

Max-Planck-Institut für Biochemie
Abteilung für Molekulare Strukturbiologie

Quantitative Analyse von Makromolekülen in
Kryoelektronentomogrammen mittels Korrelationsmethoden

Friedrich Förster

Vollständiger Abdruck der von der Fakultät für Physik
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. M. Kleber
Prüfer der Dissertation: 1. Hon.-Prof. Dr. W. Baumeister
2. Univ.-Prof. Dr. M. Rief

Die Dissertation wurde am 30.03.2005 bei der Technischen Universität München eingereicht
und durch die Fakultät für Physik am 20.06.2005 angenommen.

Inhaltsverzeichnis

Inhaltsverzeichnis	i
Zusammenfassung	v
1 Einleitung	1
2 Tomographie	6
2.1 Projektionstheorem	7
2.2 Dosisaufteilungstheorem	9
2.3 Dreidimensionale Rekonstruktion	12
3 Elektronenmikroskopie	17
3.1 Streuprozesse im Elektronenmikroskop	18
3.1.1 Kinematische- und Bornsche Näherung	19
3.1.2 Elastische Streuung	20
3.1.3 Inelastische Streuung	22
3.1.4 Streuamplitude komplexer Proben	24
3.1.5 Strahlschädigung vitrifierter Proben	25
3.2 Kontrastmechanismen	26
3.2.1 Phasenkontrast	28
3.2.2 Amplitudenkontrast	31
3.2.3 Streuabsorptionskontrast	33
3.3 Energiegefilterte Elektronenmikroskopie	35
4 Material und Methoden	38
4.1 Experimenteller Aufbau	38
4.1.1 Das Transmissionselektronenmikroskop	38
4.1.2 Der Energiefilter	40
4.1.3 Die CCD Kamera	40
4.2 Kryo Präparation	43
4.3 Automatisierte Datenakquisition	44
4.4 Tomographische Rekonstruktion	45

5	Lokalisierung von Makromolekülen	47
5.1	Korrelationsbasierte Algorithmen zur Identifikation makromolekularer Komplexe in Kryoelektronentomogrammen	50
5.1.1	Korrelationsfunktionen	50
5.1.2	Weiterentwickelte Korrelationsfunktionen für die KET	54
5.2	Erzeugung struktureller Signaturen aus atomaren Strukturen	56
5.3	Parallelisierte Implementierung – MOLMATCH und OMNIMATCH	58
5.4	Partikelidentifikation in Phantomzellen	61
5.4.1	Material und Methoden	62
5.4.2	Ergebnisse	64
5.5	Identifikation von Ribosomen im Prokaryoten <i>Spiroplasma melliferum</i>	68
5.5.1	Material und Methoden	69
5.5.2	Ergebnisse	71
5.6	Partikelidentifikation im Eukaryoten <i>Dictyostelium discoideum</i>	74
5.6.1	Material und Methoden	74
5.6.2	Ergebnisse	75
5.7	Diskussion	78
6	Dreidimensionale Mittelung von Makromolekülen	81
6.1	Algorithmus zur Mittelung von Subtomogrammen	83
6.1.1	Alignierung in einem begrenzten Winkelbereich	84
6.1.2	Initialisierung der Orientierungen	88
6.1.3	Alignierung des Polarwinkels φ und polare Symmetrieanalyse	90
6.2	Struktur der Kernporen aus <i>Dictyostelium discoideum</i>	91
6.2.1	Material und Methoden	93
6.2.2	Ergebnisse	95
6.3	Diskussion	99
7	Struktur von MoMuLV Env	101
7.1	Einführung in die Struktur des Env Protein Komplex'	104
7.2	Material und Methoden	106
7.2.1	Herstellung der Viren	106
7.2.2	Präparation und Datenaufzeichnung	106
7.2.3	Bildanalyse und Mittelung	107
7.2.4	Fitting der RBD	107
7.3	Ergebnisse	108
7.3.1	Morphologie der Viruspartikel	108
7.3.2	Struktur des Env Komplexes	109
7.4	Diskussion	114

8	Verallgemeinerte PCA für KET	117
8.1	Constrained Principal Component Analysis (CPCA)	119
8.1.1	Principal Component Analysis	119
8.1.2	Constrained PCA	120
8.2	Material und Methoden	121
8.2.1	CPCA und Klassifizierung	121
8.2.2	Simulationen	122
8.2.3	Experimentelle Daten	124
8.3	Ergebnisse	126
8.3.1	Anwendung auf simulierte Daten	126
8.3.2	Anwendung auf experimentelle Daten	131
8.4	Diskussion	134
9	Diskussion und Ausblick	136
	Abkürzungen	140
	Physikalische Konstanten	142
	Literaturverzeichnis	143
	Danksagung	163

Zusammenfassung

Die Elektronentomographie (ET) ist hervorragend geeignet, um die dreidimensionale (3D) Struktur pleomorpher biologischer Strukturen wie etwa Zellen oder Organellen abzubilden. Die Weiterentwicklungen der Technik in den letzten 10 Jahren machen es heutzutage möglich, in vitrifiziertem Eis eingebettete biologische Objekte dreidimensional abzubilden. Diese Probenpräparation gewährleistet optimale Strukturhaltung und beläßt die biologischen Proben in einer annähernd nativen Umgebung. Demnach liefert Kryo-ET (KET), die ET vitrifizierter biologischer Proben, ein naturgetreues 3D Abbild z.B. ganzer Zellen, und sie ist derzeit die einzige Technik, die einzelne makromolekulare Komplexe in ihrer natürlichen Umgebung aufzulösen vermag.

Das vorrangige Ziel dieser Arbeit war es, Methoden zur quantitativen Auswertung von Kryotomogrammen zu entwickeln und anzuwenden. Die größte Herausforderung bei der Analyse von Kryoelektronentomogrammen ist ihr außerordentlich niedriges Signal-zu-Rauschverhältnis, das in der Strahlenempfindlichkeit biologischer Proben begründet liegt. Ein weiteres Hindernis bei der Interpretation ist ein Keil fehlender Strukturdaten im Fourierraum, der vom eingeschränkten Kippbereich der Probe im Elektronenmikroskop herrührt. Während die erste Tatsache zur Wahl korrelationsbasierter Techniken zur Datenanalyse führte, wurde der zweiten Rahmenbedingung durch die Entwicklung neuer Korrelationsfunktionen, die die Analyse auf den experimentell zugänglichen Bereich einschränken, Rechnung getragen.

Im Detail wurden Methoden zum Einsatz in drei zentralen Aufgabenstellungen der KET entwickelt und angewendet: (i) Ein Ziel der Tomographie ganzer Zellen ist es, einen zellulären Atlas makromolekularer Komplexe zu erhalten. Hierfür wurde ein Algorithmus zur Detektion spezifischer Makromoleküle anhand ihrer strukturellen Signatur entwickelt und zur Lokalisierung von Komplexen in Phantomzellen, pro- und eukaryotischen Zellen angewandt. (ii) KET ist in der Lage, sehr fragile supramolekulare Komplexe abzubilden. Durch Mittelung von Subtomogrammen können diese in höherer Auflösung strukturell charakterisiert werden, wozu ein iteratives Verfahren implementiert wurde. Dieses wurde zur *in situ* Strukturbestimmung eines retroviralen Env Proteinkomplex' und des Kernporenkomplex' bis zu einer Auflösung von $(3\text{nm})^{-1}$ bzw. $(8\text{-}9\text{nm})^{-1}$ benutzt. (iii) Eine systematische Analyse der in einer biologischen Probe enthaltenen makromolekularen Komplexe erfordert eine statistische Analyse, die homogene Subpopulationen identifiziert. Es wurde eine Klassifikationsmethode entwickelt, die Artefakte der üblichen *Principal Component Analysis* bei der Anwendung auf KET-Daten deutlich verringerte.

Abstract

Electron Tomography (ET) is uniquely suited to obtain three-dimensional (3D) reconstructions of pleomorphic structures, such as cells or organelles. The technical advances of the instrumentation within the last 10 years enable us nowadays to image biological objects which are embedded in vitreous ice three-dimensionally. This preparation technique provides the best structural preservation and maintains the biological sample in its natural environment. Thus, Cryo-ET (CET), the ET of vitrified biological specimens, provides a faithful representation of the 3D structure of specimens such as cells, and it is currently the only technique being able to resolve single macromolecular complexes in their native environment.

The major objective of this work was the development and application of methods for a quantitative evaluation of cryoelectron tomograms. The major challenge in the analysis of cryoelectron tomograms is their extremely low signal-to-noise ratio arising from the radiation sensitivity of biological materials. A further hindrance in the interpretation is a missing wedge of structural data in Fourier space which is due to the limited tilt range of the sample in the electron microscope. While the former condition lead to the choice of correlation based data analysis techniques, the latter constraint was considered by the development of novel correlation functions limiting the analysis to the experimentally sampled data range.

In detail, the methods developed and applied in this work address three central tasks of CET: *(i)* It is one goal of tomography of whole cells to derive a cellular atlas of macromolecular complexes. Therefore, an algorithm was developed which detects individual macromolecules based on their structural signature. It was applied in order to localize complexes in phantom cells, pro-, and eukaryotic cells. *(ii)* CET is able to image very fragile supramolecular complexes. These can be characterized structurally by employing averaging of subtomograms. Therefore, an iterative averaging procedure was designed and employed in order to determine the structures of the retroviral Env protein complex and the nuclear pore complex to a resolution of $(3 \text{ nm})^{-1}$ and $(8\text{-}9 \text{ nm})^{-1}$, respectively. *(iii)* A systematic analysis of the macromolecular complexes populating a biological sample requires the statistical analysis of the data in order to identify homogeneous subpopulations. A classification procedure was developed that overcomes the artifacts of the wide-spread Principal Component Analysis when applied to CET data.

1. Einleitung

Die Untersuchung der Struktur biologischer Materialien auf unterschiedlichen Ebenen, die von molekularen Dimensionen bis zu denen ganzer Zellen reichen, hat das Verständnis ihrer Funktion deutlich erleichtert. Zellen besitzen eine äußerst komplexe Architektur, die viele Größenordnungen umspannt, von der Organisation auf molekularer Ebene bis hin zu Netzwerken ganzer Zellen. Das Interesse der modernen Molekularbiologie richtet sich vor allem auf das Verständnis sog. Molekularer Maschinen (*Molecular Machines* (Alberts 1998)). Molekulare Maschinen sind aus Makromolekülen, vor allem Proteinen, zusammengesetzte Komplexe, die einzeln oder im Ensemble spezifische Funktionen in der Zelle übernehmen. Nur die wenigsten Proteine können allein eine spezifische Funktion erfüllen, und nach heutigem Kenntnisstand interagiert jedes Protein durchschnittlich mit etwa 9 anderen Proteinen (Sali et al. 2003). Einige dieser Komplexe wie z.B. das Ribosom, dessen Funktion die Synthese von Proteinen ist, sind robust genug, um mittels biochemischen Verfahren isoliert werden zu können, was eine strukturelle und funktionelle Analyse *in vitro* ermöglicht. Andere, wie z.B. G-Protein assoziierte Komplexe, assemblieren nur vorübergehend und sind nur sehr schwach gebunden, was eine Aufreinigung mit konventionellen biochemischen Methoden unmöglich macht. Solche Komplexe sind meist strukturell besonders variabel, da sie durch signifikante Rearrangements und Konformationsänderungen auf Signale antworten, d.h. sie sind wesentlicher Bestandteil von Signalwegen, die in einer Zelle existieren. Die Struktur solcher besonders fragilen Komplexe sollte idealerweise in der nativen Umgebung erforscht werden. Darüber hinaus besteht in einer Zelle ein höherer Grad der Organisation verschiedener Komplexe untereinander. Diese supramolekulare Architektur (*Supramolecular Architecture* (Baumeister 2005)) "verzahnt" die Komplexe untereinander: So wird z.B. der hohe intrazelluläre Organisationsgrad

bei der Chemotaxis von Bakterien sogar mit der Organisation des Gehirns verglichen (Webre et al. 2003).

Das derzeit wohl wichtigste Verfahren zur Aufklärung der dreidimensionalen Strukturen biologischer Makromoleküle ist die Röntgenkristallographie. Dieses Verfahren eignet sich vor allem zur Strukturaufklärung einzelner Proteine und wird zunehmend als Hochdurchsatzverfahren angewandt, um die zu ganzen Genomen korrespondierenden Proteinstrukturen aufzuklären (*Structural Genomics*). Andere etablierte Verfahren zur Strukturaufklärung in atomaren Dimensionen sind Kernspinresonanz (*Nuclear Magnetic Resonance*, NMR), die für relativ kleine ($\leq 30\text{kDa}$) in Lösung befindliche Proteine angewandt werden kann, und die Elektronenkristallographie, die sich vor allem zur Strukturbestimmung membrangebundener Makromoleküle eignet. Darüberhinaus hat sich die Kryoelektronenmikroskopie kombiniert mit der Einzelpartikelanalyse (*Single Particle Analysis*) zur Strukturaufklärung großer ($\geq 300\text{kDa}$) makromolekularer Komplexe etabliert. Die erreichbare Auflösung erlaubt nur in Ausnahmefällen eine direkte Bestimmung atomarer Koordinaten. In vielen Fällen sind aber atomare Strukturen der den Komplex konstituierenden Proteine bekannt, was mittels Einpassen dieser Daten in den Gesamtkomplex (*Fitting* und *Docking*) pseudoatomare Strukturen liefert. Allen genannten Techniken der Strukturbiologie ist gemein, daß sie biochemische Aufreinigung der Makromoleküle erfordern. Auf der anderen Seite sind vor allem in der Zellbiologie verschiedenste Formen der Lichtmikroskopie verbreitet, die es ermöglichen, zelluläre Strukturen und ihre Dynamik in ihrer natürlichen Umgebung zu studieren. Die Auflösung lichtmikroskopischer Techniken ist jedoch nicht ausreichend, um die Struktur makromolekularer Komplexe aufzulösen, was auch für Varianten der Lichtmikroskopie gilt, die darauf abzielen, das durch die Lichtbeugung gesetzte Auflösungslimit zu übertreffen.

Kryoelektronentomographie (KET) hat das Potential, die Kluft struktureller Studien, die derzeit zwischen der molekularen und der zellulären Ebene besteht, zu überbrücken. Diese Technik erlaubt die Strukturuntersuchung fragiler Komplexe sowie die Visualisierung der supramolekularen Architektur in ihrer natürlichen Umgebung und

ist daher hervorragend geeignet, die Funktion einzelner Komplexe anhand ihrer Struktur zu erkunden. Traditionell wurde die Elektronenmikroskopie in der Zellbiologie angewendet, um biologische Proben abzubilden, die chemisch fixiert, getrocknet und mittels Schwermetallsalzen kontrastiert wurden. Während diese strukturverändernde Probenpräparation jegliche Interpretation auf molekularer Ebene unmöglich machte, erlaubt es die Kryopräparation, die Einbettung biologischer Proben in vitrifiziertem Eis, aufgrund der damit verbundenen Strukturhaltung, Daten auf molekularer Ebene zu deuten.

Die Elektronentomographie (ET) basiert auf einem Prinzip, das zuerst von Radon (Radon 1917) entdeckt wurde: Eine Funktion, die z.B. eine Dichteverteilung repräsentiert, kann auf der Basis ihrer Projektionen rekonstruiert werden. Dieses Prinzip wird in allen Verfahren der dreidimensionalen (3D) Elektronenmikroskopie genutzt und wurde zuerst von DeRosier *et al* (DeRosier und Klug 1968) und Hart (Hart 1968) realisiert. Um Projektion des Objekts aus unterschiedlichen Richtungen zu erlangen, wird die Probe bei der ET und Elektronenkristallographie mechanisch gekippt, während bei der Einzelpartikelanalyse ausgenutzt wird, daß sich einzelne Partikel in unterschiedlichen Orientierungen auf einem EM-Grid befinden. Die fundamentale Limitierung der KET ist letztendlich die Elektronendosis, die zur Abbildung eines eiseingebetteten biologischen Präparats verwendet werden kann; die Wechselwirkung der abbildenden Elektronen mit der biologischen Materie führt zu inelastischen Streueignissen, die die Probenstruktur verändern und schließlich zerstören können. Ein technisches Problem bei der Aufnahme eines Tomogramms ist die mechanische Instabilität des Probenhalters: Während des Kippvorgangs bewegt sich die Probe aus ihrer euzentrischen Position, da eine mechanische Genauigkeit im Nanometerbereich nicht zu erreichen ist. Bei der Aufnahme einer Kippserie muß deshalb die Position des Elektronenstrahls und der Fokus relativ zur Probe konstant gehalten werden. Durch die Entwicklung von Automatisierungsprogrammen, die die auf die Probe eingestrahlte Elektronendosis zur Durchführung der notwendigen Korrekturen auf weniger als 3% der Gesamtdosis minimiert (Typke et al. 1991; Dierksen et al. 1992; Dierksen et al.

1993) wurde es schließlich ermöglicht, erste Tomogramme eiseingebetteter Proben aufzunehmen (Dierksen et al. 1995). Die zusätzliche Einführung von Energiefiltern zur Verbesserung der Abbildungsqualität führte schließlich zu den ersten Tomogrammen prokaryotischer Zellen (Grimm et al. 1998). Schließlich zeigte die Arbeit von Medalia *et al* (Medalia et al. 2002) eindrucksvoll, daß KET es ermöglicht einzelne molekulare Maschinen wie z.B. das 26S Proteasom im Kontext ganzer eukaryotischer Zellen darzustellen. Aufgrund der sehr niedrigen Elektronendosis, die zur Abbildung eines Objekts verwendet werden darf, ist das Signal-zu-Rausch-Verhältnis (*Signal-to-noise ratio*, SNR) von Kryoelektronentomogrammen sehr niedrig und bestimmt die erreichbare Auflösung. Studien an Viruskapsiden haben gezeigt, daß die Strukturinformation eines einzelnen Makromoleküls bis etwa $(7 \text{ nm})^{-1}$ signifikant ist (Grunewald et al. 2003).

Die technischen Entwicklungen im Bereich der KET der letzten Jahre ermöglichen es jetzt, qualitativ hochwertige Tomogramme pleomorpher biologischer Proben fast routinemäßig aufzunehmen. Somit bietet diese Abbildungstechnik nun die Möglichkeit, biologisch signifikante Fragestellungen zu behandeln, die auf der quantitativen Auswertung von Tomogrammen fußen. Ziel dieser Arbeit war die Entwicklung und Anwendung von Methoden zur quantitativen Auswertung von Tomogrammen. Es wurden Methoden zum Einsatz in folgenden biologischen Aufgabenstellungen entwickelt: (i) Die Erforschung der supramolekularen Architektur erfordert grundsätzlich Algorithmen, die Makromoleküle zuverlässig lokalisieren. (ii) Eine hochauflösende Strukturbestimmung von makromolekularen Komplexen in ihrer nativen Umgebung erfordert die kohärente Mittelung von Einzelpartikeln; das niedrige SNR begrenzt die Auflösung eines einzelnen Partikels, *in silico* kann die Elektronendosis und somit Auflösung jedoch gesteigert werden. (iii) Makromolekulare Komplexe können i.A. Konformationsänderungen und Interaktionen mit anderen Makromolekülen unterliegen. Eine systematische Untersuchung der Fluktuationen spezifischer Makromoleküle erfordert eine statistische Analyse der in einem oder mehreren Tomogramm befindlichen Einzelpartikel. Im einzelnen wurden in den Kapiteln folgende Themen behandelt:

- In Kap. 2 werden die Grundlagen der Tomographie erklärt.
- Das Kap. 3 behandelt den Abbildungsvorgang bei der Elektronenmikroskopie eiseingebetteter biologischer Proben. Es wurde besonderer Wert darauf gelegt, den komplexen Abbildungsvorgang speziell dicker Proben zu erläutern, da biologische Proben mit einem Durchmesser bis zu einem μm oder mehr, wie etwa ganze Zellen, in der Elektronenmikroskopie sehr ungewöhnlich sind, und die resultierende Abbildung in Lehrbüchern nicht erschöpfend behandelt ist.
- Kap. 4 umfaßt Material und Methoden, die die Grundlage der Arbeit bildeten.
- Das Thema von Kap. 5 sind Algorithmen zur Erkennung von Makromolekülen anhand ihrer strukturellen Signatur, die im Rahmen der Arbeit entwickelt wurden. Sie wurden zur Detektion verschiedener Komplexe in tomographischen Daten von Phantomzellen, pro- und eukaryotischen Zellen.
- In Kap. 6 wird ein Verfahren zur Mittelung von Subtomogrammen makromolekularer Komplexe, insbesondere membrangebundener Moleküle, vorgestellt. Als Anwendungsbeispiel der Prozeduren wird hier die Strukturbestimmung anhand des Kernporenkomplexes dargestellt.
- Eine andere Anwendung der Verfahren aus Kap. 6 stellt die Strukturbestimmung des Env Proteinkomplexes des Moloney Murine Leukemia Virus in Kap. 7 dar. Die erhaltene Struktur wird vor dem biologischen Hintergrund diskutiert.
- Der Inhalt von Kap. 8 ist die Entwicklung einer multivariaten statistischen Analyse, die sich speziell zur Analyse elektronentomographischer Daten eignet. Ziel des Verfahrens ist die Klassifikation heterogener Partikel in homogene Subpopulationen anhand einer systematischen Analyse der strukturellen Unterschiede der Einzelpartikel.
- Die Diskussion in Kap. 9 faßt die erhaltenen Ergebnisse zusammen und behandelt theoretische und experimentelle Arbeitsgebiete, die das Potential der in dieser Arbeit entwickelten Verfahren in der Zukunft weiter vergrößern könnten.

2. Tomographie

Tomographische Abbildungstechniken zielen darauf ab, Objekte dreidimensional (3D) abzubilden. Die ersten Versuche, biologische Objekte dreidimensional abzubilden, waren Tomogramme im Sinne des griechischen Wortes *Témnêin* (*Témnêin* = "schneiden"): Die Proben wurden in Serien physikalisch geschnitten (*serial sectioning*). Die Abbildung dieser Schnittserien z.B. mit einem Licht- oder Elektronenmikroskop ergaben dann ein dreidimensionales Bild des Objekts, wobei die Schnittdicke die Auflösung in der Richtung senkrecht zur Schnittebene bestimmte.

Neuere Techniken zur 3D Strukturanalyse bilden das Objekt nach einem ganz ähnlichen Prinzip ab, jedoch nicht-invasiv. Die wohl wichtigste 3D Abbildungstechnik ist die konfokale Lichtmikroskopie, die das Objekt *virtuell* schneidet: Das Objekt wird in verschiedenen Fokusebenen abgebildet. Somit kann es auch in der Strahlrichtung aufgelöst werden, wobei die in dieser Richtung erreichbare Auflösung im wesentlichen von der Tiefenschärfe (*depth of focus*) abhängt.

Alle Abbildungsmethoden, die heute gemeinhin als tomographisch bezeichnet werden, beruhen auf einem anderen Prinzip, das von Radon entdeckt wurde (Radon 1917): Eine Dichteverteilung kann auf der Basis ihrer Projektionen rekonstruiert werden. In Kap. 3 wird gezeigt, daß eine elektronenmikroskopische Abbildung in guter Näherung eine Projektion des elektrostatischen Potentials der Probe darstellt, weshalb sich die Elektronenmikroskopie als Abbildungstechnik für Tomographie eignet.

2.1 Projektionstheorem

Die Tatsache, daß eine Dichteverteilung durch ihre Projektionen eindeutig bestimmt ist, soll im Folgenden für eine Parallelprojektion erläutert werden. Eine Parallelprojektion \mathcal{P}_z einer Dichte $\rho(\mathbf{r})$ entlang der Projektionsrichtung z ist mathematisch definiert als:

$$\mathcal{P}_z(\rho(\mathbf{r})) = \int_{-\infty}^{\infty} dz \rho(\mathbf{r}). \quad (2.1)$$

Die Fourier Transformation \mathcal{F} von (2.1) ergibt sich zu:

$$\begin{aligned} \mathcal{F}\mathcal{P}_z(\rho(\mathbf{r})) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \exp(-ik_x x) \exp(-ik_y y) \int_{-\infty}^{\infty} \rho(\mathbf{r}) dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \exp(-ik_x x) \exp(-ik_y y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-ik_z z) \delta(z) \rho(\mathbf{r}) dz dk_z \\ &= \int_{-\infty}^{\infty} \hat{\rho}(k_x, k_y, k_z) \delta(k_z) dk_z = \hat{\rho}(k_x, k_y, 0). \end{aligned} \quad (2.2)$$

Hierbei bezeichnet $\hat{\rho}(k_x, k_y, k_z)$ die Fouriertransformierte der Dichte $\rho(\mathbf{r})$. Demzufolge ergibt sich aus (2.2), daß die Fouriertransformierte der Projektion von ρ einen zentralen Schnitt durch die Fouriertransformierte der dreidimensionalen Dichte normal zu \mathbf{k}_z beschreibt, was auch als Projektionstheorem bezeichnet wird. Variiert man die Projektionsrichtung, so können weitere zentrale Schnitte der Fouriertransformierten des Objektes erhalten werden, die zusammengenommen die dreidimensionale Information der Dichte beinhalten. Allerdings wäre die Zahl der benötigten Projektionen für ein unbegrenztes Objekt unendlich groß, da jeweils nur δ -förmige Schnitte erhalten werden können. In der Praxis sind abzubildende Objekte jedoch endlich, d.h. man kann sie mathematisch durch das Produkt einer stetigen Funktion $\rho^*(\mathbf{r})$ und einer Funktion $\Theta(\mathbf{r})$ beschreiben, die die äußere Begrenzung des Objekts beinhaltet,

z.B. eine Heaviside-Funktion. Für eine solche endliche Dichte ändert sich Gl. (2.2) zu:

$$\begin{aligned} \mathcal{F}(\mathcal{P}_z(\rho^*(\mathbf{r}) \cdot \Theta(\mathbf{r}))) &= \iiint_{-\infty}^{\infty} dx dy dz \exp(-ik_x x) \exp(-ik_y y) \rho^*(\mathbf{r}) \cdot \Theta(\mathbf{r}) \quad (2.3) \\ &= \hat{\rho}(k_x, k_y, 0) \otimes \hat{\Theta}(k_x, k_y, k_z). \end{aligned}$$

Die Fouriertransformierte der Projektion von \mathbf{r} ist die Faltung des Schnittes der Fouriertransformierten der kontinuierlichen Dichte ρ mit der Fouriertransformierten der Formfunktion $\hat{\Theta}$. Interpretiert man das Objekt als Kristall mit einer Einheitszelle, so werden die Datenpunkte $\hat{\rho}(k_x, k_y, k_z)$ Strukturfaktoren genannt, und $\hat{\Theta}(k_x, k_y, k_z)$ ist der Formfaktor des Kristalliten. Als Konsequenz von Gl. (2.3) bestimmt die Projektion im Falle endlicher Objekte einen Schnitt endlicher Dicke durch die Fouriertransformierte von ρ . Crowther *et al* (Crowther et al. 1970) haben aus der Dicke dieser Schnitte ein Kriterium für die erreichbare Auflösung abgeleitet (siehe Abb. 2.1): Nimmt man ein senkrecht zur Kippachse kreisförmiges Objekt an, d.h. einen Zylinder oder eine Kugel, mit einem Durchmesser d und wählt als Maß für die Dicke die Nulldurchgänge von $\hat{\Theta}$ (hier eine *sinc* Funktion), so erhält man für die Auflösung r als Funktion der Objektdicke d :

$$r \approx d \cdot \Delta\alpha. \quad (2.4)$$

Dabei wird angenommen, daß das Objekt mit einem konstanten Kippwinkelinkrement $\Delta\alpha$ (gemessen in rad) abgebildet wird. Dieses sog. Crowther-Kriterium ist eine konservative Abschätzung der Auflösung: Bei genauerer Kenntnis der Formfunktion Θ kann diese prinzipiell zur Interpolation oder sogar Extrapolation benutzt werden (Hoppe und Hegerl 1980). Durch Extrapolation der Daten kann das Crowther-Kriterium natürlicherweise übertroffen werden, da es eine Grenze angibt, bis zu der *interpoliert* werden kann. Numerische Verfahren, die darauf abzielen, die bekannte Form des Objekts zur Auflösungssteigerung zu nutzen, sind z.B. der Gerchberg-Saxton-Fienup Algorithmus und diverse davon abgeleitete Varianten, (Gerchberg und Saxton 1972; Fienup und Wackerman 1986; Miao et al. 1999). Aufgrund der ungenau definierten Formfunktion in der biologischen 3D Elektronenmikroskopie sind diese Verfahren

hier jedoch kaum verbreitet. Das Crowther-Kriterium zur Auflösungsbestimmung ist vor allem in der Elektronenkristallographie von Bedeutung: Bei dieser Technik werden zweidimensionale Proteinkristalle abgebildet, deren Ausmaße in z genau definiert sind.

Die Signifikanz dieses Kriteriums für die Tomographie von Zellen ist jedoch begrenzt: Diese Objekte sind komplex aufgebaut, sie können mathematisch als Superposition verschiedener Objekte wie Eis, Zellmembran oder Proteinen beschrieben werden, weshalb der Durchmesser d nicht präzise definiert ist. Das Crowther-Kriterium stellt jedoch trotzdem eine wichtige Kenngröße zur Interpretation von Tomogrammen dar, allerdings muß spezifiziert werden, worauf sich der Objektdurchmesser bezieht. So ist der maßgebliche Diameter bei der Beurteilung molekularer Auflösung der Diameter eines Makromoleküls und nicht der der Zelle, wie z.B. von McIntosh *et al* herausgestellt wurde (McIntosh et al. 2005).

Die ET hat unter einer technischen Einschränkung zu leiden: Aufgrund der Geometrie der Proben (*planares* EM-Netzchen) und des zugehörigen Probenhalters können Proben im TEM maximal von $+70^\circ$ bis -70° gekippt werden. Als Konsequenz davon sind beim Kippen um eine Achse Strukturfaktoren der Probe in einer keilförmigen Region im Fourierraum senkrecht zur Richtung der Kippachse nicht zugänglich (Abb. 2.1). Diese Region wird *Missing Wedge* genannt. Diese Daten können i.A. auch rechnerisch nicht rekonstruiert werden (siehe Abschn. 2.3), so daß die resultierenden Tomogramme im Realraum in z -Richtungen typischerweise künstliche elongiert sind (z.B. Abb. 2.3). Dieser Artefakt wird als *Missing Wedge* Effekt bezeichnet.

2.2 Dosisaufteilungstheorem

Die strahlungsinduzierte Schädigung der biologischen Proben ist die grundlegende Beschränkung der Kryoelektronentomographie (siehe Abschn. 3.1.5). Aufgrund der Strahlempfindlichkeit biologischen Präparate müssen diese mit äußerst geringen Elektronendosen abgebildet werden. Die Auflösungsabschätzung des Crowther-Kriteriums

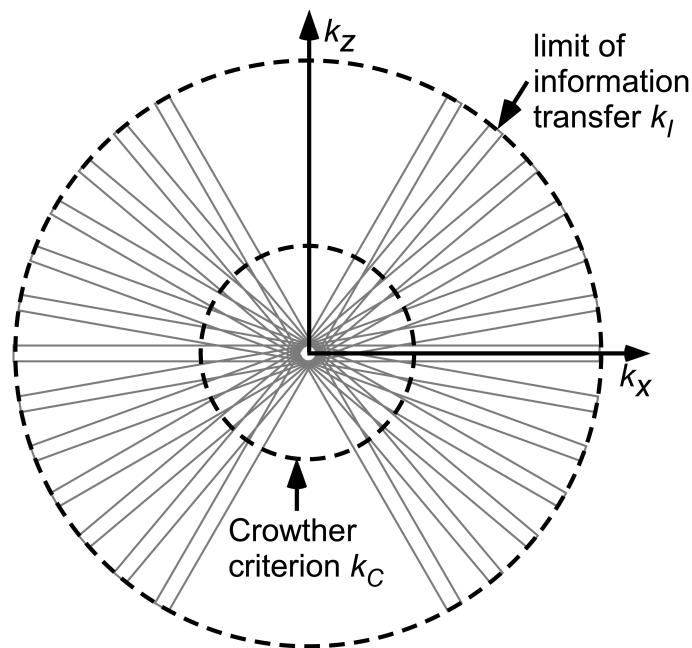


Abbildung 2.1: Zweidimensionale Veranschaulichung des Crowther Kriteriums. Die Fouriertransformierte jeder Projektion, die unter einem Kippwinkel von -60° , -50° , ... 60° aufgenommen wurden, beinhaltet die Information eines Schnitts der Dicke $1/d$ im 3D Fourierraum. Entlang k_y bleiben Strukturfaktoren aufgrund des eingeschränkten Kippwinkelbereichs unzugänglich. Von diesem Bereich abgesehen werden die Strukturfaktoren des Objekts bis zur Frequenz k_C homogen aufgenommen, die darüber hinausgehende Information ist unvollständig. Dennoch kann Objektinformation bis zu einer Auflösung aufgenommen werden, die über das Crowther Maß k_C hinausgeht und z.B. durch das Elektronenmikroskop begrenzt ist.

berücksichtigt jedoch nicht, daß der Informationsgehalt der einzelnen Projektionen durch Rauschen erheblich reduziert wird. Quantitative Behandlungen des Rauschens beruhen i.A. auf additivem Rauschen, und in diesem Modell ist das SNR der einzelnen Projektionen eine wichtige Kenngröße. Es ist definiert als:

$$SNR = \frac{\sigma_S^2}{\sigma_N^2}. \quad (2.5)$$

Dabei sind σ_S^2 und σ_N^2 die Varianzen des Signals bzw. Rauschens. Das SNR ist im Allgemeinen stark frequenzabhängig, da das Signal als Funktion der Frequenz schnell abnimmt.

Vor diesem Hintergrund erscheint es nicht besonders sinnvoll, anstatt eines recht signalschwachen Bildes eine Vielzahl von Bildern mit der gleichen Gesamtdosis aufzunehmen. Ein fundamentales Theorem, das zuerst von Hegerl und Hoppe (Hegerl und Hoppe 1976) aufgestellt wurde besagt jedoch, daß der Signalgehalt in beiden Fällen der gleiche ist. D.h. ein Tomogramm besitzt die gleiche Signalstärke σ_S wie eine einzelne Projektion, die mit der selben Dosis aufgenommen ist. Die Gültigkeit dieses Theorems wurde experimentell von McEwen *et al* bestätigt (McEwen et al. 1995). Dieses Theorem der Dosisaufteilung (*Dose Fractionation Theorem*) besagt, daß die Gesamtdosis auf beliebig viele Einzelbilder verteilt werden kann, die zusammen die gleiche Signal- oder Informationsmenge enthalten.

Als Folge dieses Theorems hat man die Wahl, eine Projektion mit hoher Dosis oder ein ganzes Tomogramm mit der gleichen Dosis aufzunehmen, um die gleiche Signalmenge zu erhalten. Das tatsächliche Signal ist jedoch in beiden Fällen ein anderes: Gemäß dem Projektionstheorem (Abschn. 2.1) ist in der Projektion das Signal bis zu einer als signifikant angesehenen Frequenz q innerhalb eines Schnitts durch den Fourierraum des Objekts enthalten, während das Tomogramm den gesamten 3D Fourierraum bis zu einer anderen Frequenz k beinhaltet. Im Falle der Tomographie benötigt man für ein Objekt mit Durchmesser D , dessen Signal spektral gleichverteilt ist, eine $(2D/k)$ mal so große Dosis um die Auflösung k zu erreichen (Saxberg und Saxton 1981).

Bei der Abbildung von Makromolekülen sollte die Tomographie von Vorteil sein, da

sie neben der Strukturinformation der Moleküle auch ihre örtliche, d.h. dreidimensionale, Verteilung enthält. Ob diese zusätzliche Information einen praktischen Nutzen hat, hängt davon ab, ob die erreichbare Auflösung ausreicht, einzelne Makromoleküle in 3D voneinander zu separieren (Saxberg und Saxton 1981).

Die ersten theoretischen Überlegungen über die erreichbare Auflösung kamen zu dem Ergebnis, daß "molekulare" Auflösung ($\sim 5\text{-}10\text{ nm}$) bei Kryopräparation nur mit Elektronendosen von über $100\text{ e}^{-\text{Å}^{-1}}$ zu erreichen ist (Saxberg und Saxton 1981). Bei diesen Überlegungen gingen Saxberg und Saxton jedoch von der Kontrastentstehung aus, wie sie mit der damaligen instrumentellen Ausrüstung zu erreichen war: Saxberg und Saxton stützen sich in dieser Hinsicht auf eine Arbeit von Misell und Burdett (Misell und Burdett 1977), die davon ausgeht, daß praktisch ausschließlich Streuabsorptionskontrast vorliegt (Abschn. 3.2.3). Durch die Einführung von Mikroskopen mit Feldemissionsquellen (*Field Emission Gun*, FEG, Abschn. 3.2.1), einer verbesserten Elektronenoptik, höheren Beschleunigungsspannungen und Energiefiltern (Abschn. 3.3) ist es jedoch heute möglich, auch dicke Proben überwiegend im Phasenkontrast abzubilden. Phasenkontrast ist ungleich ausgeprägter als Streuabsorptionskontrast, da er in erster Ordnung von der (sehr kleinen) Streuamplitude abhängt, während letzterer von zweiter Ordnung ist (Abschn. 3.2). Die Arbeit von Medalia *et al* (Medalia et al. 2002) hat erstmals experimentell demonstriert, daß es möglich ist, Makromoleküle mit einem Durchmesser von etwa $10\text{-}20\text{ nm}$ innerhalb eines zellulären Kryoelektronentomogramms aufzulösen. Demzufolge ist mit der Tomographie komplexer Objekte wie Zellen ein enormer Informationsgewinn verbunden gegenüber einer einzelnen Projektion mit der selben Dosis.

2.3 Dreidimensionale Rekonstruktion

Der erste Schritt jeder 3D Rekonstruktion ist die Alignierung der Projektionen: Die mechanischen Ungenauigkeiten des Drehmechanismus' führen zu Verschiebungen der

Probe, die *a posteriori* aus den Projektionen bestimmt werden müssen. Eine Möglichkeit dazu sind korrelationsbasierte Algorithmen, die die Projektionen aufeinander alignieren (Guckenberger 1982; Liu et al. 1995; Taylor et al. 1997). Allerdings sind diese Methoden im Allgemeinen sehr rauschempfindlich. Weiterhin erfordern sie letztlich Kenntnis über die äußere Form des zu rekonstruierenden Objekts, da von benachbarten Projektionen auf die zu jeweils zu alignierende Projektion extrapoliert werden muß.

Insbesondere bei eiseingebetteten Proben ist es deshalb üblich, die einzelnen Projektionen mit Hilfe von Markerpunkten zu alignieren, die Fixpunkte des zu bestimmenden 3D Koordinatensystems darstellen (Lawrence 1992). Die Markerkoordinaten (x_i, y_i) werden in den einzelnen Projektionen i bestimmt und mit den Koordinaten eines Modells verglichen. Der Fehler zwischen vorhergesagten und gemessenen Koordinaten wird minimiert. Das Standardmodell der Abbildung ist wie folgt: Der Ortsvektor \mathbf{r} des Markers wird abhängig von der jeweiligen Projektion durch die Fokusslage evtl. geringfügig unterschiedlich vergrößert, um den jeweiligen Kippwinkel α_i um die y -Achse gedreht, entlang der z -Achse projiziert und weiterhin wird das Bild im TEM durch die Elektronenoptik um einen Winkel β gedreht. Abschließend wird das Bild aufgrund der mechanischen Ungenauigkeit der Drehvorrichtung um einen Vektor $(\Delta x, \Delta y)$ verschoben. Demnach gilt für den Fehler eines Markerpunkts:

$$E = \sum_i |[\mathcal{R}_{\beta_i} (\mathcal{P}_z (\mathcal{R}_{\alpha_i} \mathcal{S}_i(\mathbf{r}))) + (\Delta x, \Delta y)] - (x_i, y_i)|^2. \quad (2.6)$$

Hierbei bezeichnen \mathcal{R} Drehungen, \mathcal{P}_z Projektion entlang z und \mathcal{S} Streckungen. Es ist zu beachten, daß die Lösung des obigen Gleichungssystems nicht eindeutig ist: Das Gleichungssystem ist für Lösungen mit entgegengesetzten Händigkeiten erfüllt, d.h. für den Winkel β gibt es die Lösungen β_0 und $\beta_0 + 180^\circ$. Um die richtige Kippachse zu ermitteln, muß die Kippachse *a priori* bekannt sein und wird deshalb für jedes TEMs bei verschiedenen Vergrößerungen kalibriert.

Anschließend werden die alignierten Projektionen mit einem geeigneten Verfahren 3D

rekonstruiert. Das Projektionstheorem läßt es intuitiv erscheinen, die dreidimensionale Rekonstruktion im Fourierraum durchzuführen. Tatsächlich wurde in der Pionierarbeit von DeRosier und Klug dieser Weg gewählt (DeRosier und Klug 1968). Allerdings werden solche Verfahren zur 3D Rekonstruktion im Fourierraum insbesondere für amorphe Proben heutzutage nur sehr selten benutzt, da Interpolationen im Fourierraum sehr kompliziert sind. Die Projektionsgeometrie erfordert jedoch Interpolation, so daß Verfahren im Realraum in ET üblich sind.

Eine Möglichkeit zur 3D Rekonstruktion im Realraum sind algebraische Rekonstruktionstechniken (*Algebraic Reconstruction Techniques, ART*) (Gordon et al. 1970). Diese Methoden verursachten anfangs heftige Kritik von Crowther und Klug (Crowther und Klug 1971), inzwischen sind diese Techniken jedoch etabliert. ART formuliert das Rekonstruktionsproblem als lineares algebraisches Gleichungssystem, das iterativ gelöst wird. ART bietet die Möglichkeit, Nebenbedingungen an die Lösung zu fordern (Carazo 1992). Dies ist häufig verlockend, da es die Möglichkeit zu bieten scheint, nichtabgetastete Information, insbesondere im Missing-Wedge, zu rekonstruieren. Etablierte Nebenbedingungen, die an die Lösung gestellt werden, sind minimale Energie (Penczek et al. 1992) oder eingeschränkte Basissätze wie "Blobs" (Marabini et al. 1997). Eine sehr mächtige, in der Signalverarbeitung verbreitete Bedingung stellt auch die maximale Entropie dar (Jaynes 1957). Es wurde berichtet, daß die Anwendung dieser Nebenbedingung zu drastischen Informationsgewinnen führt (Sandin et al. 2004), die jedoch noch zu beweisen sind.

Die meistverbreitete Methode zur 3D Rekonstruktion ist in der ET zweifellos die gewichtete Rückprojektion (Abb. 2.2). Bei dieser Methode werden die einzelnen Projektionen gemäß ihrer Projektionsrichtung in ein gemeinsames 3D Volumen zurückprojiziert. Die Rückprojektion erfolgt dabei numerisch im Allgemeinen mit trilinearer Interpolation. Wie in Abb. 2.1 illustriert ist, werden bei der einfachen Superposition der einzelnen Projektionen in der 3D-Rekonstruktionen die niedrigen Frequenzen

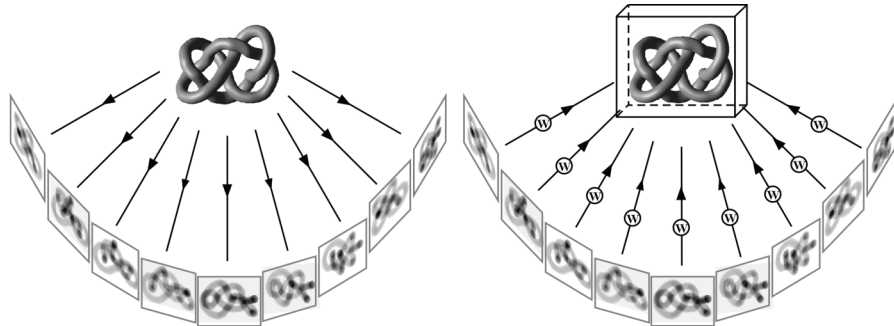


Abbildung 2.2: Prinzip der gewichteten Rückprojektion. Links: Es werden Projektionen des Objekts (hier ein DNA-Knoten) aus verschiedenen Richtungen aufgenommen. Rechts: Die einzelnen Projektionen werden zuerst gewichtet (w) und anschließend gemäß ihrer inversen Projektionsrichtung in ein gemeinsames Volumen zurückprojiziert.

überbetont: Es ist ein "Überlapp" an tieffrequenter Information vorhanden. Aus diesem Grund werden die Projektionen vor der Rückprojektion im Fourierraum gewichtet, um eine realitätsgetreue Gewichtung der einzelnen Frequenzbereiche in der 3D Rekonstruktion zu gewährleisten. Dabei unterscheidet man analytisches oder exaktes Gewichten: Analytisches Gewichten gewichtet die Fouriertransformation der Projektion in der Richtung senkrecht zur Kippachse mit dem Betrag der Frequenz in dieser Richtung, d.h. hohe Frequenzen bekommen deutlich mehr Gewicht. Bei der exakten Gewichtung nimmt man eine zylindrische bzw. kugelige Form des zu rekonstruierenden Objekts an und berechnet gemäß (2.3) die exakte Abtastungsdichte im Fourierraum (Hoppe und Hegerl 1980; Harauz und van Heel 1986). Allerdings erfordert dies die Angabe des Objektdurchmessers, da die Formfunktion bzw. ihre Fouriertransformation approximiert werden muß. Die Auswirkung verschiedener Gewichtsfunktionen ist in Abb. 2.3 gezeigt.

ART und gewichtete Rückprojektion sollten ähnliche Rekonstruktionen liefern. Allerdings besitzt ART neben der Möglichkeit Nebenbedingungen einzuführen auch einen Vorteil bzgl. der Gewichtung: ART sollte die richtige Gewichtung selbstkonsistent

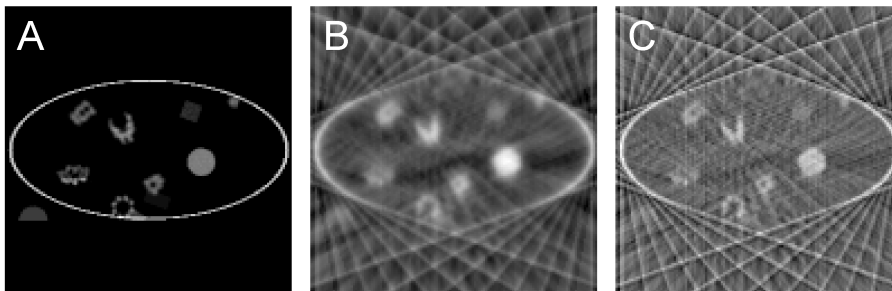


Abbildung 2.3: Effekt der Gewichtung auf die Rekonstruktion. A: xz -Schnitt einer synthetischen Zelle. B: Rekonstruktion der synthetischen Zelle aus 15 Projektionen von -70° bis $+70^\circ$ mit exakter Gewichtung, wobei als Durchmesser des Objekts der Vesikeldurchmesser in z gewählt wurde. C: Gewichtete Rückprojektion aus den selben Projektionen mit analytischer Gewichtung.

bestimmen, während dies bei der gewichteten Rückprojektion ein Parameter ist. Weiterhin sind biologisch interessante Objekte wie Zellen aus einer Vielzahl von Objekten mit unterschiedlichen Formen und charakteristischen Durchmessern aufgebaut, so daß die Rekonstruktion mit nur einer Formfunktion hier strenggenommen nicht mehr gültig ist. Allerdings ist ART sehr viel rechenintensiver, so daß die gewichtete Rückprojektion derzeit die Standardtechnik ist. Durch wachsende Rechenleistung und die Möglichkeit, den Rekonstruktionsprozeß zu parallelisieren kann ART insbesondere in der Tomographie komplexer Objekte wichtiger werden, allerdings ist der zu erwartende Gewinn eher als graduell einzustufen.

3. Elektronenmikroskopie

Die Wirkungsweise eines Transmissionselektronenmikroskops (TEM) ist mit der eines Lichtmikroskops (LM) vergleichbar. Im Gegensatz zum LM werden bei einem TEM Elektronen als Sonden zur Abbildung der Probe benutzt. Wie bei einem LM wird der Strahl der abbildenden Teilchen an der Probe gestreut und mittels eines Linsensystems wieder zu einem Abbild des Objekts zusammengeführt. Da Elektronen zur Abbildung benutzt werden, müssen die Proben im TEM im Hochvakuum gehalten werden.

Aufgrund der sehr viel kleineren Wellenlänge von Elektronen (z.B. $\lambda = 1.97$ pm bei 300 keV Elektronen) ist die erreichbare Auflösung eines TEMs prinzipiell deutlich größer als die eines LMs. Allerdings besitzt ein TEM noch einige wesentliche Unterschiede gegenüber einem gewöhnlichen LM: *i* Die Beleuchtung ist örtlich und zeitlich hochgradig kohärent, was bei einem gewöhnlichen LM nicht der Fall ist. Als Folge werden auch Objekte scharf abgebildet, die nicht exakt im Fokus sind. Die Abbildung wird durch Interferenzen der gestreuten Elektronen mit dem Primärstrahl ermöglicht, weshalb Hoppe (Hoppe 1981) das TEM gern mit einem Interferometer verglich (siehe Abschn. 3.2.1). *ii* Die Tiefenschärfe (*depth of focus*) eines TEMs ist aufgrund des sehr kleinen Streuwinkels der Elektronen sehr groß¹. Demzufolge werden identische Objekte an bzgl. der Strahlrichtung unterschiedlichen Positionen praktisch gleich abgebildet. Als Folge insbesondere der letzten Eigenschaft ist ein TEM Bild eines Objekts in erster Näherung eine Projektion der Probe in Richtung des Elektronenstrahls.

¹Die Tiefenschärfe T gibt die Distanz an, innerhalb der eine Objekt mit einer Auflösung δ abgebildet wird. T hängt vom Streuwinkel α bzw. der Apertur ab, und es gilt $T < \delta/\alpha$.

3.1 Streuprozesse im Elektronenmikroskop

Um den Zusammenhang zwischen der Probe und ihrem Abbild im TEM genauer zu erläutern, sollen im Folgenden die auftretenden Streuprozesse beschrieben werden. Grundsätzlich muß zwischen inelastischen und elastischen Streuvorgängen unterschieden werden. Die quantitative Behandlung dieser Streuvorgänge muß quantenmechanisch erfolgen.

Nimmt man die Elektronenquelle als perfekt kohärent an (zeitlich und örtlich), so kann die Wellenfunktion des einfallenden stationären Elektronstrahls als ebene Welle beschrieben werden:

$$\psi_0(\mathbf{r}, t) = |\Psi_0| \exp(i(\mathbf{k}_0 \cdot \mathbf{r} + \frac{E_0}{\hbar}t)). \quad (3.1)$$

Dabei ist E_0 die Energie des einfallenden Elektrons und \mathbf{k}_0 sein Wellenvektor. Als Flugrichtung der einfallenden Elektronen soll hier die z -Achse angenommen werden (Richtungsvektor \mathbf{e}_z), d.h. $\mathbf{k}_0 = \frac{2\pi}{\lambda} \cdot \mathbf{e}_z$. Die Wellenlänge λ der Elektronen muß relativistisch berechnet werden:

$$\lambda = \frac{h}{\sqrt{2em_0U(1 + \frac{eU}{m_0c^2})}}. \quad (3.2)$$

Hierbei ist h das Plancksche Wirkungsquantum, m_0 die Ruhemasse eines Elektrons, e die Ladung eines Elektrons und c die Lichtgeschwindigkeit. Bei einer Beschleunigungsspannung U von 300 kV beträgt die Wellenlänge eines Elektrons 1.97 pm.

Die Wechselwirkung der Elektronen mit der Materie führt zu Streuereignissen. Nimmt man an, daß die Wellenfunktion des Systems, das aus der Probe und dem Elektronenstrahl besteht, durch zwei Wellenfunktionen beschrieben werden kann, d.h. eine ebene Welle für den Elektronenstrahl und eine Wellenfunktion ξ für die streuenden Materie, so kann der Streuvorgang folgendermaßen beschrieben werden:

$$\left(\exp(i(\mathbf{k}_0 \cdot \mathbf{r} + \frac{E_0}{\hbar}t)), \xi_0 \right) \rightarrow \left(\exp(i(\mathbf{k} \cdot \mathbf{r} + \frac{E}{\hbar}t)), \xi \right). \quad (3.3)$$

Es werden also beim Streuvorgang der Wellenvektor \mathbf{k} und die Energie E des Elektrons gegenüber ihren Anfangswerten \mathbf{k}_0 und E_0 geändert, und der Ausgangszustand der

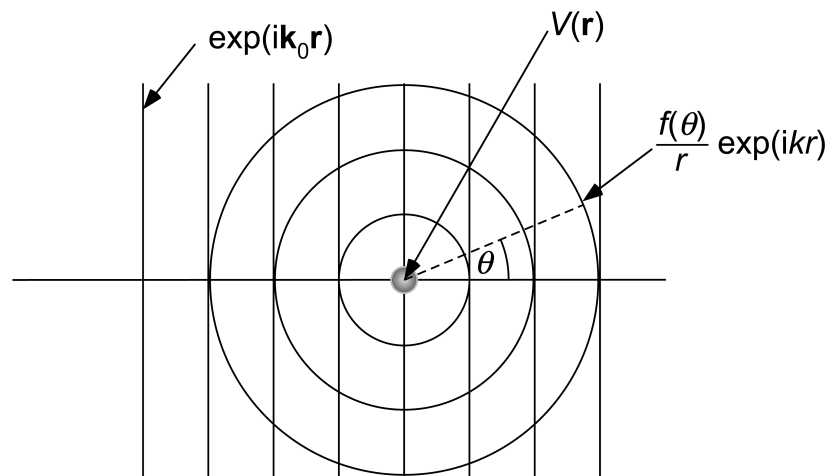


Abbildung 3.1: Streuung am Zentralpotential in erster Näherung. Es sind die Wellenfronten der ungestreuten und der gestreuten Welle eingezeichnet.

Probe ξ_0 geht durch den Streuvorgang in einen anderen Zustand ξ über. Dabei unterscheidet man elastische und inelastische Streuvorgänge: Streuvorgänge ohne oder mit Energieübertrag auf die Materie. Für elastische Streuvorgänge gilt:

$$E = E_0 \quad , \quad |\mathbf{k}| = |\mathbf{k}_0|. \quad (3.4)$$

3.1.1 Kinematische- und Bornsche Näherung

Bei der Streuung von Elektronen in dünnen amorphen Proben werden zwei wichtige Annahmen gemacht: (i) Die Elektronen werden jeweils nur einmal gestreut und die Gesamtstreuung ergibt sich somit einfach aus der Summe der Einzelereignisse. Dieses einfache Superpositionsprinzip der Streuereignisse wird kinematische Näherung genannt. (ii) Die Streuung ist eine schwache Störung und kann somit störungstheoretisch behandelt werden, in diesem Fall mit der Bornschen Näherung.

Für die Streuung eines Elektrons am elektrostatischen Potential V eines Atoms ergibt sich mit der Bornschen Näherung für die Ausgangswellenfunktion der Elektronen

(z.B. (Schwabl 1990)):

$$\psi(\mathbf{r}, t) = |\Psi_0| \left(\exp(i(\mathbf{k}_0 \cdot \mathbf{r} + \frac{E_0}{\hbar}t)) + if(\vartheta) \frac{\exp(ikr + \frac{E'}{\hbar}t)}{r} \right). \quad (3.5)$$

Die Wellenfunktion ist also in erster Näherung gleich der Eingangswellenfunktion Ψ_0 , die überlagert ist mit der (sehr viel schwächeren) gestreuten Wellenfunktion (Abb. 3.1). Diese ist eine Kugelwelle, deren Intensität durch die Streuamplitude $f(\vartheta)$ bestimmt wird. Zwischen Streuamplitude und dem differentiellen Wirkungsquerschnitt $d\sigma$ im Raumwinkelement $d\Omega$ besteht der fundamentale Zusammenhang (Schwabl 1990):

$$d\sigma = |f(\vartheta)|^2 d\Omega = 2\pi |f(\vartheta)|^2 \sin(\vartheta) d\vartheta. \quad (3.6)$$

Die Größe der Streuamplitude $f(\vartheta)$ in Gl. (3.5) ist in erster Näherung gemäß Fermis Goldenener Regel durch die Übergangswahrscheinlichkeiten der jeweiligen Streueignisse bestimmt.

3.1.2 Elastische Streuung

Die für die Bildentstehung wichtigsten Streueignisse sind elastische Streuungen, wie in Abschn. 3.2 deutlich wird. Für elastische Streuung ergibt sich für die Streuamplitude in Richtung $\vartheta = \arccos(\mathbf{k}\mathbf{k}_0/|k_0^2|)$ nach erster Ordnung Störungsrechnung:

$$f(\vartheta) = \iiint \frac{m}{2\pi\hbar^2} \exp(-i(\mathbf{k} - \mathbf{k}_0)\mathbf{r}) V(\mathbf{r}) d^3r, \quad |\mathbf{k}| = |\mathbf{k}_0|. \quad (3.7)$$

Da es sich bei den atomaren Streupotentialen um Zentralpotentiale handelt, sind die Streuamplituden in erster Näherung real (zur weitergehenden Diskussion siehe Abschnitt 3.2.2). Führt man weiterhin den Differenzvektor $\mathbf{K} = \mathbf{k} - \mathbf{k}_0$ ein, so ergibt sich für die Streuamplitude:

$$f(\vartheta) = \iiint \frac{m}{2\pi\hbar^2} \exp(-i\mathbf{K} \cdot \mathbf{r}) V(\mathbf{r}) d^3r. = \hat{V}(\mathbf{K}). \quad (3.8)$$

Demnach ist also die elastische Streuamplitude in erster Näherung proportional der Fouriertransformierten des Potentials $\hat{V}(\mathbf{K})$. Dabei liegt der Wellenvektor \mathbf{K} auf der

sog. Ewaldkugel (Abb. 3.3). Sei \mathbf{q} der Anteil des Streuvektors \mathbf{k} senkrecht zur Richtung des einfallenden Strahl, so gilt aufgrund der kleinen Wellenlänge λ und dem in der Elektronenmikroskopie beobachteten Streuwinkelbereich:

$$\mathbf{K} = \mathbf{k} - \mathbf{k}_0 \approx \mathbf{q}. \quad (3.9)$$

Im Folgenden soll statt des Streuwinkels ϑ weitgehend die Frequenz \mathbf{q} verwendet werden, und für die Umrechnung gilt in obiger Näherung:

$$\mathbf{q} \approx \frac{1}{\lambda} \vartheta. \quad (3.10)$$

Theoretisch kann also durch Messung der gestreuten, *komplexen* Wellenfunktion das Potential der Probe senkrecht zur Richtung des einfallenden Strahls ermittelt werden. Da diese jedoch keine Observable ist, kann in der Praxis (z.B. in der Röntgenkristallographie) nur die Intensität gemessen werden, weshalb das Potential nicht direkt aus den Streudaten erhalten werden kann. In der Elektronenmikroskopie wird das Potential im Prinzip durch die Rücktransformation der gestreuten Welle in den Realraum erhalten: Die Objektivlinse bewirkt dabei prinzipiell die Rücktransformation in den Realraum (Abschn. 3.2.1).

Das Coulombpotential eines Atoms der Ordnungszahl Z wird meist durch das sog. Wentzel-Modell approximiert:

$$V(r) = -\frac{e^2 Z}{4\pi\epsilon_0 r} \exp(-r/R). \quad (3.11)$$

In dieser Näherung nimmt man an, daß der erste Term, das Potential des Kerns, durch die umgebenden, gebundenen Elektronen exponentiell geschwächt wird. Für die radiale Abschwächung gilt dabei $R = a_B Z^{-1/3}$. Für kleine Z , d.h. insbesondere für biologische Materialien, wird häufig approximiert (Frank 1996):

$$V(r) \sim Z. \quad (3.12)$$

Gemäß Gl. (3.6) hat dabei Lenz den elastischen Wirkungsquerschnitt abgeschätzt zu (Lenz 1954):

$$\frac{d\sigma_{el}}{d\Omega} = \frac{4Z^2 R^4 (1 + \frac{E}{m_0 c^2})^2}{a_B^2 (1 + (\vartheta/\vartheta_0)^2)^2} \quad (3.13)$$

Dabei gilt für den charakteristischen Streuwinkel ϑ_0 :

$$\vartheta_0 = \frac{\lambda}{2\pi R} = \frac{\lambda Z^{1/3}}{2\pi\mu_B}. \quad (3.14)$$

Dieser charakteristische Streuwinkel läßt sich mit (3.9) in eine charakteristische Streufrequenz Q_0 umrechnen, so daß z.B. für Kohle ($Z=6$) gilt $Q_0 \approx (0.2 \text{ nm})^{-1}$.

3.1.3 Inelastische Streuung

Zu inelastischer Streuung kommt es bei Energieübertrag vom einfallenden Elektronenstrahl auf die Probe, d.h. die Probenatome werden in einen angeregten Zustand versetzt. Hierzu müssen theoretisch die möglichen Anregungszustände und Übergangswahrscheinlichkeiten in Gl. 3.3 berücksichtigt werden. Die wichtigsten Anregungen sind im einzelnen:

1. Phononen und Molekülschwingungen. Im Festkörper werden kooperative Schwingungen der Atomkerne bzw. Moleküle angeregt, d.h. diese Anregungen sind delokalisiert. Außerdem kommt es zu Anregungen von Molekülschwingungen (lokalisiert). Die Anregungsenergien ΔE betragen 20 meV-1eV. Diese Energiedifferenz ist zu klein, um sie mit einem herkömmlichen TEM mit Energiefilter aufzulösen, da die Energiebreite des Primärstrahls auch bei einem FEG Instrument ($\approx 1 \text{ eV}$) größer ist als diese Energiedifferenz.
2. Anregungen der äußeren Elektronen, d.h. der Valenz- und ggf. Leitungselektronen. Diese Anregungen umfassen Intra- und Interbandanregungen sowie kollektive Schwingungen der äußeren Elektronen. Letztere werden gemäß ihrer Ausbreitung an der Oberfläche oder im Volumen als Exzitonen oder Plasmonen bezeichnet. Der Energieverlust ΔE der einfallenden Elektronen liegt im Bereich von etwa 1-50 eV. Plasmonen sind die häufigsten Anregungen, weshalb in diesem Energiebereich auch vom *Plasmon Peak* gesprochen wird. Bei diesen Streuungen werden hochenergetische Elektronen an anderen Elektronen inelastisch gestreut, weshalb diese Streueignisse in Analogie zur inelastischen Streuung von Röntgenstrahlen an Elektronen auch Compton-Streuung genannt wird.

3. Ionisation durch Anregung von Kernelektronen. Elektronen der *inneren* Elektronenschalen werden auf ein höheres Energieniveau angeregt. Hierzu ist eine charakteristische Minimalenergie nötig, weshalb das Elektronenenergieverlustspektrum (*Electron Energy Loss Spectrum*, EELS) charakteristische Kanten aufweist, die mit Anregungen bestimmter Kernelektronen korrelieren. Diese Verlustkanten sind elementspezifisch und können deshalb zur Detektion bestimmter Elemente verwendet werden (*Elemental Mapping*). Die übertragene Energie ist dabei größer als die Bindungsenergie des jeweiligen Kernelektrons. Für K-Elektronen von Kohlenstoff beträgt diese charakteristische Energie z.B. etwa 280 eV. Die Energieniveaux der Kernelektronen werden unter Erzeugung von Auger-Elektronen oder Röntgenstrahlung wieder besetzt.

Unter den oben genannten Streumechanismen sind die elektronischen Anregungen der äußeren Elektronen am wichtigsten, d.h. sie besitzen mit Abstand den höchsten Streuquerschnitt. Der differentielle Streuquerschnitt kann prinzipiell wieder mit der Goldenen Regel approximiert werden, jedoch würden hierzu die Wellenfunktionen der angeregten- und Grundzustände der Elektronen benötigt werden, was außerordentlich komplex ist. Eine wiederum auf Lenz zurückgehende Näherung erlaubt jedoch eine qualitative Beschreibung des differentiellen inelastischen Wirkungsquerschnitts (Lenz 1954): Unter der Annahme des Wentzel Modells (3.11) und eines einzigen Energiezustands aller äußeren Elektronen gilt:

$$\frac{d\sigma_{inelast}}{d\Omega} = \frac{\lambda^4 \left(1 + \frac{E}{m_0 c^2}\right)^2}{4\pi^4 \mu_B^2} Z \frac{\left(1 - \frac{1}{[1 + (\vartheta^2 + \vartheta_E^2)/\vartheta_0^2]^2}\right)}{(\vartheta^2 + \vartheta_E^2)^2}. \quad (3.15)$$

Hierbei gilt für den charakteristische Winkel der inelastischen Streuung ϑ_E :

$$\vartheta_E = \frac{\Delta E}{2E}. \quad (3.16)$$

Dabei ist ΔE der Mittelwert des Energieübertrags während der Streuung. Im Vergleich zur elastischen Streuung Gl. (3.13) ist die Streuung sehr viel stärker nach vorne gerichtet. Dies ist zu erklären durch den stark delokalisierten Charakter der Streueignisse; die äußeren Elektronen werden *kollektiv* angeregt. Demzufolge enthalten inelastisch gestreute Elektronen sehr viel weniger Information über die Strukturdetails

einer Probe als elastische gestreute e^- . Der charakteristische inelastische Streuwinkel ϑ_E beträgt z.B. bei 300 kV für eiseingebettete biologische Proben ($\Delta E \approx 20 \text{ eV}$) etwa 0.03-0.04 mrad. Dies bedeutet, daß inelastische Streuung vor allem für Frequenzen $q < k_0 \cdot \vartheta_E \approx 0.02 \text{ nm}^{-1}$ stark ausgeprägt ist. Demzufolge ist die inelastische Streuung besonders wichtig für Bildinformation im Bereich von $\sim 20\text{-}50 \text{ nm}$.

Allerdings hat die Lorentz-Funktion in Gl. 3.15 auch noch nennenswerte nichtverschwindende hochfrequente Anteile, da sie als Funktion der Frequenz nur sehr langsam gegen Null abfällt. Dies führt zu einem diffusen Hintergrund in der hochfrequenten-Information wie im Abschnitt 3.2.2 erläutert wird.

Für die relative Größe von elastischem und inelastischem Wirkungsquerschnitt kann eine $1/Z$ Abhängigkeit abgeleitet werden (Lenz 1954) und experimentell wurde gezeigt, daß gilt (Egerton 1976):

$$\frac{\sigma_{inel}}{\sigma_{el}} \approx \frac{20}{Z}. \quad (3.17)$$

Demzufolge ist bei biologischen Proben die inelastische Wechselwirkung etwa dreimal so häufig wie die elastische. Trotzdem ist die elastische Wechselwirkung für die Kontrastentstehung biologischer Proben dominant, wie im Abschn. 3.2 erklärt wird.

3.1.4 Streuamplitude komplexer Proben

Gemäß der kinematischen Näherung ergibt sich die Streuamplitude eines aus vielen Streuzentren, d.h. Atomen, aufgebauten Objekts durch Summation der einzelnen Streuamplituden. Befinden sich die Atome i an Orten \mathbf{r}_i so ergibt sich für die Streuamplitude des gesamten Objekts:

$$F(\mathbf{K}) = \sum_i f_i(\mathbf{K}) \exp(i\mathbf{r}_i \cdot \mathbf{K}). \quad (3.18)$$

Die beiden Faktoren der Summanden werden meist atomarer Formfaktor und Strukturfaktor genannt. Wiederum kann die gesamte Streuamplitude in eine elastische und eine inelastische Streuamplitude aufgeteilt werden. Dabei muß jeweils der elastische bzw. inelastische Streufaktor f_i eingesetzt werden. Für biologische Objekte ist die

Näherung (3.12) der elastischen atomaren Formfaktoren für viele Anwendungen, insbesondere für die Tomographie, ausreichend, so daß gilt:

$$F(\mathbf{K}) \sim \sum_i Z_i \exp(i\mathbf{r}_i\mathbf{K}). \quad (3.19)$$

3.1.5 Strahlschädigung vitrifierter Proben

In der Probe kann inelastische Streuung zu Veränderungen der Probe führen. Der allgemeine Konsens besagt, daß es aufgrund inelastischer Wechselwirkungen zur Bildung von Radikalen kommt. Durch Diffusion dieser Radikale und anschließende Sekundärreaktionen kommt es schließlich sogar zu makroskopisch sichtbaren Strahlschäden wie z.B. zur Bildung von Blasen im amorphen Eis (Heide und Zeitler 1985).

Um eine möglichst hohe Ausbeute an Signal pro auftretendem Strahlschaden zu erreichen, ist ein hohes Verhältnis von elastischem zu inelastischem Streuquerschnitt der Elektronen in der Probe notwendig. Beide Wirkungsquerschnitte sinken als Funktion der Beschleunigungsspannung, ihr Quotient bleibt jedoch etwa konstant (Henderson 1995). Obwohl die zum Strahlschaden führenden Primärprozesse nicht temperaturabhängig sind, kann der letztlich sichtbare Strahlschaden jedoch durch Kühlung eingedämmt werden, da er auf Sekundärreaktionen beruht. Diese benötigen jedoch die Diffusion freier Radikale, die durch Kühlung auf die Temperatur flüssigen Stickstoffs (77 K) eingeschränkt werden kann².

Die unter Kryobedingungen tolerierbare Elektronendosis, d.h. die eingestrahlte Elektronenzahl pro Fläche oberhalb derer signifikante Strukturveränderungen der Probe auftreten, hängt von der Zielsetzung der Strukturuntersuchung und den chemischen

²Durch Kühlung mit flüssigem Helium können diese Prozesse noch stärker eingeschränkt werden, allerdings kommt es bei diesen Temperaturen möglicherweise zu einem Kontrastverlust, insbesondere für mittlere Auflösungsbereiche ($> 2 \text{ nm}^{-1}$). Diese Beobachtungen lassen sich durch eine Phasenumwandlung von Wasser bei etwa 20 K erklären, was zuerst von Heide (Heide und Zeitler 1985) beobachtet wurde und durch neue Messungen bestätigt wird (Schweikert 2004).

Eigenschaften der spezifischen Probe ab. So können auf höchste Auflösung abzielende Studien mittels Elektronenkristallographie nur minimalen Strahlschaden tolerieren, während bei anderen Anwendungen Schäden in der Größenordnung von einzelnen Atomen oder Aminosäuren vereinzelt auftreten dürfen. Weiterhin agieren z.B. Zucker als Radikalfänger, weshalb höhere Strahlresistenz von Proben in Trehaloselösung (De Carlo et al. 1999) oder Makromolekülen wie Ribosomen (Frank 2002), die mehr als 50% RNA enthalten, möglich ist.

Für typische Proteinkomplexe sollte die Elektronendosis bei Elektronenkristallographieprojekten weniger als $10 \text{ e}\text{\AA}^{-2}$ betragen (Henderson 1995), während die Grenze für *Single Particle* Projekte bei $10\text{-}20 \text{ e}\text{\AA}^{-2}$ bei 120 kV gesehen wird (Frank 2002), was etwa $20\text{-}30 \text{ e}\text{\AA}^{-2}$ bei 300 kV entspricht. Für Projekte in der Kryoelektronentomographie sind etwas höhere Dosen tolerierbar, abhängig von Probe und angestrebter Auflösung sollten etwa bis zu $50\text{-}100 \text{ e}\text{\AA}^{-2}$ zulässig sein.

3.2 Kontrastmechanismen

Eine ideale Linse wirkt folgendermaßen: Die an einem Punkt \mathbf{r} gestreute Kugelwelle wird mittels der Linse auf einen Punkt fokussiert, und es ergibt sich eine Vergrößerung M (Abb. 3.2). Die Wellenfunktion in der Bildebene \mathbf{r}' ergibt sich durch Integration über die innerhalb des Aperturbereichs gestreuten Elektronen. Sei ein differentielles Element in der Fokusebene $d^2\mathbf{q}$ so ist der zusätzlich von den gestreuten Elektronen zurückgelegte Weg $\lambda\mathbf{q}\mathbf{r}'$. Dies entspricht einer Phasendifferenz von $\mathbf{q}\mathbf{r}'$. Demnach gilt für die Wellenfunktion im Fokus:

$$\psi(\mathbf{r}') = \frac{1}{M}\psi_0 \left(1 + i \iint_{\mathbf{q}\in Q} F(\mathbf{q}) \exp(i2\pi\mathbf{q}\mathbf{r}') d^2\mathbf{q} \right). \quad (3.20)$$

Demnach bildet die Linse also im Wesentlichen die Fouriertransformierte der Streuamplitude ab. Allerdings müssen weitere Phasendifferenzen W berücksichtigt werden,

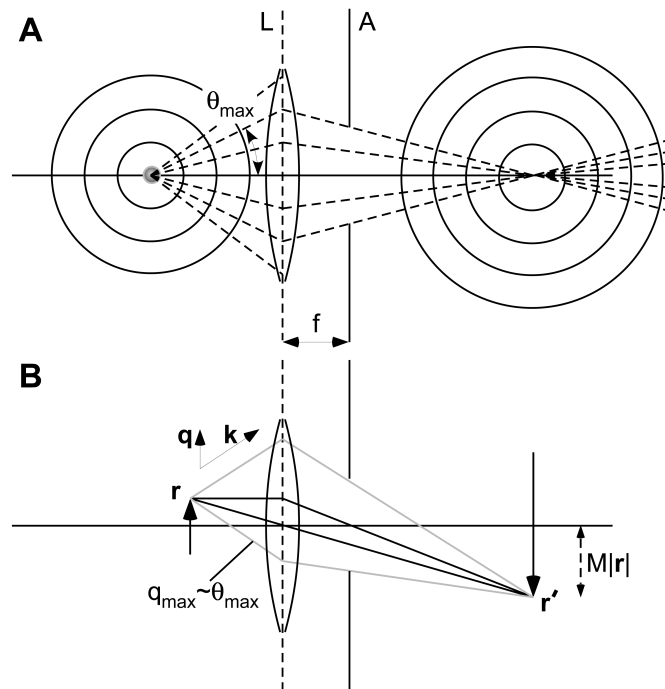


Abbildung 3.2: A. Abbildung einer idealen Linse für einen Punkt auf der optischen Achse. Durchgezogene Linien: Wellenfront, gestrichelte Linien: Trajektorien. L: Linse, A: Apertur, f: Brennweite. B. Für einen Punkt außerhalb der optischen Achse \mathbf{r} sind die Strahlengänge der gemäß \mathbf{q} (zweidimensionale Projektion des dreidimensionalen Streuvektors \mathbf{k}) gestreuten Elektronen eingezeichnet. Ein Punkt \mathbf{r} wird auf einen Punkt \mathbf{r}' abgebildet, so daß eine Vergrößerung M auftritt.

die aufgrund von Linsenfehlern (Aberrationen) und möglicher Defokussierung auftreten.

$$\psi(\mathbf{r}') = \frac{1}{M}\psi_0 \left(1 + i \iint_{\mathbf{q} \in Q} F(\mathbf{q}) e^{-iW(q)} e^{i2\pi\mathbf{q}\mathbf{r}'} d^2\mathbf{q} \right) \quad (3.21)$$

$$= \frac{1}{M}\psi_0 \left(1 + i \iint \Theta(|\mathbf{q}| - q_{max}) F(\mathbf{q}) e^{-iW(q)} e^{i2\pi\mathbf{q}\mathbf{r}'} d^2\mathbf{q} \right). \quad (3.22)$$

Hierbei beschreibt die Heaviside Funktion $\Theta(|\mathbf{q}| - q_{max})$ die Wirkung der Apertur und wird in der Optik meist als *Pupil function* bezeichnet. Für die Phasendifferenz W gilt die fundamentale Beziehung (Reimer 1997):

$$W = \frac{\pi}{2}(C_S \lambda^3 q^4 - 2\Delta z \lambda q^2). \quad (3.23)$$

Diese Beziehung berücksichtigt neben dem Gangunterschied der unterschiedlich gestreuten Strahlen in Abhängigkeit vom Defokus Δz den Einfluß der sphärischen Aberration der Objektivlinse. Dieser Linsenfehler wird mit der sphärischen Aberrationskonstante C_S beschrieben und liegt z.B. für ein modernes TEM bei etwa 2 mm.

Somit gilt für die Ausgangswellenfunktion in erster Näherung:

$$\psi(\mathbf{r}') = \frac{1}{M}\psi_0 \left(1 + i \iint \Theta(|\mathbf{q}| - q_{max}) F(\mathbf{q}) [\cos(W(q)) - i \sin(W(q))] e^{i2\pi\mathbf{q}\mathbf{r}'} d^2\mathbf{q} \right). \quad (3.24)$$

In erster Näherung, d.h. bei Vernachlässigen aller nicht-linearen Glieder, ergibt sich für die gemessene Intensität I :

$$I(\mathbf{r}') \approx |\psi\psi^*| = \left(|\psi_0|^2 + \iint 2\psi_0 F(\mathbf{q}) \Theta(|\mathbf{q}| - q_{max}) \sin(W(q)) e^{i2\pi\mathbf{q}\mathbf{r}'} d^2\mathbf{q} \right). \quad (3.25)$$

3.2.1 Phasenkontrast

Der Kontrast eines Bildes ist definiert als:

$$C_{x'y'} = I_0 - I_{x'y'}. \quad (3.26)$$

Nach Gl. (3.25) gilt für den Kontrast:

$$C_{x'y'} = \mathcal{F}^{-1}(CTF(q)\Theta(|\mathbf{q}| - q_{max}) \cdot F(\mathbf{q})). \quad (3.27)$$

Zur Vereinfachung der Beschreibung des Abbildungsprozess' wurde hier die sog. Kontrasttransferfunktion (*Contrast-transfer function*, *CTF*) eingeführt, für die gilt:

$$CTF(q) = \sin\left(\frac{\pi}{2}(C_S\lambda^3q^4 - 2\Delta z\lambda q^2)\right). \quad (3.28)$$

Der durch Gl. (3.27) beschriebene Kontrastmechanismus heißt Phasenkontrast, da er rein durch Interferenz der gestreuten Lichtstrahlen mit dem ungestreuten Elektronenstrahl entsteht. Er beruht auf der Annahme, daß die Streuzentren schwach interagieren und lediglich einfach streuen (kinematische Näherung). Nimmt man Einfachstreuung der Elektronen an und brücksichtigt die große DOF des TEMs, so muß zur Berechnung der Intensität für Objekte endlicher Dicke lediglich über die Objektdicke integriert werden:

$$I_{x'y'} = \frac{1}{M} \mathcal{F}^{-1} \left(CTF(q) \Theta(|\mathbf{q}| - q_{max}) \cdot \mathcal{F} \left(\int dz V(\mathbf{r}) \right) \right). \quad (3.29)$$

Phasenkontrast ist dominant für dünne, amorphe Proben bestehend aus Atomen niedriger Ordnungszahlen Z , d.h. insbesondere für dünne (~ 50 - 100 nm) vitrifizierte biologische Präparate. Der Anteil des Phasenkontrasts am Gesamtsignal beträgt für solche Proben im Auflösungsbereich oberhalb von etwa $(5 \text{ nm})^{-1}$ ca. 90% (Toyoshima und Unwin 1988; Langmore und Smith 1992; Zhu et al. 1997). Obwohl der inelastische Wirkungsquerschnitt dieser Atome größer ist als der elastische, ist Phasenkontrast dominant, da er linear zum Potential, d.h. der Störung ist, während Kontrastmechanismen zu denen die inelastische Streuung beiträgt von höherer Ordnung sind.

Die CTF bietet eine elegante Möglichkeit, die Abbildung des Mikroskops in der hier vorgestellten linearen Näherung zu behandeln. Weiterhin kann sie erweitert werden, um die Auswirkungen der begrenzten örtlichen und zeitlichen Kohärenz zu beschreiben. Für die CTF ergibt sich dann:

$$CTF(q) = K_s(q) \cdot K_t(q) \cdot \sin(W(q)). \quad (3.30)$$

Dabei bestimmen K_s und K_t die aus der begrenzten örtlichen bzw. zeitlichen Kohärenz resultierenden Dämpfungen des Kontrastübertrags als Funktion der Frequenz q .

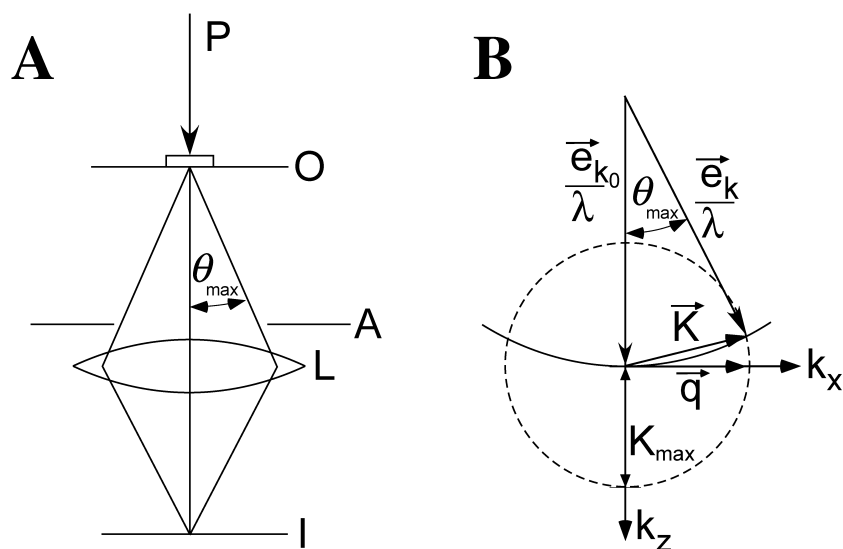


Abbildung 3.3: Abbildung im TEM in Real- und Fourierraum (adaptiert aus (Hoppe und Hegerl 1980)). A) Schema der Abbildung im Realraum. Der Primärstrahl P trifft auf das Objekt O, die Elektronen werden gestreut, und die Elektronen innerhalb der Apertur A, d.h. die maximal unter einem Winkel θ_{max} gestreuten Elektronen, werden mittels der Objektivlinse L zum Bild I zusammengeführt. B) Aufgrund der großen Tiefenschärfe entspricht das Abbild einem Schnitt durch den Fourierraum. Hier ist ein zweidimensionaler Schnitt des Fourierraums gezeigt mit den Basisvektoren k_x , der die Richtung senkrecht zu Kippachse und Strahlrichtung angibt, sowie k_z , der in Richtung von P zeigt. Im Detail liegen die erfassten Strukturfaktoren auf der Ewaldkalotten (durchgezogene Linie). Dabei werden Datenpunkte innerhalb der erreichbaren Auflösung K_{max} erfasst, die letztlich von der Objektivapertur A abhängt. Aufgrund der sehr kleinen Wellenlänge ist der Radius der Ewaldkalotten für die in der ET zu erreichenden Auflösungen praktisch unendlich, d.h. die Krümmung der Ewaldkalotten ist zu vernachlässigen.

Für die sog. Einhüllenden der CTF gilt (Frank 1976):

$$K_s(q) = \exp\left(-\frac{\pi(C_s\lambda^2 \cdot q^3 - \Delta z \cdot q)^2\alpha^2}{\log(2)}\right). \quad (3.31)$$

Hierbei ist α die Beleuchtungsapertur, die ein Maß für die Punktförmigkeit der Elektronenquelle ist. Verwendet man ein konventionelles thermisches LaB₆-Filament als Elektronenquelle (siehe Abschn. 4.1.1), so beträgt α etwa 0.2 mrad. Benutzt man eine moderne Feldemissionskathode (*Field Emission Gun, FEG*), so beträgt α nur etwa ein Zehntel, 0.02 mrad.

Für den zeitlichen Anteil der Einhüllenden gilt (Reimer 1997):

$$K_t(q) = \exp\left(-\left(\frac{C_c\pi\lambda\Delta E q^2}{4\sqrt{\log(2)}}\right)^2\right). \quad (3.32)$$

Die Halbwertsbreite ΔE der Energieverteilung beträgt etwa 3 eV für eine LaB₆ und 0.8 eV für eine FEG. Die chromatische Aberrationskonstante C_c ist eine Eigenschaft der Objektivlinse und liegt bei einem modernen TEM zwischen 1 und 5 nm.

In Abb. 3.4 sind verschiedene CTFs für verschiedene Defokuswerte Δ und Beschleunigungsspannungen U skizziert. Es wird deutlich, daß insbesondere die erhöhte räumliche Kohärenz einer FEG auch bei großen Defokuswerten einen nennenswerten hochfrequenten Kontrastübertrag ermöglicht. Aus diesem Grunde ist es bei FEG Instrumenten möglich, zur Kontrasterhöhung ungewöhnlich stark zu defokussieren, was bei der Tomographie eiseingebetteter Proben notwendig ist.

3.2.2 Amplitudenkontrast

Die Annahme, daß die Streuamplitude $f(\vartheta)$ rein real sind, stellt nur eine Näherung dar. Insbesondere für höhere Ordnungszahlen Z muß die Näherung ergänzt werden. Behandelt man die Streuung über die erste Bornsche Näherung hinaus, so ergibt sich eine komplexe Streuamplitude:

$$f(\vartheta) = f_{real}(\vartheta) + if_{imag}(\vartheta). \quad (3.33)$$

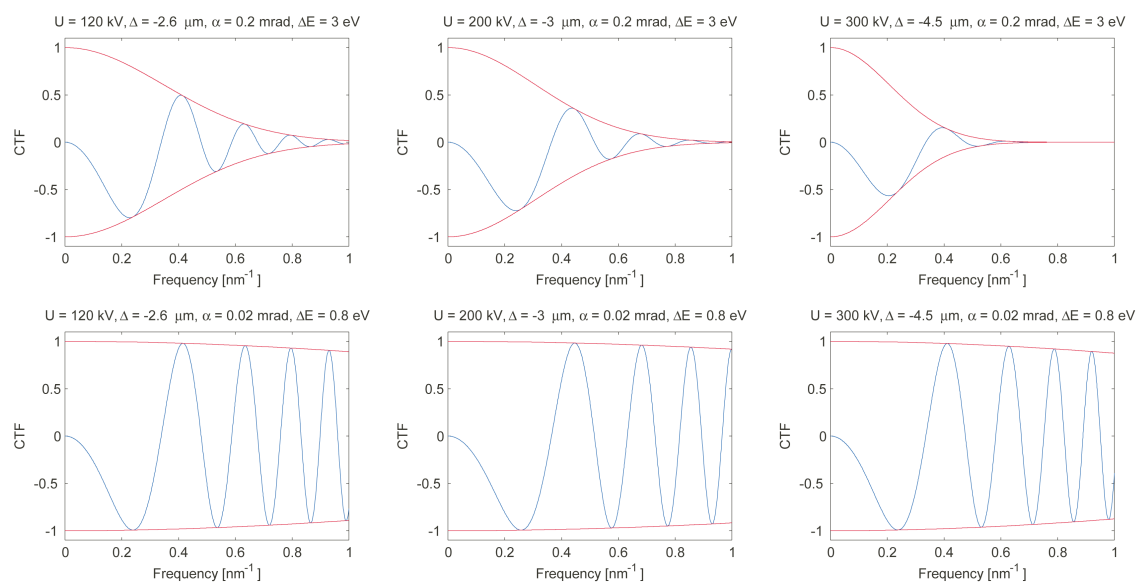


Abbildung 3.4: Kontrasttransferfunktionen (blau) in Abhängigkeit von Beschleunigungsspannung und Kohärenz. Die jeweiligen Produkte aus räumlicher und zeitlicher Einhüllender der CTF sind rot gezeichnet. Die oberen Parameter α und ΔE sind typisch für ein LaB₆ Instrument und die unteren für ein FEG-TEM. Der Defokus Δ ist immer so gewählt, daß die erste Nullstelle der CTF bei $(3 \text{ nm})^{-1}$ liegt.

In der Elektronenmikroskopie biologischer Materialien geht man meist davon aus, daß der Imaginärteil der Streuamplitude zu vernachlässigen ist (*Weak phase approximation*). Der Name dieser Approximation wird deutlich durch Umschreiben von (3.33) in:

$$f(\vartheta) = |f(\vartheta)| \exp(i\eta(\vartheta)). \quad (3.34)$$

Die imaginäre Streuamplitude führt also zu einer weiteren Phasenverschiebung η der gestreuten Welle. Das Verhältnis von Real- zu Imaginärteil ist im für die Elektronenmikroskopie relevanten Frequenzbereich in guter Näherung konstant (Reimer 1997). Setzt man die komplexe Streuamplitude in (3.25) ein, so ergibt sich die modifizierte Intensität:

$$I(\mathbf{r}') = |\psi\psi^*| = \left(|\psi_0|^2 + \iint 2\psi_0 f_{real}(\mathbf{q}) \Theta(|\mathbf{q}| - q_{max}) \sin(W(q)) e^{i2\pi\mathbf{q}\mathbf{r}'} d^2\mathbf{q} \right. \\ \left. + \iint 2\psi_0 f_{imag}(\mathbf{q}) \Theta(|\mathbf{q}| - q_{max}) \cos(W(q)) e^{i2\pi\mathbf{q}\mathbf{r}'} d^2\mathbf{q} \right). \quad (3.35)$$

Für biologische eiseingebettete Proben beträgt der Anteil des Amplitudenkontrasts etwa 10 % des Phasenkontrasts (Toyoshima und Unwin 1988). Unter der Annahme eines konstanten Verhältnisses von f_{real} zu f_{imag} kann der Amplitudenkontrast durch Berücksichtigung einer zusätzlichen Phasenverschiebung in der CTF beschrieben werden.

3.2.3 Streuabsorptionskontrast

Beim Phasen- und Amplitudenkontrast wird der Intensitätsverlust durch außerhalb der Apertur gestreute Elektronen vernachlässigt. Auch durch diesen Effekt kommt es aufgrund des lateral variierenden Wirkungsquerschnitts zu Kontrast, dem sog. Streuabsorptionskontrast (*Scattering Contrast*). Zum quantitativen Verständnis des Streuabsorptionskontrasts werden üblicherweise folgende vereinfachende Annahmen gemacht (Misell und Burdett 1977; Hawkes 1992): Die Streuquerschnitte sind translationsinvariant, somit variiert lateral lediglich die Dichte ρ und Dicke t der Probe. Der Anteil der pro Volumeneinheit a in der Apertur absorbierten Elektronen $1/\mu_{abs}$

beträgt somit:

$$\frac{1}{\mu_{\text{abs}}} = N \int_{q>q_{\text{max}}} |f(\mathbf{q})|^2 d^2q = N \int_{q>q_{\text{max}}} (|f_{\text{elast}}(\mathbf{q})|^2 + |f_{\text{inelast}}(\mathbf{q})|^2) d^2q. \quad (3.36)$$

Dabei ist N die Zahl der Streuzentren pro Volumen. Unter der Annahme einfacher Streuereignisse kann gezeigt werden, daß gilt (Hawkes 1992):

$$I_{x'y'} = \frac{I_0}{M^2} \exp\left(-\frac{\mu(x,y)}{\mu_{\text{abs}}}\right). \quad (3.37)$$

Hierbei ist μ die Massendicke der Probe, d.h. $\mu = \rho \cdot t$. In obiger Ableitung wurden die Linsenfehler der Einfachheit halber vernachlässigt und die Abbildung im Fokus angenommen.

Der Streuabsorptionskontrast ist insbesondere bei stark streuenden Proben wichtig. Obwohl eiseingebettete biologische Proben schwach streuen, ist der Streuabsorptionskontrast in der KET wichtig, da die Proben sehr dick sind, und die Massendichte aus diesem Grunde groß ist. Die Näherung eines Primärstrahls, der in der Probe keine Intensität verliert, ist in der KET nicht mehr gültig.

Unter Verwendung der Beziehung (3.6) zwischen differentiellem Streuquerschnitt und Streuamplitude ergibt sich aus (3.37): Durch Logarithmieren der Bildintensität kann die Projektion der dreidimensionalen Struktur erhalten werden. Streuabsorptionskontrast wird bei der Abbildung dicker, schwermetallkontrastierter biologischer Proben angenommen (Misell und Burdett 1977; Hawkes 1992). Da hier die Mikroskopie im Gaußschen Fokus erfolgt, ist diese Näherung weitgehend erfüllt. Eine dreidimensionale Rekonstruktion beruht dann auf den logarithmierten Bildintensitäten.

Berücksichtigt man auch Phasen- bzw. Amplitudenkontrast, so kann die Bildentstehung beliebig kompliziert werden, und eine detaillierte Behandlung erfordert Simulationen. Simulationen sind in der Materialwissenschaft für sehr dünne Proben etabliert, jedoch ist auch hier die Diskrepanz zwischen gemessenen und vorhergesagten Intensitäten sehr groß ("Stobbs Faktor" von etwa 4-5, zusammengefaßt in (Howie 2004)). Die Simulation dickerer Proben wird durch vielfache inelastische Wechselwirkungen weiterhin aufwendiger, so daß Simulationen, wie z.B. in (Dinges et al. 1995) beschrieben, bei der Abbildung komplexer biologischer Proben bisher praktisch nicht zum

Einsatz kommen.

3.3 Energiegefilterte Elektronenmikroskopie

Der Kontrast einer TEM Aufnahme kann mittels eines Energiefilters erhöht werden. Die Funktion eines solchen Filters ist die Filterung der bildgebenden Elektronen innerhalb eines bestimmten Energieintervalls. Demnach kommt es gegenüber dem ungefilterten Fall zu deutlich größerer Absorption, und es liegt somit ein größerer Anteil an Streuabsorptionskontrast vor. In der Mikroskopie eiseingebetteter biologischer Proben wird praktisch ausschließlich im *Zero-Loss* Modus gearbeitet: Hier werden die Elektronen innerhalb eines Energieintervalls von etwa 10 eV um die Energie des Primärstrahls herum gefiltert, d.h. es tragen hauptsächlich die elastisch gefilterten Elektronen zur Ausgangsintensität bei.

Der Streuabsorptionskontrast im *Zero-Loss* Modus kann unter Berücksichtigung unterschiedlicher Aperturen für elastische und inelastische Streuung gemäß (3.36) quantitativ beschrieben werden (Reimer und Ross-Messemer 1989): Formal ist die Apertur für inelastisch gestreute Elektronen Null, während sie elastisch gestreute e^- nur oberhalb eines gewissen Streuwinkels absorbiert, der durch die benutzte Objektivapertur vorgegeben ist. Somit gilt für den mittleren freien Weg analog zu (3.36):

$$\frac{1}{\mu_{\text{abs}}} = N \left(\int_{q > q_{\text{max}}} |f_{\text{elast}}(\mathbf{q})|^2 d^2q + \int |f_{\text{inelast}}(\mathbf{q})|^2 d^2q \right). \quad (3.38)$$

Der zusätzliche Streuabsorptionskontrast sollte sich jedoch vor allem im niederfrequenten Bereich bemerkbar machen: Die inelastische Streuereignisse sind stark delokalisiert (Abschn. 3.1.3), so daß Streuabsorptionskontrast vor allem im Frequenzbereich unterhalb von $(20\text{-}40 \text{ nm})^{-1}$ auftritt.

Eine weitere Folge der Energiefilterung ist eine deutliche Verbesserung der Abbildungsqualität: Insbesondere die Objektivlinse von Elektronenmikroskopen erzeugen eine starke chromatische Aberration, die zu einem diffusen Hintergrund führt. Da die inelastische Streuamplitude auch einen nichtverschwindenden hochfrequenten Anteil hat (Abschn. 3.1.2), kommt es so bei konventionellen TEM Aufnahmen zu einem

starken Untergrundsignal. Mittels Energiefilterung kann dieser Effekt umgangen werden, da hier die Abbildung nur in einem kleinen spektralen Bereich erfolgt. In der Elektronenmikroskopie eiseingebetteter Proben hat Grimm als erster den Gewinn an Signalqualität untersucht (Grimm et al. 1996; Grimm 1997). Diese Studien ergaben bereits bei etwa 100 nm dicken Proben eine Verbesserung des SNR um einen Faktor von zwei im Frequenzbereich von $(5-10 \text{ nm})^{-1}$.

Ein weiterer Effekt der Energiefilterung sei hier der Vollständigkeit halber erwähnt: Das Verhältnis von Phasen- und Amplitudenkontrast (Abschn. 3.2.2) ändert sich bei energiegefilterter Elektronenmikroskopie geringfügig gegenüber dem ungefilterten Fall und beträgt etwa 14 % für eiseingebettete Proben (Langmore und Smith 1992; Zhu et al. 1997) anstatt 9 % im ungefilterten Fall (Toyoshima und Unwin 1988). Angert *et al* haben eine lineare Beschreibung der Abbildung geliefert, die Phasen- und Amplitudenkontrast allgemeiner behandelt (Angert et al. 2000). Diese verallgemeinerte Beschreibung ist jedoch nur bei höchsten Auflösungen ($\sim 1 \text{ nm}$) von Bedeutung (Holmes et al. 2003).

Um die Auswirkung des Energiefilters auf den Kontrast eiseingebetteter biologischer Proben zu untersuchen, wurden eiseingebettete Phantomzellen bei verschiedenen Defokuswerten mit und ohne Energiefilter (*Zero-Loss* Filterung) aufgenommen (Abb. 3.5, A und B). Beim visuellen Vergleich der Bilder fällt vor allem der höhere Kontrast der energiegefilterten Bilder auf, z.B. erscheinen die Vesikel mit Energiefilterung deutlich 'dunkler'. Zum qualitativen Vergleich der Bilder wurden die Powerspektren der jeweiligen Bilder berechnet und anschließend rotationsgemittelt (Abb. 3.5, C und D). Das rotationsgemittelte Powerspektrum gibt Aufschluß darüber, welche örtlichen Frequenzen des Bildes übertragen werden. Es zeigt sich, daß vor allem das tieffrequente Signal bei Energiefilterung unterhalb einer Frequenz Q_E von etwa $(20-40 \text{ nm})^{-1}$ sehr viel stärker ist, als im ungefilterten Fall. Diese Frequenz korrespondiert zum charakteristischen inelastischen Streuwinkel von etwa 0.2 mrad bei 300 kV (Abschn. 3.1.3).

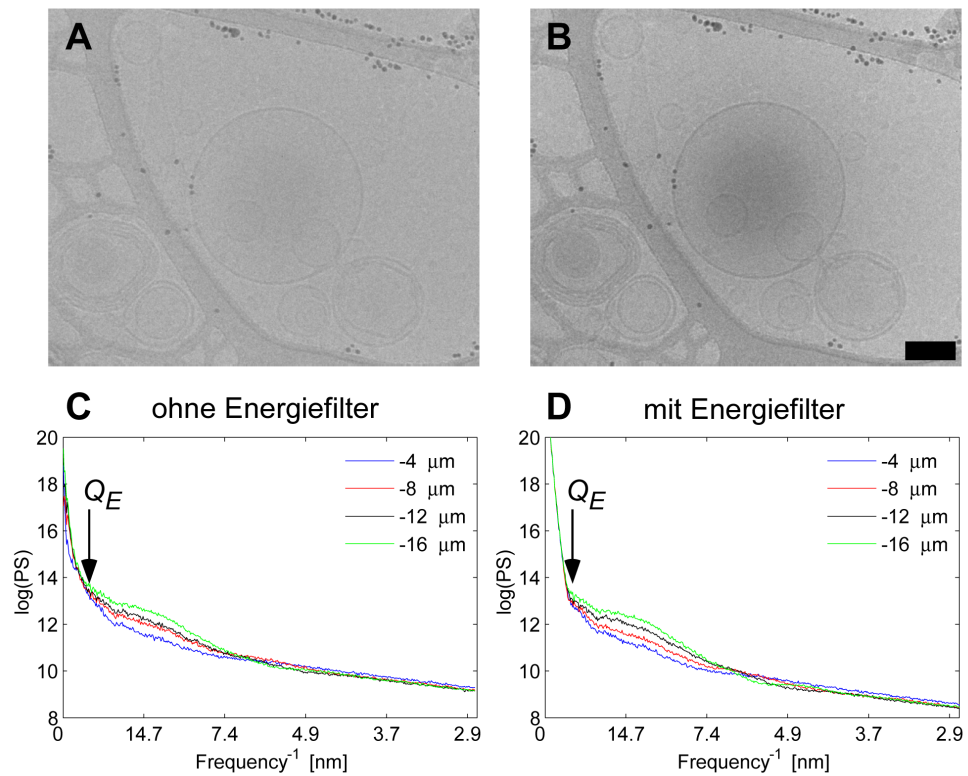


Abbildung 3.5: Auswirkung der Energiefilterung auf den Kontrast. Obere Reihe: Phantomzellen bei $8 \mu\text{m}$ Defokus ohne (A) und mit Energiefilterung (B). Der Maßstab entspricht 100 nm . Die Phantomzellen wurden nach Ref. (Frangakis et al. 2002) präpariert ausgenommen dem finaler Entfernung der Proteinkomplexe aus dem umgebenden Puffer mittels Ni-NTA Säule. C und D: Korrespondierende rotationsgemittelte Powerspektren. Pfeil: Frequenz Q_E unterhalb der das energiegefilterte Bild ein besonders starkes Signal enthält.

4. Material und Methoden

4.1 Experimenteller Aufbau

4.1.1 Das Transmissionselektronenmikroskop

Die Optik eines modernen TEMs, wie es in der vorliegenden Arbeit benutzt wurde, beinhaltet im Wesentlichen folgende Elemente (Abb. 4.1): Feldemissionsquelle (FEG), Kondensor-, Objektiv- und Projektorlinsensystem. Eine FEG ist eine Elektronenquelle, die Elektronen höchster zeitlicher und örtlicher Kohärenz erzeugt, die somit höchste Auflösung ermöglicht (Abb. 3.4). Die FEG besteht aus einer orientierten Wolframspitze als Kathode und zwei Anoden. Aufgrund des geringen Radius der Spitze ($<0.1 \mu\text{m}$) entsteht an ihrer Oberfläche ein sehr hohes elektrisches Feld ($E > 10^9 \text{ V/m}$), wenn an der gegenüberliegenden Extraktorkathode ein positives Potential (3.5-3.8 kV) angelegt wird. Aufgrund des starken Potentialgradienten können Elektronen aus der Kathode tunneln. Der so erzeugte Elektronenstrahl besitzt gegenüber LaB₆ Instrumenten eine sehr viel höhere örtliche und zeitliche Kohärenz, da die Spitze der Elektronenquelle klein ist und sehr viel weniger geheizt werden muß.

Der Elektronenstrahl wird mittels des Kondensorlinsensystems (C1 und C2 in Abb. 4.1) auf die Probe geleitet, die annähernd parallel beleuchtet wird. Die an der Probe gestreuten Elektronen werden mittels des Objektivlinsensystems wieder zu einem Zwischenbild zusammengeführt, insofern ihr Streuwinkel innerhalb der Objektivapertur liegt. Dieses Bild wird mittels des Projektivlinsensystems vergrößert.

Die in dieser Arbeit verwendeten Daten stammten überwiegend von einem CM300 TEM (Philips, Niederlande), das mit einem Gatan Post-Column Energiefilters GIF2002 ausgerüstet war (Abschn. 4.1.2). Das Instrument wurde bei 300 kV betrieben. Lediglich die Daten des 26S Proteasoms in Kap. 8 wurden an einem CM20 TEM ohne

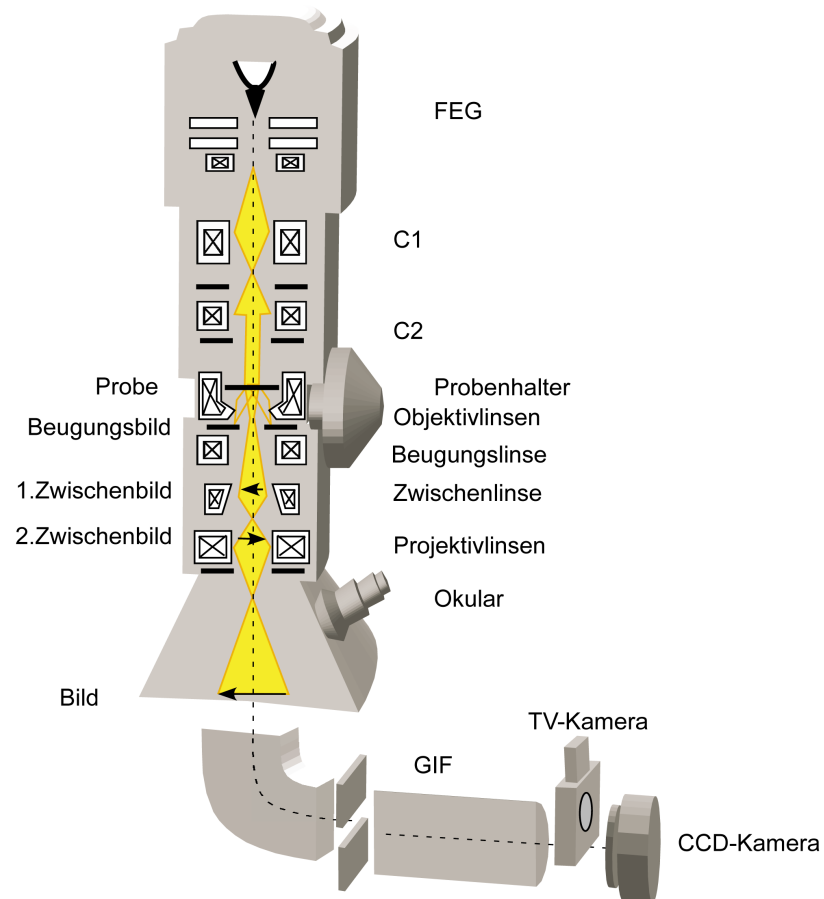


Abbildung 4.1: Schematischer Aufbau des CM300 TEMs in Martinsried. Das Gerät ist mit einem *Post-Column* Energiefilter und einer CCD Kamera ausgerüstet (aus (Schweikert 2004)).

Energiefilter aufgenommen, das bei 160 kV betrieben wurde. Die eiseingebetteten Proben wurden mit Hilfe eines Kryo-Probenhalter (Gatan, Pleasanton, Kalifornien) im TEM untersucht.

4.1.2 Der Energiefilter

Zur Verbesserung des SNR wurde am CM300 ein Energiefilter benutzt, der im *Zero-Loss* Modus betrieben wurde (Nickell 2001). Im Detail handelte es sich bei diesem um einen nachträglich an das TEM, außerhalb der eigentlichen Mikroskopsäule angebrachten *Post-column filter* (Gatan Imaging Filter, GIF, (Krivanek et al. 1995)).

In einem Energiefilter wird mit der Elektronenstrahl Hilfe eines magnetischen Prismas spektral aufgespalten. Durch Einführen einer Blende können Elektronen eines bestimmten Energiebereichs herausgefiltert werden. Anschließend wird der Strahl wieder fokussiert, und es ergibt sich ein Bild, das nur von Elektronen eines ausgewählten Energieintervalls enthält. In der vorliegenden Arbeit wurde ausschließlich *Zero-Loss* Filterung verwendet, d.h. es wurden nur die Elektronen aus dem Energiebereich des Primärstrahls verwendet, wobei die Breite des Intervalls stets 10 eV betrug.

4.1.3 Die CCD Kamera

Die in der Elektronenmikroskopie heutzutage verwendeten CCD-Kameras detektieren nicht direkt Elektronen. Stattdessen werden die Elektronen mittels eines Szintillators erst in Photonen umgewandelt, die dann mit einem CCD-Chip detektiert werden. Je nach Kopplung des Szintillators unterscheidet man in der Elektronenmikroskopie zwischen glasfiber- und linsengekoppelten Kameras (Überblick z.B. in (Fan und Ellisman 2000)). In der vorliegenden Arbeit wurden ausschließlich die sehr viel verbreiteteren fibergekoppelten CCD-Kameras verwendet: Am CM300 wurde eine Gatan GIF 2002 CCD (Gatan, Pleasanton, Kalifornien) und am CM20 eine TVIPS F214 (TVIPS, Gauting, Deutschland) verwendet.

Der schematische Aufbau einer solchen CCD-Kamera ist in Abb. 4.2 dargestellt.

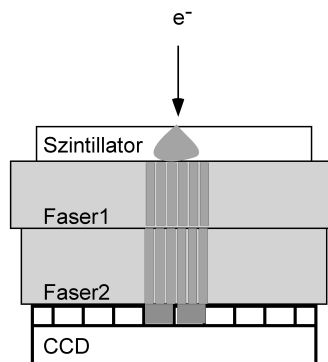


Abbildung 4.2: Aufbau einer glasfasergekoppelten CCD Kamera.

Die Primärelektronen werden zunächst durch Streuprozesse in einem Szintillatorschirm abgebremst. Dieser besteht in den verwendeten Kameras aus P43 (Gadoliniumoxysulfide GOS) Phosphor. Das Licht wird durch Felder aus Glasfaserbündeln auf den CCD Chip weitergeleitet. Aus technischen Gründen besteht die Faseroptik dabei aus zwei Schichten. Sie sind durch einen dünnen Film, bestehend aus einem Immersionsöl, gekoppelt. Die Platten bestehen aus Faserbündeln, die wiederum aus etwa $5 \mu\text{m}$ dicken Einzelfasern zusammengesetzt sind. An den Grenzen der Bündel befinden sich 2-4 Fasern breite Bereiche geringerer Empfindlichkeit. Der eigentliche CCD-Chip setzt sich aus einer Matrix aus Metall-Oxid-Halbleiter (MOS-) Kondensatoren zusammen. Durch Einstrahlung von Photonen entstehen in den einzelnen Kondensatoren sog. *Well*-Elektronen, die das letztlich ausgelesene Signal darstellen. Um das Störsignal durch thermisch generierte Elektronen zu reduzieren, wird der CCD-Chip durch ein Peltier-Kühlelement auf ungefähr -35°C gekühlt.

Das registrierte Bild einer CCD Kamera weist Artefakte auf, die korrigiert werden müssen, um in sehr guter Näherung die tatsächliche Intensität I_{xy} zu erhalten. Die Artefakte kommen zum einen durch die Eigenstruktur des CCD Chips zustande: Durch thermische Anregung werden stets einige *Well* Elektronen registriert, wobei der Erwartungswert dieser Fluktuationen aufgrund unterschiedlicher Dotierungen jedoch lokal deutlich variieren kann. Weiterhin glüht die Elektronik einiger Chips schwach, was

auch zu einem systematischen Untergrundsignal führt. Zum anderen werden strukturelle Artefakte durch die fiberoptische Kopplung eingeführt, da diese wie oben angedeutet eine laterale Struktur besitzt.

Zur Korrektur des aufgenommenen "Roh"-Bildes R_{xy} werden zwei Kalibrationsbilder benötigt:

- *Dark-Field Image* D_{xy} . Das Bild D wird mit gleicher Belichtungszeit wie R aufgenommen, jedoch bei geschlossenem Shutter. Dieses Bild gibt Auskunft über das Untergrundsignal, d.h. das Eigenrauschen des Chips.
- *Flat-field Image* oder *Gain Reference* F_{xy} . Das Bild F ist die Mittelung mehrerer "Roh"-Bilder, die bei etwa den gleichen Beleuchtungsbedingungen wie bei der Aufnahme von R aufgenommen wurden, jedoch ohne eine Probe im Bildfeld. Diese Bilder enthalten die Struktur der Fiberoptik.

Mit Hilfe dieser beiden Kalibrationsbilder errechnet sich die korrigierte Bildintensität zu:

$$I_{xy} = \frac{R_{xy} - D_{xy}}{F_{xy} - D_{xy}}. \quad (4.1)$$

Die Akquisition der digitalen Bilder wurde in der vorliegenden Arbeit mit zwei unterschiedlichen Softwarepaketen durchgeführt. Am CM300 Mikroskop wurde das CCD Bild mit der DigitalMicrograph Software (Gatan) angesteuert und ausgelesen. Das Dark-Field D wurde automatisch nach jeder Aufnahme bestimmt, und das Flat-Field wurde vor Beginn jeder Datenaufzeichnung am TEM aufgenommen. Die CCD Kamera (TVIPS, Gauting, Deutschland) des CM20 wird mit EM-MENU ausgelesen (TVIPS). Beide Programme korrigierten die Rohdaten gemäß (4.1).

Die digitale Detektion von Elektronen bei 300kV ist mit den heute verfügbaren optisch gekoppelten Systemen nur mit relativ geringer lateraler Auflösung machbar (Fan und Ellisman 2000). Die Abbildungseigenschaften der am CM300 verwendeten Ultrascan Kamera wurde bereits von G. Schweikert charakterisiert (Schweikert 2004). Die laterale Auflösung der Kamera wird im Fourierraum durch die sog. *Modulation Transfer*

Function (MTF) beschrieben: Die GIF 2002 CCD Kamera übertrug noch etwa 19% des Signals bei Nyquist/2 und 2% bei Nyquist.

4.2 Kryo Präparation

Die Fixierung biologischer Proben durch rapides Abkühlen ist heutzutage die geeignetste Methode, um Strukturhaltung zu gewährleisten (Dubochet et al. 1988). Die Probe wird schnell genug abgekühlt um eine Temperatur unterhalb von -140°C zu erreichen, bevor sich Eiskristalle bilden können. Das Eis befindet sich dann in einem Glaszustand. Die Bildung und das Wachstum von Eiskristallen, die bei langsamer Abkühlung unterhalb von 0°C einsetzt, verursacht strukturelle Schäden der Probe. Ein weiterer Effekt ist die Dehydrierung der Probe: Da die Salzkonzentration im verbleibenden flüssigen Wasser während des Einfrierprozesses ansteigt, diffundieren Wassermoleküle aus dem biologischen Material. Insbesondere bei Proteinen kann die Dehydration zu Strukturveränderungen führen. Durch rasches Einfrieren oder "Vitrifizieren" der Probe werden alle Konstituenten immobilisiert, bevor es zu Rearrangements kommen kann. In diesem Sinne befinden sich eiseingebettete biologische Proben in einem quasi-physiologischen Zustand.

Proben bis zu einer Dicke von etwa $10\ \mu\text{m}$ können durch 'Einschießen' der Probe in flüssiges Ethan vitrifiziert werden. Die für die vorliegende Arbeit verwendeten Proben, hauptsächlich aufgereinigte Proteinlösung, Vesikel oder Viren, wurden folgendermaßen präpariert:

Ein EM Netzchen (Lacey Carbon Grid, Ted Pella, Kalifornien) wurde durch etwa halbminütige Behandlung im Plasma-Cleaner (Harrick) hydrophilisiert und anschließend in eine Pinzette eingespannt. Auf das Grid wurden etwa $3-5\ \mu\text{l}$ der Probenlösung mittels einer Pipette aufgebracht. Diesem Tropfen sind $3-5\ \mu\text{l}$ kolloidal gelöste Goldkügelchen (Durchmesser $10\ \text{nm}$) zugesetzt worden, um die Kippserie später alignieren zu können. Durch etwa dreiminütiges Warten wurde eine teilweise Durchmischung

der Lösungen erreicht, und anschließend wurde die Pinzette samt Grid in eine Einschußvorrichtung (Eigenkonstruktion, MPI) eingespannt. Mit Hilfe eines Filterpapiers wurde überschüssige Flüssigkeit abgesaugt (etwa 1-2 Sekunden lang) und unmittelbar anschließend wurde die Pizettenspitze samt Grid in flüssiges Ethan 'eingeschossen'. Die Kühlung des Ethans wurde durch ein Reservoir flüssigen Stickstoffs gewährleistet. Die beschriebene Prozedur lieferte i.A. vitrifizierte Filme von einer Dicke zwischen 0.1 und 1 μm .

4.3 Automatisierte Datenakquisition

Mechanische Ungenauigkeiten des Probenhalters verursachen eine Bewegung der Probe während des Kippvorgangs, die ausgeglichen werden muß. Ein manuelles Ausgleichen ist zu zeitaufwendig und setzt die Probe unnötig lange dem Elektronenstrahl aus, weshalb eine Automatisierung der Datenakquisition unumgänglich ist. Die Entwicklung solcher automatisierter Datenakquisitionsprozeduren ermöglicht es heute, Tomogramme mit niedrigsten Elektronendosen (*Low Dose*) aufzunehmen (Typke et al. 1991; Dierksen et al. 1992; Dierksen et al. 1993). Insbesondere die Kryo-ET Untersuchungen sind erst durch die Entwicklung solcher *Low Dose* Datenaufzeichnungsprogramme möglich (Dierksen et al. 1995; Grimm et al. 1996). Die automatisierte Datenakquisition erfordert drei wesentliche Schritte (Typke et al. 1991; Koster et al. 1992):

1. *Tracking*. Bei diesem Schritt wird ein Bild mit sehr niedriger Elektronendosis aufgenommen. Anhand dieses Bildes wird die Auswanderung im Vergleich zum vorhergehenden Bild ermittelt und mittels lateraler Verschiebung des Elektronenstrahls korrigiert.
2. *Auto-focus*. Dieser Schritt kompensiert die Fokusänderung die aufgrund der Probenbewegung in z auftritt durch eine Anpassung der Objektivlinse. Der Defokus wird durch den Vergleich zweier Bilder bei unterschiedlicher Kippung des Strahls ermittelt (Koster et al. 1989; Ziese et al. 2003).

3. *Exposure*. Nachdem der Elektronenstrahl im *Tracking* Schritt auf die Probe zentriert wurde und der gewünschte Defokus mittels der Autofokussierung eingestellt wurde, wird ein Bild der Probe aufgenommen. Diese Bilder konstituieren die Kippserie.

In der KET wird die Datenaufzeichnung ausschließlich im *Low Dose* Modus durchgeführt: Hier findet *Tracking* und *Auto-focus* an einer Position statt, die bzgl. des Probenorts entlang der Kippachse verschoben ist, um die am Probenort akkumulierte Elektronendosis ausschließlich für die Kippserie zu verwenden.

In der vorliegenden Arbeit wurde am CM300 das von Grimm (Grimm 1997; Grimm et al. 1997) in der Skriptsprache `Digital Micrograph` entwickelte Programmpaket zur Aufnahme von Kippserien verwendet. Dieses Programm ermöglichte neben den oben erwähnten Schritten auch eine Steuerung des Energiefilters während der Akquisition: Nach jeweils 10 Projektionen wurde der Energiefilter neu justiert, so daß die Abbildung im *Zero-Loss Mode* (Abschn. 3.3) gesichert wurde. Die Datenakquisition am CM20 erfolgte mit der `EM-Menu` Software (TVIPS).

4.4 Tomographische Rekonstruktion

Die Alignierung der Projektionen wird anhand von Goldmarkern, ähnlich wie in Ref. (Lawrence 1992) beschrieben, durchgeführt. Allerdings wurde das Modell (2.6) vereinfacht, indem angenommen wurde, daß die Vergrößerung konstant blieb, und innerhalb der Kippserie keine Bilddrehungen auftraten¹. Somit reduziert sich (2.6) zu:

$$E = \sum_i |[\mathcal{R}_\beta(\mathcal{P}_z(\mathcal{R}_{\alpha_i}(\mathbf{r}))) + (\Delta x, \Delta y)] - (x_i, y_i)|. \quad (4.2)$$

Die Orientierung der Kippachse β und die Verschiebungsvektoren $(\Delta x, \Delta y)$ wurden durch Lösen des resultierenden linearen Gleichungssystems bestimmt, wie in Ref. (Walz 1997) beschrieben. Die Alignierung wurde in `MATLAB` implementiert und

¹Die Bilddrehungen und Vergrößerungsänderungen sind in der Elektronentomographie vernachlässigbar. Hegerl und Typke, persönliche Mitteilung.

war Teil des TOM-Toolbox (Nickell et al. 2004).

Die Lokalisierung der Goldmarker in den Projektionen erfolgte manuell mittels einer graphischen Benutzeroberfläche, die Teil des TOM-Pakets war. Es wurden jeweils mindestens drei Goldmarker pro Bild lokalisiert, nach Möglichkeit wurden identische Goldmarker in den Projektionen ausgewählt. Als Hinweis auf die Qualität der Alignierung diente der in (4.2) definierte Fehler, der über alle Markerpunkte m summiert wurde. Zur Alignierung wurden möglichst weit voneinander entfernte Markerpunkte gewählt, und der mittlere Fehler betrug dann stets zwischen 1-3 Pixeln.

Vor der Rückprojektion wurde der Mittelwert jedes Bildes von diesem subtrahiert und anschließend wurde das mittelwertfreie Bild durch den zuvor bestimmten Mittelwert dividiert, um die Bilder auf gleiche Intensität zu normieren: Durch die unterschiedliche Belichtungszeit und unterschiedliche effektive Massendicke bei verschiedenen Kippwinkeln war die detektierte Elektronenmenge abhängig vom Kippwinkel, was durch die Normierung behoben wurde. Anschließend wurden die Bilder so rotiert, daß die y -Achse die Kippachse repräsentiert, und die bestimmte Translation wurde kompensiert. Rotation und Translation wurden in einem Schritt mittels bilinearer Interpolation durchgeführt.

Die alignierten Projektionen wurden nun gewichtet, wie in (Hoppe und Hegerl 1980; Harauz und van Heel 1986) beschrieben. In der vorliegenden Arbeit wurde i.A. analytische Gewichtung durchgeführt (siehe Abschn. 2.3). Die gewichteten Projektionen wurden anschließend in ein gemeinsames 3D-Volumen projiziert, wobei trilineare Interpolation verwendet wurde. Die Vorbereitung der Projektionen wurde stets im EM-System (Hegerl 1996) durchgeführt, die Rückprojektion wurde sowohl in EM als auch mit der TOM-Toolbox durchgeführt.

5. Lokalisierung von Makromolekülen in Tomogrammen

Ein Kryoelektronentomogramm einer Zelle enthält eine gewaltige Informationsmenge: Es beinhaltet ein dreidimensionales Abbild des gesamten zellulären Proteoms und damit einen 'Schnappschuß' des Netzwerks der makromolekularen Interaktionen. Um diese in einem Tomogramm enthaltene Information zugänglich zu machen, werden geeignete Analysemethoden benötigt, die primär bestimmte Makromoleküle spezifisch lokalisieren. Die Detektion von Makromolekülen wird im Wesentlichen durch zwei Eigenschaften von Kryoelektronentomogrammen erschwert: (i) Kryoelektronentomogramme besitzen ein sehr niedriges SNR. Es existieren zwar bildverarbeiterische Techniken zur Verstärkung des Signals, sog. *Denoising* Techniken, die das SNR erhöhen, was jedoch auf Kosten eines i.A. nichtlinear veränderten Signals erreicht werden kann und eine systematische Signalerkennung problematisch macht (Überblick in (Frangakis und Förster 2004)). (ii) Das Zytoplasma besitzt eine sehr hohe Konzentration an Makromolekülen, einzelne Komplexe berühren sich sogar. Diese sehr hohe Moleküldichte verringert die Signalqualität, da sie auf einer signifikanten Dichtedifferenz beruht.

Eine interaktive, d.h. letztlich manuelle Analyse der Bildinformation ist praktisch nur in Ausnahmefällen sinnvoll: Objektivität und Reproduzierbarkeit kann kaum erreicht werden, da die manuelle Analyse durch die Schwierigkeit, dreidimensionale Objekte in verschiedenen Orientierungen visuell als gleich zu erkennen, und die große Anzahl der Makromoleküle in einem Tomogramm erschwert wird.

Prinzipiell erscheint es auch denkbar, elektronenoptisch dichte Marker, z.B. goldmarkierte Antikörper, zu benutzen, um Molekülpositionen zu markieren. Diese wären

vergleichsweise einfach spezifisch zu detektieren. Ein solches Vorgehen hat jedoch mehrere Probleme: Das Einbringen der Marker in die Zelle erfordert das Öffnen der Zellmembran, d.h. es ist kaum nicht-invasiv zu bewerkstelligen. Es können effektiv nur höchstens zwei bis drei unterschiedliche Makromoleküle durch Gold verschiedener Größe markiert werden, und nur ein Teil der jeweiligen Makromoleküle wird durch den Marker markiert.

Aus diesem Grunde wird in der vorliegenden Arbeit ein anderer Ansatz zur quantitativen Auswertung der Tomogramme gewählt, nämlich die Lokalisierung der Makromoleküle anhand ihrer strukturellen Signatur. Ist ein strukturelles Muster (*Template*) des zu detektierenden Makromoleküls vorhanden, so kann dies zur systematischen Analyse der Daten verwendet werden. Als Muster kommen sowohl atomare Strukturen, wie sie in der *Protein Data Base* (PDB) hinterlegt sind als auch Kryo-EM Daten im mittleren Auflösungsbereich von 1-3 nm in Frage. Simulationen haben gezeigt, daß ein solches Vorgehen bei den derzeit in der KET erreichbaren Auflösung und SNR die Identifikation größerer (> 10 nm Durchmesser) Makromoleküle ermöglichen kann (Böhm et al. 2000). Ein geeigneter automatisierter Suchalgorithmus sollte vor allem hinreichend spezifisch sein, um ein zu detektierendes Makromolekül mit einer ausreichenden Sicherheit zu detektieren. Weiterhin muß die numerische Realisierung es erlauben, auch große Datensätze in überschaubaren Zeiträumen von Stunden oder Tagen zu analysieren.

Im Rahmen dieser Arbeit wurden korrelationsbasierte Algorithmen, sog. *Matched Filter* (VanderLugt 1964), zur Detektion von Makromolekülen in Kryoelektronentomogrammen anhand ihrer strukturellen Signatur entwickelt. Das Prinzip des Verfahrens ist in Abb. 5.1 skizziert. Das Tomogramm wird mit strukturellen Mustern potentiell vorhandener Makromoleküle gefiltert, was eine Korrelationsfunktion liefert, die die Ähnlichkeit eines Tomogrammelements zum Muster angibt. Die Auswertung dieser Funktion sollte dann idealerweise eine eindeutige Identifikation von bestimmten Makromolekülen ermöglichen. Die hier entwickelten Korrelationsfunktionen berücksichtigen insbesondere zwei Besonderheiten von Kryoelektronentomogrammen: die im

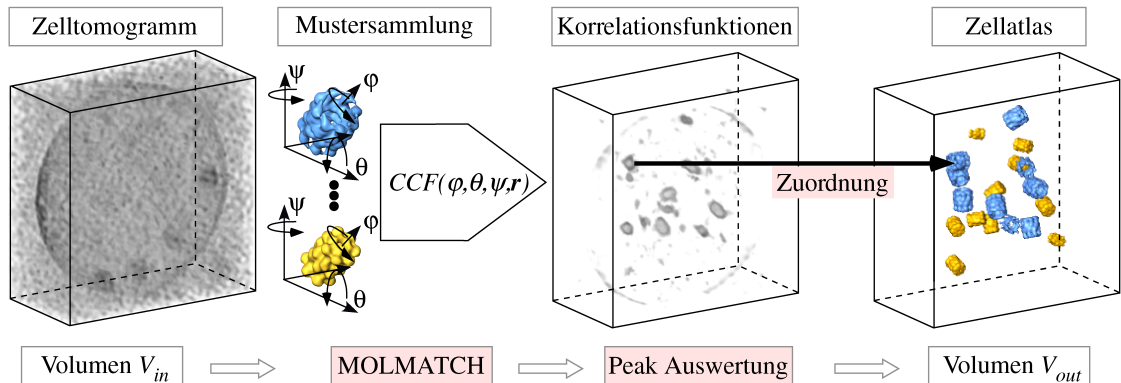


Abbildung 5.1: Schema zur entwickelten Strategie zur Identifikation von Makromolekülen. Die Komplexe werden anhand ihrer *a priori* bekannten und in einer Mustersammlung erfaßten Struktur detektiert. (Adaptiert aus (Frangakis et al. 2002))

Tomogramm variierenden Kontrastverhältnisse und den *Missing Wedge* Effekt (Abschn. 2.1).

Im ersten Abschnitt dieses Kapitels werden die entwickelte Korrelationsfunktion nach einer Einführung in die wichtigsten *Matched Filters* erklärt. Anschließend werden die Implementierungen als parallelisierte Programme erklärt. Ein weiterer Abschnitt behandelt die Generierung von Mustern aus röntgenkristallographischen Daten. Als Kontrolle der Algorithmen und um sie miteinander zu vergleichen, wurden sie zur Detektion von 20S Proteasomen in Phantomzellen, mit definierten Makromolekülen gefüllte Liposomen, benutzt. Das verlässlichste Verfahren wurde schließlich zur Identifikation von Ribosomen und Aktin in Tomogrammen ganzer Zellen angewandt.

5.1 Korrelationsbasierte Algorithmen zur Identifikation makromolekularer Komplexe in Kryoelektronentomogrammen

5.1.1 Korrelationsfunktionen

Die Korrelationsfunktion ist ein Maß für die Ähnlichkeit zweier Funktionen. Die einzelnen Funktionswerte der Korrelationsfunktion sind die Korrelationskoeffizienten (*Correlation Coefficient, CC*). Mathematisch beschreiben sie eine Projektion einer Funktion auf eine andere. Eine Dichteverteilung eines Tomogramms läßt sich als Vektor im Hilbertraum beschreiben, und der CC gibt die Projektion eines Vektors auf einen anderen an. Je nachdem, welche Norm zur Projektion verwendet wird, unterscheidet man unterschiedliche Korrelationsfunktionen, die im Folgenden kurz dargestellt werden sollen.

Gewöhnliche Korrelationsfunktion. Das Skalarprodukt im Hilbertraum ist als eine Integration des Produkts der Funktionen über den gesamten Raum definiert. Demnach gilt für den CC zweier Funktionen $I_{\mathbf{r}'}$ und $T_{\mathbf{r}'}$:

$$CC = \sum_{\mathbf{r}'} I_{\mathbf{r}'} \cdot T_{\mathbf{r}'}. \quad (5.1)$$

In der Bild- und Signalverarbeitung sind der Mittelwert und die Stärke des Signals i.A. willkürlich. Aus diesem Grund werden die Funktionen I und T meist normiert gemäß:

$$F = \frac{f - \bar{f}}{\sqrt{\sum_{\mathbf{r}'} (f_{\mathbf{r}'} - \bar{f})^2}}. \quad (5.2)$$

Hierbei bezeichnet $f_{\mathbf{r}'}$ die nicht normierte Funktion und \bar{f} den Mittelwert der Funktion f .

Bei der Korrelationsfunktion wird eine Translation der Funktion T bzgl. der Funktion I berücksichtigt. Sie wird häufig auch als Kreuzkorrelationsfunktion (*Cross-Correlation Function*, XCF) bezeichnet und ist definiert als eine Funktion des Translationsvektors \mathbf{r} :

$$XCF_{\mathbf{r}} = \sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot T_{\mathbf{r}'} = I_{\mathbf{r}} \odot T_{\mathbf{r}}. \quad (5.3)$$

Die Korrelationsfunktion läßt sich besonders effektiv im Fourierraum beschreiben, da das Faltungstheorem angewandt werden kann:

$$XCF_{\mathbf{r}} = I_{\mathbf{r}} \odot T_{\mathbf{r}} = \mathcal{F}^{-1}(\mathcal{F}(I) \cdot \mathcal{F}(T)^*). \quad (5.4)$$

Phase-only Filter. In der Literatur wurden diverse modifizierte Korrelationsfunktionen vorgeschlagen, die eine gegenüber der XCF (5.3) etwas veränderte Norm benutzen. Der Grund hierfür liegt darin, daß insbesondere bei EM-Bildern die niedrigen Ortsfrequenzen überproportional viel Gewicht besitzen. So wurde vorgeschlagen, die Amplituden der Funktionen durch eins zu ersetzen (Horner und Gianino 1964). Eine so modifizierte Korrelationsfunktion wird *Phase-only filter* (POF) genannt und wird berechnet als:

$$POF_{\mathbf{r}} = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(I)}{|\mathcal{F}(I)|} \cdot \frac{\mathcal{F}(T)^*}{|\mathcal{F}(T)|} \right). \quad (5.5)$$

Diese Funktion gewichtet jedoch hohe, i.A. rauschbehaftete Frequenzen sehr stark, weshalb sie in der Elektronenmikroskopie nur bei sehr hohen SNRs angewendet wird.

Mutual Correlation Function. Eine andere Funktion wurde von van Heel *et al* vorgeschlagen (Van Heel et al. 1992): Bei der sog. *Mutual Correlation Function*, MCF werden die Amplituden durch ihre Quadratwurzeln ersetzt.

$$MCF_{\mathbf{r}} = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(I)}{\sqrt{|\mathcal{F}(I)|}} \cdot \frac{\mathcal{F}(T)^*}{\sqrt{|\mathcal{F}(T)|}} \right). \quad (5.6)$$

Auch bei dieser Funktion werden die sehr energiereichen Frequenzen weniger stark gewichtet als bei der XCF, jedoch ist die Funktion deutlich weniger rauschempfindlich

als die POF. Beide Korrelationsfunktionen, POF und MCF, führen zu einem "Schärfen" der Korrelationsmaxima, d.h. ihre räumliche Ausdehnung wird kleiner, was bei *Single Particle* Anwendungen vorteilhaft sein kann (Van Heel et al. 1992). Allerdings ist die Ortstreue der Maxima sowohl bei der MCF als auch bei der POF geringer als bei der konventionellen XCF (Vijaya Kumar et al. 1992), was auch experimentell an EM Daten gezeigt wurde (Stoschek und Hegerl 1997). Im Hinblick auf ihre Detektionsspezifität wurden diese beiden Korrelationsfunktionen jedoch noch nicht untersucht.

Lokal normalisierte Korrelationsfunktion. Eine andere Korrelationsfunktion wurde vorgeschlagen, um kleine Muster in räumlich stärker ausgedehnten Umgebungen zu lokalisieren. Ein solches Problem liegt auch bei der Lokalisierung von Makromolekülen in Tomogrammen vor: So beinhaltet ein typisches Tomogramm eine laterale Fläche von etwa $1 \mu\text{m}$, während ein zu detektierendes Makromolekül etwa 10 nm im Durchmesser mißt. Ein analoges Problem liegt bei der Lokalisierung von Proteinen in Kryo-EM Dichten vor, und in diesem Zusammenhang wurde von Roseman eine lokale Normalisierung vorgeschlagen (Roseman 2000):

$$LCF_{\mathbf{r}} = \frac{\sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot T_{\mathbf{r}'}}{\sqrt{\sum_{\mathbf{r}'} \Theta_{\mathbf{r}'} \cdot (I_{\mathbf{r}+\mathbf{r}'} - \bar{I}_{\mathbf{r}})^2}}. \quad (5.7)$$

Bei dieser Korrelationsfunktion wird die Funktion I lokal in dem durch Θ definierten Gebiet normiert. Unabhängig davon wurde diese Korrelationsfunktion von Frangakis und Böhm entwickelt, um Makromoleküle in Tomogrammen zu identifizieren (Böhm 2001). Die Wahl der Region Θ sollte dabei zur äußeren Form des Musters korrelieren. Kann das Muster in verschiedenen Orientierungen vorliegen, so ist es vorteilhaft eine sphärische Maske zu wählen, wie dies in den erwähnten Applikationen (Böhm 2001; Frangakis et al. 2002) in der Tomographie geschehen ist. Der Nenner in Gl. 5.7 wurde in diesen Arbeiten explizit im Realraum berechnet, was jedoch rechnerisch sehr langwierig ist.

Korrelationsfunktionen für Muster in verschiedenen Orientierungen. Alle oben definierten Korrelationsfunktionen sind nicht rotationsinvariant gegenüber Rotationen der Referenz T . Demzufolge hängt die Korrelationsfunktion nicht nur von den Ortsvariablen \mathbf{r} sondern auch von der räumlichen Orientierung ab. Diese kann durch die Eulerwinkel φ , ψ , und ϑ beschrieben werden. In der vorliegenden Arbeit sind die Winkel folgendermaßen definiert: Das Objekt wird um ψ um die eigene z -Achse gedreht, um ϑ um die eigene x -Achse und um φ um die eigene z -Achse¹. So kann z.B. die XCF geschrieben werden als:

$$XCF_{\mathbf{r},\varphi\psi\vartheta} = \sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot \mathcal{R}_{\varphi\psi\vartheta}(T_{\mathbf{r}'}) = \sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot T_{\mathbf{r}',\varphi\psi\vartheta}. \quad (5.8)$$

Das Symbol $\mathcal{R}_{\varphi\psi\vartheta}$ bezeichnet dabei den Rotationsoperator, der um die Eulerwinkel $(\varphi, \psi, \vartheta)$ rotiert. Die Winkelabhängigkeit der Korrelationsfunktion ist mathematisch schwer effizient handhabbar. Existiert nur ein Rotationsfreiheitsgrad, z.B. der Winkel φ , so kann die Rotationsabhängigkeit sehr effektiv behandelt werden: Die Referenz kann nach polar harmonischen Funktionen entwickelt werden, die invariante Untergruppen unter Drehungen darstellen (Kunath und Sack-Kongehl 1989). Im allgemeinen Fall von Drehungen kann jedoch keine solche irreduzible Darstellung gefunden werden, die Untergruppen mit nur einem Element bilden.

Lokal normalisierte Korrelationsfunktionen für Muster in verschiedenen Orientierungen. Die LCF (5.7) läßt sich für orientierungsabhängige Muster schreiben als:

$$LCF_{\mathbf{r},\varphi\psi\vartheta} = \frac{\sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot T_{\mathbf{r}',\varphi\psi\vartheta} \cdot \Theta_{\mathbf{r}',\varphi\psi\vartheta}}{\sqrt{\sum_{\mathbf{r}'} \Theta_{\mathbf{r}',\varphi\psi\vartheta} \cdot (I_{\mathbf{r}+\mathbf{r}'} - \bar{I}_{\mathbf{r},\varphi\psi\vartheta})^2}}. \quad (5.9)$$

Allerdings ist die Berechnung des Nenners der LCF im Realraum, wie dies in Ref. (Böhm et al. 2001; Frangakis et al. 2002) geschah, rechnerisch sehr zeitintensiv ($t \sim$ Wochen),

¹In der Literatur und einigen Bildverarbeitungsprogrammen wird die Rotation ϑ gelegentlich um die y -Achse ausgeführt, was häufig zu Verwirrung führt. Weiterhin weicht die Nomenklatur der Winkel in den in der 3D Elektronenmikroskopie verbreiteten Programme i.A. von der aus dem EM Programm entnommenen ab: Bei Programmen wie `spider` oder `EMAN` ist die Benennung von ψ und ϕ umgekehrt.

was die Berechnung von (5.9) nicht praktikabel macht. In den meisten Fällen sind die zu lokalisierenden Makromoleküle zytosolische Proteinkomplexe. Diese haben i.A. eine globuläre Form, weshalb die Maske kugelförmig gewählt werden kann. Die rotationsymmetrische Maske Θ macht den Nenner in (5.9) rotationsinvariant, so daß er im Realraum in überschaubaren Zeiträumen (\sim Tage) berechnet werden konnte. Die LCF lautet dann:

$$LCF_{\mathbf{r}\varphi\psi\vartheta} = \frac{\sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot T_{\mathbf{r}',\varphi\psi\vartheta} \cdot \Theta_{\mathbf{r}'}}{\sqrt{\sum_{\mathbf{r}'} \Theta_{\mathbf{r}} \cdot (I_{\mathbf{r}+\mathbf{r}'} - \bar{I}_{\mathbf{r},\varphi\psi\vartheta})^2}}. \quad (5.10)$$

Die Verwendung der LCF (5.10) zeigte bei den Arbeiten von Böhm eine Verbesserung der Detektionsqualität gegenüber der XCF (5.8) (Böhm 2001); es wurde berichtet, daß ohne lokale Normalisierung anstatt von Makromolekülen häufiger Objekte mit starkem Signal wie Goldkugeln oder Membranen detektiert wurden.

5.1.2 Weiterentwickelte Korrelationsfunktionen für die KET

In der vorliegenden Arbeit wurden neue Korrelationsfunktionen zur Erkennung von Makromolekülen in Kryoelektronentomogrammen entwickelt. Diese beinhalten zum einen den *Missing Wedge* Effekt (Abschn. 2.1). Zum anderen wurde die Berechnung der Funktion möglichst effektiv gestaltet, um den Algorithmus bei der Interpretation von Tomogrammen routinemäßig einsetzen zu können.

Eingeschränkte Korrelationsfunktionen. Alle bisher diskutierten Korrelationsfunktionen berücksichtigen in keiner Weise den *Missing Wedge*. Die hier entwickelte Korrelationsfunktion beschränkt die Korrelation auf den experimentell zugänglichen Winkelbereich. Dies wird erreicht durch Faltung der Referenz T mit der experimentellen *Pointspread function* (PSF). Die PSF beschreibt den fehlenden Frequenzbereich im Realraum in Abhängigkeit vom zugänglichen Kippwinkelbereich $\alpha_{min}-\alpha_{max}$:

$$PSF = \mathcal{F}^{-1} [(\mathcal{F}\delta(\mathbf{r})) \cdot W_{\alpha_{min}\alpha_{max}}]. \quad (5.11)$$

Dabei ist W die Keilfunktion im Fourierraum, die einen Wert von 1 innerhalb des experimentell zugänglichen Winkelbereichs besitzt und 0 außerhalb. Die PSF ist für

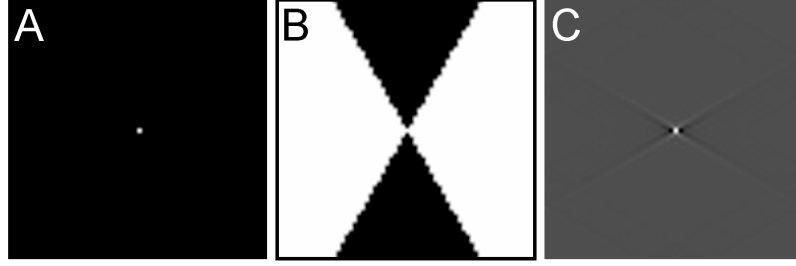


Abbildung 5.2: Konzept der PSF in 2D. Eine δ -Funktion (A) wird aufgrund des nicht zugänglichen Winkelbereichs im Fourierraum (B: hier $\alpha_{min} = -60^\circ$ und $\alpha_{max} = +60^\circ$) bei der Tomographie auf einen elongierten Punkt abgebildet, was durch die PSF (C) beschrieben wird. In allen Bildern korrespondiert weiß zu 1 und schwarz zu 0.

den zweidimensionalen Fall in Abb. 5.2 illustriert.

Unter Zuhilfenahme der PSF kann die zu Gl. 5.3 korrespondierende *global* normalisierte, eingeschränkte CCF geschrieben werden als:

$$CCF_{\mathbf{r}\varphi\psi\vartheta} = \frac{\sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot (T_{\mathbf{r}',\varphi\psi\vartheta} \otimes PSF_{\mathbf{r}'})}{\sqrt{\sum_{\mathbf{r}} (T_{\mathbf{r},\varphi\psi\vartheta} \otimes PSF_{\mathbf{r}})^2} \cdot \sqrt{\sum_{\mathbf{r}} (I_{\mathbf{r}})^2}}. \quad (5.12)$$

Entscheidend ist hierbei vor allem die Normalisierung der Referenz T , d.h. des ersten Terms im Nenner. Im Zähler beinhaltet das Tomogramm I bereits den *Missing Wedge*, weshalb hier die Faltung strenggenommen nicht nötig ist. Die CCF kann wiederum analog zu (5.7) bei Verwendung einer rotationssymmetrischen Maske Θ lokal normalisiert werden:

$$LCCF_{\mathbf{r}\varphi\psi\vartheta} = \frac{\sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot (T_{\mathbf{r}',\varphi\psi\vartheta} \otimes PSF_{\mathbf{r}'} \cdot \Theta_{\mathbf{r}'})}{\sqrt{\sum_{\mathbf{r}} (T_{\mathbf{r},\varphi\psi\vartheta} \otimes PSF_{\mathbf{r}} \cdot \Theta_{\mathbf{r}})^2} \sqrt{\sum_{\mathbf{r}'} \Theta_{\mathbf{r}'} \cdot (I_{\mathbf{r}+\mathbf{r}'} - \bar{I}_{\mathbf{r}})^2}}. \quad (5.13)$$

Berechnet man die LCCF für anisotrope Masken Θ so ergibt sich:

$$LCCF_{\mathbf{r}\varphi\psi\vartheta} = \frac{\sum_{\mathbf{r}'} I_{\mathbf{r}+\mathbf{r}'} \cdot (T_{\mathbf{r}',\varphi\psi\vartheta} \otimes PSF_{\mathbf{r}'} \cdot \Theta_{\mathbf{r}',\varphi\psi\vartheta})}{\sqrt{\sum_{\mathbf{r}} (T_{\mathbf{r},\varphi\psi\vartheta} \otimes PSF_{\mathbf{r}} \cdot \Theta_{\mathbf{r},\varphi\psi\vartheta})^2} \sqrt{\sum_{\mathbf{r}'} \Theta_{\mathbf{r}',\varphi\psi\vartheta} \cdot (I_{\mathbf{r}+\mathbf{r}'} - \bar{I}_{\mathbf{r}})^2}}. \quad (5.14)$$

In gleicher Weise kann auch eine lokal normierte, eingeschränkte MCF definiert werden, eine LMCF: Vor der Berechnung der Korrelation müssen die Amplituden von I und T lediglich durch ihre Quadratwurzeln ersetzt werden.

Effiziente Berechnung der lokalen Varianz. Roseman schlug als erster die Berechnung der Norm von (5.7) bzw. (5.13) und (5.14) im Fourierraum vor (Roseman 2003):

$$\sum_{\mathbf{r}'} \Theta_{\mathbf{r}'} \cdot (I_{\mathbf{r}+\mathbf{r}'} - \bar{I}_{\mathbf{r}})^2 = \mathcal{F}^{-1} (\mathcal{F}(I^2) \cdot \mathcal{F}(\Theta)) - N [\mathcal{F}^{-1} (\mathcal{F}(I) \cdot \mathcal{F}(\Theta))]^2. \quad (5.15)$$

In obiger Gleichung bezeichnet N die Zahl der in der Maske Θ enthaltenen Voxel. Die Berechnung der Norm im Fourierraum beschleunigt die Berechnung der Korrelationsfunktion (5.7) um ein Vielfaches, weshalb diese Berechnung der Norm in der vorliegenden Arbeit implementiert wurde. Auf diese Weise ist es möglich, lokal normalisierte Korrelationsfunktionen mit beliebigen Masken Θ in überschaubarer Zeit zu berechnen.

5.2 Erzeugung struktureller Signaturen aus atomaren Strukturen

Um Strukturdaten, die mit Hilfe anderer Abbildungstechniken erhalten wurden, mit elektronenmikroskopischen Daten zu korrelieren, muß deren elektronentomographische Abbildung approximiert werden. In Abschn. 3.2 wurde die Bildentstehung im Elektronenmikroskop erklärt. Demnach ist im Frequenzbereich oberhalb von etwa $(10 \text{ nm})^{-1}$ Phasen- und Amplitudenkontrast dominant, während unterhalb dieses Frequenzbereichs insbesondere bei energiegefilterter Elektronenmikroskopie verstärkt Streuabsorptionskontrast auftritt. Als Abbildungsmechanismus wird deshalb ein linearer Prozeß angenommen, der vor allem auf Phasenkontrast und einem 12%igen Anteil an Amplitudenkontrast basiert (Abschn. 3.2).

Die meisten Strukturen biologischer Makromoleküle sind mit Hilfe der Röntgenkristallographie aufgeklärt worden. Die erhaltenen Daten sind hier die Koordinaten der das Molekül konstituierenden Atome, die mit einer Auflösung im Ångströmbereich bestimmt werden. Weitere experimentelle Techniken zur Bestimmung atomarer Strukturen sind NMR, Elektronenkristallographie und *Single Particle* Rekonstruktionen

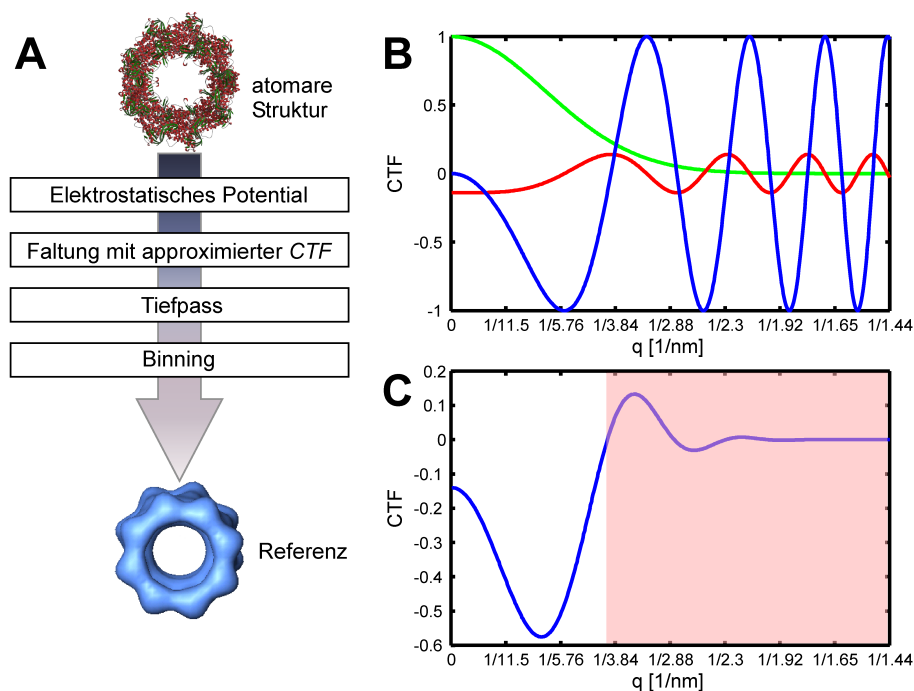


Abbildung 5.3: A: Mustererzeugung für ET anhand atomarer Strukturdaten. B: Funktionen, aus denen die effektive CTF zusammengesetzt ist: Phasenkontrasttransferfunktion (blau), Amplitudenkontrasttransferfunktion (rot) und MTF (grün). C: Effektive CTF. Oberhalb der ersten Nullstelle (rot schattiert) wird die Funktion abgeschnitten.

(meist durch *Fitting* bekannter Proteinstrukturen in größere Komplexe). In zunehmenden Maße gewinnen auch rechnerisch abgeleitete Strukturen an Bedeutung: Vor allem mit Hilfe von *Homology modelling* können in einigen Fällen Modelle von ähnlicher Qualität wie mit der Röntgenkristallographie erhalten werden.

Alle erwähnten Techniken liefern atomar aufgelöste Strukturen, die i.A. in der *Protein Data Bank* (PDB) (www.rcsb.org) archiviert sind. In Abb. 5.3A ist die Methode skizziert, mit der die 3D-EM Dichte bei einer Pixelgröße p approximiert wurde. Es wird zuerst aus den atomaren Koordinaten ein elektrostatisches Potential approximiert, indem die jeweiligen Ordnungszahlen Z auf den Gitterpunkten eines Gitters

mit der Pixelgröße p aufsummiert werden, was der Näherung (3.12) entspricht². Diese Näherung ist für den Auflösungsbereich der KET ausreichend, d.h. es müssen nicht die exakten atomaren Formfaktoren berücksichtigt werden. Das so erhaltene Volumen wird mit einer approximierten CTF gefaltet. Diese CTF soll folgenden Faktoren Rechnung tragen: Phasen- und Amplitudenkontrast sowie die MTF der CCD-Kamera. Die einzelnen Bestandteile der effektiven CTF sind in Abb. 5.3B skizziert: Die CTF besteht aus einem reinen Phasenanteil, einem 14 %igen Amplitudenanteil und einer Dämpfungsfunktion, mit der beide Funktionen multipliziert werden müssen. Die sich so ergebende effektive CTF wird oberhalb der ersten Nullstelle tiefpaßgefiltert (Abb. 5.3C), da der Defokus typischerweise nur unzureichend genau bestimmt ist.

5.3 Parallelisierte Implementierung – MOLMATCH und OMNIMATCH

Die Berechnung der Korrelationsfunktion basierte auf einem von S.Nickell entwickelten parallelisierten Programm (OSCAR³), das die nicht-normalisierte XCF aus Gl. (5.3) berechnete (Nickell 2001). Das Rotationsproblem wurde bei OSCAR durch explizite Abtastung der XCF gelöst: Die Referenz wurde in einem vorgegebenen Intervall der Eulerwinkel mit einem konstanten Inkrement jeweils rotiert, bevor die korrespondierende Korrelationsfunktion berechnet wurde. Das Speichern der vollständigen, 6-dimensionalen, Korrelationsfunktion überstieg die Kapazität heutiger Rechner, da die Datenmenge für ein typisches Tomogramm hunderte Giga- oder gar Terabytes betrug. Aus diesem Grund wurde bereits in OSCAR lediglich der maximale Korrelationskoeffizient und der zugehörige Winkel an jedem Punkt \mathbf{r} gespeichert. Die Parallelisierung wurde mittels der MPI Library (*Message Passing Interface* <http://www.lam-mpi.org/>)

²Um *Aliasing* Effekte zu verhindern, wurde die Struktur erst mit der Pixelgröße $p/2$ gerastert, bei der halben Nyquist tiefpaßgefiltert und anschließend einfach gebinnt.

³OSCAR steht dabei für *Optimized SCanning Algorithm*. Die Optimierung bestand darin, daß die XCF im Fourierraum mit Hilfe der FFTW-Library (*Fastest Fourier Transformation in the West*, <http://www.fftw.org/>) berechnet wird, die die Fouriertransformation besonders schnell bewältigt.

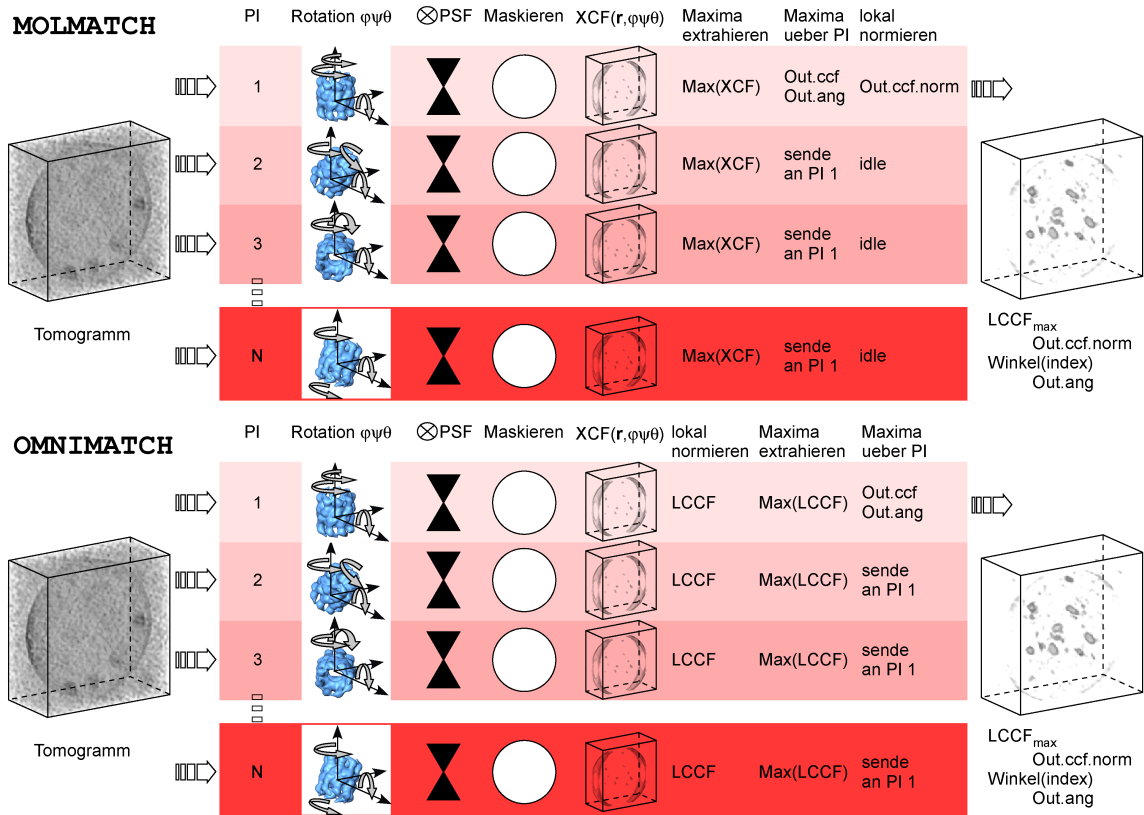


Abbildung 5.4: Oben: Parallelisierte Berechnung der lokal normalisierten Korrelationsfunktion mit MOLMATCH: Tomogramm und Referenz wurden von den benutzten Prozessoren eingelesen. Der abzutastende Winkelbereich der LCCF wurde gleichmäßig auf die N Prozessoren PI aufgeteilt. Auf jedem Rechner wurde die Referenz mittels trilinearere Interpolation in die verschiedenen Orientierungen rotiert. Jeder Orientierung wurde dabei ein spezifischer Index zugewiesen, der dem Schleifenindex entsprach (Schleifenreihenfolge: 1. φ , 2. ψ , 3. ϑ). Die rotierte Referenz wurde mit der PSF gefaltet und anschließend mit der rotationssymmetrischen Maske multipliziert. Die so präparierte Referenz wurde normiert und anschließend die XCF berechnet. Für jeden Raumpunkt wurden die Korrelationswerte mit dem des vorhergehenden Winkels verglichen und der jeweils größere Wert wurde für den nächsten Winkel übernommen. Der Schleifenindex der korrespondierenden Eulerwinkel wurde ebenfalls gespeichert. Da die lokale Norm des Tomogramms für rotationsinvariante Masken Θ nicht von der Orientierung abhing, wurde die lokale Norm noch nicht benötigt. Nachdem alle Prozessoren den ihnen zugewiesenen Winkelbereich bearbeitet hatten, wurden wiederum die Korrelationsmaxima aus den XCFs der jeweiligen Prozessoren und die korrespondierende Rotationsindizes extrahiert. Der letzte Schritt bestand in der lokalen Normalisierung, die nur der erste Prozessor ausführte. Die von MOLMATCH erzeugten Dateien "Out.ccf", "Out.ccf.norm" und "Out.ang" enthielten die Maxima der XCF, der LCCF und die korrespondierenden Laufindizes der Eulerwinkel. Unten: Berechnung der LCCF für anisotrope Masken mit OMNIMATCH: Die Berechnung erfolgte prinzipiell ähnlich wie mit MOLMATCH, jedoch wurde jede XCF zuerst lokal normalisiert bevor die Maxima extrahiert wurden. Da das Speichern der XCF hier zusätzlichen Rechenaufwand verursacht hätte, wurde nur die LCCF als "Out.ccf" gespeichert.

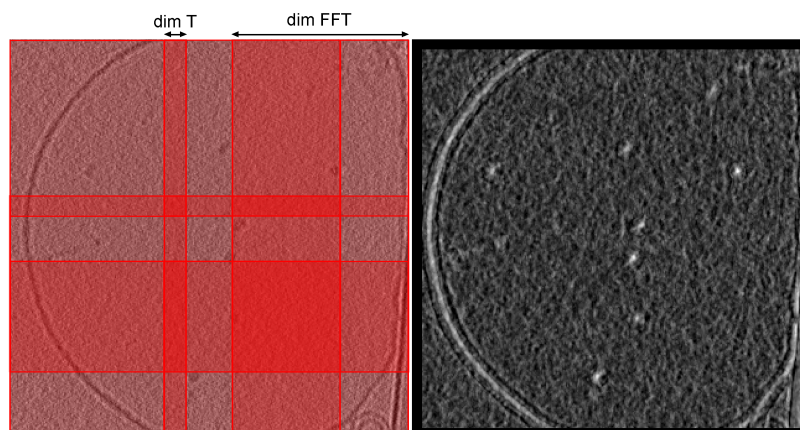


Abbildung 5.5: Links: Aufteilung der Berechnung der Korrelationsfunktion als 2D Darstellung. Die Korrelationsfunktion wurde in Würfeln der Seitenlänge $dim\ FFT$ berechnet. Dabei mußten die durch weniger Transparenz gekennzeichneten Regionen häufiger berechnet werden: Um zu verhindern, daß Moleküle an den Schnittträgern effektiv zerteilt wurden, mußte an den Grenzbereichen ein Streifen einer Breite die gleich der Kantenlänge der Referenz $dim\ T$ doppelt und an einigen Stellen sogar dreifach berechnet werden. Rechts: Die resultierende Korrelationsfunktion besaß keinerlei Stufen an den Schnittstellen.

implementiert. Die Parallelisierung erfolgte hierbei über die abzutastenden Eulerwinkel (Abb. 5.4).

Ein modifiziertes C-Programm, *MOLMATCH* (Abb. 5.4), berücksichtigte die in Abschn. 5.1.2 erläuterten Weiterentwicklungen der LCCF gegenüber der XCF speziell für sphärische Masken gemäß Gl. (5.13). Die Berechnung der LCCF für anisotrope Masken (5.14) war rechnerisch langwieriger: Da neben der Faltung des Tomogramms mit der Referenz auch das Tomogramm sowie dessen Quadrat für jede berechnete Orientierung mit der Maske gefaltet werden mußte, verdreifachte sich der Rechenaufwand etwa. Aus diesem Grunde wurde die Berechnung in einem separaten Programm implementiert, dem Programm *OMNIMATCH* (Abb. 5.4 unten).

Ein weiteres Problem bei der Berechnung der LCCF war der im Vergleich zur Datenmenge geringe Arbeitsspeicher selbst leistungsfähiger Computer. Tomogramme wurden auf $2k \times 2k$ CCD-Kameras aufgezeichnet. Selbst nach zweimaligen Binnens hatten die Tomogramme dann typischerweise eine Größe von $512 \times 512 \times 256$ Voxeln, die

jeweils als Fließkommazahl (*float*) gespeichert wurden, so daß der Speicherbedarf 256 Mb betrug. Bei der Berechnung der LCCF mußten insgesamt 4 Volumina dieser Größe gleichzeitig im Speicher vorhanden sein, so daß der gesamte Speicherbedarf mehr als 1 Gb betrug. Dieser Speicherbedarf überstieg den verfügbaren Arbeitsspeicher vieler Rechner, und zusätzlich waren bei den überwiegend verwendeten 32bit Rechnern nicht mehr als 1 Gb ansprechbar. Aus diesem Grunde berechneten MOLMATCH und OMNIMATCH die LCCF jeweils für kleinere Subvolumina. Diese einzelnen Teile der LCCF wurden anschließend zur gesamten LCCF zusammengesetzt. Es mußten jedoch die Ränder der einzelnen Unter-LCCFs doppelt berechnet werden, wobei das doppelt berechnete Volumen die Breite des Musters haben mußte (Abb. 5.5).

5.4 Partikelidentifikation in Phantomzellen

Um zu überprüfen, ob der in den vorangegangenen Abschnitten erklärte korrelationsbasierte Algorithmus in der Lage ist, Komplexe innerhalb von Kryoelektronentomogrammen ganzer Zellen spezifisch zu detektieren, muß seine Leistungsfähigkeit zunächst an Tomogrammen wohldefinierter Modellsysteme überprüft werden. Ein System, das in Struktur und Größe kleinen prokaryotischen Zellen ähnelt aber chemisch definiert ist, stellt eine Phantomzelle dar. Dies sind mit spezifischen Proteinkomplexen gefüllte Liposomen.

Die Proteinkomplexe Proteasom und Thermosom sind sehr große Makromoleküle (700 bzw. 900 kDa), die zu den größten bekannten Komplexen im Archäon *Thermoplasma acidophilum* gehören, ihre biochemische Aufreinigung ist etabliert und ihre atomare Struktur ist mittels Röntgenkristallographie bzw. *Single Particle* Analyse aufgeklärt (Lowe et al. 1995; Nitsch et al. 1998). Die Unterscheidung der beiden strukturell ähnlichen (Abb. 5.6) Komplexe in Phantomzellen stellt demnach ein ausgezeichnetes Modellsystem dar, um die Partikellokalisierung anhand ihrer strukturellen Signatur zu testen (Böhm et al. 2001). Unter Benutzung einer vorläufigen Version von MOLMATCH konnte gezeigt werden, daß es möglich ist, Proteasomen und Thermosomen spezifisch

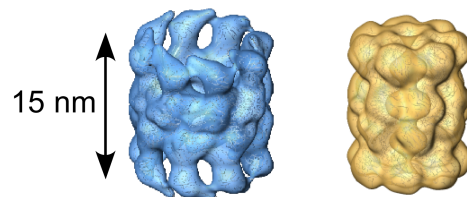


Abbildung 5.6: Links: Thermosom in offener Konformation. Es hat eine zylindrische Struktur mit 8-facher Rotationssymmetrie um die Zylinderachse. Die Dichte ist als Oberflächendarstellung mit etwa 1.4 nm Auflösung abgebildet. Rechts: 20S Proteasom (Zylinder mit 7-facher Rotationssymmetrie) in gleicher Darstellung. Der Maßstab korrespondiert zu 10 nm.

in Phantomzellen zu detektieren (Frangakis et al. 2002).

In der vorliegenden Arbeit wurde ein rigoroser Vergleich verschiedener Korrelationsfunktionen hinsichtlich ihrer Detektionsqualität vorgenommen. Der Vergleich wurde vorgenommen anhand der Detektion von 20S Proteasomen in Phantomzellen aus Ref. (Böhm 2001; Frangakis et al. 2002), die mit 20S Proteasomen gefüllt waren. Der *a priori* bekannte Aufbau der Phantomzellen erlaubte eine manuelle Klassifikation korrekter und offensichtlich falscher Detektionen. Die Korrelationsfunktionen CCF (5.12), MCF (5.6), LCCF (5.13) und LMCF wurden anhand der Häufigkeit falscher Detektionen bewertet.

5.4.1 Material und Methoden

Herstellung der Phantomzellen und Tomographie. Das entwickelte Detektionsverfahren wurde auf ein von Böhm (Böhm 2001; Frangakis et al. 2002) aufgenommenes Kryoelektronentomogramm von proteasomengefüllten Phantomzellen angewandt. Zur genauen Präparation und Datenakquisition sei auf die genannten Arbeiten verwiesen, sie sollen hier nur knapp wiedergegeben werden: Die Phantomzellen wurden hergestellt aus reinen Liposomen, in die mittels der sog. *Freeze-and-Thaw* Technik (Monnard et al. 1999) Proteasomen eingebracht wurden. Die rekombinant

hergestellten Proteinkomplexe waren mit einem "His-Tag" versehen, das benutzt wurde, um nicht in die Vesikel hineindiffundierte Komplexe mittels einer Ni-Säule aus der umgebenden Lösung zu entfernen. Die Kippserien der vitrifizierten Phantomzellen wurden mit 1.5° Kippwinkelinkrement von -70° bis $+70^\circ$ aufgenommen. Der Defokus betrug $8 \mu\text{m}$ (erste Nullstelle der CTF bei etwa $(4 \text{ nm})^{-1}$), und bei der verwendeten Vergrößerung von etwa 43k entsprach ein Pixel 0.72 nm^2 . Die 3D-Rekonstruktion erfolgte mit k -gewichteter Rückprojektion im EM-System.

Mustererkennung. Die Rekonstruktion wurde einfach gebinnt (Voxelgröße 1.44 nm^3), bevor die Mustererkennung angewandt wurde, da das gesamte Signal bis zur ersten Nullstelle der CTF durch die Datenreduktion nicht beeinträchtigt wurde. Als Referenz für das 20S Proteasom wurde die in Ref. (Lowe et al. 1995) publizierte Struktur verwendet. Für das Thermosom in seiner offenen Konformation wurde die Struktur aus (Nitsch et al. 1998) verwendet. Die EM-Dichte wurde erzeugt wie in Abschn. 5.2 beschrieben.

Das Tomogramm wurden mit beiden erzeugten Mustern auf Proteasomen und Thermosomen mit MOLMATCH abgetastet, was die CCF und LCCF lieferte. Zusätzlich wurde die MCF und LMCF von Volumen und Proteasomen mittels einer hierfür modifizierten Version von MOLMATCH berechnet. Bei allen Berechnungen wurden die Referenzen mit 10° Winkelinkrement abgetastet. Aufgrund der 7- bzw. 8-fachen Rotationssymmetrie entlang der Zylinderachse war es ausreichend, φ von Null bis $\phi_{\max} = 50^\circ$ abzutasten. Die restlichen Winkel wurden komplett abgetastet, d.h. $\psi_{\max} = 350^\circ$ und $\vartheta_{\max} = 180^\circ$.

Die berechneten Korrelationsvolumen wurden in MATLAB ausgewertet: Die Maxima wurden bestimmt und die Korrelationswerte um jedes Maximum herum innerhalb eines Radius' von 15 Voxeln Null gesetzt, was der Ausdehnung eines Partikels entsprach. Die Koordinaten und Orientierungen der potentiellen Partikel wurden in einer Motivliste gespeichert. Anhand dieser Motivliste wurden die Partikel und ihre Lokalisierung im Tomogramm als xy -Schnitt dargestellt. Detektionen im Innern der Phantomzellen wurden als korrekt angenommen, Lokalisierungen innerhalb der Lipidmembran oder

außerhalb des Vesikels wurden als Falsch-Positive angenommen, da die biochemische Präparation Proteinkomplexe außerhalb der Vesikel entfernte. Aus der Höhe der Korrelationskoeffizienten wurden Histogramme gebildet, wobei die Korrelationswerte insgesamt 25 äquidistante Werte zusammengefaßt wurden.

5.4.2 Ergebnisse

Detektionsergebnisse in Abhängigkeit von der Korrelationsfunktion. Die berechneten Korrelationsfunktionen sind in Abb. 5.7 als exemplarische xy -Schnitte dargestellt. Die verschiedenen Korrelationsfunktionen (CCF, MCF, LCCF und LMCF) von Tomogramm und 20S Proteasom (Abb. 5.7a-d) zeigen alle einen qualitativ ähnlichen Verlauf: Alle Funktionen zeigen Maxima an den Positionen der Partikel (siehe zum Vergleich das Tomogram Abb. 5.7f), jedoch auch an der Vesikelwand. Die nicht-lokal normalisierten Korrelationsfunktionen CCF und MCF scheinen dabei ein höheres "Signal" im folgenden Sinne zu besitzen: Die Maxima erscheinen sehr hell im Vergleich zum Hintergrund. Die lokal normalisierten Funktionen LCCF und LMCF (Abb. 5.7c-d) sind vergleichsweise verrauscht, d.h. das Hintergrundsignal ist stärker. Diese Eigenschaft ist Resultat der lokalen Normierung: Während z.B. bei der konventionellen CCF N_{Tom} Voxel in die Korrelationsfunktion eingehen, wobei N_{Tom} die Zahl der Voxel des gesamten Tomogramms ist, gehen z.B. bei der LCCF nur N_{Ref} Voxel (Voxel der Maske Θ) ein. Der Rausch- oder Signifikanzlevel ist $\sim 1/\sqrt{N}$, so daß die lokal normalisierte LCCF zwangsläufig verrauschter erscheint. Zum Vergleich ist in Abb. 5.7e die LCCF für eine "falsche" Referenz, hier das Thermosom gezeigt. Sie zeigt ein qualitativ ähnliches Verhalten, jedoch sind die Maxima der Partikel weniger ausgeprägt.

Um die Qualität der Detektion zu beurteilen, ist es jedoch wichtig zu bestimmen, inwieweit die Korrelationsmaxima spezifisch sind, d.h. in welchem Ausmaß es zu Falsch-Positiven kommt. Hierzu wurden die Falsch-Positiven innerhalb der ersten 160 potentiellen Partikel bestimmt. Die Ergebnisse wurden als Histogramm ausgewertet (Abb. 5.8). Es ergeben sich gravierende Unterschiede der Detektionsqualität der

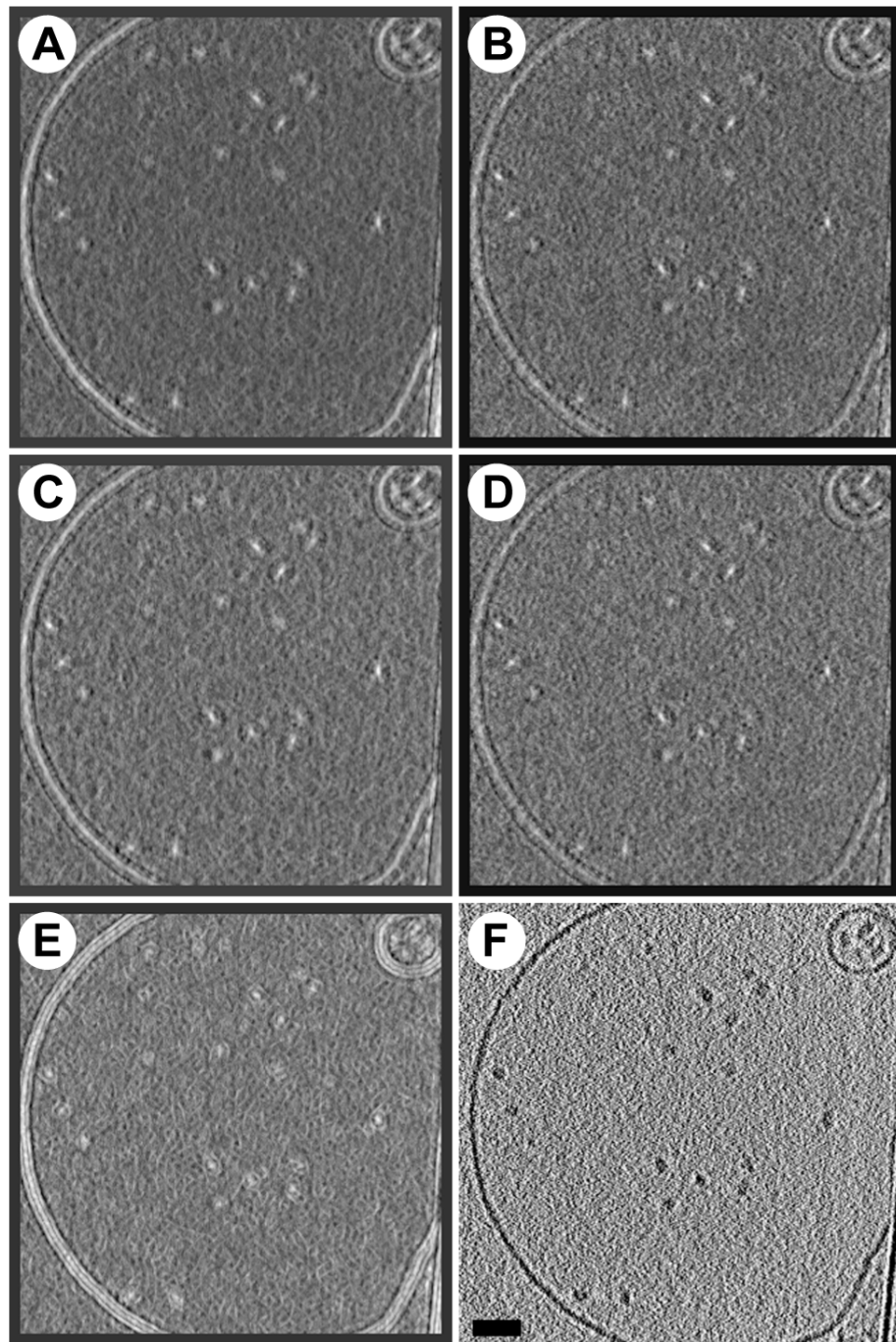


Abbildung 5.7: Verschiedene Korrelationsfunktionen. A: CCF von Tomogram und Proteasom. B: MCF mit Proteasom als Muster. C: LCCF mit Proteasom. D: LMCF mit Proteasom. E: LCCF mit Thermosom. F: Korrespondierender xy -Schnitt des Tomogramms. Maßstab: 50 nm.

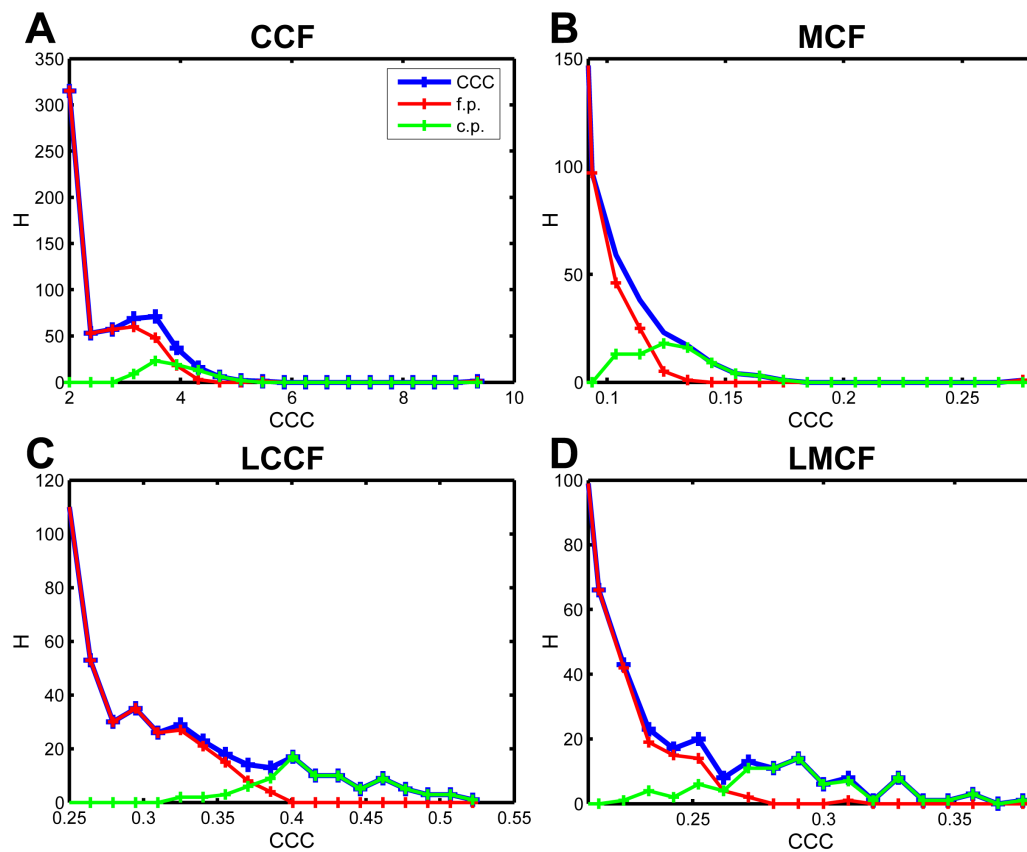


Abbildung 5.8: Histogramme der erhaltenen Korrelationskoeffizienten (CCCs) bei Verwendung unterschiedlicher Korrelationsfunktionen. Es sind Häufigkeiten (H) gegen die korrespondierenden Korrelationswerte (CCC) aufgetragen, sowie eine Unterteilung in falsch-positive (f.p.) und korrekt-positive (c.p.) Detektionen. Ausgewertet wurden CCF (A), MCF (B), LCCF (C) und LMCF (D). Die Absolutwerte von CCF und MCF sind willkürlich, da auf eine Normalisierung verzichtet wurde.

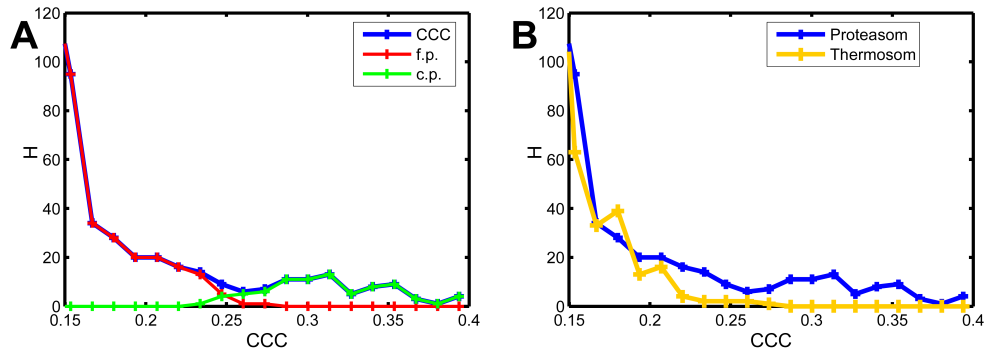


Abbildung 5.9: A: Detektionsergebnisse bei Hochpaßfilterung LCCF bei einer Frequenz von etwa $(20 \text{ nm})^{-1}$. B: Histogramme der erhaltenen Korrelationskoeffizienten (CCCs) bei Verwendung des 20S Proteasoms und des Thermosoms als Referenzen.

unterschiedlichen Korrelationsfunktionen: Die nicht-lokal normalisierten Korrelationsfunktionen (Abb. 5.8 A und B) führen offenbar zu einem inakzeptabel hohem Maß an Falsch-Positiven. Weder die Histogramme der CCF noch der MCF ermöglichen es nur anhand des Korrelationskoeffizienten falsche von korrekt positiven Detektionen zu unterscheiden. Die Detektionsergebnisse der lokal normalisierten Funktionen (Abb. 5.8 C und D) zeigen ein anderes Bild: Für die höchsten Korrelationskoeffizienten (CCs) gibt es ausschließlich korrekte Detektionen. Das Histogramm der LCCF würde es durch Beschränkung auf Korrelationswerte oberhalb von etwa $CC = 0,38$ ermöglichen, etwa 80 % der Partikel praktisch ohne Falsch-Positive zu lokalisieren. Die LMCF zeigt ein qualitativ ähnliches Verhalten, durch Abschneiden bei etwa $CC=0,26$ wäre es ebenfalls möglich einen ähnlich hohen Prozentsatz der Partikel ohne nennenswerte Falsch-Positive zu lokalisieren. Allerdings fällt auf, daß die LCCF auch noch vereinzelte korrekte Detektionen bei sehr niedrigen CCs liefert. Als Grund hierfür ist die höhere Rauschempfindlichkeit der LMCF anzunehmen: Die einzelne Partikel haben ein unterschiedliches SNR, insbesondere im Bereich höherer Frequenzen, die bei der LMCF stärker betont werden. Als Folge davon variieren die Korrelationskoeffizienten stärker.

Die Detektionsqualität der LCCF wird vor allem durch den auftretenden Streuabsorptionskontrast beeinträchtigt: Es ist quantitativ nicht bekannt, in welchem Maße die Makromoleküle zum Streuabsorptionskontrast beitragen, in jedem Fall bleibt er bei den erzeugten Referenzen diesen Anteil jedoch unberücksichtigt. Weiterhin werden die Falsch-Positiven vor allem aufgrund des niederfrequenten, starken Signals von Lipidmembran und Goldmarkern erzeugt. Beide Effekte können durch Hochpaßfiltern der LCCF relativ einfach umgangen werden. Ein Hochpaßfiltern der LCCF bei etwa $(20 \text{ nm})^{-1}$ erlaubt es, die Proteasomen praktisch ohne Falsch-Positive zu detektieren, wenn Werte von $CC < 0.25$ ausgeschlossen werden (Abb. 5.9 A).

Detektionsergebnisse in Abhängigkeit von der Referenz. Als weitere Kontrolle der Spezifität der Detektion wurde das Tomogramm mit einer strukturell ähnlichen aber falschen Referenz, dem Thermosom, korreliert. Die berechnete LCCF wurde wiederum oberhalb von (20 nm^{-1}) hochpaßgefiltert und die Detektionsergebnisse wurden als Histogramm aufgetragen (Abb. 5.9 B). Es zeigt sich, daß die höchsten Korrelationswerte etwa bei 0.25 liegen, d.h. unterhalb des Bereichs in dem die Proteasom-Korrelation spezifisch ist. Die parallele Auftragung der Korrelationshistogramme von Proteasom und Thermosom zeigt, daß die Thermosom-Korrelation nicht das für die Proteasom-Korrelation typische Maximum bei etwa $CC=0.3$ aufweist.

5.5 Identifikation von Ribosomen im Prokaryoten *Spiroplasma melleiferum*

Ganze Zellen haben zwei Kriterien zu erfüllen, um für *in toto* KET geeignet zu sein: (i) Sie müssen ausgesprochen dünn sein, um mit einem hinreichendem SNR abgebildet zu werden. (ii) Die Proteindichte im Innern darf nicht zu groß sein, da im Wesentlichen die Dichtedifferenz zwischen Zytoplasma und Makromolekülen abgebildet wird. Ist das Zytoplasma zu dicht, so ist das Signal insbesondere im Auflösungsbereich der KET ($\sim 5 - 10\text{nm}$) sehr schwach, was ein niedriges SNR bedingt. Es hat sich

herausgestellt, daß die Konzentration der Makromoleküle (*Molecular Crowding*) bei prokaryotischen Zellen besonders groß ist (Grünewald et al. 2003). Das Bakterium *Spiriplasma melliferum*, die zu den Mykoplasmen gehört, besitzt einen außerordentlich kleinen Durchmesser ($> 200\text{nm}$). Somit eignet es sich zur elektronentomographischen Untersuchung, und die Studien von Kürner *et al* (Kürner et al. 2005) konnten neue Einsichten in das Zytoskelett des Bakteriums auf molekularer Ebene geben. Ribosomen sind sehr große (70S Ribosom von *Escherichia coli* ~ 2.3 MDa) makromolekulare Komplexe, deren Aufgabe es ist, Proteine zu synthetisieren. Ihre schiere Größe und ihr zahlreiches Auftreten in einer Zelle machen sie zu geeigneten Kandidaten, um erstmalig ein Makromolekül innerhalb einer ganzen Zelle zu kartographieren.

Thema dieses Abschnitts ist die Lokalisierung von Ribosomen in ganzen *S. melliferum* Zellen. Die Ergebnisse wurden mittels einer Histogrammanalyse auf ihre Zuverlässigkeit überprüft⁴. Die erreichte Vertrauenswürdigkeit erlaubt es, einen zellulären Atlas eines makromolekularen Komplexes zu bestimmen.

5.5.1 Material und Methoden

Tomogramme. Es wurden insgesamt 3 von Kürner aufgenommene Tomogramme analysiert, die nicht zur Studie des Zytoskeletts in Ref. (Kürner et al. 2005) verwendet wurden. Hierbei handelte es sich um Kippserien, die in einem Winkelbereich von -60° bis $+60^\circ$ mit einem Inkrement von 1.5° bei einem Defokus von $8\ \mu\text{m}$ aufgenommen wurden. Zur Identifikation von Ribosomen wurden die Daten 2-fach gebinnt, so daß ein Voxel etwa $2.2\ \text{nm}$ entsprach.

Partikelidentifikation. Als Muster zur Korrelationssuche wurde die Kryo-EM Struktur des Ribosoms von *Escherichia coli* benutzt (Gabashvili et al. 2000). Die Struktur wurde auf oberhalb der zur ersten Nullstelle der CTF des Tomogramms korrespondierenden Auflösung tiefpaßgefiltert und mittels trilinearer Interpolation mit einer Voxelgröße von $2.2\ \text{nm}$ dargestellt. Das Tomogramm wurde unter Verwendung von

⁴Die Datenanalyse wurde im Rahmen dieser Arbeit entwickelt und von J. Ortiz durchgeführt.

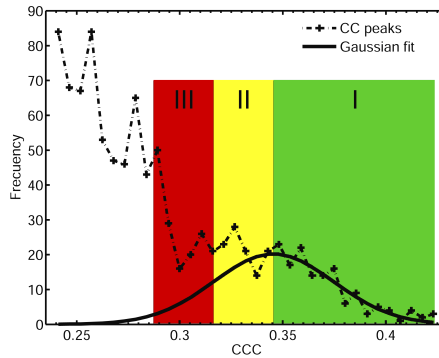


Abbildung 5.10: Histogramm der Korrelationskoeffizienten. Das Histogramm wurde gemäß der Gaußfunktionen in drei Bereiche unterschiedlicher Detektionsgüte aufgeteilt.

MOLMATCH mit diesem Muster korreliert. Dabei wurde eine Maske mit einem Durchmesser von etwa 27 nm verwendet (Durchmesser des Ribosoms ≥ 20 nm). Die Eulerwinkel wurden innerhalb des ganzen Winkelraums mit einem Inkrement von 10° abgetastet.

Die berechnete LCCF wurde mit einer Maske multipliziert, die die Zelle innerhalb des Tomogramms grob begrenzt. Die Korrelationskoeffizienten, Koordinaten und zugehörigen Orientierungen der höchsten 1000 Korrelationswerte wurden in einer Motivliste gespeichert. Dabei wurde die Ausdehnung einer Autokorrelationsfunktion (=Ausdehnung der Korrelationsmaxima) mit 22 nm approximiert. Die Korrelationskoeffizienten wurden als Histogramm aufgetragen (Abb. 5.10). Die Werte oberhalb eines lokalen Maximums wurden in MATLAB an eine Gaußkurve G angepaßt:

$$G(CC) = A_0 \exp\left(-\frac{(CC - CC_0)^2}{2\sigma^2}\right). \quad (5.16)$$

Gemäß der bestimmten Gaußfunktion wurde das Histogramm in 3 Bereiche aufgeteilt:

- I $CC \geq CC_0$,
- II $CC - \sigma \leq CC < CC_0$,
- III $CC_0 - 2\sigma \leq CC < CC_0 - \sigma$.

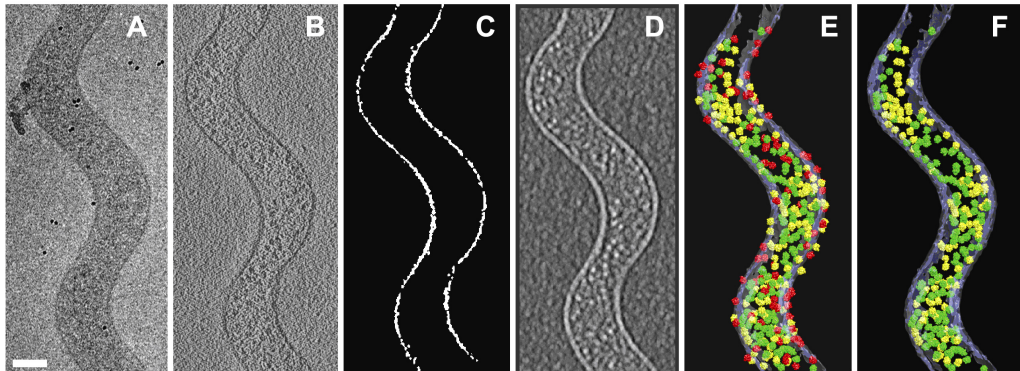


Abbildung 5.11: a: TEM Bild einer *S. melliferum* Zelle (0° Projektion). Der Maßstab korrespondiert zu 100 nm b: xy -Schnitt des rekonstruierten Tomogramms. c: xy -Schnitt der segmentierten Zellmembran. d: xy -Schnitt der CCF. e: Lokalisierungen der vermeintlichen Ribosomen. Die Farbe korrespondiert zur Höhe des CCC (siehe Abb. 5.10). Die segmentierte Membran ist in hellblau dargestellt. f: Atlas der Ribosomen. Zur Membran korrespondierende Partikel wurden als falsch-positiv angesehen und entfernt. (Abb. aus (Ortiz et al. 2005))

Segmentierung und Darstellung. Die Membran der Zelle wurde von Linaroudis auf Basis des *Scaling Index* segmentiert (Linaroudis und Hegerl 2005). Die 3D Daten wurden als Oberflächen mit 3D Studio Max dargestellt.

5.5.2 Ergebnisse

In den *S. melliferum* Tomogrammen stellen die Membranen des Bakteriums die am deutlichsten zu unterscheidenden Strukturen dar (Abb. 5.11b). Im Innern lassen sich einige globuläre Strukturen erkennen; allerdings ist die makromolekulare Dichte im Zytoplasma anscheinend sehr hoch, so daß das resultierende SNR niedrig ist. Die automatische Segmentierung der Membran korrespondierte sehr gut zur visuellen Identifikation der Membran (Abb. 5.11c).

Zur Identifikation der Ribosomen wurde die CCF eines *S. melliferum* Tomogramms und des 70S Ribosoms aus *E. coli* berechnet. Die Struktur des 70S Ribosom aus *E. coli* wurde mittels Kryo-EM bestimmt (Gabashvili et al. 2000). Da die Translationsmaschinerie evolutionär weitgehend konserviert ist, ist davon auszugehen daß diese

Struktur dem Ribosom aus *S. melliferum* im Auflösungsbereich von KET hinreichend ähnlich ist. Die CCF besitzt distinkte Maxima im Zytoplasma, es treten jedoch auch sehr hohe Korrelationswerte innerhalb der Membran auf (Abb. 5.11).

Die Korrelationsmaxima wurden bestimmt und als Histogramm aufgetragen (Abb. 5.10). Idealerweise würde man eine Gaußverteilung hoher Korrelationswerte erwarten, die oberhalb eines Hintergrunds von unspezifischen Detektionen liegt. Im ungünstigsten Fall sollte man im Histogramm eine vollkommene Überlagerung von unspezifischen mit korrekten Detektionen erhalten. Das Histogramm hat ein qualitativ ähnliches Aussehen wie die Korrelationshistogramme der Phantomzellen (Abb. 5.9a): Es liegt ein lokales Maximum der Korrelationswerte bei höheren Werten ($CC \approx 0.34$) vor, und die Häufigkeit steigt in Richtung niedriger Korrelationswerte ($CC < 0.3$) stark an. Das Histogramm oberhalb des lokalen Maximums wurde an eine Gaußverteilung angepaßt, und die vermeintlichen Ribosomen wurden gemäß dieser Verteilung in drei Bereiche I, II und III aufgeteilt (Abb. 5.10): Der Bereich I oberhalb des Maximums der Gaußfunktion (CC_0), dessen Histogramm sehr gut mit der angepaßten Gaußform übereinstimmt, der Bereich II zwischen CC_0 und $CC_0 - \sigma$ aus (5.16), der mäßig mit dem Fit übereinstimmt, und der Bereich III zwischen $CC_0 - \sigma$ und $CC_0 - 2\sigma$, der kaum zur angepaßten Kurve korrespondiert. Die graphische Darstellung der so aufgeteilten Ribosomen im Tomogramm ergibt, daß die höchsten Korrelationen (I und II) überwiegend im Zytoplasma lokalisiert sind, während Korrelationstreffer des Bereichs III vor allem in der Membran auftreten (Abb. 5.11e).

Um die Detektionsqualität zu beurteilen, wurden die Detektionsergebnisse mit der angepaßten Gaußfunktion verglichen. Die Ergebnisse sind in Tab. 5.1 zusammengefaßt: In Bereich I ist die Übereinstimmung zwischen der Gaußfunktion und den gemessenen Werten praktisch vollständig ($\sim 96\%$). Im Bereich II, beträgt das Verhältnis C von erwarteter Ribomenzahl zu Detektionen 83%, während es in Bereich III nur noch 27% beträgt. Durch Vergleich von Treffern mit der segmentierten Membran konnten sichere falsch-positive Detektionen ausgeschlossen werden, was in den Bereichen I und II durchgeführt wurde. Hierdurch wurde C im Bereich II auf 95% angehoben, d.h.

	III	II	I
R_{th}	16.5%	33%	50%
N_{th}	43	89	137
N_{exp}	162	107	143
C	27%	83%	96 %
N_{Membr}	n.b.	13	1
N_{korr}	n.b.	94	142
C_{korr}	n.b.	95%	97 %

Tabelle 5.1: Erkennungsstatistik der Ribosomen in *S. melliferum*. Aufgetragen sind detektierte und erwartete Anzahl der Ribosomen in den Korrelationsbereichen I, II, und III (Abb. 5.10). R_{th} : relativer Anteil der tatsächlich vorhandenen Ribosomen, die im jeweiligen Intervall detektiert werden sollten gemäß der angepaßten Gaußfunktion. N_{th} : absolute Zahl der theoretisch im Intervall detektierten Ribosomen. N_{exp} : tatsächlich im Intervall detektierte Ribosomen. C : Anteil der potentiell korrekten Detektionen (*Confidence level*). N_{Membr} : Detektionen in der Membran, die als falsch positiv angesehen werden können. N_{korr} : Detektion nach Abzug der Membrandetektionen ($N_{exp} - N_{Membr}$). C_{korr} : resultierender Vertrauenslevel nach dieser Korrektur.

die Übereinstimmung von Gaußfunktion und gemessenen Werten ist in den Bereichen I und II sehr gut. Die Ribosomdetektionen in den Bereichen I und II sollten rund 80% der gesamten Ribosomen beinhalten. Diese repräsentieren gemäß der Histogrammanalyse mit einer Vertrauenswürdigkeit von etwa 95% Ribosomen, die den Ribosomenatlas aus Abb. 5.11f konstituieren. Die einzelnen Ribosomen sind weitgehend homogen in der Zelle verteilt. Gemäß Tab. 5.1 sind etwa 5% des Zytoplasmas von Ribosomen besetzt. Schätzungen für die Ribosomenkonzentration existieren vor allem für *E. coli*, und liegt in der Größenordnung der hier bestimmten (Neidhardt 1996).

5.6 Partikelidentifikation im Eukaryoten *Dictyostelium discoideum*

Der Schleimpilz *Dictyostelium discoideum* besitzt in einigen Zellbereichen eine Dicke von wenigen 100 nm, was ihn für KET *in toto* geeignet macht. Die an diesem Organismus durchgeführten tomographischen Studien von Medalia *et al* (Medalia et al. 2002) ermöglichten es erstmalig, individuelle Makromoleküle wie das 26S Proteasom im Kontext ihrer annähernd physiologischen Umgebung aufzulösen.

In diesem Abschnitt wurden in Tomogrammen aus Ref. (Medalia et al. 2002) systematisch Ribosomen und das Aktinskelett detektiert.

5.6.1 Material und Methoden

Tomogramme. Es wurde das in Ref. (Medalia et al. 2002) publizierte Tomogramme von *D. discoideum* analysiert. Die zugrundeliegenden Kippserien umfaßten einen Winkelbereich von -45.5° bis $+51^\circ$, es wurde ein Inkrement von 1.5° und einem Defokus von $15 \mu\text{m}$ gewählt. Zur Identifikation von Ribosomen und Aktin wurden die Daten 2-fach gebinnt, so daß ein Voxel etwa 3.3 nm entsprach.

Partikelidentifikation. Die Identifikation der Ribosomen wurde analog Abschn. 5.5.1 durchgeführt. Bei der Detektion von Aktin wurde folgendermaßen vorgegangen: Es wurde ein Tomogrammelement mit eindeutig filamentöser, gerader Morphologie extrahiert. Die Hauptachse des Filaments wurde entlang z -gewählt, das Filament wurde entlang der Hauptachse rotationsgemittelt, entlang z projiziert und wieder zurückprojiziert, um das SNR der Referenz zu erhöhen. Anschließend wurde eine LCCF dieser Referenz und des Tomogramms mittels OMNIMATCH berechnet. Dabei wurde ein Halbzylinder als Maske verwendet: Von einem in z -Richtung orientierten Zylinder wurde die untere Hälfte zu Null gesetzt.

5.6.2 Ergebnisse

Ein xy -Schnitt des *D. discoideum* Tomogramms ist in Abb. 5.12A gezeigt. Es lassen sich Filamente und sehr kontrastreiche, globuläre Objekte mit einem Durchmesser von etwas mehr als 20 nm erkennen.

Die Korrelation des Tomogramms mit 70S Ribosomen ergibt besonders hohe Werte an den Orten der globulären Komplexe (Abb. 5.12 B und C). Außerdem treten es vor allem an Orten der Goldmarker sehr hohen Korrelationswerte auf. Es erscheint plausibel, daß es sich bei der Mehrzahl der zu den ca. 400 höchsten Korrelationswerten korrespondierenden Partikel tatsächlich um Ribosomen handelt, da sich Ribosomen neben ihrer Größe von über 20 nm vor allem durch den relativ großen Kontrast (im Vergleich zu Proteinen) aufgrund des hohen RNA Anteils auszeichnen. Manuelle Untersuchung einzelner Partikel ergab, daß die Mehrzahl der detektierten Partikel etwas größer als die 70S Ribosomen Referenz zu sein schien, was dadurch erklärt werden kann, daß in Eukaryoten die massereicheren 80S Ribosomen vorliegen. Das Auftragen der ribosomalen Korrelationskoeffizienten als Histogramm ergibt einen qualitativ ähnlichen Verlauf wie bei Phantomzellen (Abb. 5.9) und der Ribosomendetektion in *S. melliferum* (Abb. 5.10). Dies ist ein weiterer Hinweis darauf, daß es sich bei der Mehrzahl der detektierten Partikel um Ribosomen handelte.

Die Detektionen wurden analog zu Abschn. 5.5 gemäß einer an das Histogramm angepassten Gaußverteilung in 3 Gebiete aufgeteilt. Nimmt man die Übereinstimmung des Histogramms mit einer Gaußverteilung als Kriterium, so kann wiederum die Zahl der Ribosomen und die Zuverlässigkeit der Lokalisierung bestimmt werden. Dies ist in Tab. 5.2 geschehen. Demnach liegen in einem Volumen von etwa $0.3 \mu\text{m}^3$ etwa 800 Ribosomen vor, was etwa 3 % des Gesamtvolumens entspricht. Im Gegensatz zu *S. melliferum* sind die Ribosomen in *D. discoideum* nicht homogen verteilt, sondern befinden sich überwiegend in zwei definierten Schichten (Daten nicht gezeigt).

Die Korrelation des als Zylinder approximierten Aktins mit dem Tomogramm ist in Abb. 5.2 C und D. Es zeigt sich starke Korrelation mit langen filamentösen Strukturen im Tomogramm. Außerdem ist bemerkenswert, daß es auch zu hohen Korrelationen

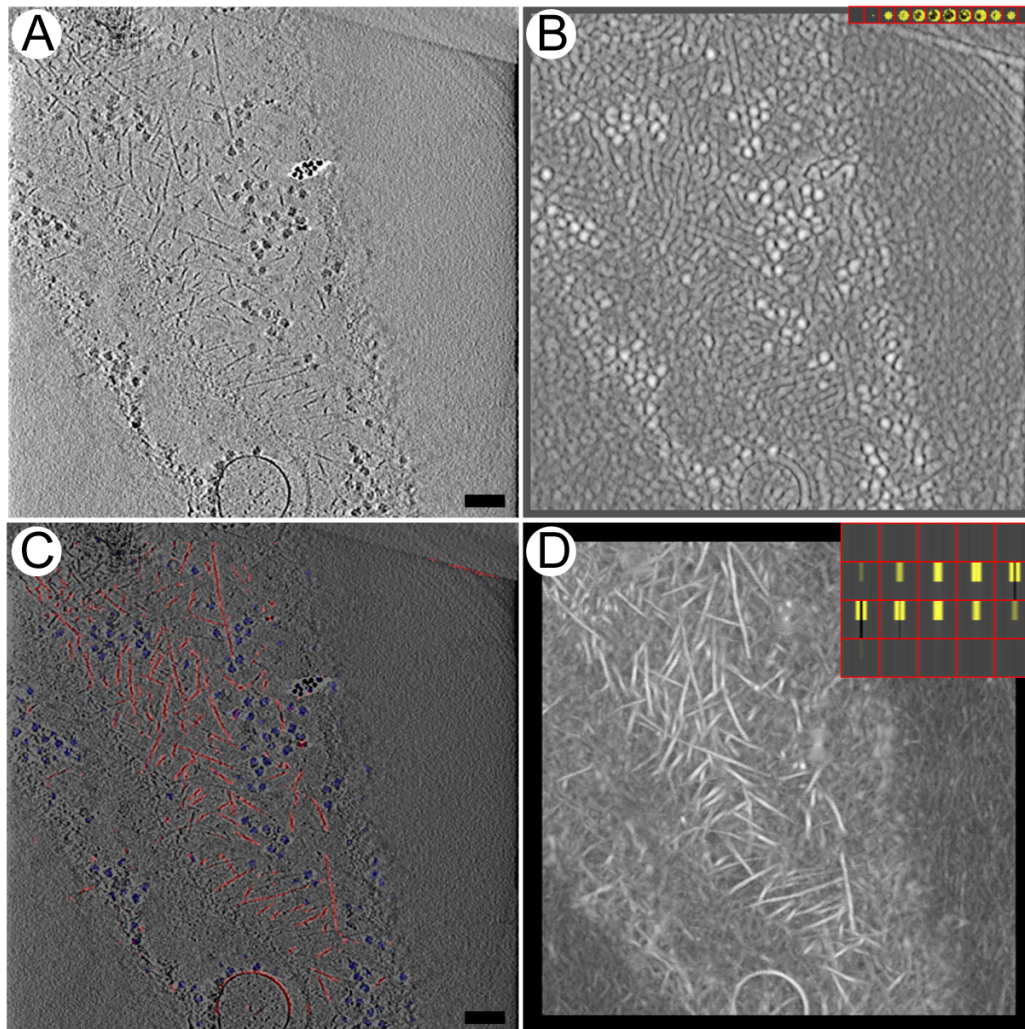


Abbildung 5.12: A: xy -Schnitt aus *D. discoideum* Tomogramm. Der Maßstab stellt 100 nm dar. B: Zu A korrespondierender Schnitt der LCCF des Tomogramms und eines Ribosoms (oben rechts: verschiedene z -Schichten des Ribosoms, in gelb ist die benutzte kugelförmige Maske dargestellt). C: Montage des Schnitts und der höchsten Werte der Korrelationsfunktionen aus B (blau) und D (rot). D: LCCF von Tomogramm und filamentöser Referenz (Einfügung oben rechts, zylindrische Maske gelb).

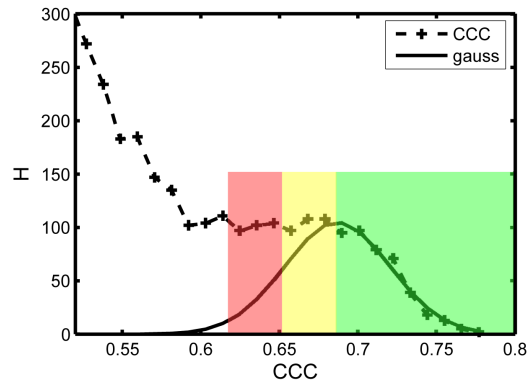


Abbildung 5.13: Histogramm der ribosomalen Korrelationskoeffizienten aus *D. discoideum*. Die Einteilung in verschiedene Zonen erfolgt anhand der angepassten Gaußfunktion analog zu Abb. 5.10.

	$CC_0 - 2\sigma < CC < CC_0 - \sigma$	$CC_0 - \sigma < CC < CC_0$	$CC_0 < CC < 1$
R_{th}	16.5%	33%	50%
N_{th}	103	263	419
N_{exp}	303	313	430
C	30%	80%	100%

Tabelle 5.2: Erkennungsstatistik der Ribosomen in *Dictyostelium*. Aufgetragen sind detektierte und erwartete Anzahl der Ribosomen in verschiedenen Intervallen des Korrelationskoeffizienten. R_{th} : relative Anteil der tatsächlich vorhandenen Ribosomen, die im jeweiligen Intervall detektiert werden sollten. N_{th} : absolute Zahl der theoretisch im Intervall detektierten Ribosomen. N_{exp} : tatsächlich im Intervall detektierte Ribosomen. C : Anteil der potentiell korrekten Detektionen (*Confidence level*).

mit Filamenten kommt, die in z -Richtung orientiert sind. Diese waren manueller Detektion mittels Visualisierungsprogrammen (z.B. *Amira*) gänzlich unzugänglich, da diese es nur erlauben, zweidimensionale Schnitte zu "segmentieren"⁵. Allerdings ist es nicht möglich, einen eindeutigen Grenzwert für die zuverlässige Detektion eines Filaments festzulegen. Dies kann vor allem mit der Eigenschaft der Filamente begründet werden: Diese bilden ein verzweigtes Netzwerk mit teilweise sehr kurzen Filamenten, und außerdem sind viele Filamente gekrümmt. Aus diesem Grunde sind nur für Teile aller filamentösen Strukturen hohe Korrelationswerte zu erwarten, für kurze oder gebogene Filamente, sind, mit fließenden Übergängen, niedrigere Werte zu erwarten.

5.7 Diskussion

In dieser Arbeit wurde ein Algorithmus zur Detektion von Makromolekülen, der auf Korrelation beruht, weiterentwickelt und angewendet. Es wurden verschiedene Korrelationsfunktionen definiert und implementiert, die anhand von proteasomgefüllten Phantomzellen auf ihre Zuverlässigkeit getestet wurden. Dabei stellte sich heraus, daß eine zuverlässige Detektion ohne eine lokale Normalisierung der Korrelation praktisch unmöglich ist. Die Partikellokalisierung mittels der CCF und MCF führt zu einem inakzeptabel hohen Maß an falsch-positiven Detektionen, wobei vor allem die starke Korrelation der Lipidmembran ein Problem darstellt. Im Gegensatz dazu liefern die lokal normierten LCCF und die LMCF deutlich weniger falsche Detektionen. Dabei ist die LMCF jedoch rauschempfindlicher, so daß die Korrelationswerte der tatsächlichen Proteasomen stärker variieren. So besitzen einzelnen Partikel einen sehr niedrigen Korrelationswert, bei dem unspezifische Detektionen überwiegen. Aus diesem Grund scheint die LCCF beim derzeit experimentell erreichbaren SNR am besten geeignet, Partikel in einem Tomogramm *vollständig* zu detektieren. Es ist jedoch möglich, daß ungewöhnliche Korrelationsmaße wie das der LMCF bei höheren SNRs vorteilhaft

⁵Sehr dünne Filamente sind manuell effektiv nur erkennbar, falls die Schnittebene den Richtungsvektor der Filamente beinhaltet; da lediglich eine geringe Zahl an Schnitten, typischerweise nur xy -, xz - und yz -Schnitte, analysiert werden können, wird nur ein Teil der Filamente erfaßt, wie in Ref. (Medalia et al. 2002).

sind.

Weiterhin wurden Ribosomen in pro- und eukaryotischen Zellen lokalisiert. Eine zentrale Rolle bei der entwickelten Auswertung der Korrelationswerte nimmt dabei das Histogramm der gefundenen Korrelationswerte ein. Es wurden stets qualitativ ähnliche Histogramme gefunden, die sich erklären lassen durch spezifische Detektionen, die gaußförmig verteilt sind, und einen unspezifischen Hintergrund. Unspezifische Detektionen traten dabei vor allem an Objekten mit sehr starken Signalen wie Membranen oder Goldkugeln auf. Im Falle der Ribosomendetektion in *S. melliferum* konnten in der Membran lokalisierte falsch-positive Identifikationen durch Kombination mit einer zweiten Segmentierungsmethode entfernt werden. So ergibt sich ein ribosomaler Atlas, der die Positionen von 80 % aller Ribosomen mit einem Vertrauensgrad von über 90 % angibt.

Eine weitere Optimierung des entwickelten Algorithmus wird vor allem darauf abzielen, die Detektion von Objekten mit sehr starken SNRs noch stärker zu unterdrücken. Hierzu sind zwei Wege denkbar: (i) Objekte mit starkem Signal können durch andere Verfahren segmentiert und somit von der Korrelationsfunktion ausgeschlossen werden. Dieser Weg wurde bei der Auswertung der *S. melliferum* Zelle gegangen, indem die Zellmembran unabhängig segmentiert wurde, und das Verfahren kann noch verallgemeinert werden. (ii) Die Korrelationsfunktion kann durch ein anderes Ähnlichkeitsmaß ersetzt werden, das eine höhere Spezifität besitzt. Ein solches Maß könnte z.B. *Maximum Likelihood* darstellen (Sigworth 1998). Allerdings wurde eine Vielzahl unterschiedlicher Detektionsverfahren zur Lokalisierung von Makromolekülen in zweidimensionalen Kryo-EM Bildern entwickelt und verglichen (Zhu et al. 2004). Dabei hat sich kein Verfahren herauskristallisieren können, das lokaler Korrelation hinsichtlich der Zuverlässigkeit deutlich überlegen gewesen wäre.

Ein weitere Faktor, der die Detektion von Makromolekülen in Tomogrammen mit dem Algorithmus in seiner jetzigen Form deutlich erleichtern könnte, ist die Verwendung anderer Algorithmen zur 3D-Rekonstruktion. In der vorliegenden Arbeit wurde hierfür ausschließlich gewichtete Rückprojektion verwendet, andere Verfahren wie ART

könnten jedoch Vorteile bieten (Abschn. 2.3). Jede quantitative Verbesserung der 3D Daten würde umgekehrt die Signifikanz der Korrelationsergebnisse erhöhen.

6. Dreidimensionale Mittelung von Makromolekülen

Die Auflösung elektronenmikroskopischer Bilder eiseingebetteter biologischer Proben ist letztendlich durch die Elektronendosis limitiert. Um diese fundamentale Limitierung in der KET zu umgehen, bietet es sich an, Aufnahmen strukturell gleicher Proben kohärent zu mitteln. Diese Lösung wird in der *Single Particle* Analyse gewählt, wo durch Mitteln von i.A. zig tausend Einzelpartikeln hochaufgelöste Strukturen von Makromolekülen erhalten werden. Im Falle asymmetrischer Makromoleküle wie dem Ribosom konnten bereits Auflösungen von über $(10 \text{ \AA})^{-1}$ erreicht werden (van Heel et al. 2000; Frank 2002) und für hochsymmetrische Objekte wie ikosaedrische Virus-kapside Auflösungen im Bereich von $(5 \text{ \AA})^{-1}$ (Zhou und Chiu 2003).

Heutzutage werden hochaufgelöste Strukturen meist folgendermaßen erhalten: Das zu untersuchende Makromolekül wird biochemisch aufgereinigt, und es werden Mikrographen, d.h. Projektionen der vitrifizierten Probe aufgenommen. Die in den Mikrographen enthaltenen Makromoleküle werden meist manuell lokalisiert. Die Hauptaufgabe besteht nun darin, den beliebig orientierten Einzelpartikeln die richtige Orientierung zuzuweisen. Liegt das zu untersuchende Molekül in allen Orientierungen vor, so kann die komplette dreidimensionale Struktur des Moleküls ermittelt werden. Dieses verfahren wird "Angular Reconstitution" genannt (van Heel 1987).

Im Detail muß jedem Partikel die Projektionsrichtung und die Rotation auf dem Grid zugewiesen werden, was durch drei Eulerwinkel φ , ψ und ϑ beschrieben werden kann. Außerdem muß die Position der Partikel verfeinert werden, da die manuelle Lokalisation der Partikel nur sehr ungenau ist. Insgesamt müssen also 5 Parameter, drei

Winkel und zwei Translationsvektoren Δx und Δy , bestimmt werden. In den meisten Fällen liegt schon eine erste Struktur des Makromoleküls vor, so z.B. bei vielen Studien über die diversen funktionellen Zustände des Ribosoms. Hier wird als Verfahren zur Strukturbestimmung meist *Projection matching* verwendet (van Heel et al. 2000; Frank 2002): Bei diesem Verfahren werden alle Einzelpartikel mit verschiedenen Projektionsrichtungen der ersten Referenz korreliert, und die maximale Korrelation erlaubt eine erste Zuordnung der Orientierungen. In der Folge wird aus den Einzelpartikeln gemäß den gefundenen Orientierungen und Translationsvektoren eine neue Referenz generiert und die Partikel werden wiederum mit verschiedenen Projektionen korreliert. Auf diese Weise wird iterativ die Auflösung gesteigert, idealerweise bis zur Grenze, die durch die Zahl der Partikel oder die Homogenität der Probe gegeben ist. Unter der Annahme struktureller Homogenität der Probe hat Henderson die erreichbare Auflösung als Funktion der Partikelanzahl abgeleitet (Henderson 1995): So werden z.B. theoretisch 12600 2D Projektionen benötigt um eine Struktur von 3 \AA^{-1} Auflösung zu erreichen, experimentell wird diese i.A. erst etwa mit der zehnfachen Teilchenzahl erreicht (z.B. in (Nogales et al. 1998)).

Prinzipiell ist es ebenso möglich, hochaufgelöste dreidimensionale Strukturen durch Mittelung von Einzelpartikeln aus Tomogrammen zu erhalten. Gemäß dem Dosisaufteilungstheorem (Abschn. 2.2) ist in einem Tomogramm die gleiche Signalmenge enthalten wie in einer Projektion, die mit der gleichen Dosis erzeugt wurde. So wurden denn auch die ersten höheraufgelösten dreidimensionalen Strukturen des Ribosoms durch Mittelung von Subtomogrammen erzeugt (Oettl et al. 1983; Knauer et al. 1983). Später ist die Mittelung von Tomogrammen zur Strukturbestimmung vor allem von Walz *et al* verwendet worden (Walz et al. 1997b; Walz et al. 1997a; Walz 1997). In diesen Arbeiten wurden aufgereinigte, vitrifizierte Komplexe tomographiert und durch Mittelung von Subtomogrammen wurden 3D Daten im mittleren Auflösungsbereich von etwa $(3 \text{ nm})^{-1}$ erhalten.

Im Rahmen dieser Arbeit wurde ein iterativer Algorithmus zur Alignierung und Mittelung von Tomogrammen entwickelt. Er stellt eine Weiterentwicklung des Algorithmus

von Walz *et al* dar (Walz et al. 1997b): So basiert die Alignierung auf der *Constrained Correlation Function* (5.12), wodurch der *Missing Wedge* berücksichtigt wird. Diese und weitere Maßnahmen wie das Maskieren bestimmter Regionen der Tomogramme ermöglichen eine *in situ* Strukturbestimmung. Der Algorithmus wird hier erklärt und das Resultat seiner Anwendung auf Daten des Kernporenkomplex' gezeigt.

6.1 Algorithmus zur Mittelung von Subtomogrammen

Die Aufgabe der Alignierung ist es, N verschiedene Subtomogramme P_i in der Weise zu addieren, daß die Summe aller Korrelationen der Partikel P_i und eines Modells A maximiert wird. Dieses Modell ist der Mittelwert der verschiedenen Partikel unter Berücksichtigung der Eulerwinkel φ , ψ und ϑ und der Translationsvektoren Δx , Δy und Δz jedes Partikels Mathematisch ist das Problem also die Maximierung des Funktionals

$$\text{Max} \left(\sum_{i=1}^N \mathcal{T}_{\Delta x_i \Delta y_i \Delta z_i} \mathcal{R}_{\varphi_i, \psi_i, \vartheta_i} A \odot P_i \right). \quad (6.1)$$

Hierbei bezeichnet \mathcal{R} wiederum den Rotationsoperator und \mathcal{T} den Translationsoperator. Dieses Funktional muß bzgl. der Variablen Δx_i , Δy_i und Δz_i sowie φ_i , ψ_i und ϑ_i maximiert werden. Dabei gilt es die Nebenbedingung der Mittelung zu erfüllen:

$$A = \frac{1}{N} \sum_{i=1}^N \mathcal{T}_{\Delta x_i \Delta y_i \Delta z_i}^{-1} \mathcal{R}_{\varphi_i, \psi_i, \vartheta_i}^{-1} P_i. \quad (6.2)$$

Ist das SNR ausreichend groß, so kann die Lösung von (6.1) durch Maximieren der einzelnen Korrelationen der Partikel P_i mit der Referenz A erhalten werden, d.h. es wird stattdessen berechnet:

$$\text{Max}(\mathcal{T}_{\Delta x_i \Delta y_i \Delta z_i} \mathcal{R}_{\varphi_i, \psi_i, \vartheta_i} A \odot P_i). \quad \text{für alle } i. \quad (6.3)$$

Im Allgemeinen ist das SNR ausreichend, um zumindest eine gute Abschätzung der tatsächlichen Struktur zu erhalten. Der Versuch der Maximierung der einzelnen Koeffizienten wird zwar nicht die bestmögliche Struktur ergeben jedoch eine gute Näherung. Ein Versuch das globale Maximierungsproblem zu lösen ist z.B. von der Penczek Gruppe kürzlich vorgeschlagen worden (Yang et al. 2004). Eine andere Möglichkeit der Korrektur der Alignierungsfehler besteht im *Sharpening*: Durch Alignierungsfehler werden hochfrequente Details der so erhaltenen EM Struktur gegenüber der "wahren" Struktur systematisch herabgewichtet, was rechnerisch auszugleichen versucht wird (Rosenthal und Henderson 2003). Alle diese Techniken sind jedoch erst dabei sich im *Single Particle* Feld zu etablieren und spielen vor allem bei höchsten Auflösungen (< 1 nm) eine Rolle, die in der Tomographie noch nicht erreicht werden.

In der vorliegenden Arbeit wurden Prozeduren entwickelt, um eine dreidimensionale Alignierung gemäß (6.3) zu realisieren. Die Maximierung von (6.3) erfordert theoretisch eine komplette Abtastung des Parameterraums, was jedoch rechnerisch illusorisch ist. Insbesondere das Abtasten des Winkelraums ist aufwendig, wie in Abschn. 5.1.1 bereits erläutert wurde. Aus diesem Grunde wurde die Alignierung im Wesentlichen in zwei Schritte gegliedert: Es wurde eine Prozedur entwickelt, die eine Alignierung in einem begrenzten Winkelraum durch Abtasten ausführt und dabei den *Missing Wedge* berücksichtigt. Weiterhin wurde eine Prozedur implementiert, die ein komplettes Abtasten zumindest eines Winkels, des Polarwinkels φ , ermöglicht. Dieses Vorgehen ist vor allem für Membranproteinkomplexe sinnvoll, da diese eine Einschränkung ihrer Rotationsfreiheitsgrade haben, so daß keine komplette Abtastung des Winkelraums benötigt wird.

6.1.1 Alignierung in einem begrenzten Winkelbereich

Ist bereits eine vorläufige Referenz bekannt, so kann die Lösung von (6.3) numerisch auf iterativem Weg folgendermaßen gefunden werden: Die Korrelationsfunktion der Partikel mit der Referenz wird als Funktion der Winkel und Translationsvektoren

berechnet, und die Δx_i , Δy_i und Δz_i sowie φ_i , ψ_i und ϑ_i , die die Korrelationsfunktion maximieren, bilden die Parameter zur Berechnung der Referenz für die nächste Iteration. Auf diese Weise werden die Parameter variiert bis die Struktur konvergiert. In der KET wurde ein solches Verfahren zur Mittelung erstmalig von Walz entwickelt und angewendet (Walz et al. 1997b; Walz et al. 1997a; Walz 1997). Es soll hier nur knapp erläutert werden: Die ortsabhängige Korrelationsfunktion der Partikel mit einer Referenz wird im Fourierraum für die Orientierung der vorherigen Iteration berechnet. So werden die Parameter Δx_i , Δy_i , Δz_i bestimmt. Anschließend wird der Korrelationskoeffizient unter Variation der Winkel ϑ und ψ um die Werte der vorherigen Iteration bestimmt. Der Winkel φ muß dabei nicht explizit abgetastet werden, da dieser Winkel durch Entwicklung der Partikel und Referenz nach Kugelkoordinaten und anschließende Korrelation bestimmt werden kann (Abschn. 6.1.3). Die Routine berechnet also Translations- und Rotationsvariablen getrennt. Dies beschleunigt die auszuführenden Rechnungen erheblich, was angesichts der damaligen Rechenkapazitäten wichtig war. Allerdings stellt die Annahme, daß sich die Translationsvektoren bei Variation der Winkel nicht ändern nur eine Näherung dar, die folglich einen gewissen Fehler einführt. Weiterhin verhindert es die Ermittlung der Polarwinkel φ mittels FFT eine Berücksichtigung des *Missing Wedge* in der Korrelationsfunktion.

In der vorliegenden Arbeit wurde die Verfeinerungsprozedur von Walz im Prinzip weiterentwickelt. Die Prozedur von Walz wurde im EM-System (Hegerl 1996) entwickelt, das jedoch nur eine begrenzte Komplexität und Übersichtlichkeit der Prozedur zulässt. Aus diesen Gründen wurde die neue Prozedur in Matlab entwickelt.

Die Verfeinerungsprozedur beruht auf der iterativen Maximierung der eingeschränkten Korrelationsfunktion (5.12) jedes Partikels P_i mit einer Referenz:

$$CCF_{i;\mathbf{r}\varphi\psi\vartheta} = \frac{P_i \odot (A_{\varphi\psi\vartheta} \otimes PSF)}{\sqrt{\sum_{\mathbf{r}} (A_{\mathbf{r},\varphi\psi\vartheta} \otimes PSF_{\mathbf{r}})^2} \cdot \sqrt{\sum_{\mathbf{r}} (P_{i;\mathbf{r}})^2}}. \quad (6.4)$$

Die gesamte iterative Prozedur ist in Abb. 6.1 skizziert. Im einzelnen sind die Schritte der Prozedur folgendermaßen: Die Referenz A (Abb. 6.1B) wird mittels trilinearere Interpolation in eine Orientierung $(\varphi, \phi, \vartheta)$ rotiert (Abb. 6.1C). Das resultierende

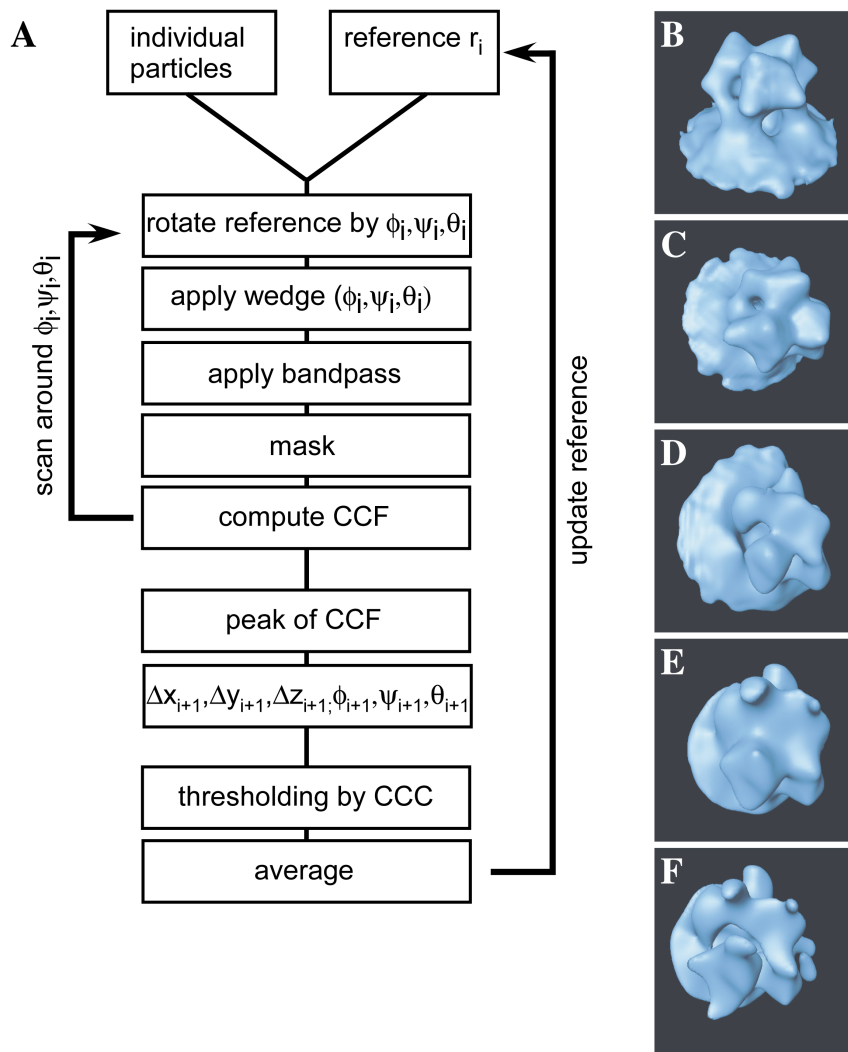
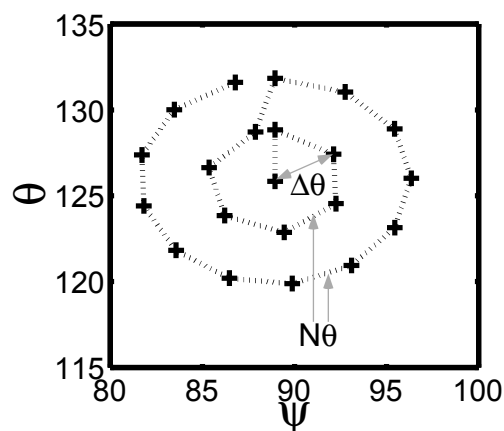


Abbildung 6.1: A: Prozedur zum iterativen Mitteln von Subtomogrammen. B-F: Verschiedene Stadien der korrelierten Referenz im Alignierungszyklus (siehe Text).

Volumen wird mit der *PSF* gefaltet, d.h. es wird der Effekt des *Missing Wedge* simuliert (Abb. 6.1D). Nun kann ein Bandpaßfilter angewendet werden (Abb. 6.1E): Aufgrund des variablen Streuabsorptionskontrasts sind die tiefen Frequenzen in einem Tomogramm meist sehr stark aber vor allem auch unterschiedlich ausgeprägt, was die Korrelationsfunktion unzuverlässig macht. Anwenden eines Hochpaßfilters, kann den Streuabsorptionskontrast weitgehend entfernen und wird in Alignierungsprozeduren häufig verwendet (Van Heel et al. 1992). Weiterhin kann die Alignierung anfangs auf tieffrequente Bereiche beschränkt werden, was durch einen entsprechenden Tiefpaßfilter bewerkstelligt werden kann. Der Tiefpaßfilter sollte in keinem Fall die erste Nullstelle der *CTF* übersteigen, da der Defokuswert nicht präzise bestimmt werden kann, so daß Information über die erste Nullstelle hinaus in einzelnen Subtomogrammen unterschiedlichen Kontrast haben kann. Der Bandpaßfilter wird stets an den Rändern mit einer Gaußfunktion geglättet, um Artefakte zu vermeiden. Anschließend wird das Partikel maskiert (Abb. 6.1C). Vor allem bei Membranproteinkomplexen ist dies wichtig, da die Membran in einiger Entfernung vom Komplex zunehmend flexibel ist, so daß dieses inhomogene Signal die Alignierung stören kann. Auch bei der Alignierung von Komplexen, die sich in sehr dichten Lösungen befinden (z.B. *Molecular Crowding* in Zellen), ist eine Maskierung wichtig. Die Maskierung berücksichtigt dabei die in der vorherigen Iteration bestimmten Verschiebungsvektoren und Winkel. Nun wird die eingeschränkte Korrelationsfunktion *CCF* bestimmt. Berücksichtigt man in (6.4) die Maskierung mit der Maske M und die Bandpaßoperation \mathcal{B}_{lo}^{hi} :

$$CCF_{i;\mathbf{r}\varphi\psi\vartheta} = \frac{(M_{\varphi\psi\vartheta;\Delta x,\Delta y,\Delta z} \cdot \mathcal{B}_{lo}^{hi}(P_i)) \odot (M_{\varphi\psi\vartheta} \cdot \mathcal{B}_{lo}^{hi}(A_{\varphi\psi\vartheta} \otimes PSF))}{\sqrt{\sum_{\mathbf{r}'} (\mathcal{B}_{lo}^{hi}(A_{\varphi\psi\vartheta} \otimes PSF)_{\mathbf{r}'} \cdot M_{\mathbf{r}',\varphi\psi\vartheta})^2 \cdot \sum_{\mathbf{r}'} (M_{\mathbf{r}',\varphi\psi\vartheta,\Delta x,\Delta y,\Delta z} \cdot \mathcal{B}_{lo}^{hi}(P_{i;\mathbf{r}'})^2}}. \quad (6.5)$$

Diese Korrelationsfunktion wird für Winkel in der Umgebung des in der vorherigen Iteration bestimmten optimalen Winkeltripels $(\varphi_{opt}, \psi_{opt}, \vartheta_{opt})$ bestimmt. Dabei müssen die Winkel so variiert werden, daß eine möglichst gleichmäßige Abtastung der Winkel um diese Orientierung herum erfolgt. Das hier verwendete Winkelschema ist in Abb. 6.2 für die Winkel ψ und ϑ skizziert. Es wird folgendermaßen konstruiert: Es


 Abbildung 6.2: Winkelabtastung von ψ und θ .

werden um den Nordpol der Einheitskugel Abtastpunkte gelegt. Der erste Punkt ist der Nordpol selbst, für den nächsten Punkte gilt:

$$\vartheta_k = k \cdot \Delta\theta \quad , \quad \psi_{kl} = \frac{l \cdot \Delta\theta}{\sin(\vartheta)} \quad (6.6)$$

Die Laufvariable k wird dabei bis zu einem Maximalwert N_θ erhöht, und l läuft von 0° bis $360 \sin(\vartheta)/(\Delta\theta)^\circ$. Die so definierten Abtastpunkte (ψ_{kl}, ϑ_k) werden so rotiert, daß die Orientierung $(\psi_{opt}, \vartheta_{opt})$ der vorherigen Iteration den neuen Pol darstellen. Insgesamt erfolgt bzg. der Winkel ψ und ϑ eine schalenweise Abtastung, wobei N_θ und $\Delta\theta$ die Parameter darstellen, die Weite und Feinheit der Abtastung definieren. Für den Winkel φ wird eine gleichmäßige Abtastung gewählt:

$$\varphi_m = m \cdot \Delta\theta \quad , \quad -N_\theta \leq m \leq N_\theta \quad (6.7)$$

Die Prozedur wurde in MATLAB als `av3_trans-rot-alig.m` implementiert.

6.1.2 Initialisierung der Orientierungen

Zur Anwendung der oben beschriebenen iterativen Prozedur sind Startwerte für die Orientierungen $(\varphi_i, \psi_i, \vartheta_i)$ nötig. Im Falle von membrangebundenen Komplexen kann

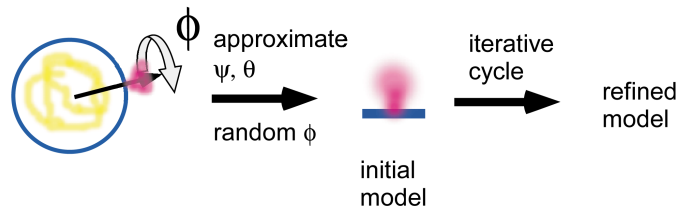


Abbildung 6.3: Ausnutzen der Geometrie zum Start des iterativen Alignierungsprozesses. Membrangebundene Komplexe (magenta) befinden sich an der Oberfläche einer Lipidmembran (blau), was näherungsweise nur einen Rotationsfreiheitsgrad (φ) zulässt.

die Einschränkung der Rotationsfreiheit durch die Membran ausgenutzt werden, um Startwerte für zumindest zwei der drei Eulerwinkel zu erhalten. In Abb. 6.3 ist skizziert, wie die Winkel ψ_i und ϑ_i für eine sphärische Membran wie den MoMuLV Virus in Kap. 7 *a priori* abgeschätzt werden können: Die Winkel ψ und ϑ beschreiben den Normalvektor der Oberfläche, d.h. für diese Winkel gilt:

$$\vartheta_{\text{start}} = \arctan\left(\frac{\sqrt{x^2 + y^2}}{z}\right) \quad \psi_{\text{start}} = 90^\circ + \arctan\left(\frac{y}{x}\right). \quad (6.8)$$

Dabei sind x , y und z die Koordinaten der einzelnen Partikel bzgl. des Zentrums der Kugel. Diese Zuweisung von Startwerten für ψ_{start} und ϑ_{start} kann verallgemeinert werden für ellipsoide Membranen. Als solche können z.B. Zellkerne approximiert werden. Hier müssen die Koordinaten x , y und z jeweils auf ihre Halbachsen normiert werden, so daß zur Berechnung der Startwinkel (6.8) verwendet werden kann. Der Polarwinkel φ ist unbestimmt und kann zur Alignierung anfangs zufällig gewählt werden. Mittels der in Abschn. 6.1.3 erläuterten Prozedur kann er anschließend genauer approximiert werden.

Eine andere Möglichkeit, geeignete Startwerte für die Verfeinerungsprozedur zu erlangen, besteht, wenn ein ungefähres Modell der zu bestimmenden Struktur existiert. Hier kann z.B. mittels MOLMATCH (Abschn. 5.3) strukturell ähnliche Objekte im gesamten Tomogramm detektiert und vorläufig aligniert werden.

6.1.3 Alignierung des Polarwinkels φ und polare Symmetrieanalyse

Die oben beschriebene Verfeinerungsprozedur tastet die Eulerwinkel nur in einem eingeschränkten Winkelbereich ab. Während dies z.B. bei Membrankomplexen ausreichend ist für die Winkel ψ und ϑ , wird φ nicht vollständig abgetastet. Ein explizites Abtasten mit der oben beschriebenen Verfeinerungsprozedur ist zeitlich deutlich zu aufwendig, weshalb eine separate Prozedur zur vollständigen Abtastung in φ geschrieben wurde. Die Alignierung geschieht auf der Basis der *Rotational Correlation Function* (RCF), die bei klassischen *Single Particle* Techniken häufig verwendet wird, um Partikel polar zu alignieren (Penczek et al. 1992). Bei dieser Korrelationsfunktion werden Partikel und Referenz zuerst in Polarkoordinaten transformiert, bzw. in 3D in Zylinderkoordinaten z , ρ und φ . Die Transformation geschieht dabei mittels trilinearärer Interpolation und wurde in der vorliegenden Arbeit in MATLAB als Teil des TOM-Pakets implementiert. Bei der RCF wird nur bzgl. des Polarwinkels φ korreliert, so daß für die RCF gilt:

$$RCF(\varphi) = \sum_z \sum_\rho \rho \mathcal{F}_\varphi^{-1} (\mathcal{F}_\varphi A_{z\rho\varphi} \cdot (\mathcal{F}_\varphi P_{z\rho\varphi})^*). \quad (6.9)$$

Hierbei bezeichnet \mathcal{F}_φ die Fouriertransformation bzgl. φ .

φ -Alignierung. Die Alignierungsprozedur wurde in der Prozedur `av3_phialig` folgendermaßen implementiert: Die Partikel sind räumlich und bzgl. ψ und ϑ vorläufig aligniert. Dann wird ein Partikel als erste Referenz gewählt. Die RCF dieser Referenz und eines zufällig ausgewählten Partikels wird berechnet, und das Maximum bestimmt. Das Partikel wird gemäß des bestimmten Winkel rotiert und zur Referenz addiert. Diese Prozedur wird mit einem anderen zufällig ausgewählten Partikel wiederholt, und der Prozeß wird fortgesetzt bis die Polarwinkel φ aller Partikel approximiert sind. Als Konsistenztest wird die Prozedur mit mehreren Startpartikeln durchgeführt, und es wird überprüft ob die sich ergebenden Mittelungen bis auf eine Rotation um φ weitgehend identisch sind.

Symmetrieanalyse. Die RCF (6.9) kann auch zur Verifizierung von Rotationssymmetrien bzgl. der z -Achse verwendet werden. Bei Vorliegen einer Rotationssymmetrie sollte sich diese in der RCF widerspiegeln: Für eine n -fache Rotationssymmetrie muß für die Auto-RCF (ARCF) in der Abwesenheit jeglichen Störsignals gelten:

$$ARCF(\varphi) = ARCF\left(\varphi + \frac{360}{n}\right). \quad (6.10)$$

Es muß beachtet werden, daß für die ARCF wie für jede Autokorrelationsfunktion Inversionssymmetrie gilt, d.h. $ARCF(\varphi) = ARCF(-\varphi)$. Die ARCF ist im Allgemeinen zu verrauscht, um alleine eine signifikante Aussage zu ermöglichen. Die Mittelung der einzelnen ARCFs jedes Partikels i eines Partikelsatzes sollte bei Vorliegen einer Symmetrie jedoch signifikant sein.

6.2 Struktur der Kernporen aus *Dictyostelium discoideum*

Kernporenkomplexe (*Nuclear Pore Complex*, NPC) sind für den Austausch von Makromolekülen zwischen Zellkern und Zytoplasma zuständig. Diese ausgesprochen großen Makromoleküle (~ 120 MDa in Metazoa (Tiere)) bestehen aus etwa 30 verschiedenen Proteinen, den Nucleoporinen. Die funktionelle und strukturelle Charakterisierung der NPCs wird durch ihre schiere Größe und durch die Verankerung in der Zellmembran erschwert. Die bisher am höchsten aufgelöste Struktur ($\sim (12\text{nm})^{-1}$) eines NPCs wurde durch Kryo-EM einer ausgebreiteten Kernmembran (*Nuclear Envelope*, NE) von *Xenopus* erhalten (Stoffler et al. 2003). Neben der Gefahr der Beeinträchtigung der Struktur durch das Ausbreiten der NE hat dieses Vorgehen den technischen Nachteil, daß nicht die volle 3D Struktur erhalten werden kann: Die einzelnen NPCs besitzen eine Vorzugsorientierung ($\vartheta \approx 0^\circ$), so daß im Fourierraum der ermittelten Struktur ein Konus unabgetastet bleibt. Das aus dieser und vorhergehenden Arbeiten Konsensmodell ist z.B. in Ref. (Fahrenkrog et al. 2004) zusammengefaßt (Abb. 6.4). Die Struktur ist im Wesentlichen ein Zylinder mit 8-facher Rotationssymmetrie, der

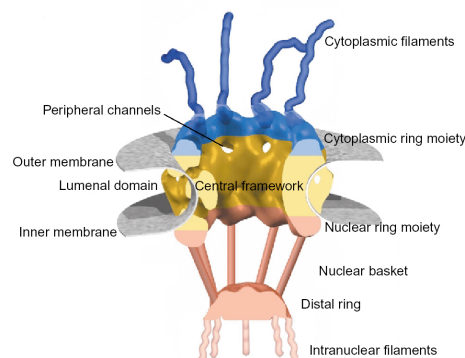


Abbildung 6.4: Konsensmodell des NPC (aus (Fahrenkrog et al. 2004)). Die Struktur aus Ref. (Stoffler et al. 2003) wurde um die in dieser Arbeit nicht oder kaum aufgelösten zyttoplasmatischen Filamente, die Kernmembran und den *Nuclear Basket* ergänzt, deren Existenz in anderen Arbeiten gezeigt wurde.

sich in der Kerndoppelmembran befindet. Auf der zytoplasmatischen Seite besitzt er sog. zytoplasmatische Filamente, und er wird auf der gegenüberliegenden Seite durch eine käfigartige Struktur abgeschlossen, den *Nuclear Basket*. Im Innern der Struktur befindet sich eine Dichte unbekannter Funktion, der *Central Transporter* (CP).

Um die Struktur von Kernporen unter möglichst physiologischen Bedingungen zu ermitteln und die kompletten dreidimensionalen Daten abzuleiten, haben wir folgende Strategie gewählt: Es wurden isolierte, intakte Nuklei des Organismus' *Dictyostelium discoideum* tomographiert, NPCs enthaltende Subtomogramme ausgeschnitten und mit der in Abschn. 6.1 beschriebenen Prozedur zu einer höheraufgelösten Struktur gemittelt. Im Bereich des zentralen Transporters wurden unter den einzelnen NPC Partikeln große Unterschiede beobachtet, die mit einer neuentwickelten Prozedur klassifiziert wurden.

6.2.1 Material und Methoden

Die Präparation und Isolation der Zellkerne aus *Dictyostelium discoideum* sowie die Aufnahme der Kippserien ist in Ref. (Beck et al. 2004) beschrieben. Es wurden insgesamt 267 individuelle NPCs aus 16 Kerntomogrammen extrahiert, die jeweils bei etwa 15 μm Defokus aufgenommen wurden. Der experimentelle Teil der Arbeit wurde von Martin Beck durchgeführt und im Rahmen dieser Arbeit wurde die Mittelung und Klassifikation der Kernporen entwickelt.

Erzeugung der ersten Referenz. Zur Approximierung der Eulerwinkel ϑ und ψ analog zu Abschn. 6.1.2 wurden die Zellkerne als Ellipsoide approximiert. Die Oberfläche wurde mathematisch beschrieben als:

$$\frac{(x - x_0)^2}{A^2} + \frac{(y - y_0)^2}{B^2} + \frac{(z - z_0)^2}{C^2} = 1. \quad (6.11)$$

Hierbei waren A , B und C die für jeden Zellkern zu bestimmenden Ellipsenhalbachsen und der Vektor (x_0, y_0, z_0) beschrieb den Mittelpunkt der Ellipse. Die Mehrzahl der Kernporen war lateral in erster Näherung rund, so daß $A = B$ gewählt wurde. Zur Bestimmung von C und z_0 wurden an drei unterschiedlichen z -Schichten interaktiv die Membran am besten beschreibenden Kreisdurchmesser A und die Kreismittelpunkte x_0 sowie y_0 bestimmt.

Die Winkel ϑ_i und ϕ_i des Partikels i ergeben sich so zu:

$$\psi_i = 90^\circ + \arctan\left(\frac{\frac{y_i - y_0}{A}}{\frac{x_i - x_0}{A}}\right) \quad , \quad \vartheta_i = \arctan\left(\frac{\sqrt{\left(\frac{x_i - x_0}{A}\right)^2 + \left(\frac{y_i - y_0}{A}\right)^2}}{\frac{z_i - z_0}{C}}\right). \quad (6.12)$$

Der Winkel φ_i wurde zufällig gewählt und die Partikel gemäß ihren zugeordneten Orientierungen gemittelt.

Alignierung. Die Prozedur `av3_trans-rot-alig.m` (Abschn. 6.1.1) wurde zu einem vorläufigen Verfeinern der Translationvektoren benutzt. Dabei wurde bei der Alignierung eine zylindrische Maske verwendet, die die Kernpore komplett umschloß, aber die Kernmembran möglichst weitgehend maskierte.

Vorläufige Konvergenz war bereits nach wenigen Iterationen erreicht (~ 5), die Prozedur wurde jedoch insgesamt knapp 30 mal iterativ wiederholt, um die Konvergenz zu überprüfen. Anschließend wurde die Rotationssymmetrie entlang der z -Achse überprüft und ein Alignierungsschritt in φ mit `av3_phialig.m` durchgeführt (Abschn. 6.1.3). Schließlich wurde die iterative Translations- und Rotationsalignierung wieder aufgenommen und nach der Konvergenz wurde die Auflösung mittels Fourierringkorrelation (FRCC) zweier Mittelungen basierend auf jeweils der Hälfte der Partikel bestimmt.

Klassifikation gemäß des Schwerpunkts des CP. Die einzelnen NPC Partikel wurden mittels einer neuentwickelten Klassifikationsprozedur in homogenere Untergruppen aufgeteilt. Die Klassifikation geschah dabei aufgrund des Schwerpunkts des CP.

Die einzelnen Partikel wurden in die durch die Alignierung bestimmten Positionen gebracht und mit einer kreisförmigen Maske gefiltert, um insignifikante, hochfrequente Information zu unterdrücken, d.h. das Signal wurde geglättet. Nach Normalisierung der Partikel gemäß ihrer Varianz wurden die außerhalb des zentralen Kanals liegenden Strukturen der Partikel durch Maskierung mit einer ellipsoiden Funktion entfernt. Die Grauwerte der Partikel wurden invertiert, so daß sie zur Masse korrespondierten. Der mittlere Grauwert des Puffers in der gemittelten Struktur wurde als Ursprung des Maßes für die Masse gewählt, auftretende negative Werte wurden zu Null gesetzt. Schließlich wurde der Schwerpunkt bestimmt, und die bestimmten z -Koordinaten wurden als Histogramm aufgetragen. An das Histogramm wurde eine doppelte Gaußverteilung angepaßt, und das Minimum der Verteilung wurde als Kriterium zur Klassifikation benutzt. Die Parameter der Gaußfunktionen wurden in MATLAB durch Minimierung der quadratischen Abweichung zum gemessenem Histogramm bestimmt. Die klassifizierten Partikel wurden erneut iterativ aligniert.

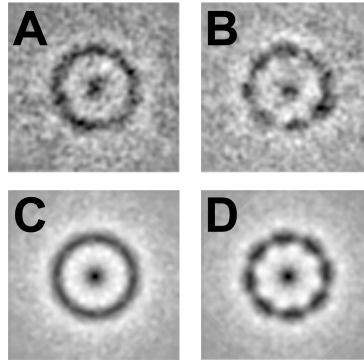


Abbildung 6.5: A: xy Schnitt der NPC Struktur vor φ -Alignierung. B: Gleicher xy Schnitt nach φ -Alignierung. C und D: Strukturen 8-fach symmetrisiert.

6.2.2 Ergebnisse

Die Tomogramme der Zellkerne ergaben, daß die tomographierten Teile der Zellkerne bis zu etwa $1 \mu\text{m}$ in z maßen. Demzufolge war das SNR der Tomogramme sehr niedrig, es reichte jedoch aus, um Kernporen manuell zu detektieren. Ohne Mittelung der Daten konnte im Wesentlichen jedoch nur der zentrale Kanal und vereinzelt der zentrale Transporter erkannt werden.

Nach der Erzeugung der initialen Referenz und anschließender iterativer Alignierung wurde eine Struktur erhalten, die entlang der z -Achse weitgehend rotationssymmetrisch war (Abb. 6.5). Parallel dazu wurde die Struktur mit der Prozedur von Walz *et al* gemittelt; die sich ergebende Mittelung war aber offensichtlich kontrastärmer und besaß eine kleinere Auflösung. Daraufhin wurde nachfolgend nur noch die in diesem Kapitel entwickelte Prozedur verwendet. Die an die vorläufige Alignierung anschließende Symmetrieanalyse deutete auf die erwarteten 8-fache Rotationssymmetrie der NPCs (Abb. 6.6). Nach Anwendung der φ Alignierung wurde diese auch in der Rekonstruktion deutlich (Abb. 6.5).

Wiederaufnahme der iterativen Alignierung mit Aufprägen der verifizierten 8-fachen

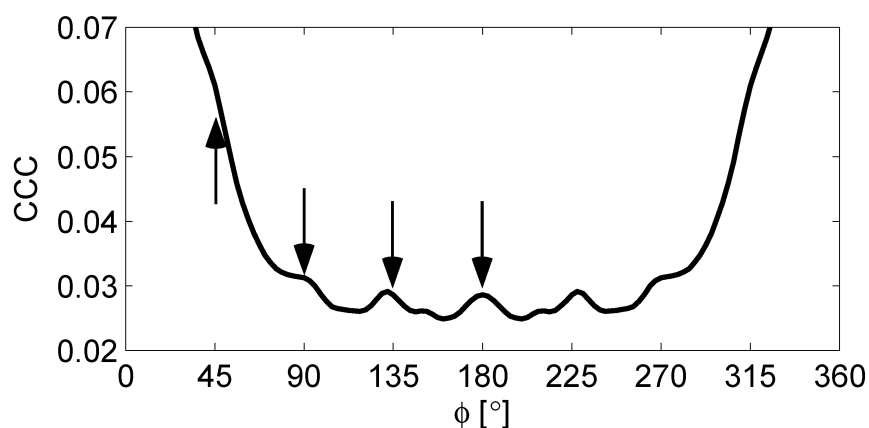


Abbildung 6.6: Symmetrieanalyse des NPC Komplexes.

Rotationssymmetrie ergab ein detaillierteres Bild des NPCs. Die finale Mittelung erreicht eine Auflösung von etwa $(8-9 \text{ nm})^{-1}$, und stellt damit die bis heute höchst aufgelöste dreidimensionale Struktur des NPCs dar. Aufgrund der Probengeometrie war es außerdem erstmals möglich, eine artefaktfreie dreidimensionale Struktur zu erhalten: Alle vorherigen EM-Strukturen (z.B. (Akey und Radermacher 1993; Stoffler et al. 2003)) konnten aufgrund des *Missing Cone* im Fourierraum in z -Richtung liegende Strukturelemente nicht auflösen. Die Mittelung ist in Abb. 6.7 gezeigt.

Visuelle Inspektion der einzelnen NPC Partikel deutete auf eine große strukturelle Fluktuation im Bereich des CP. Aus diesem Grunde wurde eine Klassifikation mittels PCA versucht. Dieses Vorgehen konnte jedoch keine charakteristischen Klassen aufzeigen, da das Verfahren durch den *Missing Wedge* der einzelnen Partikel gestört wird (siehe auch Kap. 8). Aus diesem Grunde wurden die einzelnen Partikel mit einem neuentwickelten Verfahren bzgl. des Schwerpunkts des CP klassifiziert (Material und Methoden). Das Histogramm der z -Koordinaten der Schwerpunkte des CP ergab eine Verteilung, die sich sehr gut durch eine Überlagerung zweier Gaußfunktionen beschreiben lässt (Abb. 6.8D).

Die Partikel wurden gemäß der Lokalisierung der Schwerpunkte in zwei Klassen aufgeteilt: die *Cytoplasmatic Filament* (CF) Klasse und die *Luminal Spoke Ring* (LR)

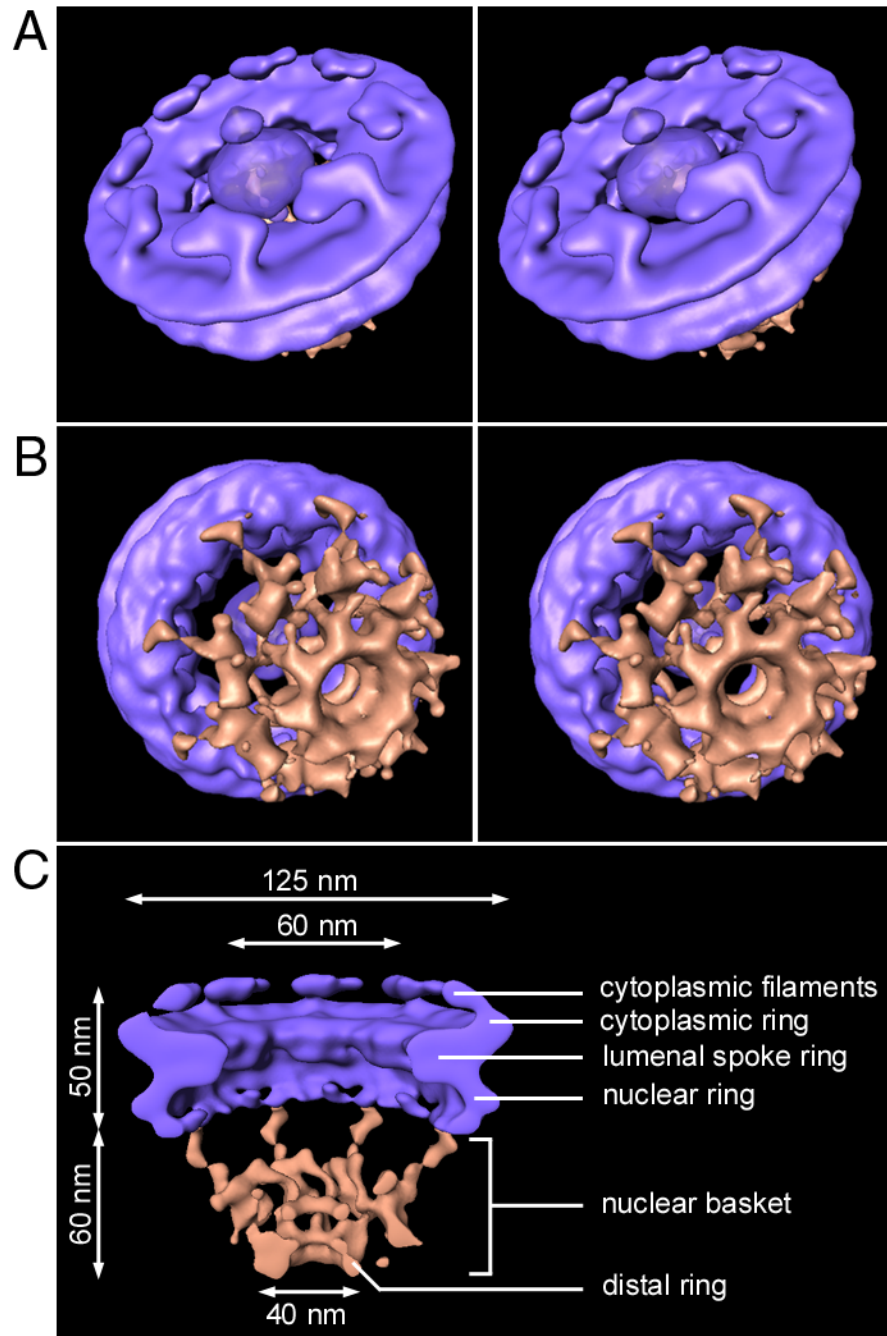


Abbildung 6.7: Ermittelte Struktur des NPC. A: Zytoplasmatische Ansicht des NPC als Stereodarstellung. B: Kernseitige Ansicht des NPC. C: Seitenansicht des halbierten NPC nach Entfernen des CPs. Die Struktur zeigt die wesentlichen Elemente des Konsensmodells (Abb. 6.4), insbesondere die zytoplasmatischen Filamente und den Nuclear Basket, die in vorhergehenden TEM-Studien unzureichend aufgelöst wurden. Die Struktur ist auch weniger ausgedehnt in z als vorher bestimmte Modelle, was wahrscheinlich daher rührt, daß diese durch den *Missing Cone* Effekt in z elongiert waren. Die ebenfalls aufgelöste Kernmembran wurde zur Darstellung manuell entfernt.

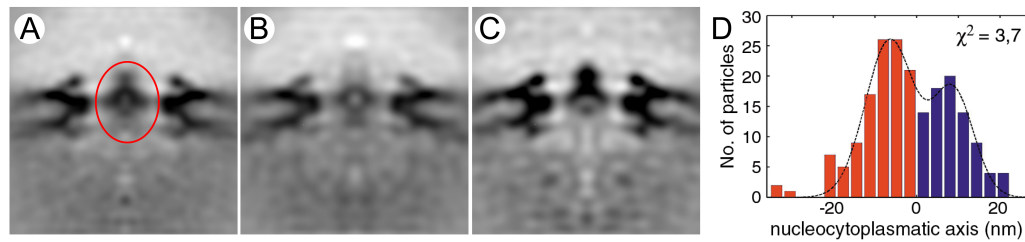


Abbildung 6.8: Klassifizierung des NPCs. A: Seitenansicht der Mittelung aller NPCs. B: LR-Klasse. C: CF-Klasse. D: Histogramm der Schwerpunkte des CPs.

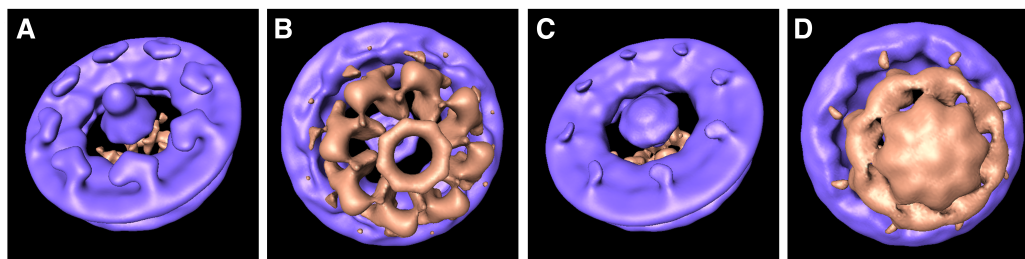


Abbildung 6.9: CF- und LR-Klasse. A und B: Oberflächendarstellung der Mittelung der CF-Klasse gesehen vom Zytoplasma (A) und vom Innern der Zellkerns (B). C und D: Gleiche Darstellungen der LR-Klasse. Der *Nuclear Basket* ist braun dargestellt.

Klasse. Die Mittelwerte der so klassifizierten Partikel weisen deutliche Unterschiede auf. So variiert erwartungsgemäß die Position des CP (Abb. 6.8 A, B, C), bei der CF Klasse ist er dem Zytoplasma näher während er in der LR Klasse dem Zellkern näher ist. Darüberhinaus sind bei der CF-Klasse die zytoplasmatischen Filamente stark ausgeprägt, was den Namen der Klasse erklärt. Der Schwerpunkt des CP ist fast auf Höhe der Filamente, und es scheint eine große Affinität zwischen den Filamenten und dem CP zu bestehen (Abb. 6.9a und b). Bei der LR Klasse befindet sich der CP etwa auf Höhe des *Luminal Spoke Rings*. Der CP erscheint bei dieser Klasse etwas massereicher und konturloser zu sein (Abb. 6.9c und d); dies deutet auf weitere strukturelle Variabilität des CP bei dieser Klasse hin. Der *Nuclear Basket* ist bei dieser Klasse ebenfalls etwas ausgeprägter.

6.3 Diskussion

Die hier entwickelte Methode zur Strukturbestimmung mittels KET und folgender Mittelung ist eine Methode um größere membrangebundene Komplexe in möglichst physiologischem Zustand zu bestimmen. Sie kombiniert die Vorteile der KET, die in der Lage ist komplexe biologische Proben in ihrer annähernd nativen Umgebung dreidimensional abzubilden mit *Single Particle* Techniken, die darauf abzielen *in silico* die Zahl der bildgebenden Elektronen pro Komplex durch Mittelung zu erhöhen. Durch diese Strategie wird eine deutlich höhere Auflösung als mit KET allein erreicht. Die Strukturbestimmung membrangebundener Komplexe wird durch mehrere Faktoren erleichtert: Der erste Schritt der Strukturbestimmung, die Lokalisierung einzelner Partikel wird dadurch erheblich vereinfacht, daß sich die Partikel nur an der Membran befinden können, die relativ einfach zu lokalisieren ist. Weiterhin können die geometrischen Randbedingungen, die durch die Assoziation mit der Membran gegeben sind, ausgenutzt werden, um eine erste Referenz zu generieren: Membrangebundene Makromoleküle wie der NPC besitzen eine Vorzugsrichtung bzgl. der Membran, so daß die Partikel in erster Näherung nur einen Rotationsfreiheitsgrad besitzen (Jiang et al. 2001). Somit kann die Beeinflussung der Struktur durch eine erste, künstliche Referenz ausgeschlossen werden. Schließlich besitzt die Alignierung von Tomogrammen zu einem 3D Modell gegenüber der Mittelung basierend auf Projektionen den Vorteil, daß die Zuordnung der Orientierungen bei 3D Daten einfacher ist. Projektionen aus verschiedenen Richtungen können einander stark ähneln, so daß es z.B. dazu kommen kann, daß Projektionen aus $\vartheta = +90^\circ$ und $\vartheta = -90^\circ$ irrtümlich gleichen Projektionsrichtungen zugeordnet werden können, was dann zu pseudo-spiegelsymmetrischen Strukturen führt (Rubinstein et al. 2003).

Die erreichbare Auflösung der Methode ist in erster Linie durch die Zahl der Partikel begrenzt. Sie kann jedoch nicht unbegrenzt gesteigert werden. Ein limitierender Faktor ist die Alignierung der Kippserien vor der 3D Rekonstruktion. Die Alignierung basiert bis heute in allen entwickelten 3D Rekonstruktionsprogrammen auf dem einfachen in Abschn. 4.4 dargestellten Modell. Die auftretenden Fehler sind von vielen Faktoren

abhängig, wie Justierung des TEMs, Vergrößerung und Dicke der Probe, und liegen bei Kryoelektronentomogrammen meist in der Größenordnung von 1-3 nm. Weiterhin wird die erreichbare Auflösung durch die CTF begrenzt: Ohne eine Korrektur des CTF Einfluß' kann das Signal nur innerhalb der ersten Nullstelle berücksichtigt werden. Weiterhin müßte der Fokusgradient bei hohen Kippwinkeln berücksichtigt werden, falls Defokuswerte in der Größenordnung der lateralen Probendimension benutzt werden. Bis zu einer Auflösung von etwa $(3 \text{ nm})^{-1}$ (Defokus $\sim 4 \mu\text{m}$) sind diese Korrekturen jedoch nicht notwendig.

Die in diesem Kapitel vorgestellte Mittelungsprozedur wurde zur *in situ* Mittelung des Kernporenkomplex' von *Dictyostelium discoideum* verwendet. Bei diesem Projekt waren die Anforderungen an die zu erreichende Auflösung deutlich unter $(3\text{nm})^{-1}$. Die Herausforderungen waren hier andere: Aufgrund der extremen Dicke der Zellkerne ($\sim 1 \mu\text{m}$) war das SNR der Tomogramme sehr niedrig. Außerdem war die Probe inhomogen in der Hinsicht, daß die an die NPCs grenzende Membran flexibel war und sich Objekte wie Teile des Endoplasmatischen Retikulums häufig in unmittelbarer Umgebung der NPCs befanden. Schließlich war die Signalinformation des NPC im Fourierraum sehr inhomogen verteilt. Die Membran besitzt am meisten Kontrast und demzufolge ist die größte Varianz entlang der k_z -Achse im Fourierraum lokalisiert. Die entwickelte Prozedur wurde diesen Anforderungen durch Maskieren der Struktur und Bandpaßfilterung signifikanter Strukturdetails während der Alignierung sowie Benutzung der *Constrained Correlation Function* gerecht. Die Auflösung der erhaltenen NPC Struktur übertrifft sogar die aller bisherigen EM Rekonstruktionen, die auf der Basis extrahierter NPCs erhalten wurden (zusammengefaßt in (Fahrenkrog et al. 2004)). Darüberhinaus hat die bestimmte Struktur im Gegensatz zu den bisherigen keinen Einfluß des *Missing Cone* auf die Struktur, was erstmalig eine genaue Bestimmung der Ausmaße des NPCs in z -Richtung ermöglichte.

Das Verfahren ist auf vielerlei massereiche Membrankomplexe anwendbar und eine Erweiterung auf wasserlösliche Komplexe ist möglich. Zum Ausblick sei auf Kap. 9 verwiesen.

7. Strukturbestimmung des Env Protein Komplex' von MoMuMLV

Murine Leukemia Viren (MuLV) sind Viren, die bei Nagern Leukämie verursachen. Das Erbgut von MuLV ist als RNA in den Viren verpackt. In der infizierten Zelle wird die RNA über ein DNA Intermediat repliziert, weshalb MuLV zur Familie der *Retroviridae* gehört. Retroviren übersetzen in der infizierten Zelle ihre RNA mittels Reverser Transkriptase (*Reverse Transcriptase*, RT) in DNA, die dann ins Genom der Wirtszelle eingebaut wird. Die virale RNA bzw. das DNA Intermediat beinhalten sehr starke Promotoren, die zur starken Transkription der viralen Proteine in der infizierten Zelle führt. Neben der Produktion der viralen Proteine kann es dadurch in den infizierten Zellen zur Überexpression von Onkogenen kommen (zur detaillierten Darstellung sei auf (Dimmock et al. 2001) verwiesen). Deshalb sind viele Retroviren karzinogen. Allerdings wird der Replikationsmechanismus in der heutigen Biotechnologie auch elegant ausgenutzt: Der Einbau von Plasmiden in das Genom von Zellen mittels Retroviren wird z.B. in der Gentherapie ausgenutzt.

Retroviren werden gemäß ihrer genetischen und strukturellen Eigenschaften in insgesamt sieben verschiedene Gattungen unterteilt (www.ncbi.nlm.nih.gov/ICTV). MuLV gehört dabei zur Gattung der Gammaretroviren. Eine andere sehr prominente Gattung der Retroviren sind Lentiviren zu denen der Human Immunodeficiency Virus (HIV) oder auch der Simian Immunodeficiency Virus (SIV) gehören. Beide Gattungen sind strukturell sehr ähnlich, und insbesondere MuLV, HIV und SIV werden häufig gemeinsam diskutiert.

MuLV war einer der ersten isolierten Retroviren und war lange einer der meiststudierten Viren. Erst nach der Entdeckung von HIV hat sich der Schwerpunkt retroviraler

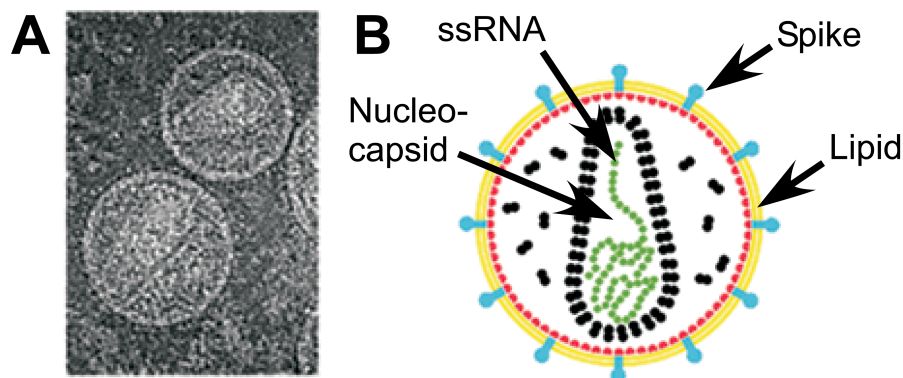


Abbildung 7.1: Aufbau von Retroviren am Beispiel von HIV (adaptierte aus (Briggs et al. 2004a)).
 A: Kryo-EM Bild eines reifen HI-Virus (Durchmesser ~ 100 nm). B: Schematischer Aufbau. Farbkodiert sind die Lipidschicht (gelb), die Env Proteine (blau), die NC (grün), CA (schwarz) und MA-Domäne (rot).

Forschung verlagert. Allerdings spielt MuLV noch immer eine sehr wichtige Rolle bei der strukturellen Erforschung von Retroviren.

Retroviren sind im allgemeinen 80-100 nm große Partikel, die mit dornenförmigen Proteinkomplexen, sog. *Spikes*, besetzt sind, wie in Abb. 7.1 bei HIV zu sehen ist. Die einsträngige RNA (*single-stranded RNA*, ssRNA) befindet sich in einem sog. *Nucleocapsid*, das symmetrisch oder auch pleomorph sein kann. Das Kapsid ist eingebettet in eine Lipidhülle die mit Spikes besetzt ist.

Das Erbgut von Retroviren enthält drei Gene, die für die Replikation notwendig sind. Dies sind Pol, Gag und Env. Das Pol-Gen kodiert drei Enzyme: Eine RT, eine Integrase und eine Protease. Alle drei Enzyme sind zur Replikation der Viren in der Wirtszelle unabdingbar: Die RNA muß in DNA übersetzt werden und ins Genom der Wirtszelle integriert werden. Da die viralen Proteine als eine einzige Aminosäurekette translatiert werden, ist die virale Protease notwendig um das Polypeptid an eindeutigen Punkten zu zerschneiden. Das Gag-Gen kodiert drei Proteine, das sog. Matrix Protein (MA), das Kapsid Protein (CA) und das Nucleocapsid Protein (NC). Im unreifen Virus bilden die Gag-Domänen einen Komplex, der dann durch die Einwirkung

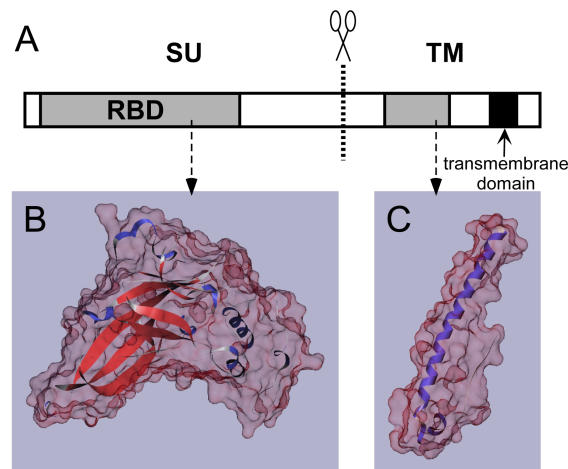


Abbildung 7.2: A: Schematischer Aufbau des Env-Gens von MuLV. B: Struktur der 30 kDa *Receptor Binding Domain* (RBD), die im Durchmesser etwa 5 nm mißt (Fass et al. 1997). C: Struktur eines 5.6 kDa Fragments des TM Proteins (Fass et al. 1996).

der Protease zerfällt. Im reifen Virus ist die Lokalisierung der Gag-Proteine wie durch die Benennung angedeutet: MA bleibt an der Lipidmatrix assoziiert, CA bildet die Wand des Nucleocapsids und NC befindet sich wie auch die Pol Proteine im Innern des Kapsids (Briggs et al. 2004a). Die Kapsid Proteine bilden das Kapsid durch ein Netzwerk von (Homo-)Hexameren. Die atomare Struktur der Hexamere der N-terminalen Domäne wurde für MuLV kürzlich aufgeklärt (Mortuza et al. 2004). Weiterhin wurde die genaue Wechselwirkung von NC mit der viralen RNA, die das "Verpacken" der Viren auslöst, strukturell mittels NMR aufgeklärt (D'Souza und Summers 2004). Das Env-Gen von MuLV kodiert zwei Proteine. Beide Domänen werden als eine Aminosäurekette translatiert, die dann durch die virale Protease zerschnitten werden (siehe Abb. 7.2). Die durch proteolytische Spaltung entstehenden Untereinheiten heißen in der Reihenfolge ihrer Translation *Surface Subunit* (SU) und *Transmembrane*

Subunit (TM). Die strukturelle Aufklärung der Env Proteine ist weit weniger weit fortgeschritten als die der Gag Proteine. Die Env Proteine bilden trimere, Typ-I Transmembranproteinkomplexe (Transmembranproteine mit nur einer Transmembrandomäne), die sich an der Oberfläche von Retroviren befinden. Diese Proteine bilden die bereits erwähnten Spikes an der Oberfläche von Retroviren. Während der viralen Infektion vermitteln die Env Proteine die Bindung des Virus' an den Rezeptor in der Wirtszelle und die anschließende Fusion von viraler und zellulärer Membran. Dabei unterliegt Env einer weitgreifenden Konformationsänderung (Eckert und Kim 2001; Colman und Lawrence 2003).

7.1 Einführung in die Struktur des Env Protein Komplex'

Obwohl hochaufgelöste Strukturen vieler Env Fragmente bestimmt wurden (Fass et al. 1996; Fass et al. 1997; Kwong et al. 2005; Kobe et al. 1999; Caffrey et al. 1998; Chan et al. 1997; Weissenhorn et al. 1997), ist die Struktur des intakten Env Komplex', wie er sich auf der viralen Membran befindet, nur auf einer groben morphologischen Ebene als trimere Projektion bekannt (Wilk et al. 2000; Center et al. 2001; Zhu et al. 2003). Die Quarternärstruktur des Env Komplexes an der Oberfläche frischer Viren ist der Ausgangspunkt für die strukturellen Umlagerungen des Komplex' im Verlaufe des Eintritts des Virus' in die Zelle, und sie ist somit essentiell, um den Infektionsvorgang strukturell zu verstehen. Die Bestimmung der Struktur retroviralen Envs mittels Röntgenkristallographie wird durch mehrere Punkte erschwert: Die Proteine sind stark glykosiliert (bis zu 50% der Gesamtmasse), was die Kristallisation behindert. Weiterhin sind Transmembranproteine mit nur einer Transmembrandomäne als notorisch schwierig zu kristallisieren. Schließlich ist es problematisch, große Mengen an löslichem Env herzustellen, wie zur Kristallisation benötigt; die Env Präparationen sind darüberhinaus aufgrund struktureller Instabilitäten der Komplexe häufig inhomogen.

Elektronenmikroskopie der Komplexe umgeht viele der Schwierigkeiten der Röntgenkristallographie und kann trotz der geringeren Auflösung nützliche Informationen liefern. Konventionelle *Single Particle* Techniken erfordern typischerweise auch Solubilisierung und Aufreinigung der Komplexe, was wiederum strukturelle Veränderungen von Env induzieren kann. Um einige dieser Probleme zu umgehen, wurde *Cross-Linking* der Env Komplexe des *Simian Immunodeficiency Virus* (SIV) vor der Solubilisierung durchgeführt, was zu einer 2-D Studie führte, die eine dreieckige, dreiflügelige Morphologie der Partikel ergab (Center et al. 2001). Aufnahmen von getrocknetem, schwermetallkontrastiertem *Human Foamy Virus* (HFV) waren die Basis für die erste *in situ* 3D Rekonstruktion von Env mittels *Single Particle* Techniken (Wilk et al. 2000). Aufgrund der Trocknung und der quantitativ schwer beschreibbaren Kontrastierung der Komplexe ist eine weitgehende Interpretation dieser Daten jedoch nicht möglich.

ET ganzer Viren ermöglicht es, auf Solubilisierung der Env Komplexe zu verzichten und die Anwendung dieser Technik auf getrockneten, schwermetallkontrastierten SIV mit nachfolgender 2D Analyse der Env Partikel gab ebenfalls einen Einblick in die Morphologie von Env im Präfusionsstadium (Zhu et al. 2003). Die Präparation verhinderte jedoch wiederum eine detaillierte Analyse der Daten. Im Gegensatz dazu ist die Auswertung von KET-Daten durch die Präparation nicht limitiert.

In dieser Arbeit wurde der trimere Env Komplex des Moloney MuLV, einer Unterart des MuLV, in seiner Präfusionsform studiert. Dabei haben wir Tomogramme ganzer Viruspartikel aufgenommen, Env Komplexe enthaltene Subtomogramme extrahiert und mittels des in Kap. 6 beschriebenen Algorithmus gemittelt. Env hat eine Masse von nur etwa 270 kDa, was die untere Grenze für die Anwendung von *Single Particle* Techniken darstellt. Die so ermittelte Struktur von MoMuLV Env hat eine Auflösung von etwa $(3 \text{ nm})^{-1}$ und gibt einen Einblick in die 3D Morphologie des Komplex'. Die ermittelte Dichte ermöglicht eine Lokalisierung der hochaufgelösten RBD Struktur (Fass et al. 1997) innerhalb des trimeren Komplex'.

7.2 Material und Methoden

7.2.1 Herstellung der Viren

Die Viren wurden von Deborah Fass (Weizmann Institut, Israel) hergestellt wie in (Förster et al. 2005) beschrieben, und die Prozedur soll hier nur kurz dargestellt werden: Der Überstand MoMuLV produzierender CL-1 Zellen wurde in einem Sucrosegradienten aufgereinigt, und anschließend wurde die Sucrose durch Dialyse mit Tris Puffer entfernt.

7.2.2 Präparation und Datenaufzeichnung

Die zu untersuchenden EM Grids wurden präpariert, indem etwa je $3 \mu\text{l}$ Virus Lösung und kolloidales Gold auf einem 'Lacey Carbon Grid' (Ted Pella, Redding, Kalifornien) pipettiert wurden. Überschüssige Flüssigkeit wurde mit Filterpapier abgesaugt und in flüssiges Ethan eingeschossen (siehe Abschn. 4.2). Kippserien wurden auf dem Philips CM300 TEM (Abschn. 4.1) mit *Zero-Loss* Filterung aufgenommen. Insgesamt 14 Kippserien wurden mit unterschiedlichen Parametern aufgenommen: Der Kippwinkelbereich betrug immer -66° bis 66° , das Kippwinkelinkrement wurde zwischen 1.5° und 3° und der nominelle Defokus zwischen 6 und $3 \mu\text{m}$ variiert. Die bei $-6 \mu\text{m}$ aufgenommenen Daten wurden für die 3D Visualisierung der einzelnen Viruspartikel verwendet, während die näher am Fokus aufgenommenen Kippserien zur Mittelung des Env Komplexes verwendet wurden. Die effektive Vergrößerung betrug bei allen Kippserien 55,000, was einer Pixelgröße von 5.5 \AA entsprach. Die Projektionen wurden anschließend mit Hilfe der Gold Marker unter Verwendung des TOM-Pakets aligniert (Abschn. 4.4), wobei hier jeweils ein Marker in der unmittelbaren Nähe des zu rekonstruierenden Viruspartikel als Referenzpunkt genommen wurde, um den Alignierungsfehler lokal zu minimieren. Die 3D Rekonstruktion erfolgte mit dem EM Programm mit k -gewichteter Rückprojektion (Abschn. 4.4).

7.2.3 Bildanalyse und Mittelung

Tomogrammelemente, die für Env Komplexe gehalten wurden, wurden manuell in Tomogrammen von Einzelviren lokalisiert. Die Detektion wurde durch starke Tiefpaßfilterung der Tomogramme auf etwa 10 nm Auflösung erleichtert. Subtomogramme mit $40 \times 40 \times 40$ Voxeln wurden extrahiert und eine erste Referenz kreiert. Diese wurde unter der Annahme generiert, daß die dreifache Symmetrieachse Env Partikel senkrecht auf der näherungsweise sphärischen Lipidmembran stehen, und es wurden zufällige Startwerte für den Polarwinkel φ gewählt (siehe Abschn. 6.1.1). Anschließend wurden die Partikel mit der in Abschn. 6.1.1 beschriebenen Prozedur `av3_trans-rot-align.m` vorläufig aligniert. Nach drei Iterationen wurden die Partikel mit der Prozedur `av3_phialign.m` aus Abschn. 6.1.3 entlang φ aligniert. In der Folge wurde die iterative Verfeinerung der Translationsvektoren und Eulerwinkel mit `av3_trans-rot-align.m` fortgeführt, bis die Prozedur konvergierte. Die Auflösung der Struktur wurde mit Fourier-Ring Correlation zweier Mittelungen, die aus je der Hälfte der Partikel generiert wurden, bestimmt (Prozedur `av3_resdet.m`), wobei ein Koeffizient von 0.5 als Schwellwert zur anschließenden Tiefpaßfilterung benutzt wurde (Abb. 7.6). Während der Alignierung wurde das Signal außerhalb der ersten Nullstelle der CTF durch einen Tiefpaßfilter entfernt.

Zur Visualisierung ganzer Viruspartikel wurde das Signal mittels nicht-linearer anisotroper Diffusion verstärkt (Frangakis und Hegerl 2002). Alle Rechnungen wurden auf Linux-PCs mit 1.8 GHz Opteron Prozessoren (AMD, Sunnyvale, Kalifornien) und 8 Gb Speicher durchgeführt.

7.2.4 Fitting der RBD

Das Einpassen der Röntgenstruktur (*Fitting*) von RBD in die Env Dichte wurde mittels Kreuzkorrelation unter Verwendung des TOM-Pakets bewerkstelligt: Das elektrostatische Potential der RBD wurde wie in Abschn. 5.2 durch Summation der Ordnungszahlen der atomaren Koordinaten der PDB-Datei 1AOL auf einem Gitter mit

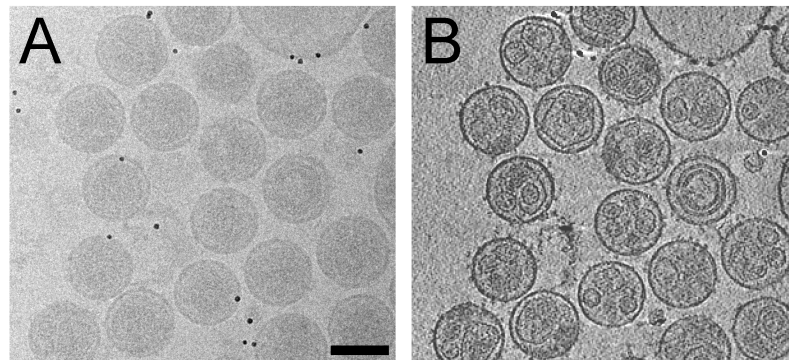


Abbildung 7.3: A: Ausschnitt aus der 0° Projektion. Der nominelle Defokus dieser Kippserie beträgt $5 \mu\text{m}$. Der Maßstab entspricht 100 nm . B: Korrespondierender Schnitt entlang der xy -Ebene des Tomogramms.

5.5 \AA Maschengröße erhalten. Diese Dichte wurde anschließend mit der angenommenen CTF gefaltet und auf eine Auflösung von 2.7 nm tiefpaßgefiltert. Schließlich wurde die 6-dimensionale Kreuzkorrelationsfunktion bestimmt, wobei die Eulerwinkel mit einem Inkrement von 5° variiert wurden. Die maximale Korrelation betrug etwa 0.5 .

7.3 Ergebnisse

7.3.1 Morphologie der Viruspartikel

Rekonstruktionen von etwa 100 Viruspartikeln, die aus 14 Tomogrammen stammen, wurden analysiert. Die Viren messen etwa 100 nm im Durchmesser (Abb. 7.3), der etwas heterogen variiert ($\sim \pm 10 \text{ nm}$) (Yeager et al. 1998). Das SNR der Tomogramme war ausreichend, um eine manuelle Segmentierung der Viren in *Core*, Lipid Doppelmembran und Env Komplexe durchzuführen (Abb. 7.4). Die Morphologie der viralen *Cores* ist der für den Mouse Mammary Tumor Virus berichteten ähnlich (Briggs et al. 2004b): Die *Cores* sind exzentrisch lokalisiert und haben unregelmäßige Formen; Partikel mit mehreren Kapsiden pro Virus sind häufig beobachtet worden.

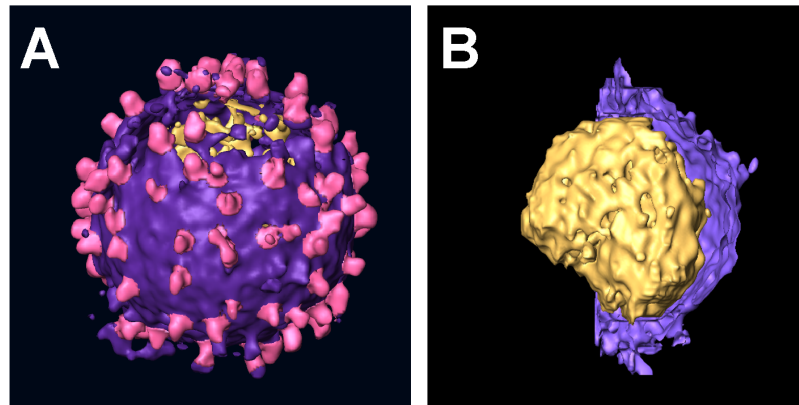


Abbildung 7.4: A: Isosurface eines Viruspartikels. Der Datensatz wurde bei einem Defokus von etwa $6 \mu\text{m}$ aufgenommen und das Signal des resultierenden Tomogramms wurde mittels nicht-linearer anisotroper Diffusion (NAD) verstärkt (Frangakis et al. 2001). Die Segmentierung in Lipidmembran (blau) und Env Komplexe (magenta) wurde manuell ausgeführt. B: Ein anderes Viruspartikel, das manuell in Lipidhülle (blau) und Kapsid (gelb) segmentiert wurde. Dieser Datensatz wurde bei einem Fokus von etwa $-3 \mu\text{m}$ aufgenommen und wiederum mittels NAD nachbehandelt.

Die Env Komplexe konnten als senkrecht aus der Virusmembran herausragend aufgelöst werden (Abb. 7.3 und 7.4). Jeweils etwa 50 ± 20 Spikes mit einem Durchmesser von ca. 10 nm wurden manuell auf insgesamt 59 einzelnen Viren aus 4 Kippserien lokalisiert. Die Zahl der Env Komplexe pro Virus unterschätzt die tatsächliche Zahl wahrscheinlich um etwa 50 %, da Komplexe an den oberen und unteren Polen durch den Missing-Wedge Effekt schwierig zu erkennen sind. Bei einigen Env Partikeln konnte schon vor der Mittelung eine dreizählige Symmetrie erahnt werden.

7.3.2 Struktur des Env Komplexes

Die vorläufige Env-Struktur besaß vor dem Aufprägen jeglicher Symmetrie eine eindeutige dreifache Rotationssymmetrie entlang der z -Achse (Abb. 7.5). Die finale Rekonstruktion des MoMuLV Env-Komplexes erreicht eine Auflösung von etwa $(2.7 \text{ nm})^{-1}$ (Abb. 7.6). In Gesamtgröße, Form und Symmetrie ist die Projektion der erhaltenen Struktur konsistent mit den publizierten 2D Daten (Center et al. 2001; Zhu

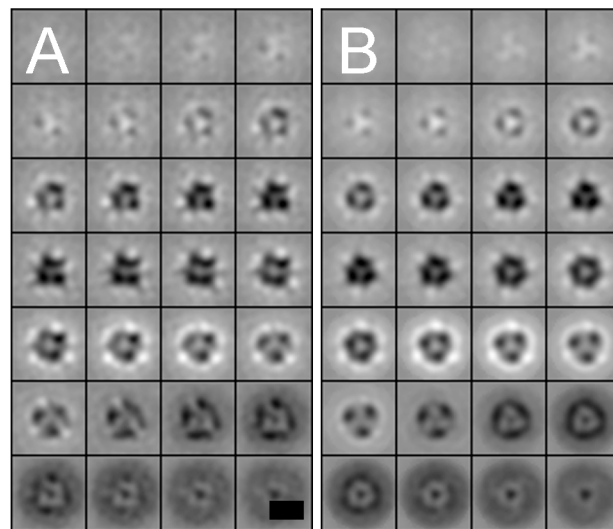


Abbildung 7.5: Env Struktur visualisiert als xy -Schnitt entlang der z -Achse (senkrecht zur Membran) vor (A) und nach (B) Aufprägen 3-facher Rotationssymmetrie entlang der z -Achse. Die Rekonstruktion ist aus 1114 einzelnen Env Subtomogrammen erhalten, aufgenommen bei $5 \mu\text{m}$ Defokus. Der Maßstab entspricht 10 nm und jeder Schnitt korrespondiert zu 0.55 nm Dicke. Die letzten z -Abschnitte (untere Reihe) korrespondieren zur Virusmembran.

et al. 2003; Chen et al. 2004). Die absolute Händigkeit der Struktur ist bestimmt, da die Kippachse des TEMs etwa bekannt ist (Abschn. 2.3).

Eine Oberflächendarstellung der 3D Struktur nach Symmetrisierung ist in Abb. 7.7a gezeigt. Env besteht aus einem extraviralen Hauptteil, der ausgeprägte Händigkeit besitzt, was auch kürzlich für SIV Env im Vor-Fusionsstadium berichtet wurde (Zhu et al. 2003). Auf der viroplasmatischen Seite des Env Komplex' wurde praktisch keine Dichte aufgelöst, was mit der Strukturvorhersage von MoMuLV anhand der Sequenz korrespondiert, die weniger als 4 kDa Masse ($\sim 4 \%$ der Gesamtmasse von MoMuLV) am C-Terminus der Transmembranregion postuliert. Der extravirale Hauptteil wird mit der Virusmembran durch drei Beine verbunden, die einen zentralen Hohlraum einschließen.

Die 3D Struktur von MoMuLV Env beinhaltet eine Reihe bemerkenswerter Strukturmerkmale, die in 2D Schnitten normal zur Membran durch die Rekonstruktion

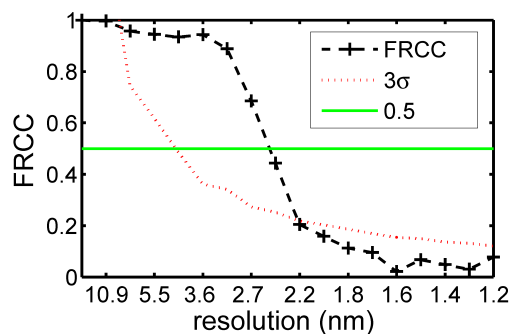


Abbildung 7.6: Fourier-Ring Korrelation von zwei Mittelungen der Env Partikel aus je 50% der Partikel als Funktion der Auflösung.

gesehen werden (Abb. 7.7c). Das diffuse Grau in den z -Schnitten 28 bis 26 korrespondiert zur Lipiddoppelschicht. Die Schichten 27 und 28, die sich etwa in der Mitte der Doppelmembran befinden, enthalten drei dichte Punkte, die etwa je 4.5 nm voneinander entfernt sind. Obwohl die Auflösung der Struktur nur etwa $(2.7 \text{ nm})^{-1}$ beträgt, suggeriert die dichte, punktförmige Morphologie, daß sie von einer etwa 1-1.5 nm im Durchmesser messenden Unterstruktur herrühren, was einer α -Helix entsprechen würde. Das Muster wird an der Oberfläche der Membran komplexer: Hier, in den Schichten 23 bis 25, beginnt eine weitere, undeutliche Struktur in größerer Distanz zur Symmetrieachse aufzutauchen. Weiter außerhalb der Membran kondensiert die Dichte in zwei kompakte Unterstrukturen. In den Schichten 18 bis 21 können die dreifach symmetrischen Untereinheiten klar in zwei Dichten aufgeteilt werden, die etwa 2.5-3 nm voneinander entfernt sind.

Folgt man der Dichte weiter in extraviraler Richtung, so besitzt sie in den Schichten 13 bis 16 wieder eine charakteristische Struktur: Jede der drei symmetrischen Untereinheit ist wieder in zwei Domänen aufgeteilt, wobei jede größer ist als die eben beschriebenen vermeintlichen α Helizes (etwa 2.5-3.5 nm im Durchmesser).

Retrovirales Env besteht wie eingangs erwähnt aus SU und TM. Berücksichtigt man das Modell für Env Aktivität, bei dem die TM Untereinheit zwischen Prä- und Postfusionszustand eine Konformationsänderung großen Umfangs durchläuft (Colman und

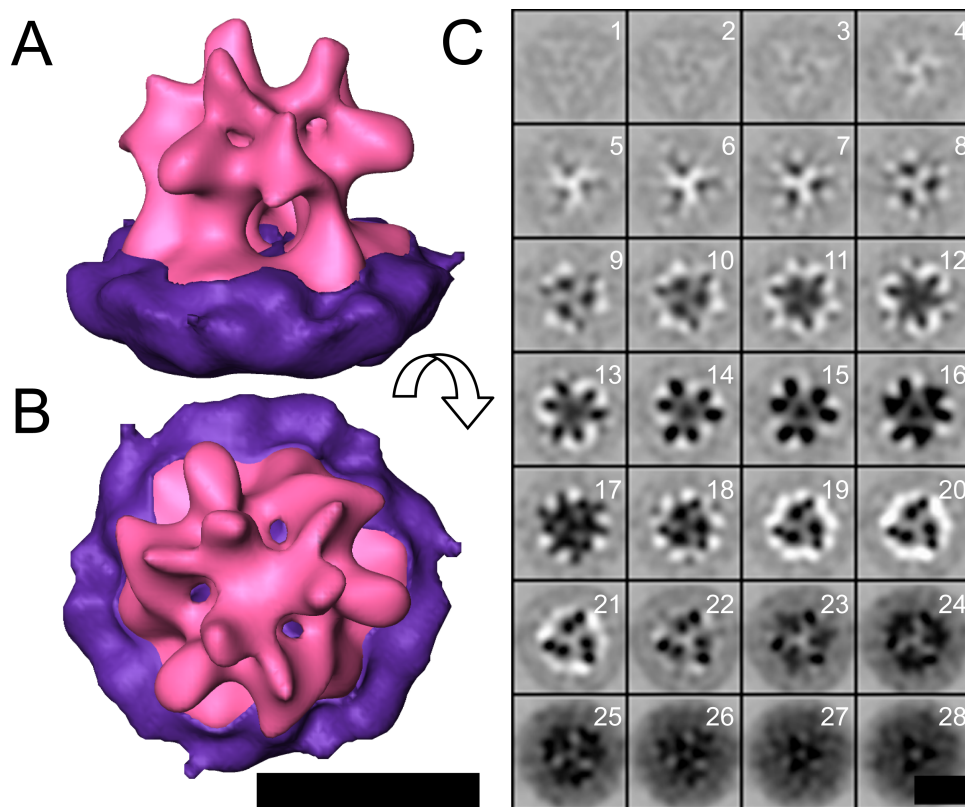


Abbildung 7.7: A and B: Isosurface Representation des ermittelten Env Mittelwerts bei $4 \mu\text{m}$ Defokus als Seitenansicht (A) and als Ansicht von oben (B). Manuelle Segmentation der viralen Membran (violett) and des Komplex' (magenta) wurde zu Visualisierungszwecken durchgeführt. C: Visualisierung der Env Dichte wie in Abb. 7.5. Der Maßstab korrespondiert zu 10 nm.

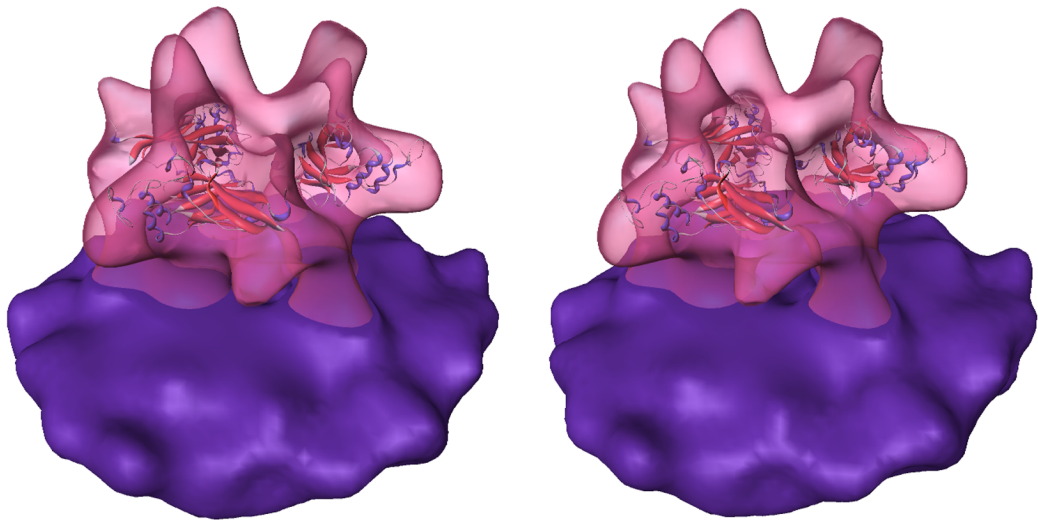


Abbildung 7.8: Stereo Ansicht der vorgeschlagenen Position der RBD (Fass et al. 1997) im trimeren Env Komplex.

Lawrence 2003; Eckert und Kim 2001), so erscheint es wenig sinnvoll, die Kristallstruktur des SU Fragments (Fass et al. 1997) im Komplex lokalisieren zu wollen: Abgesehen von der geringen Masse des Fragments macht die potentielle Flexibilität das Fitten praktisch unmöglich. Jedoch unterliegt die RBD von SU wahrscheinlich keiner Konformationsänderung, da sie sowohl *in vitro* als auch im assemblierten Virus an den Rezeptor bindet (Davey et al. 1999). Die RBD mißt etwa 6 nm im Durchmesser entlang ihrer größten Ausdehnung (Abb. 7.2B). Sie ist L-förmig aufgebaut und besteht aus einem β -Sandwich von etwa $4 \times 3 \times 2$ nm und einer helikalen Unterdomäne von etwa 3 nm Durchmesser. Die erhaltene Struktur des Env Komplexes läßt eine Lokalisierung der RBD nur im distalen Ring möglich erscheinen. Um die RBD genauer im Komplex zu lokalisieren, wurde die Röntgenstruktur mittels Kreuzkorrelation in die EM Dichte gefittet. Die so gefundene Position (Abb. 7.8) erscheint plausibel: Die RBD wird von der Einhüllenden der EM-Dichte umschlossen, die vermeintliche Bindungsstelle von RBD ist exponiert während evolutionär besonders gut erhaltene

Residuen ins Innere der Env Struktur zeigen, und es ist zusätzliche Dichte am C-Terminus der RBD vorhanden, so daß das für die Röntgenstrukturanalyse C-terminal gekürzte SU hier seine Fortsetzung finden kann. Die erhaltene Auflösung der Env Struktur und die Signifikanz der erhaltenen Maxima der Korrelationsfunktion lassen es jedoch nicht zu, die vorgeschlagene Position als eindeutig und sicher zu bezeichnen.

7.4 Diskussion

Das Studium von MoMuLV Env mittels KET *in situ* hat einige Vorteile gegenüber anderen Verfahren. Die Env Struktur ist bestmöglich erhalten, denn die Gefahr einzelne Env Komplexe durch den Einsatz von Detergenzien zu zerstören ist nicht vorhanden. Weiterhin ist Env in seiner natürlichen (Lipid-) Umgebung, so daß die Quaternärstruktur des Komplexes erhalten bleibt. In der praktischen Arbeit hat KET den Vorteil, daß die Env Komplexe nicht überexprimiert und aufgereinigt werden müssen. Schließlich birgt diese Technik auch das Potential, Env Komplexe in unterschiedlichen Zuständen des Infektionszyklus' aufzulösen (Heymann et al. 2003).

Die Env Struktur von MoMuLV, die in dieser Arbeit bestimmt wurde, besteht aus sehr großen, massereichen Domänen, die etwa 5 nm von der Membranoberfläche entfernt sind. Diese Masse ist mit der Membran durch insgesamt drei Beine pro Trimer verbunden. Jedes Bein kann weiterhin in zwei unterscheidbare, aber eng zusammenliegende Unterstrukturen aufgelöst werden. Die Morphologie von MoMuLV unterscheidet sich etwas von der einzigen bisher hochaufgelösten Struktur eines Spike Protein Komplexes eines eingehüllten Virus' in seiner potentiell nativen Prä-Fusionsform: dem Hemagglutinin (HA) Komplex des Orthomyxovirus' Influenza (Wilson et al. 1981) und dem F Protein des Paramyxovirus' Newcastle Disease Virus (Chen et al. 2001). Diese Proteinkomplexe sind relativ kompakte Strukturen, die etwa die Form auf dem Kopf stehender Bowlingkegel besitzen und keinerlei innere Aushöhlungen besitzen. Die Aushöhlung entlang der Symmetrieachse ist jedoch schon bei anderen Fusionsproteinen gefunden worden, nämlich bei Alphaviren (Lescar et al. 2001) und Flaviviren

(Zhang et al. 2002), wenn diese im Gegensatz zu MoMuLV auch Typ II Fusionsproteine sind. Das Fehlen einer Dichte entlang der Trimer Achse in der Nähe der Membran unterstützt nicht die Relevanz des *Spring-loaded* Modell (Carr und Kim 1993) für retrovirale Env Proteine. Dieses Modell wurde ursprünglich entwickelt, um die Konformationsänderungen von Influenza HA nach Aktivierung des Mechanismus' zur Fusion der Membranen zu erklären. Sollte es auch für Retroviren gültig sein, so sollten die *Coiled Coil* Strukturen, die in TM Peptiden *in vitro* auftreten, im nativen Komplex in seiner Präfusionsform nicht vorkommen. Stattdessen sollten sich diese *Coiled Coil* Strukturen erst nach der Initiierung der Fusion mit der Zellmembran ausbilden. Da die *Coiled Coil* Struktur eine kompakte, stäbchenförmige Form hat (Fass et al. 1996), müßte sie sich genau auf der dreizähligen Symmetrieachse des Env Komplexes befinden. Die von uns bestimmte Env Struktur ist inkompatibel mit solch einer trimeren *Coiled Coil* Struktur in unmittelbarer Membrannähe (~ 5 nm). Es wäre zwar möglich, daß diese *Coiled Coil* Strukturen weiter von der Membran entfernt lokalisiert sind, dies ist jedoch unwahrscheinlich, da nur etwa 40 Aminosäurereste die *Coiled Coil* Struktur mit der Transmembranregion verbinden. Stattdessen erscheint es wahrscheinlich, daß die TM Polypeptide zu den "Beinen" des Env Komplexes beitragen und erst nach einer Konformationsänderung in die dreisträngige *Coiled Coil* Struktur übergehen. Die Ausmaße der zwei Unterstrukturen, die die Beine aufbauen und die Aushöhlung an der Basis des Env Komplexes umschließen (z -Schichten 18 bis 22 in Abb. 7.7C), sind den in der Membran auftretenden Strukturen ähnlich (z -Schichten 27 und 28), bei denen wir annehmen, daß es sich um α -Helizes handelt. Strukturvorhersagen anhand der Sequenz von retroviralen TM postulieren einen hohen Anteil von Helizes und hochaufgelöste Strukturen von TM Fragmenten, denen das Fusionspeptid und die Transmembrandomäne fehlt, sind helikal aufgebaut (Fass et al. 1996; Kobe et al. 1999; Caffrey et al. 1998; Chan et al. 1997; Weissenhorn et al. 1997). Ein mögliches Modell ist, daß die Beine aus einem Paar antiparallel angeordneter Helizes besteht, wobei die äußere Helix aus der Sequenz unterhalb des Fusionspeptids

bestehen könnte. Diese Helix, die nach der Fusion die dreisträngige *Coiled Coil* Struktur bildet, würde sich mit ihrem N-Terminus nahe der Membran befinden. In dieser Orientierung würde das Cys-X6-Cys-Cys Motiv (X bezeichnet eine beliebige Aminosäure außer Cystein) am C-Terminus der Helix (Fass et al. 1996) günstig positioniert sein, um eine kovalente Bindung mit einem Cystein in der C-terminalen Region von SU einzugehen, das sich wahrscheinlich über den Beinen befindet. An diesem Punkt würde die TM Struktur ihre Richtung ändern und sich in Richtung Membran gemäß der Strukturvorhersage von TM als eine zweite Helix fortsetzen. Es muß jedoch erwähnt werden, daß der Abstand der potentiellen Helizes in den z -Schichten 18 bis 22 etwa 2.5-3 nm beträgt anstatt 1.2-1.5 nm, die man für enggepackte Helizes erwarten sollte.

Etwa 5-6 nm oberhalb der Membran können zwei sehr massereiche Domänen in der Env Struktur aufgelöst werden. Die relativen Größen dieser Domänen, bestimmt anhand der Durchmesser der z -Schichten 14 und 15 in Abb. 7.7C, korrespondieren etwa zur Zahl der Aminosäurereste von RBD (~ 240) und zum C-terminalen Abschnitt von SU (~ 180). Da MoMuLV RBD sich selbstständig faltet und eigenständig funktional ist (Davey et al. 1997), kann erwartet werden, daß es Bestandteil einer der beiden Domänen ist. Die C-terminale Region von SU besitzt diese Eigenschaften nicht und Strukturvorhersagen postulieren keine kompakte Domäne. Der hohe Anteil an vorhergesagten β -Faltblattstrukturen (www.embl-heidelberg.de/predictprotein, Ref. (Rost 1996)) läßt es jedoch möglich erscheinen, daß die C-terminale SU Region im nativen Komplex eine globuläre Struktur einnimmt. Die RBD wurde mittels Kreuzkorrelation in die Env Dichte hineingepaßt, es muß jedoch beachtet werden, daß die Lösung mit maximaler Korrelation nicht notwendigerweise die korrekte Lokalisierung der Domäne angibt: Die Korrelationslösung berücksichtigt nur eine Domäne und kann durch nichtberücksichtigte Nachbardomänen fehlgeleitet werden. Jedoch ist die erhaltene Lokalisierung für RBD plausibel: Die an den Rezeptor bindende helikale Unterdomäne (Fass et al. 1997) befindet sich an einer exponierten Auswölbung und im massereichsten Teil der Env Struktur.

8. Verallgemeinerte Principal Component Analysis für KET

Die Anwendung von Mittelungstechniken setzen strukturelle Homogenität der einzelnen Partikel voraus. Diese ist jedoch häufig nicht gegeben: Auch bei hoher chemischer Reinheit der Probe, können Proteinkomplexe evtl. verschiedene Konformationen annehmen, die eine kohärente Mittelung in hoher Auflösung unmöglich machen. In der Tomographie ist diese Problematik i.A. noch gravierender, da die Proben hier kaum aufgereinigt werden, sondern in einem möglichst nativen Milieu abgebildet werden. Die Erzeugung hochauflösender Mittelungen aus inhomogenen Daten erfordert eine Klassifikation der einzelnen Partikel in strukturell homogene Untergruppen.

Die in der statistischen Analyse inhomogener Datensätze üblichen Verfahren zur Klassifizierung basieren auf Multivariater Statistischer Analyse (*Multivariate Statistical Analysis, MSA*). Bei diesen Verfahren werden die Datensätze anhand der Varianzen zwischen den unterschiedlichen Aufnahmen analysiert. Es können dabei im Wesentlichen zwei Verfahren unterschieden werden: *Principal Component Analysis* (PCA) und Korrespondenzanalyse (*Correspondence Analysis, CA*). Beide Verfahren sind sehr ähnlich, sie basieren auf Eigenvektorzerlegung der Kovarianzmatrix, und unterscheiden sich nur in ihren Normierungen. Im Bereich der Elektronenmikroskopie makromolekularer Komplexe wurde zuerst CA angewandt (van Heel und Frank 1981), wobei die Implementierung von Hill verwendet wurde (Hill 1974). Heutzutage wird in den meistverbreiteten *Single Particle* Programmen sowohl CA wie auch PCA zur Analyse und folgenden Klassifikation von Bildern benutzt (Frank et al. 1996; van Heel et al. 1996); PCA hat speziell in der Tomographie den Vorteil, daß sie auch auf nicht-positive Daten anwendbar ist; als Folge der gewichteten Rückprojektion beinhalten

3D Rekonstruktionen stets auch negative Dichtewerte.

In der ET besteht jedoch das Problem, daß Daten systematische Unterschiede aufgrund des *Missing Wedge* aufweisen: Gleichen Partikeln, die in unterschiedlichen Orientierungen bzgl. des Koordinatensystems des Tomogramms vorliegen, fehlt ein jeweils anderer Teil des Signals im Fourierraum. Da die konventionelle PCA den *Missing Wedge* nicht berücksichtigt, neigt sie dazu, die unterschiedlichen Effekte des *Missing Wedge* zu detektieren statt vorliegenden strukturellen Inhomogenitäten (Walz et al. 1997b).

Als Teil dieser Arbeit wurde eine PCA unter Berücksichtigung des *Missing Wedge* entwickelt, die *Constrained Principal Component Analysis* (CPCA). Um die entwickelte CPCA zu testen, wurde sie zuerst zur Klassifikation simulierter Daten benutzt und mit der konventionellen PCA verglichen und anschließend auf experimentelle Tomogramme angewendet.

Im nächsten Abschnitt wird zunächst die konventionelle PCA kurz erklärt, um anschließend die CPCA darzustellen. Daran anschließend wurden anhand einer Simulation die Auswirkungen des *Missing Wedge* auf beide Klassifikationsmethoden in Abhängigkeit des SNR untersucht. Simuliert wurden Tomogramme des Proteinkomplex' Thermosom, der in einer offenen und einer geschlossenen Konformation auftreten kann. Basierend auf den atomar aufgelösten Strukturen der beiden Konformationen (Ditzel et al. 1998; Nitsch et al. 1998) wurden Tomogramme mit unterschiedlichem *Missing Wedge* und SNR simuliert, und es wurde getestet, ob die Klassifikation der Daten im Eigenfaktorraum von PCA bzw. CPCA mittels *K-means Clustering* identische Konformationen gruppiert. Es konnte gezeigt werden, daß die CPCA deutlich weniger vom *Missing Wedge* beeinflusst wird. Schließlich wurde die PCA noch auf experimentelle Tomogramme vom Kernporenkomplex (NPC) und vom Proteinkomplex Proteasom 26S angewendet.

8.1 Constrained Principal Component Analysis (CP-CA)

8.1.1 Principal Component Analysis

Für MSA ist die Dimensionalität der Daten prinzipiell gleichgültig, da die Daten als Vektor in einem multidimensionalen Raum behandelt werden. Sei z.B. ein Volumen definiert als:

$$V = x(i, j, k), \quad i, j, k \in N, \quad (1, 1, 1) \leq i \leq (N_x, N_y, N_z). \quad (8.1)$$

Dies ist in der Untersuchung mittels MSA gleichbedeutend mit:

$$x(i, j, k) \rightarrow x(l), \quad \text{mit } 1 \leq l \leq N_x \cdot N_y \cdot N_z = N. \quad (8.2)$$

In der PCA ist die statistische Untersuchung von P verschiedenen Datensätzen das Ziel. Dies sind im konkreten Fall P verschiedene Tomogramme $x(l, p)$, wobei p der Index des jeweiligen Tomogramms ist. In der PCA wird nun die Kovarianz- oder Korrelationsmatrix des Datensatzes aufgestellt, und die Analyse dieser Matrix sollte statistische Gesetzmäßigkeiten des Datensatzes erkennen lassen. Prinzipiell muß dabei unterschieden werden ob die Beschreibung im *Partikelraum* oder im *Pixelraum* erfolgen soll. In der klassischen *Single Particle* Analyse ist dabei der Pixelraum vorteilhaft, da die Zahl der einzelnen Partikel meist deutlich höher ist als die Zahl der Pixel und eine unmittelbare Interpretation an Pixeln anschaulicher ist (für eine weitergehende Diskussion sei auf (Frank 1996) verwiesen). Dies ist z.B. im EM Programm der Fall (Hegerl 1996). Im Falle tomographischer Daten ermöglicht diese Basiswahl jedoch keine Berücksichtigung des Missing-Wedges, weshalb hier für eine PCA für Tomographie der - sonst unübliche - Partikelraum gewählt wird.

Die Kovarianzmatrix eines Partikelsatzes ist definiert als:

$$C(p, p') = \sum_l x(l, p) \cdot x(l, p'). \quad (8.3)$$

Die Korrelationsmatrix $CC(p, p')$ unterscheidet sich lediglich dadurch, daß in Gl. (8.3) die Tomogramme auf die Standardabweichung normiert werden, d.h. der Mittelwert wird von den einzelnen Partikeln abgezogen und sie werden durch die Standardabweichung dividiert.

Der entscheidende Schritt der PCA ist die Bestimmung der Eigenwerte und Eigenvektoren der Korrelations- oder Kovarianzmatrix. Die Eigenvektoren e_i geben die statistisch wichtigsten Merkmale des Datensatzes an und die Eigenwerte λ_i geben ihren Anteil an der Gesamtvarianz bzw. Gesamtkorrelation an. Die Eigenvektoren oder Hauptkomponenten können in den Pixelraum transformiert werden mittels:

$$e_i(l) = e_{i1} \cdot x(l, 1) + e_{i2} \cdot x(l, 2) + \dots + e_{iN} \cdot x(l, N). \quad (8.4)$$

Sinn der PCA ist die Datenreduktion auf wenige, charakteristische Merkmale. Das geschieht in der Praxis durch Beschränkung der Datenanalyse auf eine Zahl von Eigenvektoren. Der zum größten Eigenwert korrespondierende Eigenvektor gibt im Wesentlichen den Mittelwert der Einzelpartikel an, der nächstkleinere Vektor gibt den wichtigsten Differenzvektor an u.s.w..

Die Grundlage für eine anschließende Klassifikation sind die Faktoren e_{ip} des jeweiligen Partikels p . Diese werden meist noch mit dem zugehörigen Eigenwert oder dessen Quadratwurzel gewichtet. Zur graphischen Beurteilung werden sie in einem sog. *Scatter plot* aufgetragen, was einen ersten Aufschluß über mögliches Clustering der Partikel in Subpopulationen gibt. Für die tatsächliche Klassifikation der in den Eigenvektorraum projizierten Daten gibt es verschiedene Verfahren. In der vorliegenden Arbeit wurde das *K-means* Verfahren benutzt (Frank 1996).

8.1.2 Constrained PCA

In der ET besteht das Problem, das die Daten nicht gleichmäßig im Fourierraum abgetastet sind, jedes Tomogramm hat einen *Missing Wedge* an Daten. Deshalb ist es vorteilhaft, die Korrelationsmatrix analog zu Abschn. 5.1.2 mit der eingeschränkten Korrelationsfunktion (5.12) zu formulieren.

Im Rahmen der PCA müssen gemäß dieser Idee die Korrelationskoeffizienten zweier Partikel auf den in beiden Volumen abgetasteten Bereich im Fourierraum eingeschränkt werden. Wir definieren den *Constrained Correlation Coefficient CCC* als die Korrelation zweier Partikel P_i und P_j , deren Daten im Fourierraum eingeschränkt werden:

$$CCC_{ij} = \frac{(\tilde{P}_i - \bar{\tilde{P}}_i) \cdot (\tilde{P}_j - \bar{\tilde{P}}_j)}{\sqrt{\sum_{xyz} (\tilde{P}_i - \bar{\tilde{P}}_i)^2} \cdot \sqrt{\left(\sum_{xyz} (\tilde{P}_j - \bar{\tilde{P}}_j)^2\right)}}. \quad (8.5)$$

Sei Ω der in den Partikeln P_i und P_j gemeinsam abgetastete Bereich im Fourierraum, so gilt für die eingeschränkten Datensätze \tilde{P}_i bzw. \tilde{P}_j :

$$\tilde{P} = \mathcal{F}^{-1}(\mathcal{F}(P) \cdot \Omega). \quad (8.6)$$

Mit Hilfe des CCC_{ij} kann die eingeschränkte Korrelationsmatrix eines Partikelsatzes berechnet werden. Die Eigenvektoren dieser Matrix geben wiederum die Hauptkomponenten der Daten an, die Eigenwerte die zugehörigen Korrelationskoeffizienten. Die Daten werden durch Beschränkung auf wenige Eigenvektoren mit den zugehörigen höchsten n Eigenwerten reduziert und in diesem n -dimensionalen Raum kann klassifiziert werden. Zu diesem Zwecke wurde die Klassifizierung nach dem *K-means* Verfahren verwendet (zusammengefaßt in (Frank 1996)), die die einfachste und damit meistverbreitete Methode darstellt.

8.2 Material und Methoden

8.2.1 CPCA und Klassifizierung

Die Berechnung der Korrelationsmatrix (8.5) wurde in der Prozedur `av3_cccmat` in MATLAB implementiert. Die Berechnung der Korrelationsmatrix ermöglicht die Angabe einer Motivliste, die die Verschiebungsvektoren und Orientierungen der Partikel enthält (Abschn. 6.1). Außerdem kann eine Maske angegeben werden, auf die die Korrelation beschränkt werden soll: So ist es z.B. sinnvoll die Korrelation kugelförmig zu begrenzen, wenn die Partikel in verschiedenen Orientierungen vorliegen. Die

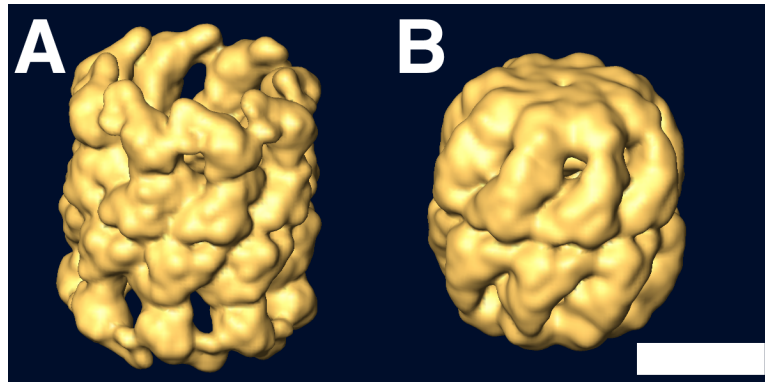


Abbildung 8.1: A: Thermosom in offener Konformation. B: Thermosom in geschlossener Konformation. Die elektrostatischen Potentiale beider Komplexe sind als *Isosurface* dargestellt, wobei die Daten auf etwa $(1.5 \text{ nm})^{-1}$ Auflösung gefiltert sind. Maßstab: 10 nm.

Eigenwerte und Eigenvektoren wurden mit MATLAB berechnet und gemäß (8.4) in 3D Daten umgerechnet. Zur Klassifikation der Eigenfaktordaten in unterschiedliche Klassen wurde das EM Programm verwendet.

8.2.2 Simulationen

Simulation der Daten. Die atomaren Koordinaten des Thermosoms von *Thermoplasma acidophilum* in der geschlossenen Konformation sind aus der PDB entnommen (Nummer 1A6D (Ditzel et al. 1998)), die Koordinaten der offenen sind in Ref. (Nitsch et al. 1998) publiziert. Beide Strukturen sind in Abb. 8.1 gezeigt.

Die Simulation der tomographischen Abbildung sollte die Auswirkung des *Missing Wedge* und Rauschens auf die Klassifikation zeigen. Um letzteres zu untersuchen war ein möglichst realistisches Rauschmodell nötig. Hier wurde folgendes angenommen: Es existierte ein Untergrundrauschen (Saad et al. 2001), und darüberhinaus lag *Shot Noise* vor, der daher rührt, daß nur vereinzelt Elektronen detektiert werden. Ersterer ist additiv zum Signal und wurde hier spektral als gaußförmig angenommen. Der *Shot Noise* wurde der Einfachheit halber ebenfalls als additiv angenommen, im Gegensatz zum Untergrundrauschen wurde er auch durch den Abbildungsvorgang moduliert.

Es wurden Daten mit einem *Missing Wedge* W von 0 bis 40° und einem SNR von 1, $1/10$ und $1/100$ berechnet. Im Detail wurde die tomographische Abbildung folgendermaßen simuliert:

1. Das elektrostatische Potential der in den PDB Dateien beschriebenen Strukturen wurde auf ein diskretes Koordinatensystem mit einer Maschengröße von 0.4 nm abgebildet (siehe auch Abschn. 5.2).
2. Die Volumina wurden jeweils um einen Winkel ϑ_i von 0° , 10° u.s.w. bis 90° rotiert.
3. Es wurde der *Shot Noise* Anteil des Rauschens addiert mit einem SNR von 2, $2/10$ und $2/100$. Das Rauschen war im Realraum gaußförmig verteilt und somit spektral konstant.
4. Das Volumen wurde mit erst einer CTF gefaltet, deren erste Nullstelle bei $(2.5 \text{ nm})^{-1}$ lag ($U=300 \text{ kV}$, $\Delta = 3\mu \text{ m}$) und anschließend mit einer approximierten MTF (Abfall des Signals bei $N_y/2$ auf 20%).
5. Aus den Daten wurde ein entlang der k_y -Achse orientierter Datenkeil mit einem Halbwinkel $\Delta\alpha$ von 0° , 10° , 20° , 30° oder 40° herausgeschnitten.
6. Auf die Volumina wurde das Untergrundrauschen mit einem SNR von 2, $2/10$ und $2/100$ (bezogen auf das Originalsignal) versetzt, wobei das im Realraum Gaußverteilte Rauschen im Fourierraum mit einer Gaußfunktion multipliziert wird (Abfall des Signals bei $N_y/2$ auf 20%).

Insgesamt wurden somit für jeden Wert von W ($W = 0, 10, \dots, 40$) und SNR (SNR= 1, 0.1, 0.01) 20 Partikel erzeugt, jeweils 10 offene und 10 geschlossene Thermosomen. Den offenen Thermosomen wurden jeweils ungerade Partikelnummern zugeordnet und den geschlossenen gerade. Die Größe der Volumina betrug jeweils $64 \times 64 \times 64$ Voxel. Die Orientierungen der Partikel wurden in einer Motivliste gespeichert.

Datenanalyse. Die Eigenwerte und Vektoren dieser Datensätze werden mittels eingeschränkter und konventioneller PCA berechnet. Die eingeschränkte CPCA erfolgte dabei wie in Abschn. 8.1 beschrieben. Dabei wurde die Korrelation auf eine Kugel mit einem Radius von 29 Pixeln eingeschränkt.

Die konventionelle PCA wurde in EM durchgeführt. Hierfür wurden die Partikel mit der gleichen sphärischen Maske multipliziert und in zweidimensionale Daten umgewandelt, da die EM Klassifikation mit zweidimensionalen Bildern arbeitete. Die EM PCA berechnete die Eigenwerte und Eigenvektoren auf Basis der Varianzmatrix.

Die graphische Darstellung der Eigenwert- und Klassifikationsanalyse erfolgt durch einen sog. *Scatter Plot*: Hierbei werden die Eigenfaktoren der Eigenvektoren k und l , d.h. die e_{ik} bzw. e_{il} in (8.4) in einem Graphen dargestellt. Zur graphischen Darstellung und zur späteren Klassifizierung werden die Eigenfaktoren mit dem zugehörigen Eigenwert gewichtet. Zur Darstellung wurden die Eigenvektoren $k = 2$ und $l = 3$ gewählt. Zur Klassifikation wurden die Eigenfaktoren beider PCA Methoden auf die Eigenvektoren 2 und 3 beschränkt, und sie wurde mit der *K-means* Methode in EM durchgeführt.

8.2.3 Experimentelle Daten

Kernporen. Es wurden die bereits in Abschn. 6.2 zur Mittelung benutzten Subtomogramme von 267 Kernporen analysiert. Nach dem Berechnen der Eigenfaktoren mittels CPCA wurden diese mittels *K-means* im EM Programm klassifiziert. Die Klassifikation wurde in den Eigenfaktoren der Eigenvektoren 2 und 3 mit verschiedenen Anzahlen von Klassen durchgeführt. Eine Klassenzahl von vier ergab das höchste Verhältnis von kleinster Inter-Klassenentfernung zu größter Intra-Klassenentfernung (~ 1.4) und wurde somit für die Klassifikation gewählt. Die vier Klassen wurden schließlich erneut für 5 Zyklen iterativ aligniert und 8-fache Rotationssymmetrie wurde aufgeprägt.

20S/26S Proteasomen. Es wurden Tomogramme eiseingebetteter 26S Proteasomen am CM20 TEM (Kap. 4) aufgenommen ¹. Die in Sukroslösung eingefrorenen 26S Proteasomen (präpariert von Hölzl (Hölzl et al. 2000)) wurden aufgetaut, und etwa 5 μl der Lösung wurden auf ein Lacy Carbon Grid pipettiert, das vorher mit 3 μl kolloidalem Gold versetzt wurde. Überschüssige Lösung wurde mit Filterpapier abgesaugt und anschließend wurde das Grid mit ATP Puffer gewaschen. Nach erneutem Absaugen überschüssigem Puffers wurden die Proben vitrifiziert. Die Kippserien wurden bei 37000-facher Vergrößerung aufgenommen, so daß ein Pixel zu 4.1 Å korrespondierte. Es wurden Kippserien in einem Kippwinkelbereich von $\pm 65^\circ$ aufgenommen, wobei ein Winkelinkrement von 5° gewählt wurde.

Die Daten wurden wie in Abschn. 4.4 beschrieben mittels k -gewichteter Rückprojektion rekonstruiert. Im Tomogramm wurden die 26S- bzw. 20S Proteasomen mittels MOLMATCH lokalisiert. Das Muster für die Suche wurde aus der Struktur des 20S Proteasoms von *Thermoplasma acidophilum* generiert (Abschn. 5.2), und es wurde eine Maske verwendet, die den Durchmesser des 20S Proteasoms besitzt. Die Orientierung der Partikel variiert von $\varphi = \psi = \vartheta = 0$ bis $\varphi = 50^\circ$, $\psi = 355^\circ$, $\vartheta = 180^\circ$ mit einem Inkrement von 5° . Die zu den 400 höchsten Korrelationswerten korrespondierenden Partikel wurden ausgeschnitten und ihre potentiellen Orientierungen in einer Motivliste gespeichert.

Es wurden die Eigenvektoren dieser Partikel mittels CPCA berechnet, wobei die Korrelation mittels einer Maske auf Regionen ober- und unterhalb der Öffnungen der vermeintlichen Proteasomen beschränkt wurde. Die Eigenfaktoren der Eigenvektoren 2 bis 4 wurden mittels K -means Clustering in 4 Klassen aufgeteilt.

¹Die Präparation der EM Grids und der Großteil der Datenakquisition wurde von O. Mihalache durchgeführt.

8.3 Ergebnisse

8.3.1 Anwendung auf simulierte Daten

Thermosomen haben ein Molekulargewicht von etwa 900 kDa, eine zylinderförmige Struktur und einen Durchmesser von etwa 15 nm. Somit sind sie Makromoleküle, die in einem Kryoelektronentomogramm aufgelöst werden können. Die Konformationsänderung zwischen offenen und geschlossenem Thermosom bewegt sich in einer Größenordnung (ca. 3-4 nm), die auch in einem Tomogramm aufzulösen sein sollte (Abb. 8.1). Von beiden Strukturen sind Strukturen in atomarer Auflösung bestimmt; die Struktur der geschlossenen Konformation wurde durch Röntgenkristallographie ermittelt (Ditzel et al. 1998) und die der offenen Konformation durch Fitten dieser Struktur an Kryo-EM Daten (Nitsch et al. 1998). Somit ist es möglich, aus beiden Strukturen in gleicher Weise tomographische Daten auf der Basis der atomaren Koordinaten zu simulieren.

Analyse der Korrelationsmatrix. Die eingeschränkte Korrelationsmatrix CCC (8.5) wurde graphisch dargestellt, wobei die Korrelationskoeffizienten farbkodiert wurden (z.B. Abb. 8.2). Bei $SNR=1$ ergibt diese Darstellung für alle Werte von W im Wesentlichen ein "Schachbrettmuster" (Daten nicht gezeigt): So führt z.B. die Korrelation der offenen Thermosomen untereinander (gerade Partikelnummern) zu höheren Korrelationswerten als zwischen offenen und geschlossenen Partikeln, was das Muster erklärt. Bei niedrigerem SNR wird dies nicht mehr erreicht: Die Autokorrelationskoeffizienten, d.h. die Hauptdiagonalelemente sind z.B. bei $SNR=1/100$ deutlich höher als alle anderen Werte (Abb. 8.2). Weiterhin ist bei höheren Werten von W wie 30° eine Abhängigkeit des Korrelationsmusters von der Partikelnummer zu erkennen (Abb. 8.2). Dies deutet darauf hin, daß die Orientierung des *Missing Wedge* bei niedrigen $SNRs$ trotz der Einschränkung der Korrelation im Fourierraum weiterhin einen gewissen Einfluß auf die Korrelationswerte besitzt.

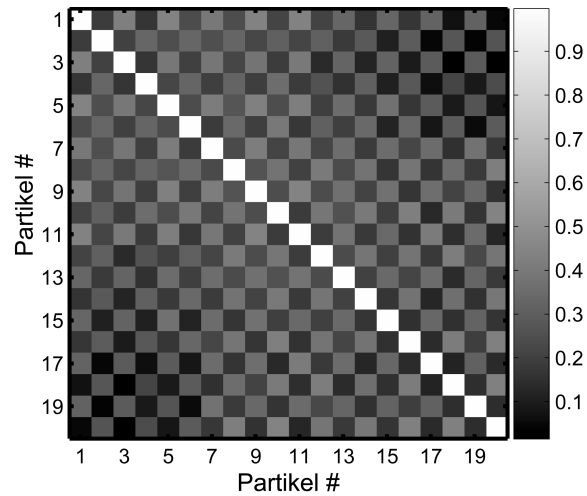


Abbildung 8.2: CCC für $SNR=1/100$ und $W = 30^\circ$. Die Korrelationskoeffizienten sind farbkodiert.

Eigenwerte. Im nächsten Schritt wurden die Eigenwerte und -Vektoren der CCC und mittels eingeschränkter (CPCA) und gewöhnlicher PCA berechnet. Die Höhe der berechneten Eigenwerte sind als Balkendiagramm in Abb. 8.3 gezeigt. Bei einem SNR von 1 und einem *Missing Wedge* von 0° sind bei PCA und CPCA nur zwei signifikante Eigenwerte vorhanden. Im Falle der PCA geben diese Eigenwerte die Anteile der korrespondierenden Eigenvektoren an der Varianz der Partikel an. Im Falle der CPCA geben sie den Anteil der jeweiligen Eigenvektoren an der gesamten Korrelation an (multipliziert mit der Gesamtzahl der Partikel - hier 20). In beiden Fällen entspricht der erste Eigenvektor dem Mittelwert aller Partikel. Da insgesamt nur zwei signifikante Eigenwerte vorliegen, können die Partikel im Wesentlichen durch zwei Eigenvektoren dargestellt werden, den Mittelwert und den zweiten Eigenvektor. Dies ist das erwartete Ergebnis, wenn Partikelsätze sich aus lediglich zwei verschiedenen Partikelarten zusammensetzen.

Wird der *Missing Wedge* jedoch bei $SNR=1$ vergrößert, so werden die Eigenwerte drei und vier, die idealerweise nicht auftreten sollten, kontinuierlich größer. Im Gegensatz dazu zeigt die CPCA eine wesentlich geringere Abhängigkeit vom *Missing Wedge*: Auch bei einer Größe des *Missing Wedge* von 40° beträgt der dritte Eigenwert nur

etwa ein Zehntel des zweiten.

Bei niedrigerem SNR ändert sich das Verhalten der Eigenwerte wenig: Der erste Eigenwert wird lediglich kleiner im Verhältnis zu den folgenden Eigenvektoren. Die CPCA zeigt jedoch eine stärkere Veränderung der Eigenwerte: So sind der dritte und vierte Eigenwert insbesondere bei $\text{SNR}=1/100$ jetzt im Verhältnis zum zweiten sehr viel größer als bei $\text{SNR}=1$; z.B. beträgt der dritte Eigenvektor knapp 50% des zweiten bei einem *Missing Wedge* von 40° .

Eigenfaktoren und Klassifikation. Die Eigenfaktoren sind wie in Abschn. 8.2 beschrieben als *Scatter Plot* aufgetragen (Abb. 8.4). Die mit PCA berechneten Eigenfaktoren zeigen unabhängig vom SNR ein qualitativ ähnliches Verhalten: Bei $W = 0^\circ$ bilden sich deutliche Wolken im *Scatter plot*, die zu offenen bzw. geschlossenen Thermosomen korrespondieren. Mit steigendem W werden diese Wolken immer stärker elongiert. Werden diese Daten mit *K-means* klassifiziert, so kommt es ab $W = 30^\circ$ zu falscher Klassifikation einiger Partikel aufgrund der starken Verformung der Wolken. Die Darstellung der CPCA Eigenvektoren zeigt für $\text{SNR}=1$ ein qualitativ anderes Verhalten: Bei $W = 0^\circ$ ergeben sich zwei weitgehend isotrope Wolken, die mit größerem W anisotrop ausdehnen. Diese Ausdehnung ist jedoch weit weniger stark ausgeprägt als bei der PCA. Die anschließende Klassifikation ist für alle Werte von W korrekt. Mit sinkendem SNR wird die Verbreiterung der Wolken in der Eigenfaktordarstellung sukzessive größer. Die Ausdehnung wird jedoch nicht zu groß, um eine korrekte Klassifikation zu verhindern. Genauere Untersuchung der Abhängigkeit der Größe der Eigenfaktoren des dritten Eigenvektors zeigt, daß diese mit der Orientierung des Kippwinkels korrelieren (Daten nicht gezeigt).

Berechnungsdauer. Die Eigenwertberechnung benötigt mit beiden Methoden deutlich unterschiedlich viel Zeit: Während die gewöhnliche PCA die Berechnung der Eigenwerte und Eigenvektoren aller Datensätze im Zeitraum von Minuten bewerkstelligt, beträgt die Zeitdauer mittels CPCA etwa 2 Stunden. Der mit Abstand zeitaufwendigste Schritt ist hierbei die Berechnung von *CCC*.

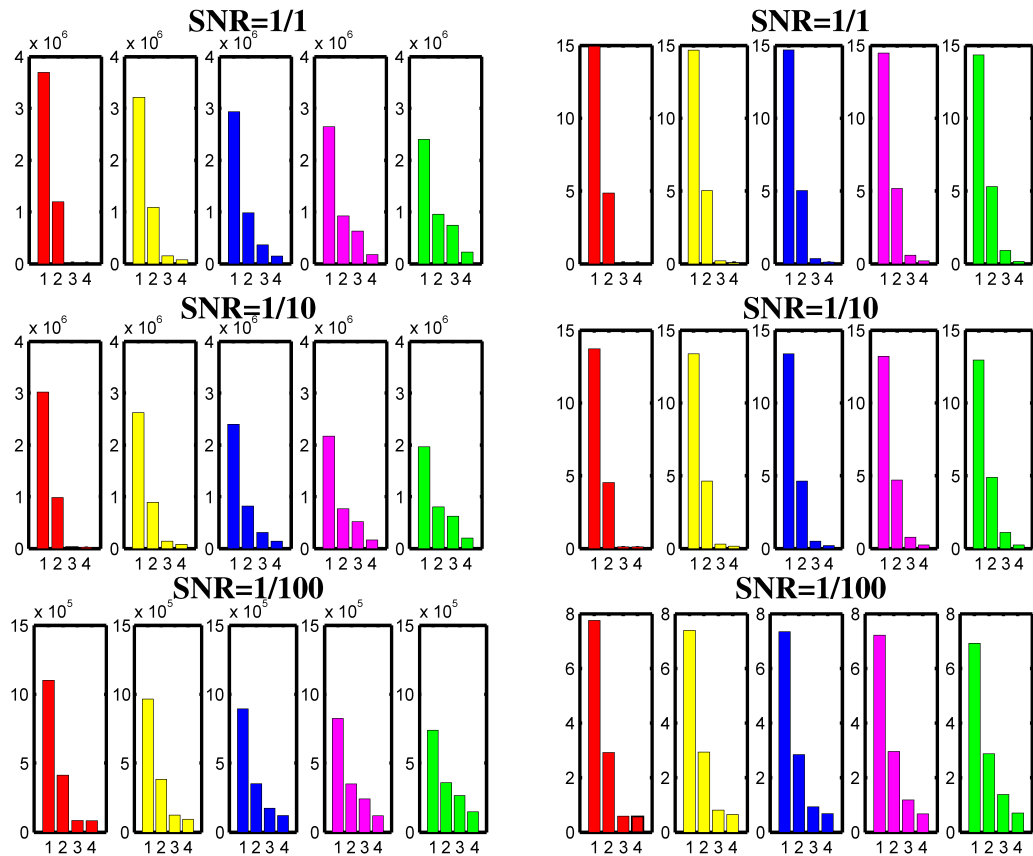


Abbildung 8.3: Die größten vier Eigenwerte des Testdatensatzes. Linke Spalte: Eigenwerte der konventionellen PCA (EM-System) bei einem SNR von 1, 1/10 und 1/100 bei verschiedenen Größen des *Missing Wedge* (rot: 0°, gelb: 10°, blau: 20°, magenta: 30°, grün: 40°). Der Eigenwertnummer wird im Balkendiagramm jeweils der Eigenwert zugeordnet. Rechte Spalte: Eigenwerte der *Constrained PCA* in gleicher Auftragung.

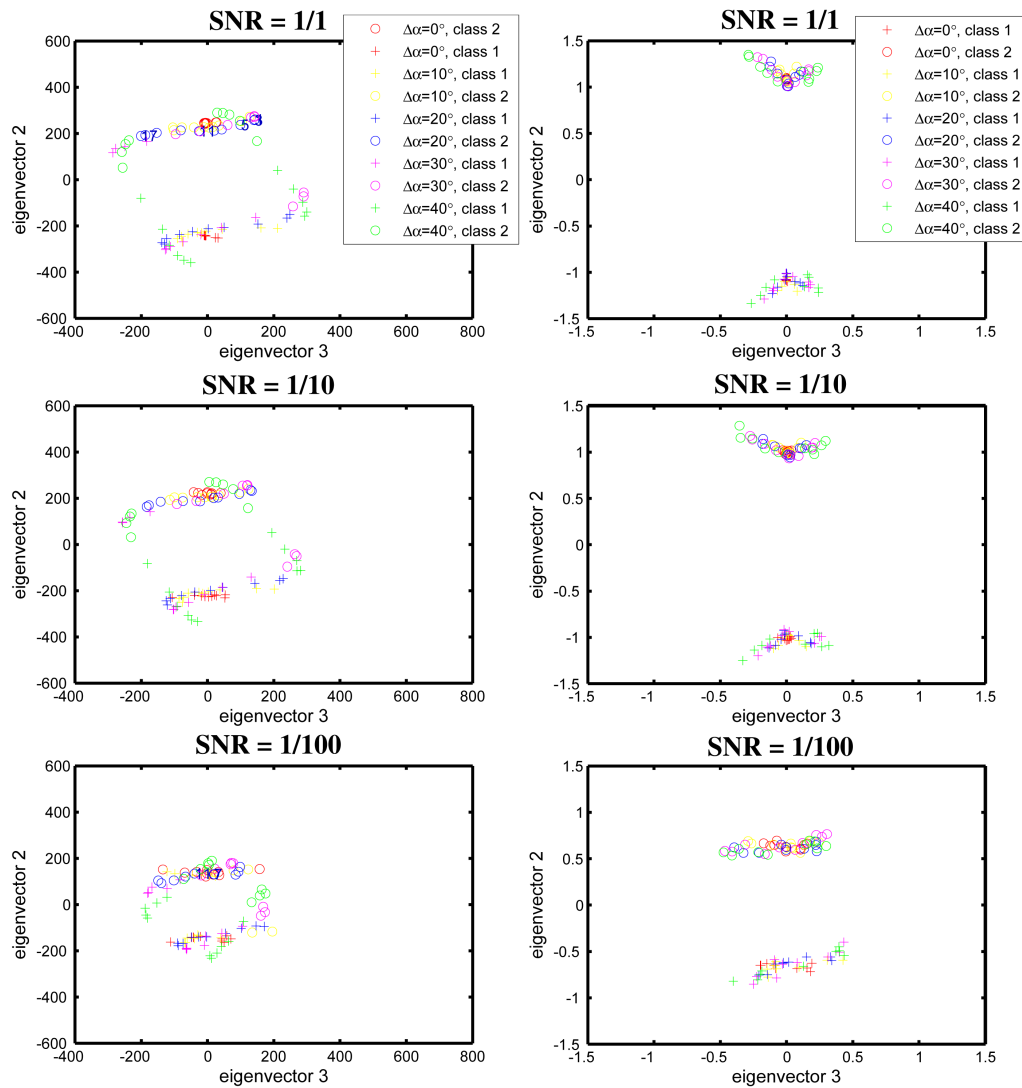


Abbildung 8.4: Scatter plot der Eigenfaktoren, die zu den Eigenvektoren 2 und 3 korrespondieren. Linke Spalte: Eigenfaktoren bei Verwendung der PCA. Die Partikel wurden im Eigenfaktorraum mittels *K-means* Klassifizierung in zwei Klassen aufgeteilt. Rechte Spalte: Ergebnisse der CPCA.

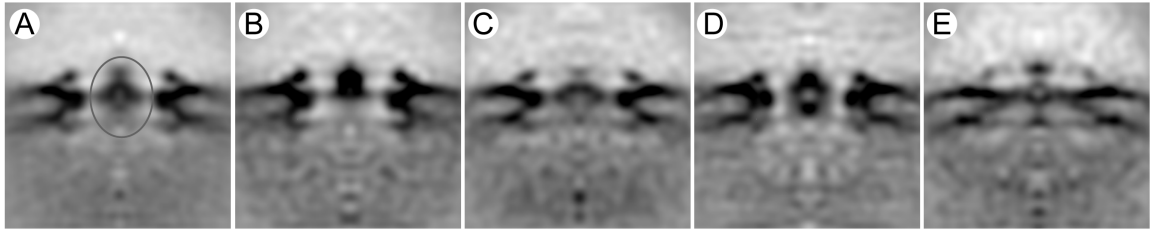


Abbildung 8.5: Klassifikationsergebnisse des NPC. A: Seitenansicht der NPC Mittelung aus Abschn. 6.2. Die CPCA wurde auf den oval markierten Bereich beschränkt. B-E: Seitenansichten der 4 ermittelten Klassenmittelungen. In Klasse 1 fallen 134 (B), in Klasse 2 44 (C), in Klasse 3 71 (D) und in Klasse 4 18 Partikel (E).

8.3.2 Anwendung auf experimentelle Daten

Kernporen. Die CPCA Klassifizierung der Partikel ergab die deutlichste Klassifizierung bei Verwendung von vier Klassen. Die Mittelungen der unterschiedlich klassifizierten Partikel sind in Abb. 8.5 gezeigt. Die Klassen zeigten deutliche Strukturänderungen im Bereich des CP wie dies auch die Klassifikation des Datensatzes anhand des Schwerpunkts des CP ergab (vgl. Abschn. 6.2). Die Klassenmittelwert 1 und 3 ähnelten der CF-Klasse aus Abschn. 6.2, bei Klasse 1 befand sich jedoch weniger Masse in der unteren Hälfte während die Klasse 3 mehr Masse im unteren Teil der Kernpore besitzt als die CF Klasse. Dies legte den Schluß nahe, daß durch die CPCA Klassifikation eine weiter Unterteilung der CF-Klasse in zwei Subpopulationen erhalten wird. Der Klassenwert 2 hatte starke Ähnlichkeit mit der LR Klasse. Der Klassenmittelwert 4 resultiert aus nur 18 Einzelpartikeln und spiegelte wahrscheinlich eine schlechtere Alignierung einiger Partikel wider.

20S/26S Proteasomen. Das 26S Proteasom ist ein labiler makromolekularer Proteinkomplex. Er besteht aus dem 20S Proteasom, das strukturell sehr gut charakterisiert ist, und den regulatorischen 19S Untereinheiten an seinen Enden (zusammengefaßt in (Voges et al. 1999)). Bislang gelang nur eine grobe 3D Charakterisierung des gesamten Komplexes anhand von getrockneten, mit Schwermetallen kontrastierten

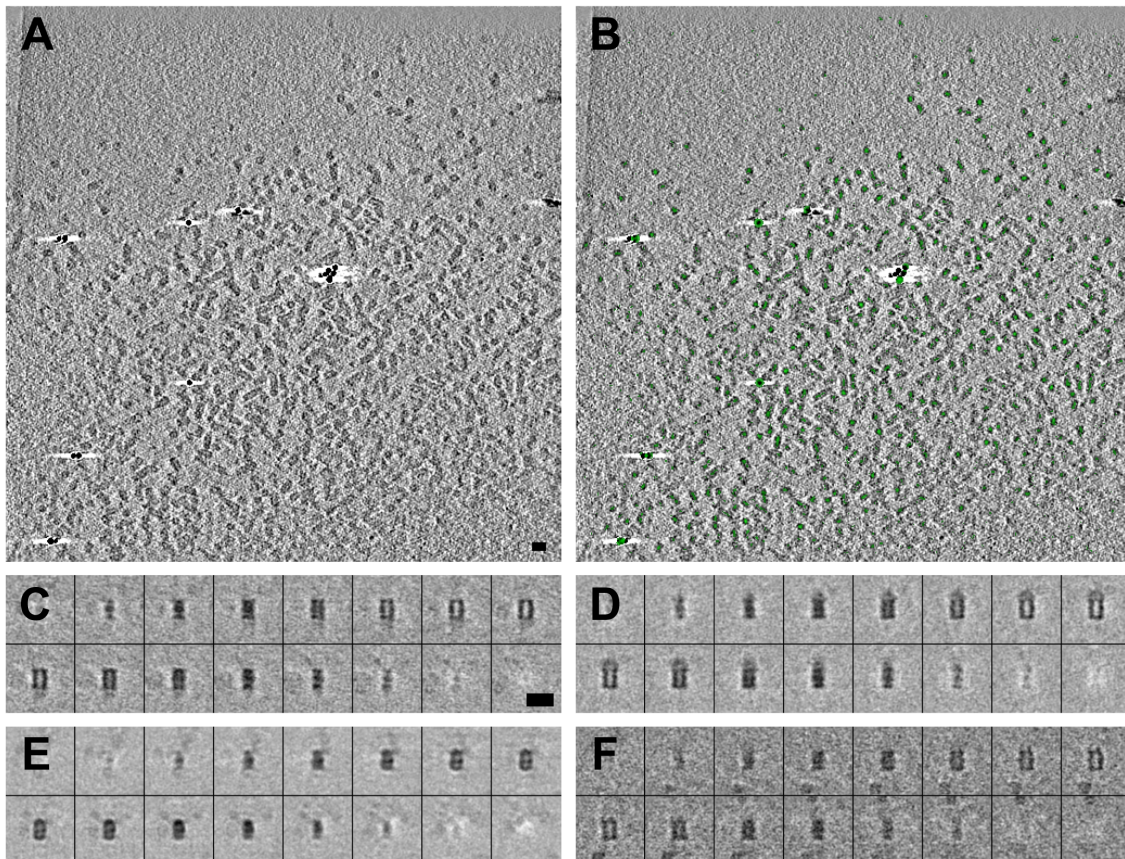


Abbildung 8.6: A: xy -Schnitt eines Tomogramms von 20S und 26S Proteasomen. B: Gleicher Schnitt mit Überlagerung der Korrelationsfunktion, die mit MOLMATCH berechnet wurde. Es sind nur die höchsten Korrelationswerte gezeigt. C-F: Klassenmittelungen der detektierten Partikel. Es sind Schnitte in y Richtung durch die 3D Daten dargestellt. Die Klassenmittelwerte resultieren aus 126 (A), 184 (B), 48 (C) und 42 (D) Partikeln. Der Maßstab in A und C korrespondiert zu 20 nm.

Partikeln (Walz et al. 1998).

Die aufgenommenen Tomogramme der Proteasomen besaßen ein relativ niedriges SNR (Abb. 8.6A). Die Erklärung hierfür liegt zum einen darin, daß das Eis trotz des Waschens des Grids noch einen hohen Anteil an Sukrose besaß, was das Signal, d.h. die Dichtedifferenz zwischen Protein und Puffer verringerte. Eine weitere Ursache ist möglicherweise der teilweise Zerfall der 26S Proteasomen zu 20S Proteasomen und verschiedenen Zerfallsprodukten der 19S Untereinheiten. Diese erhöhten auch die Hintergrunddichte, was zu einer Verschlechterung der Signalqualität führte. Die visuelle Inspektion der Tomogramme ergab, daß ein großer Teil der erkennbaren Komplexe, schätzungsweise über 50%, nur noch als 20S Proteasomen vorlagen.

Eine Lokalisierung der 20S Proteasomen mittels MOLMATCH sollte in der Lage sein, sowohl isolierte 20S Partikel als auch mit einer oder zwei 19S Untereinheiten verbundene 20S Proteasomen zu lokalisieren. Die visuelle Evaluation der Detektionsergebnisse ergab, daß Makromoleküle und teilweise auch die zur Alignierung verwendeten Goldkugeln detektiert wurden (Abb. 8.6B).

Die CPCA Klassifikation der zu den 400 höchsten Korrelationskoeffizienten korrespondierenden Partikel führte zu 4 Klassen (Abb. 8.6C-F). Die Klasse 1 korrespondiert augenscheinlich zu reinen 20S Proteasomen. Die Klasse 2 besitzt eine zusätzliche Dichte oberhalb des 20S Partikels. Die Klasse 3 weicht deutlich ab von der Form des 20S Musters: Manuelle Untersuchung der einzelnen Partikel ergab, daß diese Klasse fast ausschließlich aus Goldkugeln gebildet wird. Die letzte Klasse besteht nur aus wenigen Partikeln, und es handelt sich vermutlich um 20S Proteasomen.

Es wurde versucht, aus den Partikeln der Klasse 2 eine Struktur zu mitteln, die dann ein 20S Proteasom mit einer 19S Untereinheit an einem Ende dargestellt hätte. Die Mittelung mit der in Kap. 6 Prozedur führte jedoch am Ende zu keiner höheraufgelösten Struktur des 26S Proteasoms. Der Grund hierfür liegt wahrscheinlich darin, daß die Partikel während des Waschens auf dem EM Grid zerfielen. Die Klasse 2 enthält somit vermutlich diverse Zerfallsprodukte des 26S Proteasoms.

8.4 Diskussion

In diesem Kapitel wurde ein neuer Algorithmus zur Klassifikation tomographischer Daten dargestellt. Anhand simulierter Daten wurde gezeigt, daß das herkömmliche Verfahren zur Klassifikation, PCA kombiniert mit *K-means Clustering*, bei Vorhandensein eines *Missing Wedge* (W) im Fourierraum, der 20° übersteigt, zu falschen Ergebnissen führt. Die neuentwickelte CPCA lieferte auch bei $W = 40^\circ$ noch korrekte Ergebnisse. Allerdings wurden die Eigenwerte und -Vektoren der CPCA durch tiefes SNR negativ beeinflusst. Bei hohem Rauschniveau traten "Phantomeigenwerte" auf, deren Eigenfaktoren zur Orientierung des *Missing Wedge* korrelierten. Der Grund hierfür liegt in der anisotropen Verteilung der Varianz des Signals im Fourierraum. In Abb. 8.7 ist dies demonstriert: Aufgetragen wurde Standardabweichung des offenen Thermosoms in Abhängigkeit von einem *Missing Wedge* im Fourierraum dessen Orientierung variiert. Es zeigt sich, daß die Amplitude der Strukturfaktoren des Thermosoms in verschiedenen Orientierungen deutlich variiert. Bei der Simulation wurde das SNR bezüglich der im Fourierraum vollständigen Struktur bestimmt, d.h. bei konstantem Rauschniveau ändert sich das effektive SNR als Funktion der Orientierung des *Missing Wedge*. Dies spiegelt sich in der CPCA wider. Der Algorithmus liefert jedoch auch für eine sehr anisotrope Struktur, d.h. eine nicht-globuläre Struktur wie das Thermosom, korrekte Klassifikationen.

Die Klassifikation experimenteller Tomogramme zeigte vielversprechende Ergebnisse. Die Klassifikation der Kernporen zeigte qualitativ ähnliche Ergebnisse, wie der in Abschn. 6.2 verwendete Algorithmus, es kommt sogar noch zu einer feineren Aufteilung. Um jedoch biologisch signifikante Ergebnisse zu erzielen, wird es nötig sein, in zukünftigen Studien den Transportzustand der Kernporen biologisch möglichst genau zu definieren und die Statistik weiter zu verbessern.

Die Anwendung der CPCA verspricht ebenfalls großes Potential bei der Strukturbestimmung des 26S Proteasoms. Um die Mittelungsprozeduren aus Kap. 6 zur Strukturbestimmung zu benutzen, ist es nötig die Proben *in silico* mittels Klassifikation homogen zu gestalten. Die bisherige biochemische Aufreinigung des 26S Proteasoms

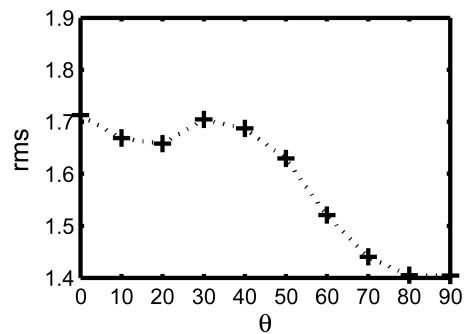


Abbildung 8.7: Standardabweichung des offenen Thermosms, das im Fourierraum einen unterschiedlich orientierten *Missing Wedge* mit einem Halbwinkel von 30° besitzt. Die Orientierung wird um den Winkel ϑ variiert.

hat sich als nicht zufriedenstellend erwiesen, um eine ausreichende Strukturhaltung der 26S Proteasomen zu gewährleisten. Sollte es jedoch gelingen, diese Probleme zu lösen, so bietet die verfolgte Strategie aus Tomographie, Lokalisierung von 20S und 26S Proteasomen mittels Mustererkennung, Klassifizierung und anschließender Mittelung der potentiell homogenen Klasse der 26S Proteasomen eine vielversprechende Möglichkeit, innerhalb kurzer Zeit eine Struktur des 26S Proteasoms im Auflösungsbereich von etwa $(3 \text{ nm})^{-1}$ zu erhalten.

9. Diskussion und Ausblick

Mustererkennung. Die Aufteilung eines Tomogramms in seine strukturellen Bestandteile (Segmentierung) beruhte in der Vergangenheit auf subjektiver Antizipation (z.B. in Ref. (Medalia et al. 2002)). Die offensichtlich mangelnde Objektivität und Arbeitsintensivität dieses Verfahrens machen die Entwicklung quantitativer Methodologien zur Analyse von Elektronentomogrammen notwendig. Prinzipiell sind hierzu zwei Vorgehensweisen denkbar: Eine Möglichkeit ist Tomogramme anhand ihrer Morphologie idealerweise parameterfrei in morphologisch ähnliche Segmente aufzuteilen (z.B. (Frangakis und Hegerl 2002; Linaroudis und Hegerl 2005)). Geht man davon aus, daß das Tomogramm aus einer begrenzten Zahl verschiedener aber strukturell wohldefinierter "Bausteine" aufgebaut ist, so kann die strukturelle Signatur der Bausteine zur Segmentierung ausgenutzt werden. Während sich ersterer Ansatz anbietet, um pleomorphe Strukturelemente wie z.B. Membranen abzugrenzen, ist der zweite Ansatz zur Identifikation makromolekularer Komplexe in Tomogrammen überlegen; der Vorteil besteht dabei darin, daß zur Detektion ein Maximum an *a priori* Information ausgenutzt werden kann, was angesichts des sehr niedrigen SNRs von Kryoelektronentomogrammen entscheidend ist.

In dieser Arbeit wurden korrelationsbasierte Verfahren zur Identifikation von Makromolekülen anhand ihrer bekannten strukturellen Signatur entwickelt. Bei der Detektion von Proteasomen in Phantomzellen hat sich die lokal normierte, eingeschränkte Korrelationsfunktion LCCF als das zuverlässigste Verfahren zur Detektion erwiesen. Es wurde auch zur Detektion von Ribosomen in pro- und eukaryotischen Zellen und zur Lokalisierung von Aktinfilamenten erfolgreich angewendet. Als momentan größtes Problem bei der Partikeldetektion mittels Korrelation hat sich die Empfindlichkeit der Methode auf unterschiedliche SNRs herausgestellt. Objekte wie Goldkugeln oder

Zellmembranen besitzen stärkeren Kontrast als Proteinkomplexe und sind aus diesem Grund häufig die Ursache falsch-positiver Detektionen. Die lokale Normierung der LCCF ist diesbezüglich am wenigsten empfindlich, jedoch leiden die Korrelationsergebnisse auch hier unter dem innerhalb eines Tomogramms inhomogenen SNR.

Die hinreichend zuverlässige Detektion von Ribosomen in Zellen gelang trotz des variierenden SNRs insbesondere deshalb, da Ribosomen zum einen sehr große molekulare Maschinen sind (Durchmesser > 20 nm), zum anderen sind sie aufgrund ihres RNA Anteils relativ dicht ($\rho \approx 1.7 \text{ gcm}^{-3}$). Um in der Zukunft zunehmend kleinere Proteinkomplexe in Zellen oder anderen biologischen Systemen zuverlässig detektieren zu können, müssen die theoretischen Methoden und der experimentelle Aufbau weiter verbessert werden. Das SNR der Tomogramme kann auf der experimentellen Seite vor allem durch die Entwicklung signifikant verbesserter CCD Kameras erhöht werden, woran derzeit intensiv geforscht wird (Übersicht in (Fan und Ellisman 2000)). Rechnerisch kann die Partikelidentifikation potentiell durch den Einsatz iterativer Rekonstruktionsmethoden oder alternativer Ähnlichkeitsmaße wie *Maximum Likelihood* zuverlässiger werden (siehe Abschn. 5.7).

***In situ* Strukturbestimmung.** Als Teil dieser Arbeit wurde ein Verfahren zur Strukturbestimmung von Makromolekülen in ihrer annähernd natürlichen Umgebung entwickelt. Die Kombination von Kryoelektronentomographie und *Single Particle* Mittelungstechniken erlaubte z.B. die Auflösung der Quarternärstruktur des Env Komplex von MoMuLV bis zu etwa $(3 \text{ nm})^{-1}$. Mit dieser Auflösung lassen sich bereits biologisch signifikante Fragen beantworten, wie die Gültigkeit eines strukturellen Modells der retroviralen Infektion.

Das hier vorgestellte Verfahren birgt jedoch noch erhebliches weiteres Potential, um höhere Auflösungen zu erzielen. Die finale Auflösung ist derzeit im Wesentlichen durch zwei Faktoren limitiert: Erstens ist die derzeit nutzbare Information der Einzelpartikeldaten im Fourierraum auf den Bereich innerhalb der ersten Nullstelle der CTF beschränkt. Eine Erweiterung dieses Bereichs würde eine Dekonvolution der Tomogramme mit der CTF erfordern, was z.B. von Winkler und Taylor (Winkler und

Taylor 2003) realisiert wurde. Eine zuverlässige CTF Korrektur ist darauf angewiesen, die CTF Parameter anhand der einzelnen EM Bilder zu bestimmen, jedoch ist das SNR der einzelnen Bilder einer Kippserie unter Kryobedingungen hierfür derzeit nicht ausreichend. Wiederum kann die Entwicklung signifikant verbesserter CCD Kameras diesen Zustand möglicherweise entscheidend verbessern. Zweitens ist die erreichbare Auflösung durch die Qualität der Alignierung der gesamten Tomogramme limitiert (siehe Abschn. 2.3). Das in dieser Arbeit verwendete, relativ einfache Abbildungsmodell, das Gl. (4.2) zugrunde liegt, ist nicht ausreichend, um höchste Präzision der Alignierung zu erreichen. Die Implementierung erweiterter Abbildungsmodelle wie Gl. (2.6) oder noch allgemeinere lineare und ggf. auch nichtlineare Transformationen können die Alignierung der gesamten Tomogramme deutlich verbessern. Auf der experimentellen Seite kann die Alignierung durch die Entwicklung mechanisch höchst stabiler Probenhalterungen verbessert werden. Sowohl in der rechnerischen "Post"-Alignierung sowie der mechanischen Alignierung liegt noch großes Potential zur Verbesserung.

Darüberhinaus besteht die Möglichkeit, die in dieser Arbeit entwickelte Alignierungsprozedur weiter zu verbessern. Prinzipiell ist es möglich, verschiedenste Alignierungsverfahren, die in nunmehr über zwei Jahrzehnten im Bereich der *Single Particle Analysis* zur Auflösungssteigerung entwickelt wurden, auf Verfahren zur Mittelung von Subtomogrammen zu übertragen. Zusammenfassend läßt sich sagen, daß eine Auflösung von *in situ* Mittelungen im Bereich von $(2 \text{ nm})^{-1}$ in der nahen Zukunft realistisch ist. Es ist besonders wichtig, diese Auflösungsbarriere zu durchbrechen, da ab dieser Auflösung ein unzweideutiges Einpaßen atomarer Strukturen, die z.B. mit Röntgenkristallographie bestimmt wurden, heute möglich ist (Wriggers und Chácon 2001). Der fortgeschrittenen Interpretation von elektronenmikroskopischen Dichten im mittleren Auflösungsbereich von $(2-4 \text{ nm})^{-1}$ wird in der Zukunft wachsende Bedeutung zukommen, da strukturelle Daten schwach und vorübergehend gebundener Komplexe auf absehbare Zeit in diesem Auflösungsbereich liegen werden. Eine detaillierte Analyse dieser Dichten wird die Kombination der EM Struktur mit weiteren

Daten, die neben experimentell bestimmten atomaren Strukturen auch anhand der Proteinsequenzen vorhergesagten Modelle oder andere biochemische Daten über Proteininteraktionen umfassen, benötigen, und erste Versuche zeigen daß dieser Ansatz sehr großes Potential birgt (Aloy et al. 2004; Topf et al. 2005).

Constrained Principal Component Analysis. Die in dieser Arbeit entwickelte Klassifikationsmethode verringert die Artefakte der klassischen PCA bei der Klassifikation tomographischer Daten signifikant. Ziel der Klassifikation ist es, einzelne makromolekulare Komplexe in strukturell homogene Populationen aufzuteilen. Die Bedeutung der Klassifikation einzelner, in einem Tomogramm enthaltener makromolekularer Komplexe wird in der Zukunft deutlich an Bedeutung gewinnen: Zum einen erfordert die höherauflösende strukturelle Studie von in Tomogrammen enthaltenen Makromolekülen die Mittelung homogener Datensätze. Zum anderen fußt eine rigore statistische Analyse der Zustände eines Makromoleküls in einer biologisch aktiven Umgebung auf einer solchen mathematischen Beschreibung.

Eine erfolgreiche Klassifikation z.B. verschiedener Zustände eines Makromoleküls erfordert jedoch zuvorderst experimentelle Daten, die signifikante Unterschiede im Auflösungsbereich der zu detektierenden Konformationsänderungen aufweisen. Mit der derzeitigen experimentellen Ausrüstung ist dies i.A. nur bei sehr dünnen Proben (~ 100 nm) möglich. Die hochauflösende Abbildung fragiler makromolekularer Komplexe erfordert deshalb entweder das Zerschneiden ganzer, eiseingebetteter Zellen (Kryoschnitten, siehe z.B. (Al-Amoudi et al. 2004)) oder Formen äußerst schonender Aufreinigung, die strukturerhaltend arbeiten und nicht notwendigerweise höchste Reinheiten erreichen wie konventionelle biochemische Techniken (Baumeister 2005). Die Entwicklung beider Techniken wird derzeit vorangetrieben, so daß die entwickelte Partikelmittelung in Verbindung mit der CPCA ein sehr effektives Werkzeug werden kann, um fragile molekulare Maschinen im Auflösungsbereich von $(2-4 \text{ nm})^{-1}$ strukturell zu charakterisieren.

Abkürzungen

2D	zweidimensional
3D	dreidimensional
Abb.	Abbildung
Abschn.	Abschnitt
ART	<i>Algebraic Reconstruction Techniques</i>
ARCF	<i>Auto Rotational Correlation Function</i>
bzgl.	bezüglich
bzw.	beziehungsweise
CA	<i>Capsid</i> (Protein) in Kap. 7 und <i>Correspondence Analysis</i> in Kap. 9
CC	<i>Correlation Coefficient</i>
CCC	<i>Constrained Correlation Coefficient</i>
CCD	<i>Charged Coupled Device</i>
CCF	<i>Constrained Correlation Function</i>
CF	<i>Cytoplasmatic Filament</i>
CP	<i>Central Transporter</i>
CPCA	<i>Constrained Principal Component Analysis</i>
CTF	<i>Contrast Transfer Function</i>
Env	<i>Envelope Protein Complex</i>
EM	Elektronenmikroskopie
ET	Elektronentomographie
FEG	<i>Field Emission Gun</i>
FFT	<i>Fast Fourier Transform</i>
ggf.	gegebenenfalls
Gl.	Gleichung
HA	Hemagglutinin
HIV	<i>Human Immunodeficiency Virus</i>
i.A.	im Allgemeinen
KET	Kryoelektronentomographie
LCCF	<i>Local Constrained Correlation Function</i>
LMCF	<i>Local Mutual Correlation Function</i>

LM	Lichtmikroskop
LR	<i>Luminal Spoke Ring</i>
MA	Matrix (Protein)
MCF	<i>Mutual Correlation Function</i>
MuLV	<i>Murine Leukemia Virus</i>
MoMuLV	<i>Moloney Murine Leukemia Virus</i>
MSA	Multivariate Statistical Analyse
NC	<i>Nucleocapsid</i>
NMR	<i>Nuclear Magnetic Resonance</i>
NPC	<i>Nuclear Pore Complex</i>
PCA	<i>Principal Component Analysis</i>
PDB	<i>Protein Data Bank</i>
RBD	<i>Receptor Binding Domain</i>
RCF	<i>Rotational Correlation Function</i>
SIV	<i>Simian Immunodeficiency Virus</i>
SNR	<i>Signal-to-Noise Ratio</i>
SU	<i>Surface Subunit</i>
Tab.	Tabelle
TEM	Transmissionselektronenmikroskop
TM	<i>Transmembrane Subunit</i>
u.s.w.	und so weiter
XCF	<i>Cross Correlation Function</i>
z.B.	zum Beispiel

Physikalische Konstanten

Lichtgeschwindigkeit	$c = 2.9979 \cdot 10^8 \text{ ms}^{-1}$
Elementarladung	$e = 1.602 \cdot 10^{-19} \text{ C}$
Ruhemasse des Elektrons	$m_0 = 9.1091 \cdot 10^{-31} \text{ kg}$
Ruheenergie des Elektrons	$E_0 = m_0 c^2 = 511 \text{ keV}$
Plancksches Wirkungsquantum	$h = 6.6256 \cdot 10^{-34} \text{ Nms}$ $= 4.136 \cdot 10^{-15} \text{ eVs}$
Dielektrizitätskonstante	$\epsilon_0 = 8.8419 \cdot 10^{-12} \text{ C}^2\text{N}^{-1}\text{m}^{-2}$ $= 1.4165 \cdot 10^{-30} \text{ C}^2\text{eV}^{-1}\text{m}^{-1}$
Bohrscher Radius	$a_H = \frac{h^2 \epsilon_0}{\pi m_0 e^2} = 0.0529 \text{ nm}$
Wellenlänge des Elektrons	$\lambda = hc(E E_0 + E^2)^{-1/2}$ $= 1.97 \cdot 10^{-12} \text{ nm bei } 300 \text{ kV}$
Atomgewicht	$\text{Da} = 1.6605387313 \cdot 10^{-24} \text{ g}$

Literaturverzeichnis

- C. W. Akey und M. Radermacher (1993). *Architecture of the Xenopus nuclear pore complex revealed by three-dimensional cryo-electron microscopy*. J Cell Biol, 122(1):1–19.
- A. Al-Amoudi, J. J. Chang, A. Leforestier, A. McDowall, L. M. Salamin, L. P. Norlen, K. Richter, N. S. Blanc, D. Studer und J. Dubochet (2004). *Cryo-electron microscopy of vitreous sections..* EMBO J, 23(18):3583–8.
- B. Alberts (1998). *The cell as a collection of protein machines: preparing the next generation of molecular biologists*. Cell, 92:291–294.
- P. Aloy, B. Bottcher, H. Ceulemans, C. Leutwein, C. Mellwig, S. Fischer, A.-C. Gavin, P. Bork, G. Superti-Furga, L. Serrano und R. B. Russell (2004). *Structure-Based Assembly of Protein Complexes in Yeast*. Science, 303(5666):2026–2029.
- I. Angert, E. Majorovits und R. R. Schroder (2000). *Zero-loss image formation and modified contrast transfer theory in EFTEM*. Ultramicroscopy, 81(3-4):203–222.
- W. Baumeister (2005). *From proteomic inventory to architecture*. FEBS Lett, 579(4):933–7.
- M. Beck, F. Förster, M. Ecke, J. M. Plitzko, F. Melchior, G. Gerisch, W. Baumeister und O. Medalia (2004). *Nuclear pore complex structure and dynamics revealed by cryoelectron tomography*. Science, 306(5700):1387–90.
- J. Böhm (2001). *Identifikation makromolekularer Komplexe in Elektronentomogrammen eiseingebetteter Phantomzellen*. Doktorarbeit, TU München.

- J. Böhm, A. S. Frangakis, R. Hegerl, S. Nickell, D. Typke und W. Baumeister (2000). *Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms*. Proc Natl Acad Sci U S A, 97(26):14245–50.
- J. Böhm, O. Lambert, A. Frangakis, L. Letellier, W. Baumeister und J.-L. Rigaud (2001). *FhuA-mediated phage genome transfer into liposomes: A cryo-electron tomography study*. Current Biology, S. in press.
- J. A. G. Briggs, M. N. Simon, I. Gross, H.-G. Krausslich, S. D. Fuller, V. M. Vogt und M. C. Johnson (2004a). *The stoichiometry of Gag protein in HIV-1*. Nat Struct Biol, 11(7):672–675.
- J. A. G. Briggs, B. E. Watson, B. E. Gowen und S. D. Fuller (2004b). *Cryoelectron Microscopy of Mouse Mammary Tumor Virus*. J. Virol., 78(5):2606–2608.
- M. Caffrey, M. Cai, J. Kaufman, S. J. Stahl, P. T. Wingfield, D. G. Covell, A. M. Gronenborn und G. M. Clore (1998). *Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41*. Embo J, 17(16):4572–84.
- J. Carazo (1992). *The Fidelity of 3D Reconstructions from Incomplete Data and the Use of Restoration Methods*. In: J. Frank, Hrsg.: *Electron Tomography*, S. 117–166. Plenum Press, New York.
- C. M. Carr und P. S. Kim (1993). *A spring-loaded mechanism for the conformational change of influenza hemagglutinin*. Cell, 73(4):823–32.
- R. J. Center, P. Schuck, R. D. Leapman, L. O. Arthur, P. L. Earl, B. Moss und J. Lebowitz (2001). *Oligomeric structure of virion-associated and soluble forms of the simian immunodeficiency virus envelope protein in the prefusion activated conformation*. Proc Natl Acad Sci U S A, 98(26):14877–82.
- D. C. Chan, D. Fass, J. M. Berger und P. S. Kim (1997). *Core structure of gp41 from the HIV envelope glycoprotein*. Cell, 89(2):263–73.

- B. Chen, Y. Cheng, L. Calder, S. C. Harrison, E. L. Reinherz, J. J. Skehel und D. C. Wiley (2004). *A chimeric protein of simian immunodeficiency virus envelope glycoprotein gp140 and Escherichia coli aspartate transcarbamoylase*. J Virol, 78(9):4508–16.
- L. Chen, J. J. Gorman, J. McKimm-Breschkin, L. J. Lawrence, P. A. Tulloch, B. J. Smith, P. M. Colman und M. C. Lawrence (2001). *The structure of the fusion glycoprotein of Newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion*. Structure (Camb), 9(3):255–66.
- P. M. Colman und M. C. Lawrence (2003). *The structural biology of type I viral membrane fusion*. Nat Rev Mol Cell Biol, 4(4):309–19.
- R. A. Crowther und A. Klug (1971). *ART and science or conditions for three-dimensional reconstruction from electron microscope images*. J Theor Biol, 32(1):199–203.
- R. Crowther, D. DeRosier und A. Klug (1970). *The reconstruction of a three-dimensional structure from its projections and its applications to electron microscopy*. Proceedings of the Royal Society, London, 317:319–340.
- R. A. Davey, C. A. Hamson, J. J. Healey und J. M. Cunningham (1997). *In vitro binding of purified murine ecotropic retrovirus envelope surface protein to its receptor, MCAT-1*. J Virol, 71(11):8096–102.
- R. A. Davey, Y. Zuo und J. M. Cunningham (1999). *Identification of a receptor-binding pocket on the envelope protein of friend murine leukemia virus*. J Virol, 73(5):3758–63.
- S. De Carlo, M. Adrian, P. Kalin, J. M. Mayer und J. Dubochet (1999). *Unexpected property of trehalose as observed by cryo-electron microscopy*. J Microsc, 196(1):40–45.

- D. J. DeRosier und A. Klug (1968). *Reconstruction of three-dimensional structures from electron micrographs*. Nature, 217:130–134.
- K. Dierksen, D. Typke, R. Hegerl und W. Baumeister (1993). *Towards automatic electron tomography. II. Implementation of autofocus and low-dose procedures*. Ultramicroscopy, 49:109–120.
- K. Dierksen, D. Typke, R. Hegerl, A. Koster und W. Baumeister (1992). *Towards automatic electron tomography*. Ultramicroscopy, 40:71–87.
- K. Dierksen, D. Typke, R. Hegerl, J. Walz, E. Sackmann und W. Baumeister (1995). *Three-dimensional structure of lipid vesicles embedded in vitreous ice and investigated by automated electron tomography*. Biophysical Journal, 68(4):1416–1422.
- N. Dimmock, A. Easton und K. Leppard (2001). *Introduction to Modern Virology*. Blackwell Publishing.
- C. Dinges, A. Berger und H. Rose (1995). *Simulation of TEM images considering phonon and electronic excitations*. Ultramicroscopy, 60(1):49–70.
- L. Ditzel, J. Löwe, D. Stock, K.-O. Stetter, H. Huber, R. Huber und S. Steinbacher (1998). *Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT*. Cell, 93:125–138.
- V. D’Souza und M. F. Summers (2004). *Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus*. Nature, 431(7008):586–90.
- J. Dubochet, M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowell und P. Schultz (1988). *Cryo-electron microscopy of vitrified specimens..* Q Rev Biophys, 21(2):129–228.
- D. M. Eckert und P. S. Kim (2001). *Mechanisms of viral membrane fusion and its inhibition*. Annu Rev Biochem, 70:777–810.

- R. Egerton (1976). *Measurement of inelastic/elastic scattering ratio for fast electrons and its use in the study of radiation damage*. Phys Status Solidi A, 37:663–668.
- B. Fahrenkrog, J. Koser und U. Aebi (2004). *The nuclear pore complex: a jack of all trades?*. Trends Biochem Sci, 29(4):175–82.
- G. Y. Fan und M. H. Ellisman (2000). *Digital imaging in transmission electron microscopy*. J Microsc, 200 (Pt 1):1–13.
- D. Fass, R. A. Davey, C. A. Hamson, P. S. Kim, J. M. Cunningham, J. M. Berger, D. C. Chan und S. C. Harrison (1997). *Structure of a murine leukemia virus receptor-binding glycoprotein at 2.0 angstrom resolution*. Science, 277(5332):1662–6.
- D. Fass, S. C. Harrison und P. S. Kim (1996). *Retrovirus envelope domain at 1.7 angstrom resolution*. Nat Struct Biol, 3(5):465–9.
- J. Fienup und C. Wackerman (1986). *Phase-retrieval stagnation problems and solutions*. Journal of the Optical Society of America, 3(11):1897–1907.
- F. Förster, O. Medalia, N. Zauberman, W. Baumeister und D. Fass (2005). *Retrovirus Envelope Protein Complex Structure in Situ Determined by Cryo-Electron Tomography*. Proc Natl Acad Sci U S A, S. submitted.
- A. S. Frangakis, J. Bohm, F. Förster, S. Nickell, D. Nicastro, D. Typke, R. Hegerl und W. Baumeister (2002). *Identification of macromolecular complexes in cryo-electron tomograms of phantom cells*. Proc Natl Acad Sci U S A, 99(22):14153–8.
- A. S. Frangakis und F. Förster (2004). *Computational exploration of structural information from cryo-electron tomograms*. Curr Opin Struct Biol, 14(3):325–31.
- A. Frangakis und R. Hegerl (2002). *Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis*. J Struct Biol., 138:105–113.
- A. Frangakis, A. Stoschek und R. Hegerl (2001). *Wavelet Transform Filtering and Nonlinear Anisotropic Diffusion Assessed for Signal Reconstruction Performance*

- on Multidimensional Biomedical Data*. IEEE Transactions on Biomedical Engineering, 48(2):213–222.
- J. Frank (1976). *Determination of source size and energy spread from electron micrographs using the method of Young's fringes*. Optik, 44(4):379–391.
- J. Frank (1996). *Three-dimensional electron microscopy of macromolecular assemblies*. Academic Press, San Diego.
- J. Frank (2002). *Single-particle imaging of macromolecules by cryo-electron microscopy*. Annu Rev Biophys Biomol Struct, 31:303–19.
- J. Frank, M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj und A. Leith (1996). *SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields*. J Struct Biol, 116(1):190–9.
- I. S. Gabashvili, R. K. Agrawal, C. M. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank und P. Penczek (2000). *Solution structure of the E. coli 70S ribosome at 11.5 Å resolution*. Cell, 100(5):537–49.
- R. Gerchberg und W. Saxton (1972). *A practical algorithm for the determination of phase from image and diffraction plane pictures*. Optik, 35(2):237–246.
- R. Gordon, R. Bender und G. T. Herman (1970). *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography*. J Theor Biol, 29(3):471–81.
- R. Grimm (1997). *Anwendung von Energiefilterung bei der automatisierten Elektrentomographie von eiseingebetteten biologischen Objekten*. Dissertation, Technische Universität München.
- R. Grimm, M. Bärmann, W. Häckl, D. Typke, E. Sackmann und W. Baumeister (1997). *Energy-filtered electron tomography of ice-embedded actin and vesicles*. Biophysical Journal, 72(N1):482–489.

- R. Grimm, A. J. Koster, U. Ziese, D. Typke und W. Baumeister (1996). *Zero-loss energy-filtering under low-dose conditions using a postcolumn energy filter*. Journal of Microscopy, 183(Juli):60–68.
- R. Grimm, H. Singh, R. Rachel, D. Typke, W. Zillig und W. Baumeister (1998). *Electron tomography of ice-embedded procaryotic cells*. Biophysical Journal, 74:1031–1042.
- K. Grunewald, P. Desai, D. C. Winkler, J. B. Heymann, D. M. Belnap, W. Baumeister und A. C. Steven (2003). *Three-dimensional structure of herpes simplex virus from cryo-electron tomography*. Science, 302(5649):1396–8.
- K. Grünewald, O. Medalia, A. Gross, A. C. Steven und W. Baumeister (2003). *Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding*. Biophys Chem, 100(1-3):577–91.
- R. Guckenberger (1982). *Determination of a common origin in the micrographs of tilt series in three-dimensional electron microscopy*. Ultramicroscopy, 9:167–174.
- G. Harauz und M. van Heel (1986). *Exact filter for general geometry three-dimensional reconstruction*. Optik, 73:146–156.
- R. Hart (1968). *Electron microscopy of unstained biological material: the polytropic montage*. Science, 159:1464–1467.
- P. Hawkes (1992). *The electron microscope as a structure projector*. In: J. Frank, Hrsg.: *Electron Tomography*, S. 17–38. Plenum Press, New York.
- M. van Heel (1987). *Angular reconstitution: a posteriori assignment of projection directions for 3D reconstructions*. Ultramicroscopy, 21:111–124.
- M. van Heel und J. Frank (1981). *Use of multivariate statistics in analysing the images of biological macromolecules*. Ultramicroscopy, 6:187–194.

- M. van Heel, B. Gowen, R. Matadeen, E. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz und A. Patwardhan (2000). *Single-particle electron cryo-microscopy: towards atomic resolution*. Q. Rev. Biophys., 33(4):307–69.
- M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt und M. Schatz (1996). *A new generation of the IMAGIC image processing system*. J Struct Biol, 116(1):17–24.
- R. Hegerl (1996). *The EM program package: A platform for image processing in biological electron microscopy*. J Struct Biol, 116:30–34.
- R. Hegerl und W. Hoppe (1976). *Influence of electron noise on three-dimensional image reconstruction*. Z. Naturforschung, 31a:1717–1721.
- H.-G. Heide und E. Zeitler (1985). *The physical behavior of solid water at low temperatures and the embedding of electron microscopical specimens*. Ultramicroscopy, 16:151–160.
- R. Henderson (1995). *The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules*. Q Rev Biophys, 28(2):171–93.
- J. B. Heymann, N. Cheng, W. W. Newcomb, B. L. Trus, J. C. Brown und A. C. Steven (2003). *Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy*. Nat Struct Biol, 10(5):334–41.
- M. Hill (1974). *Correspondence analysis: A neglected multivariate method*. Appl Stat, 23:340–354.
- K. C. Holmes, I. Angert, F. J. Kull, W. Jahn und R. R. Schroder (2003). *Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide*. Nature, 425(6956):423–7.
- H. Hölzl, B. Kapelari, J. Kellermann, E. Seemuller, M. Sumegi, A. Udvardy, O. Medalia, J. Sperling, S. A. Müller, A. Engel und W. Baumeister (2000). *The regulatory*

- complex of Drosophila melanogaster 26S proteasomes. Subunit composition and localization of a deubiquitylating enzyme.* J Cell Biol, 150(1):119–30.
- W. Hoppe (1981). *Three-dimensional electron microscopy.* Annu Rev Biophys Bioeng, 10:563–92.
- W. Hoppe und R. Hegerl, Hrsg. (1980). *Three-dimensional structure determination by electron microscopy (nonperiodic specimens)*, Bd. 13 d. Reihe *Topics in Current Physics Computer Processing of Electron Microscope Images*. Springer-Verlag, Berlin Heidelberg New York.
- J. Horner und P. Gianino (1964). *Phase-only matched filtering.* Appl Opt, 23:812–816.
- A. Howie (2004). *Hunting the Stobbs factor.* Ultramicroscopy, 98(2-4):73–79.
- E. Jaynes (1957). *Information Theory and Statistical Mechanics.* Phys. Rev., 106(4):620–30.
- Q. X. Jiang, D. W. Chester und F. J. Sigworth (2001). *Spherical reconstruction: a method for structure determination of membrane proteins from cryo-EM images.* J Struct Biol, 133(2-3):119–31.
- V. Knauer, R. Hegerl und W. Hoppe (1983). *Three-dimensional reconstruction and averaging of 30 S ribosomal subunits of Escherichia coli from electron micrographs.* J Mol Biol, 163(3):409–30.
- B. Kobe, R. J. Center, B. E. Kemp und P. Pountourios (1999). *Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins.* Proc Natl Acad Sci U S A, 96(8):4319–24.
- A. J. Koster, H. Chen, J. W. Sedat und D. A. Agard (1992). *Automated microscopy for electron tomography.* Ultramicroscopy, 46(1-4):207–27.

- A. J. Koster, W. J. d. Ruijter, A. v. d. Bos und K. D. v. d. Mast (1989). *Autotuning of a TEM using minimum electron dose*. Ultramicroscopy, 27:251–72.
- O. Krivanek, S. Friedman, A. Gubbens und B. Kraus (1995). *An imaging filter for biological applications*. Ultramicroscopy, 59:267–282.
- W. Kunath und H. Sack-Kongehl (1989). *Circular harmonic averaging of noisy single-molecule images*. Ultramicroscopy, 27:171–184.
- J. Kürner, A. S. Frangakis und W. Baumeister (2005). *Cryo-electron tomography reveals the cytoskeletal structure of Spiroplasma melliferum*. Science, 307(5708):436–8.
- P. D. Kwong, R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski und W. A. Hendrickson (2005). *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody*. Nature, 393.
- J. Langmore und M. Smith (1992). *Quantitative energy-filtered electron microscopy of biological molecules in ice*. Ultramicroscopy, 46:349–373.
- M. Lawrence (1992). *Least-squares method of alignment using markers*. In: J. Frank, Hrsg.: *Electron tomography*, S. 197–204. Plenum Press, New York.
- F. Lenz (1954). *Zur Streuung mittelschneller Elektronen in kleinste Winkel*. Zeitschrift für Naturforschung, 9(3):185–204.
- J. Lescar, A. Roussel, M. W. Wien, J. Navaza, S. D. Fuller, G. Wengler und F. A. Rey (2001). *The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH*. Cell, 105(1):137–48.
- A. Linaroudis und R. Hegerl (2005). *Segmentation of cryoelectron tomograms based on the scaling index*. in Vorbereitung.
- Y. Liu, P. Penczek, B. McEwen und J. Frank (1995). *A marker-free alignment method for electron tomography*. Ultramicroscopy, 58:393–402.

- J. Lowe, D. Stock, B. Jap, P. Zwickl, W. Baumeister und R. Huber (1995). *Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 Å resolution*. Science, 268(5210):533–9.
- R. Marabini, E. Rietzel, R. Schroeder, G. Herman und J. Carazo (1997). *Three-Dimensional Reconstruction from Reduced Sets of Very Noisy Images Acquired Following a Single-Axis Tilt Schema: Application of a New Three-Dimensional Reconstruction Algorithm and Objective Comparison with Weighted Backprojection*. J Struct Biology, 120:363–371.
- B. McEwen, K. Downing und R. Glaeser (1995). *The relevance of dose-fractionation in tomography of radiation-sensitive specimens*. Ultramicroscopy, 60:357–373.
- R. McIntosh, D. Nicastro und D. Mastronarde (2005). *New views of cells in 3D: an introduction to electron tomography*. Trends Cell Biol, 15(1):43–51.
- O. Medalia, I. Weber, A. S. Frangakis, D. Nicastro, G. Gerisch und W. Baumeister (2002). *Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography*. Science, 298(5596):1209–13.
- J. Miao, P. Charalambous, J. Kirz und D. Sayre (1999). *Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens*. Nature, 400(6742):342–344.
- D. Misell und I. Burdett (1977). *Determination of the mass thickness of biological sections from electron micrographs*. J Microsc, 109:171–182.
- P. Monnard, N. Berclaz, K. Conde-Frieboes und T. Oberholzer (1999). *Decreased solute entrapment in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine liposomes prepared by freeze/thaw in the presence of physiological amounts of monovalent salts*. Langmuir, 15(22):7504–7509.

- G. B. Mortuza, L. F. Haire, A. Stevens, S. J. Smerdon, J. P. Stoye und I. A. Taylor (2004). *High-resolution structure of a retroviral capsid hexameric amino-terminal domain*. Nature, 431(7007):481–5.
- F. C. Neidhardt (1996). *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*.
- S. Nickell (2001). *Elektronentomographische Abbildung eiseingebetteter prokaryotischer Zellen*. Dissertation, Technische Universität München.
- S. Nickell, F. Förster, A. Linaroudis, W. Del Net, F. Beck, R. Hegerl, W. Baumeister und J. M. Plitzko (2004). *TOM toolbox Acquisition and Analysis for Electron Tomography*. J Struct Biol, in press.
- M. Nitsch, J. Walz, D. Typke, M. Klumpp, L. O. Essen und W. Baumeister (1998). *Group II chaperonin in an open conformation examined by electron tomography*. Nat Struct Biol, 5(10):855–7.
- E. Nogales, S. G. Wolf und K. H. Downing (1998). *Structure of the alpha beta tubulin dimer by electron crystallography*. Nature, 391(6663):199–203.
- H. Oettl, R. Hegerl und W. Hoppe (1983). *Three-dimensional reconstruction and averaging of 50 S ribosomal subunits of Escherichia coli from electron micrographs*. J Mol Biol, 163(3):431–50.
- J. Ortiz, F. Förster, J. Kurner, A. Linaroudis und W. Baumeister (2005). *Mapping 70S ribosomes in intact cells by cryo-electron tomography and pattern recognition*.
- P. Penczek, M. Radermacher und J. Frank (1992). *Three-dimensional reconstruction of single particles embedded in ice*. Ultramicroscopy, 40:33–53.
- J. Radon (1917). *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten*. Berichte Sächsische Akademie der Wissenschaften, Leipzig, Math.-Phys. Kl., 69:262–277.

- L. Reimer (1997). *Transmission electron microscopy*, Bd. 36 d. Reihe *Springer Series in Optical Sciences*. Springer Verlag, Heidelberg, 4th Aufl.
- L. Reimer und M. Ross-Messemer (1989). *Contrast in the electron spectroscopic imaging mode of a TEM. I. Influence of zero-loss filtering on scattering contrast*. *Journal of Microscopy*, 155(2):169–182.
- A. M. Roseman (2003). *Particle finding in electron micrographs using a fast local correlation algorithm*. *Ultramicroscopy*, 94(3-4):225–36.
- A. Roseman (2000). *Docking structures of domains into maps from cryo-electron microscopy using local correlation*. *Acta Crystallographica D*, 56:1332–1340.
- P. B. Rosenthal und R. Henderson (2003). *Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy*. *J Mol Biol*, 333(4):721–45.
- B. Rost (1996). *PHD: predicting one-dimensional protein structure by profile-based neural networks*. *Methods Enzymol*, 266:525–39.
- J. L. Rubinstein, J. E. Walker und R. Henderson (2003). *Structure of the mitochondrial ATP synthase by electron cryomicroscopy*. *Embo J*, 22(23):6182–92.
- A. Saad, S. J. Ludtke, J. Jakana, F. J. Rixon, H. Tsuruta und W. Chiu (2001). *Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination*. *J Struct Biol*, 133(1):32–42.
- A. Sali, R. Glaeser, T. Earnest und W. Baumeister (2003). *From words to literature in structural proteomics*. *Nature*, 422(6928):216–25.
- S. Sandin, L.-G. Ofverstedt, A.-C. Wikstrom, O. Wrangé und U. Skoglund (2004). *Structure and Flexibility of Individual Immunoglobulin G Molecules in Solution*. *Structure*, 12(3):409–415.

- B. Saxberg und W. Saxton (1981). *Quantum noise in 2D projections and 3D reconstructions*. Ultramicroscopy, 6:85–90.
- F. Schwabl (1990). *Quantenmechanik*. Springer Verlag, Berlin, Heidelberg.
- G. Schweikert (2004). *Quantitativer Vergleich der Strahlschädigung biologischer Proben im Transmissions-Elektronenmikroskop bei Stickstoff und Helium Temperatur*. Diplom, Technische Universität München.
- F. J. Sigworth (1998). *A maximum-likelihood approach to single-particle image refinement*. J Struct Biol, 122(3):328–39.
- D. Stoffer, B. Feja, B. Fahrenkrog, J. Walz, D. Typke und U. Aebi (2003). *Cryo-electron tomography provides novel insights into nuclear pore architecture: implications for nucleocytoplasmic transport*. J Mol Biol, 328(1):119–30.
- A. Stoschek und R. Hegerl (1997). *Automated detection of macromolecules from electron micrographs using advanced filter techniques*. J Microsc, 185:76–84.
- K. A. Taylor, J. Tang, Y. Cheng und H. Winkler (1997). *The use of electron tomography for structural analysis of disordered protein arrays*. J Struct Biol, 120(3):372–86.
- M. Topf, M. L. Baker, B. John, W. Chiu und A. Sali (2005). *Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy*. Journal of Structural Biology, 149(2):191–203.
- C. Toyoshima und N. Unwin (1988). *Contrast transfer for frozen-hydrated specimens. I. Determination from pairs of defocused images*. Ultramicroscopy, 25:279–292.
- D. Typke, K. Dierksen und W. Baumeister (1991). *Automatic electron tomography*. In: *49th Annual Meeting EMSA*, S. 544–545. San Francisco Press.
- M. Van Heel, M. Schatz und E. Orlova (1992). *Correlation function revisited*. Ultramicroscopy, 46:307–316.

- A. VanderLugt (1964). *Signal detection by complex spatial filtering*. IEEE Trans. Inf. Theory, IT-10:139–145.
- B. Vijaya Kumar, F. Dickey und J. DeLaurentis (1992). *Correlation filters minimizing peak location errors*. J. Opt. Soc. Am. A, 9(5):678–682.
- D. Voges, P. Zwickl und W. Baumeister (1999). *The 26S proteasome: a molecular machine designed for controlled proteolysis*. Annu Rev Biochem, 68:1015–68.
- J. Walz (1997). *Elektronentomographische Abbildung und dreidimensionale Mittelung von biologischen Makromolekülen*. Dissertation, Technische Universität München.
- J. Walz, A. Erdmann, M. Kania, D. Typke, A. J. Koster und W. Baumeister (1998). *26S proteasome structure revealed by three-dimensional electron microscopy*. J Struct Biol, 121(1):19–29.
- J. Walz, T. Tamura, N. Tamura, W. Baumeister, R. Grimm und A. Koster (1997a). *Tricorn protease exists as an icosahedral supermolecule in vivo*. Mol Cell, 1(1):59–65.
- J. Walz, D. Typke, M. Nitsch, A. J. Koster, R. Hegerl und W. Baumeister (1997b). *Electron Tomography of Single Ice-Embedded Macromolecules: Three-Dimensional Alignment and Classification*. J Struct Biol, 120(3):387–95.
- D. J. Webre, P. M. Wolanin und J. B. Stock (2003). *Bacterial chemotaxis*. Curr Biol, 13(2):R47–9.
- W. Weissenhorn, A. Dessen, S. C. Harrison, J. J. Skehel und D. C. Wiley (1997). *Atomic structure of the ectodomain from HIV-1 gp41*. Nature, 387(6631):426–30.
- T. Wilk, F. de Haas, A. Wagner, T. Rutten, S. Fuller, R. M. Flugel und M. Lochelt (2000). *The intact retroviral Env glycoprotein of human foamy virus is a trimer*. J Virol, 74(6):2885–7.

- I. A. Wilson, J. J. Skehel und D. C. Wiley (1981). *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution*. Nature, 289(5796):366–73.
- H. Winkler und K. A. Taylor (2003). *Focus gradient correction applied to tilt series image data used in electron tomography*. J Struct Biol, 143(1):24–32.
- W. Wriggers und P. Chácon (2001). *Modeling Tricks and Fitting Techniques for Multiresolution Structures*. Structure, 9:779–788.
- C. Yang, E. G. Ng und P. A. Penczek (2004). *Unified 3-D structure and projection orientation refinement using quasi-Newton algorithm*. J Struct Biol, In Press, Corrected Proof.
- M. Yeager, E. M. Wilson-Kubalek, S. G. Weiner, P. O. Brown und A. Rein (1998). *Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms*. Proc Natl Acad Sci U S A, 95(13):7299–304.
- W. Zhang, S. Mukhopadhyay, S. V. Pletnev, T. S. Baker, R. J. Kuhn und M. G. Rossmann (2002). *Placement of the structural proteins in Sindbis virus*. J Virol, 76(22):11645–58.
- Z. H. Zhou und W. Chiu (2003). *Determination of icosahedral virus structures by electron cryomicroscopy at subnanometer resolution*. Adv Protein Chem, 64:93–124.
- J. Zhu, P. A. Penczek, R. Schröder und J. Frank (1997). *3-dimensional reconstruction with contrast transfer-function correction from energy-filtered cryoelectron micrographs - procedure and application to the 70s escherichia-coli ribosome*. Journal of Structural Biology, 118(N3):197–219.
- P. Zhu, E. Chertova, J. Bess, J., J. D. Lifson, L. O. Arthur, J. Liu, K. A. Taylor und K. H. Roux (2003). *Electron tomography analysis of envelope glycoprotein*

- trimers on HIV and simian immunodeficiency virus virions*. Proc Natl Acad Sci U S A, 100(26):15812–7.
- Y. Zhu, B. Carragher, R. M. Glaeser, D. Fellmann, C. Bajaj, M. Bern, F. Mouche, F. d. Haas, R. J. Hall und D. J. Kriegman (2004). *Automatic particle selection: results of a comparative study*. Journal of Structural Biology, 145(1-2):3–14.
- U. Ziese, W. J. Geerts, T. P. Van Der Krift, A. J. Verkleij und A. J. Koster (2003). *Correction of autofocusing errors due to specimen tilt for automated electron tomography*. J Microsc, 211(Pt 2):179–85.

Veröffentlichungen, die Teile der Dissertation enthalten

F. Förster, O. Medalia, N. Zauberman, W. Baumeister und D. Fass (2005). *Retrovirus Envelope Protein Complex Structure in Situ Determined by Cryo-Electron Tomography*. Proc. Natl. Acad. Sci. USA, im Druck.

V. Lucic, F. Förster und W. Baumeister (2005). *Structural Studies by Electron Tomography: From Cells to Molecules*. Ann. Rev. Biochem., im Druck.

S. Nickell, F. Förster, A. Linaroudis, W. Del Net, F. Beck, R. Hegerl, W. Baumeister und J.M. Plitzko (2005). *TOM toolbox Acquisition and Analysis for Electron Tomography*. J. Struct. Biol. 149, 227-34.

J. Ortiz, F. Förster, J. Kürner, A. Linaroudis und W. Baumeister (2005). *Mapping 70S ribosomes in intact cells by cryo-electron tomography and pattern recognition*. in Vorbereitung.

M. Beck, F. Förster, M. Ecke, J.M. Plitzko, F. Melchior, G. Gerisch, W. Baumeister und O. Medalia. *Nuclear pore complex structure and dynamics revealed by cryoelectron tomography*. Science 306, 1387–90.

A.S. Frangakis und F. Förster (2004). *Computational exploration of structural information from cryoelectron tomograms*. Curr. Opin. Struct. Biol. 14, 325-31.

A.S. Frangakis, J. Böhm, , F. Förster, S. Nickell, D. Nicastro, D. Typke, R. Hegerl und W. Baumeister (2002). *Identification of macromolecular complexes in cryoelectron tomograms of phantom cells*. Proc. Natl. Acad. Sci. USA 99, 14153-8.

J.M. Plitzko, A.S. Frangakis, S. Nickell, F. Förster, A. Gross und W. Baumeister (2002). *In vivo veritas: electron cryotomography of cells*. Trends in Biotechnology 20, S40-S44.

Weitere Veröffentlichungen

J. Briggs, K. Grünewald, B. Glass, F. Förster, H.G. Krausslich und S. Fuller (2005). *Cryo-electron tomography of authentic HIV-1: implications for the mechanism of core assembly*. in Vorbereitung.

V. Lucic, T. Yang, G. Schweikert, F. Förster und W. Baumeister (2005). *Cryo-electron Tomography of Synaptosomal Fractions*. Structure, im Druck.

P. Kurz, F. Förster, L. Nordström, G. Bihlmayer und S. Blügel (2004). *Ab initio treatment of noncollinear magnets with the full-potential linearized augmented plane wave method*. Phys. Rev. B. 6902, 4415-4429.

Danksagung

Die vorliegende Arbeit wurde von Juni 2001 bis März 2005 in der Abteilung Molekulare Strukturbiologie des Max-Planck-Instituts für Biochemie, Martinsried, durchgeführt. Ich möchte mich bei allen Mitarbeitern für das gute Arbeitsklima bedanken. Danken möchte ich besonders:

Herrn Prof. Wolfgang Baumeister für die Betreuung der Arbeit und ständige Unterstützung. Insbesondere bin ich dankbar, einen zweimonatigen Gastaufenthalt am Californian Institute of Technology verbracht haben zu dürfen.

Ohad Medalia für die Initiierung des MoMuLV Projekts, das Vorantreiben des NPC Projekts und diverse anregende Diskussionen.

Martin Beck für die angenehme Zusammenarbeit bei der Strukturbestimmung des NPCs und sein Verständnis beim Auftauchen diverser Fehler meiner Programmierung.

Deborah Fass (Weizmann Institut, Israel) für die Kollaboration bei der Bestimmung der Env Struktur und die Hilfe beim Schreiben des Manuskripts.

Achilleas Frangakis für die intensive Zusammenarbeit während der ersten zwei Jahren meiner Doktorarbeit am Thema der Lokalisierung von Komplexen.

Julio Ortiz für die gemeinsame Arbeit bei der Identifikation der Ribosomen in *Spiroplasma melliferum*.

Oana Mihalache für nimmermüden Einsatz beim Versuch, Kryopräparationen intakter 26S Proteasomen zu erzeugen.

Jürgen Plitzko für die Hilfestellung am TEM, das Korrekturlesen der Arbeit und die Gewährleistung des Mikroskopiebetriebs im "Bunker".

Vladan Lucic für viele stimulierende Diskussionen bei der Entstehung des "großen" Reviews.

Stephan Nickell für die Initiierung und Koordination der TOM-Toolbox und die Einweisung in den ersten Monaten meiner Doktorarbeit.

Gabriele Schweikert für das Korrekturlesen der Arbeit und das Interesse an meinen Vorschlägen für ihre eigene Arbeit.

Alexandros Linaroudis für die Hilfe bei Problemen der Visualisierung und Segmentierung.

meinen langjährigen Mitdoktorandinnen aus der Tomographiegruppe, Ariane Briegel, Christine Kofler und Julia Kürner, für das immer angenehme und kollegiale Verhältnis.

Reiner Hegerl für zahllose Diskussionen über die Mathematik der Abbildung, Tomographie und allgemeinen Bildverarbeitung sowie das Korrekturlesen dieser Arbeit, insbesondere der darin enthaltenen Gleichungen.

Weiterhin danke ich der Stiftung Stipendien-Fonds der Chemischen Industrie und dem Bundesministerium für Bildung und Forschung für die Vergabe eines Kékulé Stipendiums für die Zeit von November 2001 bis Oktober 2003.

Besonderen Dank schulde ich meiner Freundin Beatrix für ihre stete Unterstützung, ihren unverzichtbaren Rat, gelegentliche Motivation und vieles mehr.

Am meisten danke ich meinen Eltern, die mir meine gesamte Ausbildung ermöglicht haben und mich im Verlauf der Arbeit unentwegt gefördert haben.