinized, diluted to a concentration of 2 cells/ml and plated in 96 well plates (100 µl/well). After 11 days, cells were further expanded in 24 well plates. Expression of Ncad-SNAP_f was analyzed by immunoblotting with an anti-N-cadherin antibody (mouse, BD Biosciences). Cell clones, which were positive in Western blot, were stained with SNAP-Surface Alexa Fluor 647 (New England BioLabs) and analyzed by confocal laser scanning microscopy using a LSM780 (Plan-Apochromat 40×/1.4 oil objective, Carl Zeiss, Jena, Germany). Cell lines displaying high-levels of Ncad-SNAP_f at the plasma membrane were chosen for further experiments.

HeLa cells were grown in complete medium (DMEM, 10% FBS, 0.29 mg/ml L-glutamine, and 100 U/ml penicillin/streptomycin). For transfections X-tremeGENE HP (Roche Applied Science) and plasmids pSNAPf-H2B, pSNAPf-ADRBeta2, pSNAPf-Cox8A (New England BioLabs) and pFN21AB5231 (HaloTag-Nucleolin; Promega, Mannheim, Germany/Kazusa DNA Research Institute) were used. To generate stable cell lines, cells transfected with pSNAPf-H2B and pSNAPf-ADRBeta2 were transferred from 24 well plates into 6 cm dishes 24 h post transfection and cultivated in selection medium (complete medium +400 µg/ml Geneticin) for 3 weeks. For dual color imaging, HeLa cells stably expressing SNAPf-ADRBeta2 were transfected with pFN21AB5231.

4.3. Cellular labeling

L cells stably expressing Ncad-SNAP_f were grown to a confluency of 60–80%, stained in 2 µM SNAP-Surface Alexa Fluor 647 solution (dilution in phenolred free complete medium) at 37 °C for 5 min and subsequently washed three times with phenolred free complete medium.

For staining of HeLa cells, 3 µM SNAP-Cell TMR Star, 2 µM SNAP-Surface Alexa Fluor 647 (New England BioLabs), 3 µM SiR-SNAP (Lukinavicius et al., 2013) and 50 µM HaloTag TMR Ligand (Promega) diluted in OptiMEM (Life Technologies) with 10% FBS were used. After 30 min of staining, samples were rinsed three times and washed for at least 30 min with OptiMEM with 10% FBS.

4.4. Cryo fixation and freeze substitution

Sapphire discs with a diameter of 3 mm and thickness of 0.05 mm (Engineering Office M. Wohlwend GmbH, Sennwald, Switzerland) coated with carbon were used as support for the cells. The deposition of carbon was performed on a Cressington 208 carbon coater (Cressington Scientific Instruments) in two steps. In order to divide the working area into sub-areas identifiable by light microscope, the carbon was first evaporated over the sapphire disc covered with a finder grid. After removal of the finder grid a second layer of carbon was deposited and the coated sapphire discs were then baked at 120 °C overnight. Prior to seeding of the cells, discs were attached to the bottom of 35 mm cell culture dishes with a small drop of BD Matrigel™ (BD Biosciences, Heidelberg, Germany).

Cryo-fixation of cells grown on carbon coated sapphire discs and labeled as described above was performed by high-pressure freezing using an HPM 010 (ABRA Fluid AG) by placing the sapphire discs between two aluminum planchettes so that the cells were protected in a 50-µm cavity in one planchette. As a cryo protectant, the OptiMEM with 10% FBS was used. For freeze substitution, a temperature-controlling device (Leica EM AFS2; Leica Microsystems) was used. High pressure frozen samples were transferred in liquid nitrogen onto the substitution medium (2% (w/v) uranyl acetate, 2% water, 10% methanol, 88% acetone) precooled to –140 °C and the temperature was raised by 20 °C/h to –90 °C. For the titration uranyl acetate concentration of 0%, 0.1%, 0.4%, 2% and 4% were used. At –90 °C the substitution medium is liquid, thus the discs were separated from planchettes, kept at –90 °C for 1 h, warmed by 20 °C/h to –45 °C, transferred into glass distilled acetone and washed 3 times for 10 min each. Afterwards samples were infiltrated with increasing concentrations (10%, 25%, 50%, and 75%) of Lowicryl HM20 in acetone while warmed at 1.2 °C/h to –25 °C. Three exchanges with pure Lowicryl HM20 were done in 10 h steps. For resin polymerization the specimens were exposed to UV light at –25 °C for 48 h, warmed at 5 °C/h to 20 °C and further UV polymerized for 48 h at 20 °C.

Sections with a nominal thickness of 300 nm were cut with a microtome (Leica EM UC7) using a diamond knife (Diatome, Biel,

Switzerland) and placed on formvar coated nickel (diameter 3.05 mm, parallel bars, 50 bars per inch) or copper (diameter 3.05 mm, slot 2 × 0.5 mm) TEM grids (Plano GmbH, Wetzlar, Germany). Some sections were post stained with 2% (w/v) uranyl acetate (in 70% methanol) and Reynolds lead citrate.

4.5. Fluorescence microscopy

Laser scanning fluorescence microscopy was performed on an upright confocal microscope (Carl Zeiss, LSM700) with a W Plan-Apochromat 63×/1.0 water objective or Plan-Apochromat 63×/1.4 oil objective and Immersol 518F (Carl Zeiss, Jena, Germany) using laser lines 488 nm, 555 nm and 639 nm. For confocal microscopy of resin embedded cells, the resin block was fixed onto a custom-made device that allows for translational and rotational movement and the pinhole was set to 1 airy unit. Prior to imaging of sections, TEM grids were dipped into glycerol and placed onto glass slides and covered with a cover slip (thickness 170 µm; Carl Zeiss, Jena, Germany). For imaging of sections with the LSM700, the pinhole was completely open.

Direct Stochastic Optical Reconstruction Microscopy (dSTORM) was performed on an inverted microscope (IX-71; Olympus, Hamburg, Germany or a Eclipse Ti; Nikon, Japan) using an oil-immersion objective (PlanApo 60×, NA 1.45, Olympus or a CFI Apo TIRF 100×, NA 1.49, Nikon) (Heilemann et al., 2008). Nickel TEM grids were placed into a LabTek II chamber (Nunc, Langensfeld, Germany) such that the surface of the section was on top and accessible to imaging buffer (0.1 M bicarbonate-carbonate, 100 mM, 0.5 mg/ml glucose oxidase, 40 µg/ml catalase, 10% w/v glucose, pH 7.5–8). The 647 nm and 488 nm laser beams of an argon-krypton laser (Innova 70C; Coherent, Santa Clara, USA) were selected by an acousto-optic tunable filter (AOTF) and an additional clean-up filter (Z488/568/647 RPC, AHF Analytentechnik) and were used in continuous mode for readout and activation. The laser beam was coupled into the microscope objective by a dichroic filter (HC-Dual 560/659, AHF Analytentechnik, Tuebingen, Germany). Fluorescence light was spectrally filtered with two filters (700/75 BP and 647 LP RazorEdge, AHF Analytentechnik) and detected with an EMCCD camera (Andor Ixon DV897). Typical laser powers used for dSTORM imaging were 0.5–5 mW (488 nm) and 30–80 mW (647 nm), ensuring a sufficiently low fraction of single activated fluorophores at any given time. 3000 Frames with a pixel size of 108 nm or 5000 frames with a pixel size of 158 nm at a frame rate of 20 Hz were recorded. Fluorescent spots identified in each image frame were analyzed with the rapidSTORM software (Wolter et al., 2012) (parameters: photon count threshold 400, reconstructed pixel size 10 nm). Experimental precision of the protein localization in the super-resolution images after drift correction was estimated at 17 nm for N-Cadherin and 16 nm for H2B (Endesfelder et al., 2014). This is in good agreement to the full width at half maximum measured across the junction (Fig. 4a lower insert).

4.6. Transmission electron microscopy

15-nm Colloidal gold particles (Aurion, The Netherlands) were adsorbed on both sides of the grids (and section) to serve as fiducial markers for the alignment of the tilt-series. Grids were placed into a Model 2040 high-tilt holder (Fischione Instruments, Pittsburgh, USA) or Single Tilt Holder for Ultra Twin (#FP 6596/05, FEI, Eindhoven, The Netherlands) and imaged using a 300 kV Tecnai F30 transmission electron microscope (FEI, Eindhoven, The Netherlands). Digital micrographs and tilt-series were recorded on a US4000 CCD camera (Gatan, USA) with a 1° increment over various tilt ranges at a pixel size of 0.9–1.32 nm. For the reconstruction of tomograms, the IMOD software package (version 4.

3.1; (Kremer et al., 1996)) was used. Sub-tomogram averaging was performed as previously described (Al-Amoudi et al., 2007).

4.7. Ion-abrasion scanning electron microscopy (IA-SEM)

Lowicryl embedded cell monolayers were loaded into a scanning electron microscope (Helios 600i, FEI company, Eindhoven, The Netherlands). The front surface of the block was aligned parallel to the scanning direction of the electron beam. Subsequently, the stage was brought to eucentric height and tilted to 52°. After the initial alignment, the area of interest was identified using backscattered electron (BSE) imaging with an In-lens detector. In order to reduce re-deposition and to improve imaging conditions, surplus block material around the area of interest was removed with a high ion beam current (30 keV, 2.5 nA). Afterwards a sacrificial platinum layer (500 nm) was deposited on the top surface of the sample using the gas injection system. As a final preparation step, the front surface was polished with a decreased ion current (30 keV, 2.8 nA). Serial milling and imaging of a stack of $33 \times 28 \times 17 \mu\text{m}^3$ was performed using the Auto Slice and View software with a pixel size of ~ 8 nm perpendicular to the beam direction, and ~ 20 nm along the beam direction (FEI company, Eindhoven, The Netherlands) with 30 keV and 0.7 nA for the milling steps and 2 keV and 1.4 nA for the imaging steps. Overlay of LSM and IA-SEM stacks and manual segmentation was performed using the Amira software package (Amira 3.0; TGS).

4.8. Automated image superposition

For the superposition of light and electron microscopic images, the auto-fluorescence, which occurs in the area of the embedded cells was exploited (Supplementary Fig. 1a). The auto-fluorescence signal provides strong and pronounced contours, in which the EM image can be matched in a template-based approach. This is comparable to template matching in electron microscopy as well as the centroid fit in super-resolution microscopy. For achieving a sub-pixel precise superposition of the LM image on the EM images the matching procedure had to be performed between the LM image and the lowest magnification EM image, and repeated for all EM images recorded in a magnification series of the region of interest. The FM image (99 nm pixel size) and the EM image of lowest magnification (13 nm pixel size) were low-pass filtered followed by Canny edge detection (Supplementary Fig. 1b and c) (Canny, 1986). In a brute force manner, affine transformations consisting of shift, rotations, magnifications, and stretching independently in x and y were globally applied to one image which was then compared to the other matching image, until the values with the highest cross-correlation score were found (Supplementary Fig. 1). The same procedure was applied to the next pair of subsequent images in the magnification series, until the global transformation parameters between the LM image and EM image with the highest magnification was found. Afterwards the FM image could be overlaid with all the EM images by combining all transformations from the magnification series. For the implementation, custom scripts were written for Matlab (The MathWorks, Inc., Natick, USA).

Contributions

A.S.F., E.S., M.H. designed the experiments. M.P., U.E., S.B., C.W., V.V.H., S.M. performed experiments. M.P., M.K., Z.Y., U.E., C.W., V.V.H., M.S., A.S. analyzed data. A.S.F., M.P., C.W., M.H., E.S., M.S., A.S. wrote the paper.

Acknowledgments

We are grateful to E.H.K. Stelzer, W. Kuehlbrandt and M. Vabulas for critical reading and discussion of the manuscript. M.H., S.M. and U.E. acknowledge funding by the BMBF (research Grant 0315262) and the CRC 902. This work was also supported by the CRC 902, ERC, and CEFII to A.S.F.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jsb.2014.03.018>.

References

- Al-Amoudi, A., Diez, D.C., Betts, M.J., Frangakis, A.S., 2007. The molecular architecture of cadherins in native epidermal desmosomes. *Nature* 450, 832–837.
- Boggon, T.J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B.M., Shapiro, L., 2002. C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* 296, 1308–1313.
- Canny, J., 1986. A computational approach to edge-detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8, 679–698.
- Denk, W., Horstmann, H., 2004. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* 2, e329.
- Ellisman, M.H., Deerinck, T.J., Shu, X., Sosinsky, G.E., 2012. Picking faces out of a crowd: genetic labels for identification of proteins in correlated light and electron microscopy imaging. *Methods Cell Biol.* 111, 139–155.
- Endesfelder, U., Malkusch, S., Fricke, F., Heilemann, M., 2014. A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment. *Histochem. Cell Biol.* (PMID: 24522395).
- Gaietta, G., Deerinck, T.J., Adams, S.R., Bouwer, J., Tour, O., Laird, D.W., Sosinsky, G.E., Tsien, R.Y., Ellisman, M.H., 2002. Multicolor and electron microscopic imaging of connexin trafficking. *Science* 296, 503–507.
- Grabenbauer, M., Geerts, W.J., Fernandez-Rodriguez, J., Hoenger, A., Koster, A.J., Nilsson, T., 2005. Correlative microscopy and electron tomography of GFP through photooxidation. *Nat. Methods* 2, 857–862.
- Harrison, O.J., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C.M., Troyanovsky, R.B., Ben-Shaul, A., Frank, J., Troyanovsky, S.M., Shapiro, L., Honig, B., 2011. The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. *Structure* 19, 244–256.
- Hawes, P., Netherton, C.L., Mueller, M., Wileman, T., Monaghan, P., 2007. Rapid freeze-substitution preserves membranes in high-pressure frozen tissue culture cells. *J. Microsc.* 226, 182–189.
- He, W., Cowin, P., Stokes, D.L., 2003. Untangling desmosomal knots with electron tomography. *Science* 302, 109–113.
- Heilemann, M., van de Linde, S., Schüttelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P., Sauer, M., 2008. Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angew. Chem. Int. Ed. Engl.* 47, 6172–6176.
- Heymann, J.A., Hayles, M., Gestmann, I., Giannuzzi, L.A., Lich, B., Subramaniam, S., 2006. Site-specific 3D imaging of cells and tissues with a dual beam microscope. *J. Struct. Biol.* 155, 63–73.
- Keppeler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., Johnsson, K., 2003. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* 21, 86–89.
- Kim, S.A., Tai, C.Y., Mok, L.P., Mosser, E.A., Schuman, E.M., 2011. Calcium-dependent dynamics of cadherin interactions at cell–cell junctions. *Proc. Natl. Acad. Sci. USA* 108, 9857–9862.
- Knott, G., Marchman, H., Wall, D., Lich, B., 2008. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *J. Neurosci.* 28, 2959–2964.
- Kremer, J.R., Mastrorade, D.N., McIntosh, J.R., 1996. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* 116, 71–76.
- Kukulski, W., Schorb, M., Welsch, S., Picco, A., Kaksonen, M., Briggs, J.A., 2011. Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *J. Cell Biol.* 192, 111–119.
- Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D.H., Bultel, R.F., Wood, K.V., 2008. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* 3, 373–382.
- Lukinavicius, G., Umezawa, K., Olivier, N., Honigsmann, A., Yang, G., Plass, T., Mueller, V., Reymond, L., Correa Jr., I.R., Luo, Z.G., Schultz, C., Lemke, E.A., Heppenstall, P., Eggeling, C., Manley, S., Johnsson, K., 2013. A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* 5, 132–139.

- Muller-Reichert, T., Srayko, M., Hyman, A., O'Toole, E.T., McDonald, K., 2007. Correlative light and electron microscopy of early *Caenorhabditis elegans* embryos in mitosis. *Methods Cell Biol.* 79, 101–119.
- Murphy, G.E., Narayan, K., Lowekamp, B.C., Hartnell, L.M., Heymann, J.A., Fu, J., Subramaniam, S., 2011. Correlative 3D imaging of whole mammalian cells with light and electron microscopy. *J. Struct. Biol.* 176, 268–278.
- Nanguneri, S., Flottmann, B., Horstmann, H., Heilemann, M., Kuner, T., 2012. Three-dimensional, tomographic super-resolution fluorescence imaging of serially sectioned thick samples. *PLoS One* 7, e38098.
- Narayan, K., Danielson, C.M., Lagarec, K., Lowekamp, B.C., Coffman, P., Laquerre, A., Phaneuf, M.W., Hope, T.J., Subramaniam, S., 2013. Multi-resolution correlative focused ion beam scanning electron microscopy: applications to cell biology. *J. Struct. Biol.* 185 (3), 278–284.
- Sartori, A., Gatz, R., Beck, F., Rigort, A., Baumeister, W., Plitzko, J.M., 2007. Correlative microscopy: bridging the gap between fluorescence light microscopy and cryo-electron tomography. *J. Struct. Biol.* 160, 135–145.
- Schwartz, C.L., Sarbash, V.I., Ataullakhanov, F.I., McIntosh, J.R., Nicastro, D., 2007. Cryo-fluorescence microscopy facilitates correlations between light and cryo-electron microscopy and reduces the rate of photobleaching. *J. Microsc.* 227, 98–109.
- Shu, X., Lev-Ram, V., Deerinck, T.J., Qi, Y., Ramko, E.B., Davidson, M.W., Jin, Y., Ellisman, M.H., Tsien, R.Y., 2011. A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS Biol.* 9, e1001041.
- van Driel, L.F., Valentijn, J.A., Valentijn, K.M., Koning, R.I., Koster, A.J., 2009. Tools for correlative cryo-fluorescence microscopy and cryo-electron tomography applied to whole mitochondria in human endothelial cells. *Eur. J. Cell Biol.* 88, 669–684.
- Verkade, P., 2008. Moving EM: the rapid transfer system as a new tool for correlative light and electron microscopy and high throughput for high-pressure freezing. *J. Microsc.* 230, 317–328.
- Wacker, I., Schroeder, R.R., 2013. Array tomography. *J. Microsc.* 252, 93–99.
- Watanabe, S., Punge, A., Hollopeter, G., Willig, K.I., Hobson, R.J., Davis, M.W., Hell, S.W., Jorgensen, E.M., 2011. Protein localization in electron micrographs using fluorescence nanoscopy. *Nat. Methods* 8, 80–84.
- Wilmes, S., Staufenbiel, M., Lisse, D., Richter, C.P., Beutel, O., Busch, K.B., Hess, S.T., Piehler, J., 2012. Triple-color super-resolution imaging of live cells: resolving submicroscopic receptor organization in the plasma membrane. *Angew. Chem. Int. Ed. Engl.* 51, 4868–4871.
- Wolter, S., Loschberger, A., Holm, T., Aufmkolk, S., Dabauvalle, M.C., van de Linde, S., Sauer, M., 2012. RapidSTORM: accurate, fast open-source software for localization microscopy. *Nat. Methods* 9, 1040–1041.

A.4. PAPER IV

Eltsov M., Dubé N., **Yu Z.**, Haselmann-Weiss U., Brunner D., Frangakis AS. “Large-volume electron tomography reveals the mechanism of epithelial tissue sealing”, submitted manuscript

Main conclusion: The closure of epidermal openings is an essential biological process that causes major developmental problems such as spina bifida in humans if it goes awry. To date, the mechanism of closure remains elusive. To address this, we reconstructed this tissue shaping process by large-volume correlative electron tomography using dorsal closure in fly embryos as a model. The molecular resolution we could obtain provides the first comprehensive and quantitative analysis of the cytoskeletal reorganization that allows separated epidermal cells to make contact and form a sealed epithelium. We find that cells establish contact through actin-driven exploratory filopodia, then generate stable “roof tile” like overlaps of single lamella, which shorten to produce a pulling force for the final tissue sealing. Using correlative live imaging, we can show that although propelled by actin, both filopodia and lamella are oriented towards the tissue opening by microtubules. The spatial distribution of growing and shrinking microtubule ends as well as actin bundles in the tomograms furthermore shows that the sealing force that shortens the lamellar overlaps and zips the tissue close is produced by microtubules and not by actin bundles as previously thought. Our work provides the first 4D picture of a developmental process with molecular resolution and thereby reveals the cytoskeletal mechanism of epidermal closure.

My contributions: Analysis of data (together with EM, ASF), data processing and figure/video preparation (together with EM, ASF).

Acknowledgements

First, I would like to thank my supervisor Prof. Dr. Achilleas Frangakis for giving me the opportunity to carry out my PhD on the exciting projects in this great scientific environment. I am very thankful for his support, for the discussions and suggestions and for his novel ideas on my projects. His optimistic view on my projects motivated me continuously.

I would also like to thank Prof. Dr. Jochen Triesch for his support and supervision.

I greatly appreciate the help from Dr. Friedrich Foester, Dr. John Briggs and Dr. Sjors Scheres who shared their valuable opinions on my projects and as well as their valuable data sets.

I am very grateful to Dimitra Keramisanou, Maria Kokkinoupolou and Sylvain Trepout for guiding me through the darkness and into the field of electron microscopy at the beginning of my PhD work.

I would like to thank Anja Seybert, Christoph Wigge, Margot Scheffer and Michael Kunz for careful reading of my manuscripts and my thesis. Their comments and corrections helped to improve the quality of my writings significantly.

I want to thank all the other members and former members of the group, Anja Habermann, Christian Geiss, Mario Perkovic, Mikhail Eltsov and Valentin Hondrinou, for creating a great working atmosphere, and all the fun we had together.

I enjoyed spending time with my very good friends Dingding Wang, Guo Li, Jin Liu and Qiuzhu Zhang who supported me throughout the my whole PhD time. They have made my time more enjoyable.

Acknowledgements

And finally, I specially and profoundly thank my parents Jianhua Yu and Yanping Li for their love and invaluable support over the years. They are always behind me whenever I need them.

Curriculum vitae

Personal data

| | |
|----------------|----------------------|
| Name | Zhou, Yu |
| Date of birth | 16 November 1983 |
| Place of birth | Beijing, P. R. China |
| Nationality | German |

Education

| | |
|---------------------|---|
| 03. 2010 – present | PhD Study at Goethe University Frankfurt Buchmann Institute for Molecular Life Sciences Cryo-electron Microscopy group Laboratory of Prof. Dr. Frangakis |
| 05. 2009 – 11.2009 | Diploma Thesis at University of Stuttgart and Sony Deutschland GmbH “Automatic intrinsic calibration of a stereo camera system” Prof. Dr. –Ing. B. Yang and K. Zimmermann |
| 10. 2004 – 11. 2009 | Dipl. –Ing. Study at University of Stuttgart |

Curriculum vitae

Faculty of Electrical engineering

Major: Signal processing and system theory

Participated research projects:

- “SuperClass - Classification of personal video recordings”
- “Analysis and classification of data of an automotive FMCW-Radar”

10. 2003 – 10. 2004

German course at University of Stuttgart

09. 2002 – 09. 2003

B. Sc. Study at University of Wuhan, China

Faculty of Electrical engineering

Major: Communication technology

09. 1996 – 08. 2002

The Affiliated High School of Peking University, China