Negative Stain Electron Microscopy and Single Particle Analysis on Myosin VI and Myosin IX

Dario Saczko-Brack

München 2017

Aus dem Physiologischen Institut der Ludwig-Maximilians-Universität München
Lehrstuhl: Zelluläre Physiologie
Vorstand: Prof. Dr. Claudia Veigel

Negative Stain Electron Microscopy and Single Particle Analysis on Myosin VI and Myosin IX

Dario Saczko-Brack

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Dario Saczko-Brack
aus Aachen

München, den 23. November 2017

Erstgutachter: Prof. Dr. Hermann E. Gaub
Zweitgutachter: Prof. Dr. Claudia Veigel
Tag der mündlichen Prüfung: 23. Januar 2018

”The electron microscope is like the monkey wrench on the garage wall; what you do with it is the important thing.”

James Hillier, 1960

Zusammenfassung

In der vorliegenden Arbeit wurde die negativ-kontrastfärbende Elektronenmikroskopie (Negativkontrast EM) angewendet, um biologische Objekte strukturell zu untersuchen[1]. Auch wenn es heutzutage bereits höher auflösende Mikroskopietechniken gibt, wie zum Beispiel Cryo-EM, hat die kontrastfärbende Elektronenmikroskopie nichts von ihrem Charme und ihren Fähigkeiten verloren. Insbesondere dann, wenn die zu untersuchenden Objekte eine bestimmte Größe unterschreiten. So würden beispielsweise die in dieser Arbeit behandelten Proteine auf Grund Ihrer Größe (unter ca. 200 kDa) einen zu geringen Kontrast liefern, um mittels Cryo-EM aufgenommen werden zu können. Im Gegensatz dazu ist die Negativkontrast Elektronenmikroskopie auch heute noch ein hervorragendes Mittel für erste strukturelle Untersuchungen von noch nicht oder wenig charakterisierten Molekülenmachen auf Grund des brillanten Kontrastes, der gleichzeitig dennoch hohen Auflösung (ca. 10-20 Å) und der Reduzierung der Freiheitsgrade durch Fixierung der Moleküle auf dem EM-Gitternetz in einer meistens geringen Anzahl von Orientierungen. Allerdings ist die Mikroskopie selber nicht der einzige Faktor, welcher für die erfolgreiche strukturelle Untersuchung von Bedeutung ist. Einen mindestens ebenso hohen Stellenwert nimmt die Analyse der Bilddaten mittels bildverarbeitenden Methoden ein (Einzelmolekülanalyse, bzw. Single Particle Analysis, SPA). Programme, welche auf solche Anwendungen spezialisiert sind, gibt es mittlerweile viele, wobei der Trend dahin geht, vorgefertigte Protokolle auch ohne tieferes Hintergrundwissen auf eigene Datensätze anwenden zu können. In den meisten Fällen ist dies auch erfolgversprechend, allerdings ist es von großem Vorteil, die hinter den Protokollen liegenden Abläufe zu kennen und zu verstehen. Denn nur so ist es möglich, eigene, auf den jeweiligen Datensatz spezialisierte Prozesse zu entwickeln.

Der zentrale Aspekt dieser Arbeit war es, mittels Negativkontrast EM im Zusammenspiel mit auf SPIDER (EM-spezifisches Bildverarbeitungstool) basierenden, maßgeschneiderten SPA Protokollen Myosine strukturell zu charakterisieren. Diese Protokolle beinhalten bestimmte Abfolgen verschiedener Prozesse, wie z. B. das gegeneinander Ausrichten,

[1] Der Begriff *strukturell* wird sich im Laufe dieser Schrift hauptsächlich auf die zwei- bzw. dreidimensionale räumliche Anordnung eines Objektes beziehen.

Klassifizieren und Überlagern der Rohdaten, sowie das Vergleichen mit entsprechenden Kristallstrukturen zur Interpretation der EM Bilder.
Myosine sind eine Form von eukaryontischen molekularen Motoren, die in der Lage sind, chemische Energie in Form von ATP in mechanische Energie umzuwandeln, um somit Kraft oder Bewegung zu erzeugen. Zusammen mit ihrer Eigenschaft, Aktinfilamente zu binden, tragen sie unter anderem maßgeblich zum intrazellulären Transport von Vesikeln und Organellen bei und sind essenziell für die Funktion von Muskeln. Darüber hinaus sind Myosine ein wichtiger Bestandteil für eine Vielzahl weiterer zellulärer Vorgänge, weshalb sich im Laufe der Evolution spezialisierte Myosin Klassen gebildet haben. Zwei dieser Klassen, Myosin VI und Myosin IX wurden in der vorliegenden Arbeit strukturell untersucht:

Myosin VI ist auf Grund seiner Eigenschaft, die einzige Myosinklasse zu sein, welche sich entlang von Aktinfilamenten in Richtung des Filament-Minus-Endes bewegt, hochinteressant. Diese Einzigartigkeit erlaubt es Myosin VI, spezialisierte Aufgaben an einerseits prozessiven Vorgängen wie Endo-, Exo- und Autophagozytose, und andererseits an strukturbildenden Funktionen wie dem Ausbilden von Stereozilien in den Sinneshaarzellen des Innenohrs zu übernehmen. Darüber hinaus wurde eine Überexpression von Myosin VI Molekülen in Prostata- und Ovarialkarzinomen gemessen. Dieses vermehrte Vorkommen kann als Marker für das Fortschreiten des Krebses betrachtet werden. Um dieses Aufgabenspektrum abdecken zu können, ist eine komplexe Regulierung und Aktivierung dieses Motors unabdingbar. Die Mechanismen dahinter allerdings sind weitestgehend unerforscht. In der folgenden Arbeit konnten durch die Anwendung der entwickelten SPA Methoden bisher unentdeckte Kalzium und Calmodulin abhängige, strukturelle Veränderungen des Myosin VI Motors zum Vorschein gebracht werden. Die Analyse der EM Daten führte zu Rekonstruktionen mit einer Auflösung von 18 Å, welche eine Neuanordnung des zweiten, am Hebelarm des Myosin bindenden Calmodulin und des daran anschließenden Drei-Helix-Bündels entschlüsselten. Mit dem Hintergrundwissen, dass Calmoduline eine stabilisierende Funktion für den Hebelarm haben und dieser essenziell für die Funktion des molekularen Motors ist, wurde im Zusammenhang mit dieser strukturellen Reorganisation ein kalziumbedingter Verlust in der Stabilität des Hebels aufgelöst. Diese Instabilität konnte anschließend mit dem Verlust von Motilität in Kontext gebracht werden.
Basierend auf unter anderem diesen Ergebnissen, ließ sich ein neuartiger zweistufiger Regulationsmechnismus von Myosin VI beschreiben: Zunächst fungiert Kalzium als zellulärer Auslöser, durch welchen der Motor von einem ruhenden, inaktivierten Zustand in einen zwar Fracht bindenden, jedoch weiterhin nicht-beweglichen Zustand versetzt wird. Erst der Rückgang von Kalzium generiert Fracht-gebundene, aktivierte bzw. be-

wegliche Motoren. Im Falle von ungebundener Fracht geht Myosin VI wieder in seinen ursprünglichen Zustand zurück, bereit für den nächsten Zyklus.

Ebenso von hohem Interesse ist das im Zytoskelett häufig vorkommende Myosin IX, da es eine zusätzliche Sequenz von ca. 200 Aminosäuren im Loop2 seiner Motordomäne besitzt. Solch ein Loop ist in den meisten Myosinklassen vorhanden, beteiligt an der Aktinbindung und in der Regel nur wenige Aminosäuren lang. Diese einzigartige Sequenz macht Myosin IX besonders und lässt hochspezialisierte Funktionen vermuten. Myosin IX kann in unterschiedlichen Geweben und Strukturen gefunden werden, wie zum Beispiel Gehirn, Testis oder Zellen des Immunsystems, und ist dabei hauptsächlich im Bereich von Zell-Zell-Kontakten lokalisiert. Obwohl auch für diese Myosinklasse eine prozessive Bewegung entlang von Aktinfilamenten nachgewiesen wurde, wird seine Hauptfunktion in der Bildung und Aufrechterhaltung von dynamischen Aktin - Myosin - Strukturen im Zytoskelett vermutet. Weiterhin wurde eine Fehlfunktion dieses Motors in Zusammenhang mit einer Beeinträchtigung von gerichteter Zellbewegung gebracht. Durch die Untersuchung von Myosin IX als eigenständiges Projekt der vorliegenden Arbeit konnte erstmals ein direkter Zusammenhang mit der Formation von Aktinnetzwerken nachgewiesen und diese mittels Negativkontrast EM und entsprechenden SPA Methoden strukturell charakterisiert werden. Die Rekonstruktionen von Aktin-Myosin-Netzwerken mit einer Auflösung von ca. 30 Å enthüllten deren zugrundeliegenden Strukturen. Es konnte gezeigt werden, dass die Myosin IX Motordomäne alleine ausreichend war, Aktinfilamente in einer streng geordneten und kooperativen Art und Weise zu verbinden. Essenziell dafür war die Myosin IX spezifische Sequenz im Loop2, welche als zweite Aktinbindestelle identifiziert wurde. Einzelne Moleküle verknüpften zwei Aktinfilamente alle 36 nm, entsprechend der helikalen Struktur von Aktinfilamenten. Die Struktur der vernetzenden Myosin IX Moleküle konnte in drei Klassen unterteilt werden. Während zwei davon tatsächliche Verbindungen zwischen zwei Filamenten ausbilden, bindet eine mit seinen beiden Aktinbindestellen an ein und dasselbe Filament. In dieser Formation wäre eine prozessive Bewegung entlang des Filamentes möglich, um eine passende Bindestelle bzw. ein in Reichweite befindliches Filament zu finden. Die Analyse der Netzwerke entschlüsselte weiterhin das Polaritätsverhältnis der verbundenen Filamente, welche fast ausschließlich parallel zueinander angeordnet waren. Wie schon bekannt bei anderen Myosinklassen, konnte eine Regulierung über Kalzium aufgelöst werden, unter dessen Einfluss sich das Polaritätsverhältnis umkehrte.

Abschließend wurden die in dieser Arbeit erlangten Erkenntnisse in Kombination mit Ergebnissen aus anderen biochemischen Methoden gebracht, welche von meinen Kollegen simultan zu den strukturellen Untersuchungen ausgeführt wurden, um ein umfassendes Bild der zugrundeliegenden Mechanismen und Prozesse zu zeichnen.

Abstract

In this work, negative stain electron microscopy was used to structurally investigate the myosin classes VI and IX[2]. Although cryo-EM is becoming the tool of choice these days due to its ability to resolve structures down to the atomic scale, the negative stain method did not lose any of its power and magic. The high contrast of stained particles, the reduction of degrees of freedom due to a usually small number of preferred orientations of molecules on the EM grid, and the yet good resolution of approx. 10-20 Å, make negative stain EM still an excellent approach for first insights into structurally un- or less well known molecules. Moreover, for relatively small objects like myosin (below 200 kDa), this technique is often the only possible method, as for instance cryo-EM produces not enough contrast to detect such particles. Likewise, newly discovered mechanisms and properties of well-studied objects can be effectively reviewed on a structural level by negative stain EM. However, a successful structural characterization using the electron microscope is not only based on the EM operation itself, but also heavily dependent on the subsequent image processing of the obtained data set (single particle analysis, SPA). While many "ready-to-use" software packages for this purpose become available, the saying by James Hillier[3] is today as apposite as it was back in 1960: "The electron microscope is like the monkey wrench on the garage wall: What you do with it is the important thing." In analogy to this quote, knowing the underlying image processing algorithms and being able to control them to a large degree, will improve the outcome significantly.

The central aim of this work was to develop and apply customized protocols for single particle analysis of negatively stained electron micrographs (including mainly alignment, averaging, classification, and crystal structure projection matching) based on the SPIDER software (EM specific image processing tool) in order to investigate the structural properties of myosin. These protocols include mainly the alignment, averaging, and classification of images, as well as crystal structure projection matching, in order to

[2]Henceforth, the expression *structure* will refer throughout the following work to the two- and three-dimensional spatial arrangement of molecules.

[3]Dr. James Hillier was heavily involved in the development of the first practical electron microscope in the late 1940's, based on the first built EM by Ernst Ruska and Max Knoll in 1933.

interpret the resulting data.
Myosins are actin activated ATPase motor proteins that are able to transform chemical energy into mechanical work, in order to generate force or movement. In the cell they are responsible for a large number of functions, such as muscle contraction, membrane anchorage, longer range transport of cargo, collective cell migration, and cell signaling. Given the wide variety of tasks, many different specialized myosin classes have been emerged throughout the evolution. Two of these, myosin class VI and class IX, were subject to the research presented in this work.

The cytoskeletal motor myosin VI is highly interesting due to its unique ability to move towards the minus end of actin filaments. This unique feature enables myosin VI to carry out important tasks in many motile and non-motile processes such as endo- and exocytosis or the formation of stereocilia in the sensory hair cells of the ear. Furthermore, being overexpressed in migrating carcinoma cells, myosin VI can be used as an early marker for the development of these kinds of cancer cells. However, the profound mechanisms behind its regulation and activation are not fully understood. By applying the developed SPA protocols to large image data sets of single myosin VI molecules, undiscovered calcium- and calmodulin-induced structural changes of myosin VI could be resolved with a resolution of up to 18 Å. In particular, the second calmodulin bound to the lever arm of myosin VI and the following three helix bundle underwent major rearrangements. Combined with the exploration of calcium dependent lever arm stiffness, these studies revealed that the structural rearrangements in the neck caused a break in its stability, which was brought into context with the loss of (motile) functionality. We further proposed a novel two-step mechanism for the regulation of myosin VI, where calcium is the cellular switch that directs the rearrangement of the motor from a dormant, inactive state at low calcium to a cargobinding non-motile state at high calcium. The return to low calcium generates either cargobound active motors, which translocate to the center of the cell, or refolded inactive motors ready for the next cellular calcium flux.

Myosin IX, on the other hand, is unique due to an approximately 200 amino acid (aa) long insert into loop2 of its catalytic domain. While the short loop2 itself (approx. 20 aa) is present in almost all myosin classes and is known to be involved in actin binding, the insert in myosin IX is not found in any other class and is thought to provide unique and highly specified functions. Myosin IX, mainly located at the sites of cell-cell-contacts, can be found for example in the brain and testis and is crucial for the epithelial differentiation and morphology of cells. Furthermore, although known to be able to move processively along actin filaments, this molecular motor is thought to have more of a structural role.

Our studies revealed for the first time, that myosin IX was heavily involved in the formation, organization and maintenance of actin networks. In particular the unique insert in loop2 was crucial for these processes. In the project presented in this work, first structural information of these actin - myosin - networks have been resolved using negative stain EM and specialized SPA methods. It could be shown that the catalytic domain of myosin IX assembled actin bundles in highly ordered and cooperative lattices at a repeat distance of precisely 36 nm, matching the helical repeat of actin. Furthermore, the underlying structural characteristics of the crosslinks could be revealed. Actin filaments with parallel polarity and in phase were linked by the myosin IX motor domain, which could adopt three distinct conformations. In two of these myosin IX actually crosslinked different actin filaments, while in one the motor only bound to two adjacent actin monomers of the same filament. This suggested that molecules in this conformation had possibly an exploring function, being able to move processively along the actin filament in order to find their preferred binding position. Furthermore, the polarity of the crosslinks was found to be regulated by calcium, switching to an anti-parallel polarity of the filaments in the presence of it.

Simultaneously to the structural investigation, both myosin were examined with various biochemical techniques and under different aspects in our lab by my colleagues. Although each result was strong in its own niche, the combination of these diverse points of view revealed novel and intriguing underlying processes of how myosins can be activated, regulated and take actively part in the formation of larger complex structures.

The following thesis is divided into several parts: Part I introduces the reader into the basic biological and in particular physical principles. Subsequently, the studies on myosin VI and myosin IX are presented in Part II, separated into two self contained chapters. Both, Chapter 5 and Chapter 6 include their own introduction, result, method, and discussion section with the purpose that they can be accessed independently. The work is concluded by a general discussion in Part III.

Contents

Zusammenfassung **vii**

Abstract **xi**

List of abbreviations **xix**

I Introduction **1**

1 Biological background **5**
1.1 Actin . . . 7
1.1.1 Preparation of F-actin . . . 8
1.2 Myosin . . . 11
1.3 Methods . . . 15

2 Electron microscopy **17**
2.1 History and principle . . . 17
2.2 Setup of a Transmission Electron Microscope . . . 21
2.2.1 Electron gun . . . 22
2.2.2 Condenser lenses and their effect on resolution . . . 22
2.2.3 Specimen stage . . . 24
2.2.4 Objective and projector lenses . . . 24
2.2.5 Projector lens and imaging devices . . . 27
2.3 Principles of image formation in the TEM . . . 27
2.3.1 Contrast Transfer Function CTF . . . 28
2.4 Negative staining: Principle and sample preparation . . . 31

3 Single particle image processing **37**
3.1 Two-dimensional alignment and averaging methods . . . 39
3.1.1 The cross correlation function CCF . . . 39
3.1.2 Reference-free and reference based alignment . . . 40
3.2 Two-dimensional classification: Multivariate data analysis . . . 43
3.2.1 Principal component analysis PCA . . . 44
3.2.2 Correspondence analysis CA . . . 45
3.2.3 Factor maps . . . 47
3.2.4 Classification by hierarchical ascendant clustering HAC . . . 50
3.2.5 Classification by K-means . . . 52

3.3 Resolution determination (2D) using FRC . . . 54
3.4 Software for image processing . . . 56
3.4.1 SPIDER / SPIRE . . . 56

4 Crystal structure: Modification and Projection matching 61
4.1 Modification . . . 62
4.2 Projection matching . . . 62

II Results 67

5 Myosin VI 69
5.1 Introduction . . . 71
5.2 Results . . . 75
5.2.1 Effect of calcium and calmodulin on myosin VI: Structure . . . 75
5.2.2 Effect of calcium and calmodulin on myosin VI: Statistics . . . 84
5.2.3 Calcium induced conformational changes of myosin VI . . . 88
5.2.4 Destabilization of the myosin VI lever arm induced by calcium . 96
5.3 Discussion . . . 101
5.3.1 Model: Two-step mechanism for calcium regulation of myosin VI 108
5.4 Outlook . . . 113
5.5 Methods . . . 115
5.5.1 Preparation, staining and image acquisition of nucleotide-free full-length myosin VI . . . 115
5.5.2 SPA workflow for full-length myosin VI . . . 117
5.5.2.1 Step 1: General shape finding . . . 117
5.5.2.2 Step 2: Multireference alignment using shapes found in step 1 . . . 121
5.5.2.3 Step 3: Detailed analysis . . . 123
5.6 Supplementary material . . . 129

6 Myosin IX 151
6.1 Introduction . . . 153
6.2 Results . . . 159
6.2.1 SPA of myosin IX - actin - crosslinks . . . 159
6.2.1.1 Myosin IX crosslinks actin filament in highly ordered lattices . . . 159
6.2.1.2 Actin filaments crosslinked by myosin IX have parallel polarity . . . 162
6.2.1.3 Myosin IX adopts three different conformations . . . 170
6.2.2 Myosin IX - actin - crosslinks in the presence of calcium . . . 174
6.2.2.1 Calcium does not influence the order, periodicity and properties of the myosin - actin - crosslinks . . . 175
6.2.2.2 Calcium influences the polarity of actin filaments crosslinked by myosin IX . . . 177
6.2.2.3 Myosin IX conformations in the presence of calcium . 183
6.3 Discussion and Outlook . . . 189

6.3.1 Model: How a monomeric myosin self-organizes actin networks 189
6.3.2 Myosin IX - actin - crosslinks in the absence and presence of calcium . . . 193
6.3.3 Why is the structure of myosin IX just a blob? . . . 197
6.4 Methods . . . 201
6.4.1 Preparation and negative staining of myosin IX bound to F-actin 201
6.4.2 SPA workflow for EM images of myosin IX bound F-actin . . . 204
6.4.2.1 Acquisition of micrographs and particle picking . . . 204
6.4.2.2 General shape of actin - myosin - crosslinks . . . 208
6.4.2.3 Polarity and phase of actin-crosslinks . . . 212
6.4.2.4 Workflow for determination of polarity and phase of actin - myosin IX - crosslinks . . . 214
6.4.2.5 Interpretation of the actin - myosin IX - crosslinks . . . 226
6.5 Supplementary material . . . 239

III Conclusion and Outlook 251

IV Appendix 259

List of publications 281

List of figures 283

List of tables 291

Acknowledgments 293

List of abbreviations

3HB	three-Helix-bundle
aa	amino acid
ACEX	actin extraction buffer
ADP	Adinosine DiPhysophate
ap/ip	anti-parallel polarity / in phase
ap/op	anti-parallel polarity / out of phase
approx.	approximately
ATP	Adinosine TriPhosphate
CA	Correspondence Analysis
Ca^{2+}	Calcium
CaM	Calmodulin
CBS	Calmodulin Binding Site
ccc	cross correlation coefficient
CCD	Charge-coupled device
CCF	Cross Correlation Function
CTF	Cross Transfer Function
EM	Electron Microscope
EMAN	Electron Micrograph ANalysis
F-actin	Filamentous actin
FL	Full-Length
FRC	Fourier Ring Correlation
G-actin	Globular actin
GUI	Graphical User Interface
HAC	Hierarchical Ascendant Classification
MD	Motor Domain
P-value	Probability-value
PCA	Principal Component Analysis
PDB	Protein Data Base
pp/ip	parallel polarity / in phase

pp/op ..... parallel polarity / out of phase
SAH ..... single α-helix
SDS-PAGE ..... Sodium Dodecyl Sulfate Polyacrylamid Gel Electrophoresis
SPA ..... Single Particle Analysis
SPIDER ..... System for Processing Image Data from Electron microscopy and Related fields
SPIRE ..... SPIDER Reconstruction Engine
TEM ..... Transmission Electron Microscopy
UA ..... Uranyl Acetate
UF ..... Uranyl Formate

Part I

Introduction

Introduction

In the recent years, intense development and technical progress were able to push the resolution limit of electron microscopy towards the sub-nanometer range and the future will likely allow for the investigation of structures below one Angstrom. In particular, the transmission electron microscopy of samples at cryogenic temperatures (cryo-EM) is nowadays close to resolving biological structures at a resolution-range that was previously reserved for crystallography only. The ability to generate reconstructions of the molecule of interest, however, is not only based on the pure specifications of the electron microscope (EM) itself. To a at least same degree of importance is the preceding sample preparation and in particular the following image processing of acquired images (single particle analysis, SPA). Single micrographs recorded with the EM usually do not contain enough contrast and comprise a relatively low signal-to-noise ratio (SNR). Only the subsequent power of single particle analysis of several hundreds of images, which includes the iterative application of alignment, classification and averaging, is able to drastically increase the SNR and thus contrast, detect homo-and heterogeneities within the data set and reveal intriguing structural insights of the studied object.

Although cryo-EM is becoming the tool of choice these days due to its ability to resolve structures down to the atomic scale, the old fashioned negative stain method did not loose any of its power and magic. The high contrast of stained particles, the usually small number of preferred orientations of molecules on the EM grid and the relatively comfortable sample preparation make the negative stain EM still an excellent - and depending on the size of the studied particle sometimes the only - approach for the first insights into structurally unknown molecules. Likewise, newly discovered mechanisms and properties of well studied objects can be effectively reviewed on a structural level by negative stain EM and the following SPA. These results might enable higher resolution studies using for instance cryo-EM.

As mentioned, a successful structural examination using the electron microscope is heavily dependent on the subsequent image processing of the obtained data set. In this

context, and parallel to the development of the EM technology itself, many different and very clever approaches were made to establish image analysis protocols that allow for the alignment, classification (multivariate statistical analysis) and (3D) reconstruction of EM data sets, which can be, on the account of the vast growing computer industry, very large (indeed up to several hundred-thousands images). Today, with electron microscopes becoming more and more commercially available and the operation more user-friendly, the software developers start to concentrate on rather out-of-the-box applications. This might lead to results relatively quick and easy, but as James Hillier said[4]: "The electron microscope is like the monkey wrench on the garage wall: What you do with it is the important thing." In analogy to this quote from 1960, knowing the underlying processes of the EM as well as being able to precisely control the image processing algorithms can improve the outcome drastically. Furthermore, depending on the studied object, custom workflows are often inevitable.

The following work concentrates on the development of protocols for the single particle analysis of negatively stained electron micrographs in order to investigate the structural properties of full-length myosin VI in dependency on calcium and calmodulin and of the myosin IXa motor domain forming crosslinks with actin filaments.

[4]Dr. James Hillier was heavily involved in the development of the first practical electron microscope in the late 1940's, based on the first built EM by Ernst Ruska and Max Knoll in 1933.

1 Biological background

A large variety of cellular processes is based on the interaction of the cytoskeleton of the cell with proteins that are able to transform chemical energy, present in the form of ATP, into mechanical energy. Examples for such directed movements are muscle contraction, cell motility, mitosis, cytokinesis, endocytosis and vesicle and organelle delivery. The cytoskeleton, a skeleton within the cytoplasm, is present in all cells. It is constructed out of three main elements, microfilaments (composed of **actin monomers**), intermediate filaments and microtubuli, which stabilize the cell membrane and act among other functions as tracks for directed movement of organelles and other elements. This movement is based on the ability of ”motor proteins” to gain mechanical energy from ATP hydrolysis, which enables them to translocate processively along filament tracks (e.g. in order to transport cell organelles) or to form a functional, cooperative unit (e.g. myosin-filaments responsible for muscle contraction). Motor proteins are divided into various classes depending on which filaments they interact with (and also depending on their functional characteristics). While ”kinesins” and ”dyneins” bind to microtubuli, motor proteins that are assigned to actin filaments are called ”**myosin**”. [1, 2]
The following section will describe the structure and functionality of actin and their corresponding molecular motor **myosin** in more detail. Furthermore, the interaction between both are presented on a basic level:

1.1 Actin

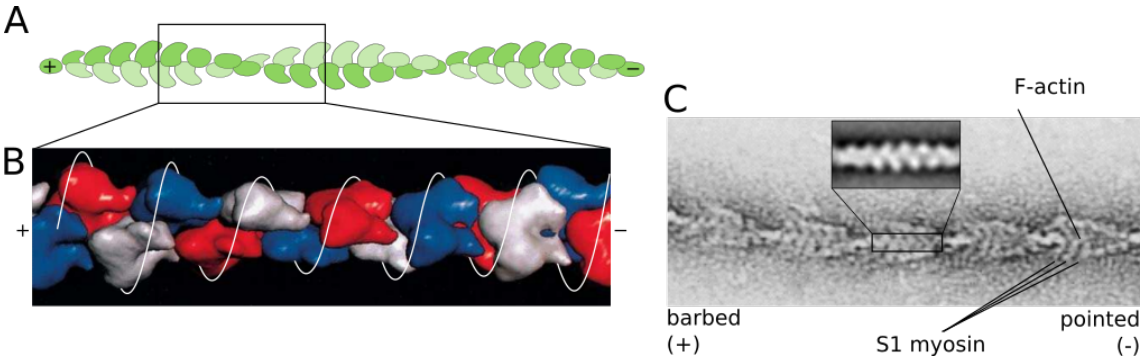


Figure 1.1: Structure of F-actin: **A** Simplified representation of an actin filament, indicating the two stranded helical appearance. **B** Model of F-actin that covers one crossover, showing a more realistic structure of the actin monomers including the ATP binding clefts (best seen on the right, gray monomer) and the assembling of the monomers with offsets of 166° per monomer. Also visible is the different appearances of each actin monomer depending on its orientation (rather big on the right side, whereas flat on the left side). Modified from [1]; **C** EM micrograph of F-actin decorated with myosin S1 fragments, leading to the notation for its polarity by "pointed" and "barbed" ends. The inset shows the underlying structure of bare F-actin.

Microfilaments are composed of globular subunits, called actin. Actin is a protein that is very abundant in every eukaryotic cell. More than 20% of cell protein is accounted for by actin. It is a single polypeptide chain with a molecular mass of 42 kDa. In its monomeric form actin is called "G-actin" (globular-actin) and has a pear-like shape with a size of approx. 4-5 nm [3], see Figure 1.1 **B**. Depending on the ionic strength and the presence of the divalent cations Mg^{2+} and Ca^{2+}, the actin monomers start to polymerize and build filaments, called "F-actin" (filamentous actin), in an helical manner as shown in Figure 1.1 **A**. The monomers of F-actin assemble with an 166° offset around the filament axis as indicated in Figure 1.1 **B**, which makes the filament appear like two strands twined around each other. In total, a filament repeats every 26 subunits, which corresponds to a distance of approximately 72 nm, termed "actin-repeat". This distance includes one crossover, which is why there is also the term "actin-pseudo-repeat", which is the distance from one to the next crossover, 36 nm. In its filamentous form actin has a diameter of approx. 8 nm [1, 4].

X-ray analysis of G-actin revealed [5], that the monomers are characterized by a cleft, which represents the ATP binding site, especially seen in the right, gray monomer in Figure 1.1 **B**. The polar structure of the monomers consequently leads to a polarity of the entire actin filament. The end, where the ATP binding site is exposed, is called the minus (-) end, while the other side is called plus (+) end. This notation originates from the different binding constants of monomers to the two ends of the filament, which results in a slow and a fast growing end. First structural details of F-actin could be revealed by

electron microscopy (EM) of heavy decorated filaments with myosin S1 fragments [6]. As shown in Figure 1.1 **C**, the myosins bind in a polar way to actin, resulting in ”pointed” and ”barbed” ends, which correspond to the (-) and (+) ends respectively [7].
The inset in Figure 1.1 **C** shows the underlying F-actin structure without molecules bound to it. Apart from an ”arrowed” structure of the monomers, another polar feature is recognizable. As seen in Figure 1.1 **B**, the actin subunits are not spherical, but rather elliptical. This has the effect, that in a projection view the monomers appear relatively large towards the minus end, while towards the plus end they appear flat. In the EM, where images are in fact projection views of the objects' electron densities, the monomers of an actin filament appear as distinct ”intensity-blobs” on the (-) side compared to a rather continuous appearance on the (+) side, as seen in Figure 1.1 **C**. This feature will be particularly important when the polarity of actin filaments is studied by EM.

1.1.1 Preparation of F-actin

The following part will describe the preparation of filamentous-actin, which was divided into two steps. The first step included the extraction of G-actin from rabbit skeletal muscles, followed by the second step, in which phalloidin was used to polymerize G-actin to obtain purified F-actin (protocols adapted from [8–10]).

Step 1 - G-Actin preparation

For the purpose of this work (see chapter 6), rabbit skeletal muscle actin was provided by the lab in the form of dried acetone powder. To extract purified G-actin, the preparation comprised steps of depolymerization, extraction, polymerization, cleaning up, removing of tropomyosin and a final depolymerization. The powder was first mixed, centrifuged and filtered in a solution of 2 mM TrisHCL (pH 8.0), 0.2 mM $CaCl_2$, 1 mM DTT and 0.2 mM Na_2ATP. This led to an extracted, depolymerized solution of actin. A following polymerization and purification was obtained by mixing the solution with 5 mM TrisHCL (pH 8.0), 50 mM KCl, 2 mM $MgCl_2$ and 1 mM ATP. At this point a small test sample was taken and later analyzed by gel electrophoresis in order to control the concentration and the degree of purity. The extracted actin contained tropomyosin, a protein bound to the filaments which regulates their function. For the EM analysis in our studies, tropomyosin was not required and therefore removed. This was done by adding KCl until a concentration of 800 mM and a pH between 8.3 and 8.5 was reached. Further centrifugation, mixing with actin extraction buffer (ACEX) and homogenizing the prep finished this step. Here a second test sample was taken for later control. Final depolymerization was

obtained by dialysis in ACEX. A third test sample was taken and the concentration of the final, purified G-actin solution was measured using spectrophotometry (see section 1.3). For storage and later usage, the solution was aliquoted and stored at -80°C by shock freezing in liquid nitrogen.

- **Concentration measurement**: To measure the concentration with the spectrophotometer, the solution was 5-fold and 10-fold diluted. After measuring a blank sample for the background, Fig. 1.2 shows the spectrum of the two samples.

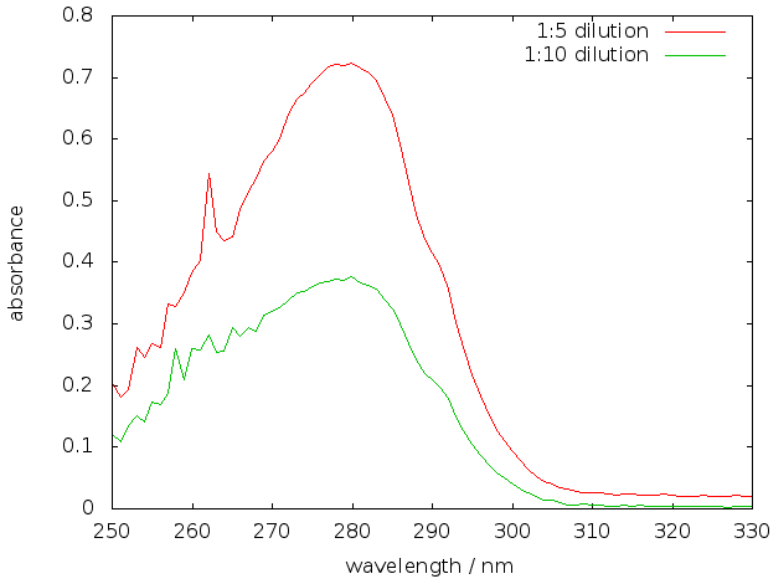


Figure 1.2: Spectrum of the absorbance measurement of G-actin in two different dilutions. The absorbance at 260 nm and 280 nm are used for calculating the protein concentration of the samples.

The absorbance values at 280 nm and 260 nm were used to calculate the protein concentration of the samples. The following table shows the absorbance A at the two wavelengths and the calculated concentration according to equation 1.1 in section 1.3. The molecular weight of G-actin is 42 kDa.

By taking the average of both results the final concentration of the G-actin used for following experiments in this work was **0.095 mM or 95 µM**.

- **Analyzing test samples**: The three samples taken during the G-actin preparation were analyzed by gel electrophoresis (see section 1.3). Figure 1.3a shows the 10%

Dilution	$A_{280\ nm}$	$A_{260\ nm}$	c in mg/ml	c in µmol/ml
1:5	0.723	0.384	4.14	0.099
1:10	0.376	0.260	3.85	0.092

Table 1.1: Concentration measurement with the spectrophotometer and the Warburg-Christian method [11].

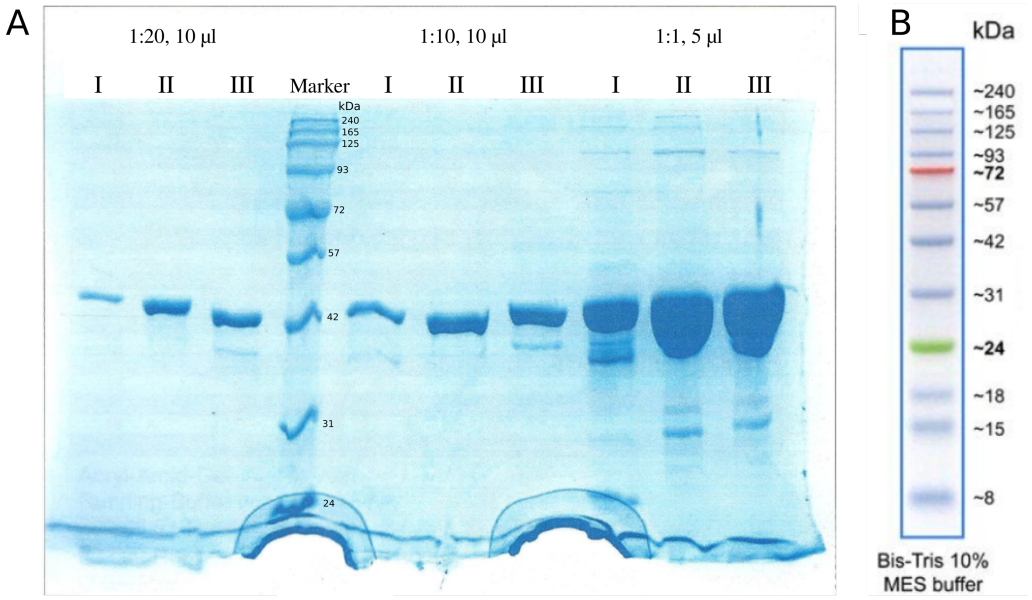


Figure 1.3: **A** SDS PAGE gel of three samples taken during the G-actin preparation process. The 42 kDa G-actin band is dominant, indicating a successful purification. **B** Protein marker to identify corresponding molecular masses. From [12]

acrylamid-gel of the SDS-PAGE for the three samples (I, II and III), each in three different dilutions, 1:10, 1:5 and 1:1. The fourth lane shows the control sample with corresponding molecular masses are shown in Fig. 1.3b. Apart from other shallow bands, the 42 kDa band, representing G-actin, is clearly visible. The thickness of these bands increased during the preparation process, indicating an increase of the concentration. The fact that no other band appears notably indicates a high degree of purification of the G-actin solution.

Step 2 - F-actin preparation

To obtain filamentous actin, the G-actin solution was first diluted to 45 µM by mixing it with 0.2 mM EGTA, 50 µM $MgCl_2$ and TEM-buffer (25 mM KCl, 50 mM Tris pH 7.5, 100 µM $MgCl_2$, 100 µM DTT) and subsequently dialyzed in the TEM-buffer. After the dialysis, 66µM phalloidin was added and the solution allowed to rest at room temperature. Phalloidin is one example of a group of toxins of the death cap mushroom that enhances and stabilizes actin polymerization, because it has a higher affinity to actin filaments than to actin monomers [13]. It has been reported [10] that a 1:1 concentration between G-actin and phalloidin is sufficient to prevent the formed F-actin to depolymerize. Successful polymerization can be observed by an increasing viscosity of the solution. Similar to the G-actin stock the **45 µM** F-actin solution was aliquoted and shock frozen to -80°C.

1.2 Myosin

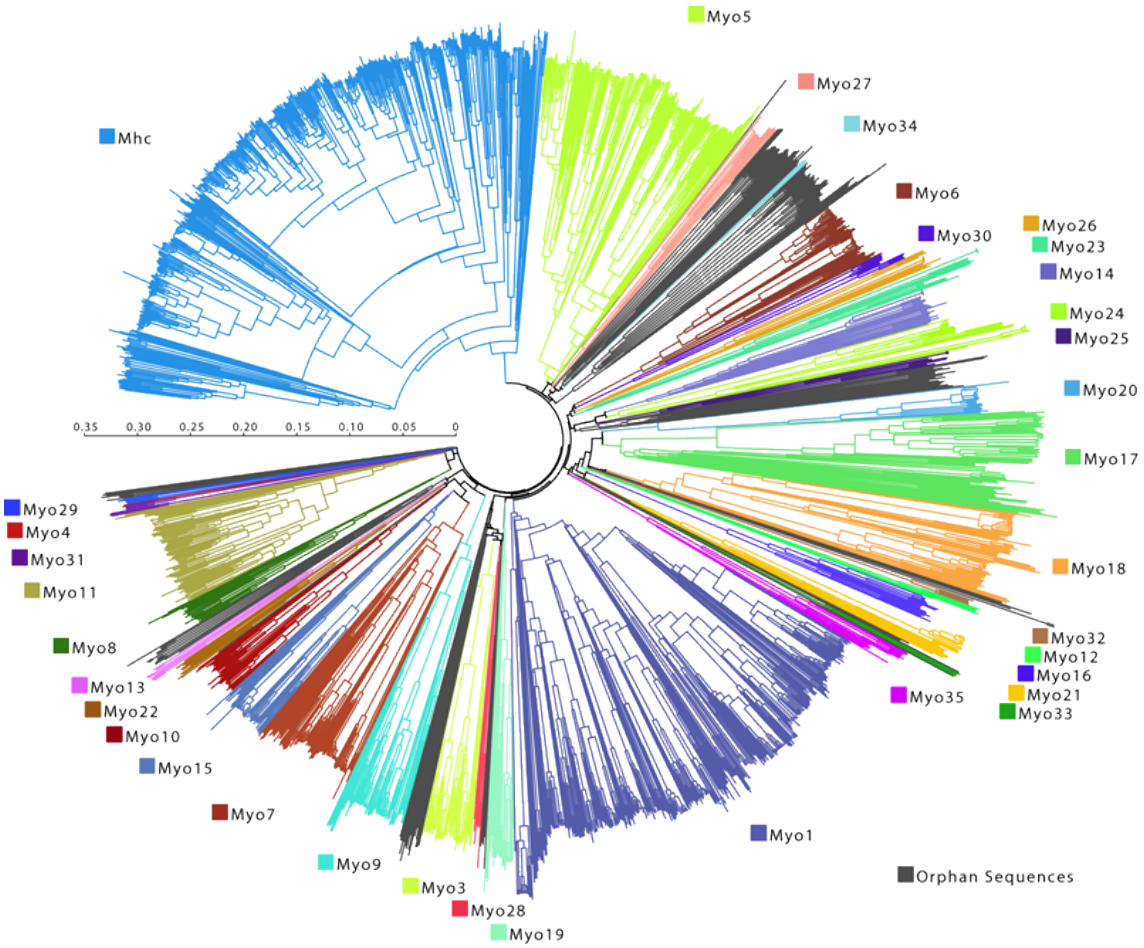


Figure 1.4: Phylogenetic tree indicating the wide range of different myosin classes. From [14]

As mentioned in the introduction, the filaments in the cytoskeleton (microfilaments, intermediate filaments and microtubuli) act, among other functions, as tracks for movements of associated motor proteins. Actin filaments, described in the previous section, are associated with myosins, a super family of molecular motors. Myosins are actin activated ATPase motor proteins that are responsible for many functions in the cell, such as muscle contraction, membrane anchorage, longer range transport of cargo and cell signaling. The generation of force and movement is established by transforming chemical energy, gained by hydrolysis of ATP, into mechanical work.

The first form of myosin was found in skeletal muscles of mammals where it is responsible for muscle contraction. During further research, various different myosin forms were found, being responsible for several different functions in the cell. Depending on their functional characteristics and their structural features, they can be grouped into different classes. Today a wide variety of 47 different classes is known [14], from

which only few are well studied (Figure 1.4). The best studied classes are myosin class I, II and V, being responsible for muscle contraction and cytokinesis (class II) and for the interaction between the cytoskeleton and the cell membrane and vesicle transport (class I and V).

Due to their functional properties and specialization in various roles they are divided into 47 different classes [14]. In order to perform the wide range of necessary tasks within the cell, usually a cooperation of many different myosins is needed. Eukaryotic cells, for example, express at least 11 different classes [15].

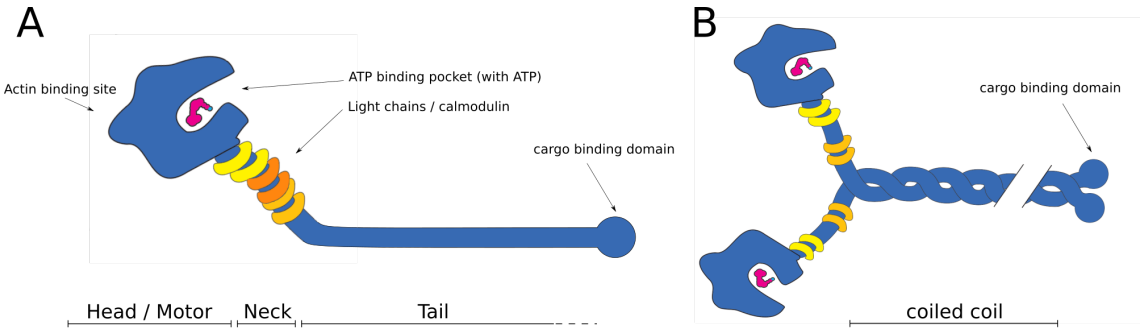


Figure 1.5: Structure of myosin molecules, which can be divided into a head / motor, a neck and tail domain. The head domain contains the actin binding site and the ATP binding pocket. Light chains of the calmodulin family, which have a mass of approx. 20 kDa, bind to the neck domain and regulate the motor activity. **B** Under certain conditions, some myosins are able to form dimers with the two tails forming a coiled coil. At the end of the tail a cargo binding domain may be present. Modified from [1]

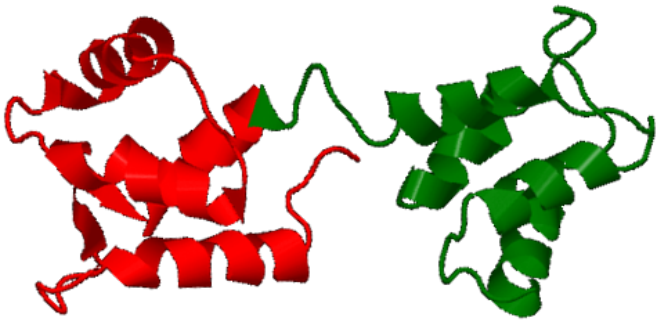


Figure 1.6: Crystal structure of calcium-free calmodulin showing its two lobes connected by a linker helix. PDB: 1CFD [16]

Although there is a wide structural and functional variety of different myosins, they share a basic, common structure, as demonstrated in Figure 1.5:
All myosins are composed of three main regions: The **head domain** or catalytic domain contains the binding site for actin and for ATP. This domain is responsible for the generation of motion, giving it the term *motor domain*. Following the head domain, the **neck**

region, or lever arm, acts as the binding site for several light chains and calmodulin (CaM), which regulate the activity of the molecule. Calmodulin is a 17 kDa protein that has the structure of two globular domains, a N-lobe and a C-lobe. Both lobes contain two calcium-binding sites each and are connected by a single helix that is flexible to some degree [17]. The crystal structure of calmodulin in Figure 1.6 (PDB: 1CFD [16]) shows the asymmetric shape of this protein. Upon binding of calcium, calmodulin undergoes structural changes which can affect its binding behavior. Apart from myosin, calmodulin is associated with a number of other target proteins including protein kinases, phosphatases, and ion channels.
The lever arm of myosin plays a key-role in transducing and amplifying the small conformational changes generated by ATP hydrolysis in the catalytic domain into nanometer displacements at the end of the **tail domain**. The tail domain contains further binding sites that regulate and determine the functional characteristics. Some classes contain another domain at the end of the tail is able binds cargo. Depending on the structure of the neck and tail domain, some myosin monomers are able to form dimers by a coiled coil structure, e.g. myosin II or myosin V, as illustrated in Figure 1.5 **B**.
The light chains within the neck region regulate the activity of the motor domain and also the formation of dimers by binding calcium ions. The number of bound light chains and their characteristic interactions with calcium can vary among different classes. Myosin class II for example contains two different light chains, called essential and regulatory light chain. Myosin I and V on the other hand contain calmodulin as light chains [1]. Another example of how calcium can regulate the molecular motor is myosin VI and will be presented later in this work (see chapter 5).
The motor domain of myosins is a specialized ATPase, whose activity is in combination with the presence of ATP activated by the binding to F-actin. In the absence of ATP, the myosin is in the *nucleotide-free / apo* state and its motor domain is strongly bound to the actin filament, a conformation called *rigor* conformation. As soon as ATP is present, a chemomechanical ”powerstroke cycle” is activated, which involves the binding of ATP to the motor domain, dissociation of the myosin from the filament, hydrolysis of ATP to ADP and P (*ADP state*) resulting in conformational changes of the myosin and the reattachment to the actin filament [18]. These conformational changes result in an angular shift of the lever arm (called ”power stroke”), which allows the myosin molecule to move filaments or, in the case of dimeric myosins, walk along filaments. These movements are usually directed from the minus to the plus end, only one class is known to work in the reversed direction, myosin VI.

From all three regions described above the motor domain is the highest conserved one. All myosin classes share the basic structure of the head domain, by which molecules

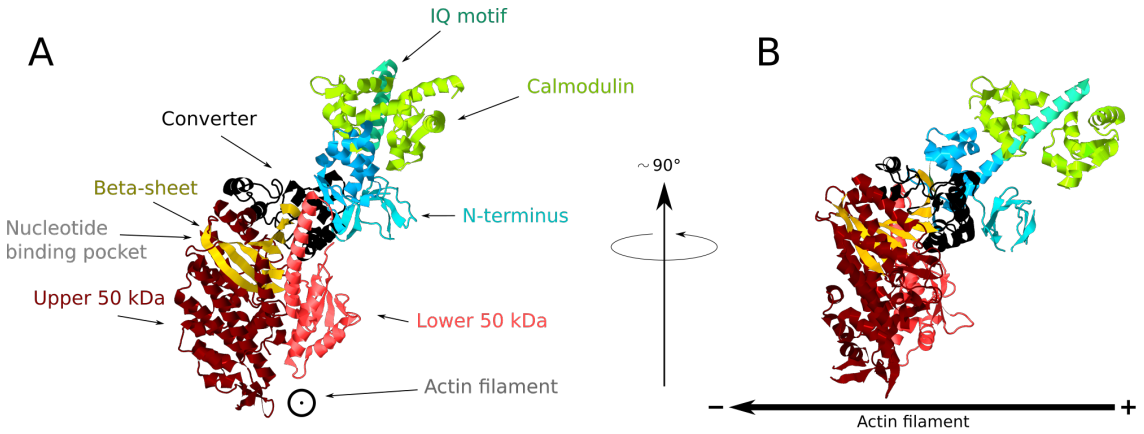


Figure 1.7: Structural composition of the myosin motor domain and part of the neck on the example of myosin V (PDB entry: 1OE9). The main elements are the upper and lower 50 kDa domains, which form the actin binding site, the nucleotide binding pocket, a beta-sheet, which connects the to 50 kDa domains, the extending N-terminus and a converter, which connects the motor domain with the protruding lever arm. Image **A** represents the usual orientation of free myosin as imaged by the EM, whereas myosin molecules bound to actin appear in a 90° rotated orientation.

are identified as myosins, while variations in the neck and tail region lead to functional variability among the classes. Figure 1.7 shows the structure of the myosin V motor in a nucleotide-free state with two calmodulin bound at the neck region in ribbon representation, PDB entry 1OE9. The motor domain can be divided into several subunits:
The **actin binding site** (the direction of an actin filament would be in this case perpendicular to the image plane as indicated) is composed of two subunits called the upper and lower 50 kDa domain, which are connected by a seven-stranded β-sheet. The beta-sheet also connects several other elements in the motor domain and is believed to be an important factor for the communication between the various functional sites of the molecular motor [19]. Apart from the cleft between the two 50 kDa domains, another cleft can be seen between the lower 50 kDa domain and the N-terminus. Close to the upper 50 kDa domain is the **nucleotide binding pocket**. Protruding from the motor domain is the **neck region (lever arm)**, formed by one or several IQ motifs[1] and bound light chains (here calmodulin). This region is linked with the motor domain by a subdomain called **converter**, allowing the lever arm to bend relative to the motor.
The appearance of myosin molecules imaged by negative stain electron microscopy is usually biased towards a few orientations. In the case of free myosin molecules, the preferred binding orientation to the surface is similar to the view in Figure 1.7 **A**. Various elements such as the two clefts between the lower, upper 50 kDa and N-terminus domains and the protruding lever arm with different calmodulins bound can be seen in the

[1]The notation is based on the amino acid sequence, which starts with the two amino acids isoleucine I and glutamine Q

EM. In the case of myosin bound to actin, the observed orientation is rotated by approx. 90° as indicated in Figure 1.7 **B**. Images of myosin bound to actin show usually a triangular shape of the motor domain, which is conserved among different myosin classes and different nucleotide states. Structural variability are mostly seen in the conformations of the lever arms and tails.

1.3 Methods

Spectrophotometry

To determine the concentration of samples, a spectrophotometer can be used, which is a device that is able to measure the amount of light that is absorbed by the proteins. This amount is dependent on the concentration and on the wavelength of the used light. The measurement results in a spectrum showing the amount of absorbed light at different wavelengths. A reference solution without protein is also measured for background correction [20].
To calculate the concentration of the protein from the absorbance spectrum one makes use of the fact that the amino acids tyrosine and tryptophan, contained in most proteins, absorb light at 280 nm. However, nucleic acids also absorb strongly at 280 nm. The Warburg-Christian method corrects for nucleic acid contamination by taking advantage of a second absorbance band of nucleic acids at 260 nm, at which wavelength proteins do not absorb light. The corrected concentration c is calculated as follows [11]:

$$c = (1.55 \cdot A_{280} - 0.76 \cdot A_{260}) \cdot d, \tag{1.1}$$

where A_{280} and A_{260} are the absorbance at 280 nm and 260 nm respectively and d is the dilution factor.
The concentration has the unit mg/ml and, divided by the molecular mass, mmol/ml or M (molar concentration).

SDS-Page

One way to analyze proteins is the PolyAcrylamid Gel Electrophoresis, short PAGE, which is an analytic method to separate mixtures of biological macromolecules, usually proteins or nucleic acids, in an electric field according to their molecular masses. In contrast to other gel electrophoresis methods, the charge of each molecule should not affect the resulting separation. Therefore sodium dodecyl sulfate (SDS) is added to the sample, which is an anionic tenside, to impart an even negative charge distribution per

unit mass of the proteins. Together with a denaturation the proteins will linearize, that allows for a fractionation by chainlength of the linearized proteins, proportional to their molecular mass. When an electric voltage is applied to the polyacylamid gel, which acts as a filter, the proteins will run trough the gel with different velocities according to their size. After a specific time, the proteins will be sorted by size and form bands that can be visualized by staining procedures, such as Comassie-Staining. To correctly interpret the bands usually a control-sample with proteins of known size is simultaneously analyzed. [21]

2 Electron microscopy

2.1 History and principle

The invention of the light microscope was a huge technological improvement that allowed for the observation of objects that are too small to be seen by eye, such as single cells. However, although intense improvements of the technique in the 19th century allowed for resolving even smaller objects, the resolution of the light microscope was naturally limited to approx. 250 nm due the wavelength of light. The desire to resolve e.g the detailed, underlying architecture of the cell with its numerous organelles down to single molecules led to the idea of using electrons instead of light. Electrons have a significant shorter wavelength and thus would make it possible to push the resolution further. With the development of electromagnetic lenses that were able to focus an electron beam in the same way light could be focused with optical lenses, the first microscope based on electrons was built in 1933 by Ernst Ruska and Max Knoll (Figure 2.1) [22, 23] and further developed for practical application by Dr. James Hiller in the late 1940's (Figure 2.2). The electron microscope and its constant improvements made the study and visualization of small structures in the nanometer range possible and today's biology would not be imaginable anymore without it [24].

During the last century, various techniques for high resolution microscopy in the sub-nanometer range have been invented that overcome the resolution of the EM, for example X-ray crystallography or scanning probe microscopy. However, electron microscopy is still a very important technique, since it spans the resolution gap between X-ray crystallography and light microscopy, allowing for the investigation of whole cells down to single proteins [24].

As mentioned, the electron microscope is based on the same principles as the light microscope but uses electrons instead of light. According to the wave-particle duality

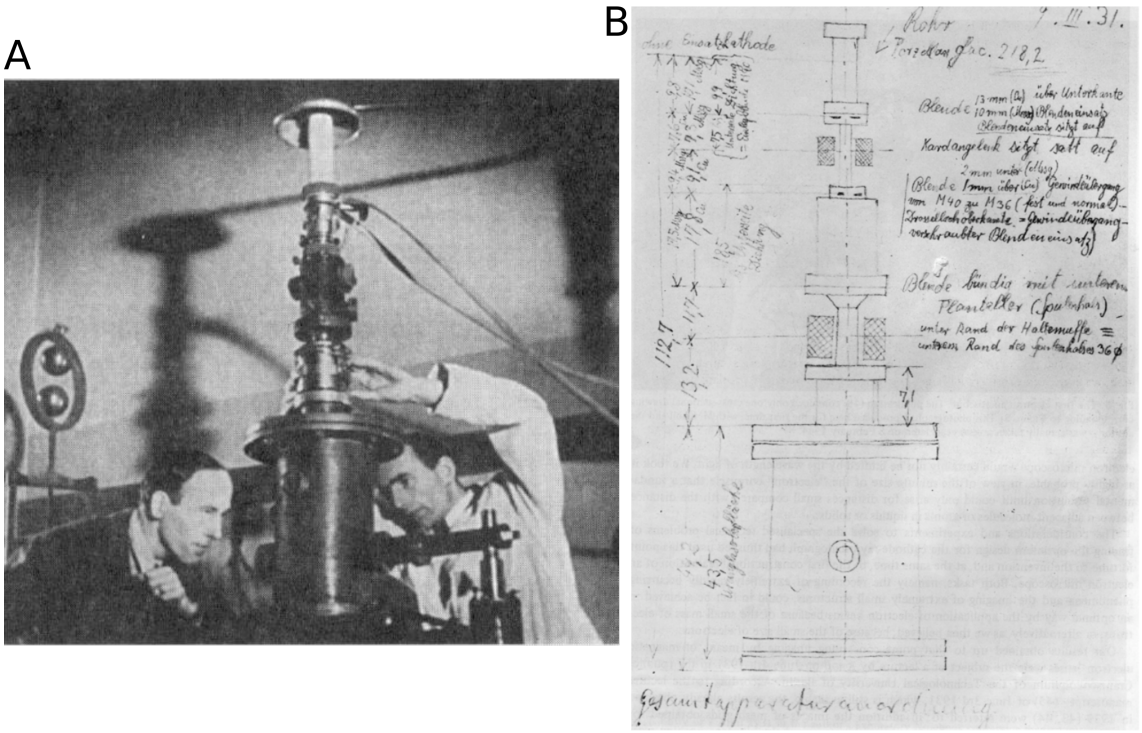


Figure 2.1: **A** First electron based microscope built by Ruska and Knoll around 1933. **B** Drawings of this first microscope. [25]

Figure 2.2: Dr. James Hillier, inventor of the modern TEM, at the controls of an RCA-EMU-series microscope in late 1940's [26]

electrons can be described as waves characterized by their de Broglie wavelength. Electrons used in EM usually have energies between 80 keV and 300 keV, which correspond to wavelength between 0.042 and 0.020 Å. The resolution of a microscope is traditionally defined by the minimum distance between two points that can be still distinguished. Due to diffraction a point of light will be imaged as a blurry disc, the extension of which depends among other things on the wavelength of the radiation used for illumination. Therefore there is a natural resolution limit described by the Abbe criterion: [27]

$$d = 0.61 \cdot \frac{\lambda}{NA}, \tag{2.1}$$

where d is the resolution, λ the wavelength and NA the numerical aperture. The NA is a dimensionless number which takes into account that the optical system can accept or emit light only within a specific range of angles. For optical microscopy NA is approximately 1 and it follows that for light with λ = 500 nm a resolution of 305 nm can be achieved. The NA for electron microscopes in contrast is approximately 0.01. The wavelength of the electrons depends on their energy, defined by the accelerating voltage V.

$$\lambda = \sqrt{\frac{h^2}{2 \cdot m \cdot e \cdot V}}, \tag{2.2}$$

where h is the Planck's constant and m the electron mass. It follows that electrons for an accelerating voltage of 100 keV have a wavelength of 0.0037 nm and a theoretical resolution of 2.3 Å could be achieved and with higher voltages such as 300kV the resolution can be pushed towards the 1 Å-range [27].
One kind of electron microscopy - which was used in this project - is the **transmission electron microscope**, short TEM. The principle of image formation in the TEM is based on collecting electrons that are transmitted through a very thin specimen. Depending on the atoms that are hit by the electrons, the interaction between the electrons and the specimen differs and the electron beam is either transmitted unscattered or scattered. Scattering occurs either without loss of energy (elastic scattering) or with an energy transfer from the electrons of the beam to electrons of the specimen (inelastic scattering), which usually leads to radiation damage, especially in biological samples. It is the way how these electrons interact with the specimen that gives rise to the image formation. Transmitted electrons are collected and focused by a large optical setup, which will be discussed in the following.

2.2 Setup of a Transmission Electron Microscope

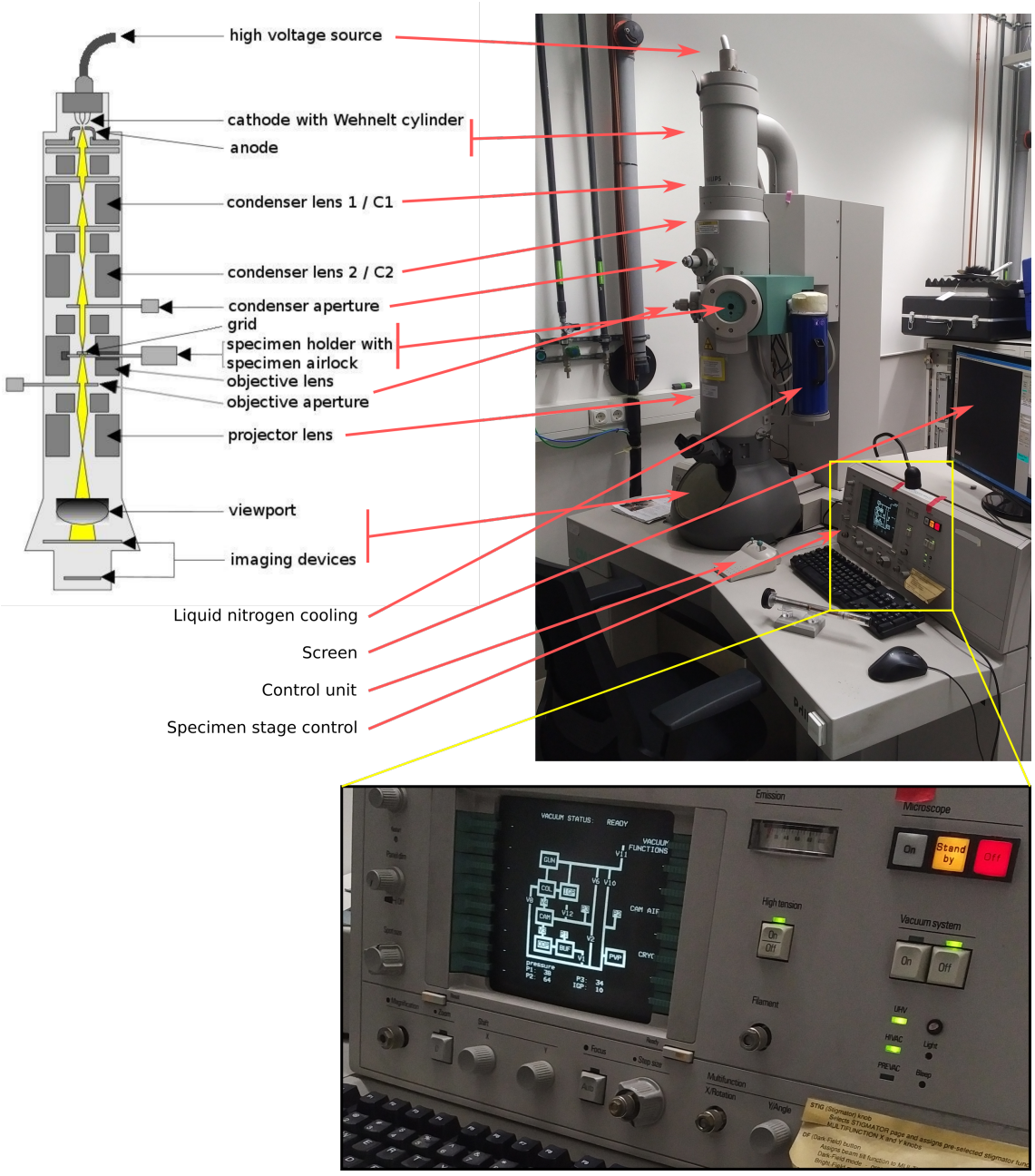


Figure 2.3: Scheme and real setup of a transmission electron microscope (Philips CM 100).

Figure 2.3 shows the schematic and real setup of the TEM used in this work (Philips CM 100). The main components are an electron gun connected to a high voltage source, followed by a system of condenser lenses that focus the electron beam on the grid, which contains the sample and is kept in place by a special holder. Objective and projector lenses magnify the transmitted beam and focus it onto image devices which can be either a fluorescent screen, photographic film or a camera (for example CCD or direct electron

detection). In order to minimize the scattering of electrons by molecules of the air, the electron microscope operates at high vacuum constantly. Most of the lenses described here are kept aligned by various shift and tilt coils. In order for maintaining the high vacuum and functions of the lenses, the EM is constantly cooled using liquid nitrogen. The magnified image shows the control unit to keep track of the various parameters (mainly pressure and valves) and to adjust important settings such as magnification, focus and spot size. The actual image acquisition is done using special softwares.

2.2.1 Electron gun

The electron beam necessary for imaging is in most cases created by thermionic emission of electrons from a filament, usually a hairpin like tungsten filament. While the whole microscope is kept at ground potential, the filament is connected to a high negative voltage, therefore called cathode. Given enough current, electrons are emitted from the filament and accelerated by the applied voltage.
In order to focus and control the electron beam, the filament is surrounded by an electrode called "Wehnelt cylinder", that has a central bore of about 1 mm at the bottom and is kept on a slightly more negative potential than the filament. The specific design of this cylinder focuses and restricts the electron beam onto the optical axis. The restriction is important in order to achieve high coherence and thus high resolution. The electrons that emerge the opening of the cylinder form the actual electron source of the TEM [24].

2.2.2 Condenser lenses and their effect on resolution

The condenser system consists of two condenser lenses, C1 and C2 and apertures of C2. These lenses control the way the specimen is illuminated by the electron beam. As illustrated in Figure 2.4, electrons coming from the filament are focused by the first lens C1 which is situated immediately after the Wehnelt cylinder. The black dots in this figure indicate the focal points of each lens. C1 is a very strong lens and creates a demagnified image of the source in space. The strength of the first condenser lens controls the size of this virtual source, called "spot size". Choosing the right "spot size" is important for high resolution electron microscopy.
The second condenser lens focuses the new virtual source onto the specimen. As demonstrated in Figure 2.5 **A**-**C**, the strength of C2 controls the illuminated area of the specimen and the convergence of the electron beam, which is also important for the resolution. In Figure 2.5 **A** the virtual electron source is focused directly onto the specimen. The angular range from which the electrons appear to come from (convergence), is at its maximum. The intensity of the illuminated area has also its maximum in this setting.

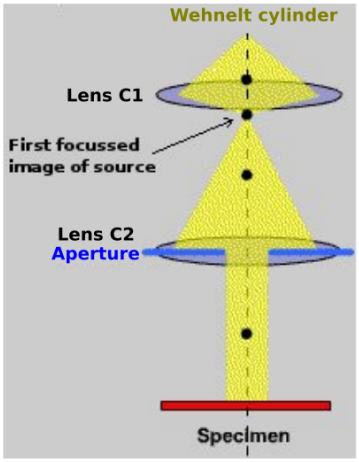


Figure 2.4: Simplified ray diagram illustrating the focal lengths of the two condenser lenses C1 and C2. The first lens C1, situated immediately after the Wehnelt cylinder, creates a first demagnified image of the source, which is subsequently focused onto the specimen plane by the second condenser lens C2. The strength of C1 determines the size of the virtual source ("source size"), which has an influence on the brightness and resolution of the final image. Modified from [28].

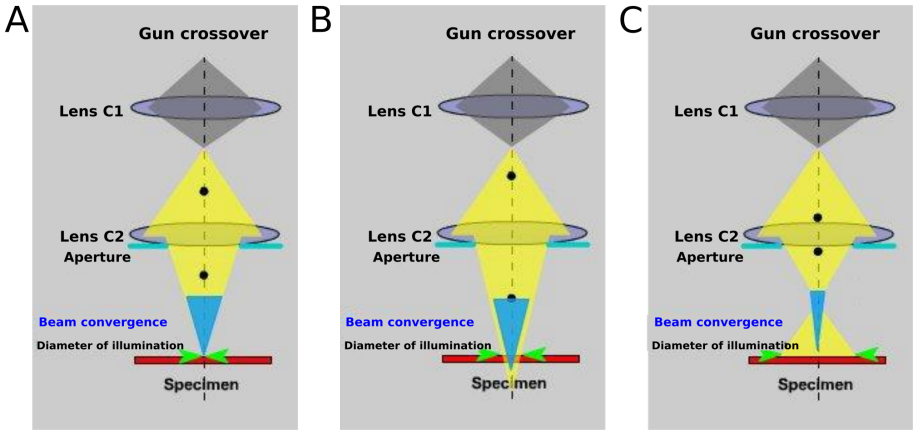


Figure 2.5: Effect of the second condenser lens C2. Depending on its strength, the virtual source is directly focused onto the specimen **A** or either under- or overfocussed, **B** and **C**. The change from focus to underfocus by weakening C2 increases the illuminated area and decreases the convergence only slightly. By strengthening C2, however, the illuminated area as well as the beam convergence are increased significantly, which improves the resolution. The lens is said to be overfocussed in this case. Modified from [28].

By weakening the strength of C2 (Figure 2.5 **B**), the focus of the source moves away from the specimen plane, the illuminated area increases and the convergence decreases slightly. In this case, C2 is said to be underfocussed. A bigger difference is seen in Figure 2.5 **C**, when C2 is overfocussed. This is done by increasing the current in the lens. The focus of the electron source moves up and the electrons appear to come from this point, which decreases the angular range substantially.
The overfocus setting of C2 in combination with a very small spot size (obtained by a very strong first condenser lens) provides the highest coherence and thus the highest resolution. A disadvantage of this setting is the decrease of intensity of illumination and it might become difficult to operate under these conditions properly. However, the effect of the spot size on the resolution is negligible when negatively stained samples are investigated, where the resolution is more limited by the stain [24, 28, 29].

2.2.3 Specimen stage

To provide an easy exchange of the specimen while maintaining the vacuum the specimen holder (shown in Figure 2.6) is inserted trough an airlock. The holder is specially designed in order to minimize vibrations that would decrease the resolution of the microscope. Furthermore it allows for shifts and rotations of the specimen, for example for the usage of tilt series. Some holders can be loaded with more than one grid at a time. Standard grids on which the samples are placed are discs of 3.05 mm in diameter, a few 100 µm thick and made out of a copper mesh with optional coatings, as shown in Figure 2.6 **C**.

2.2.4 Objective and projector lenses

The most important lens in the system of the TEM is the objective lens, which does the first step of image forming and -magnification. It has a very small focal length of only a few millimeters, which makes it necessary for the specimen to be positioned almost within the lens. The objective lens creates a diffraction pattern of the transmitted electrons on the back focal plane, where an objective aperture is usually inserted. The aperture controls the portion of scattered electrons that contribute to the image and therefore its size affects the obtained contrast. A small aperture screens out widely scattered electrons and the contrast is mainly achieved by the amplitude of the transmitted electron beam. On the other hand, a very large objective aperture allows diffracted electrons to participate in forming the image. Scattered and unscattered electrons undergo different phase shifts as they pass through the sample and interfere with each other. The information of this phase shifts in addition to the amplitude contrast are used to recon-

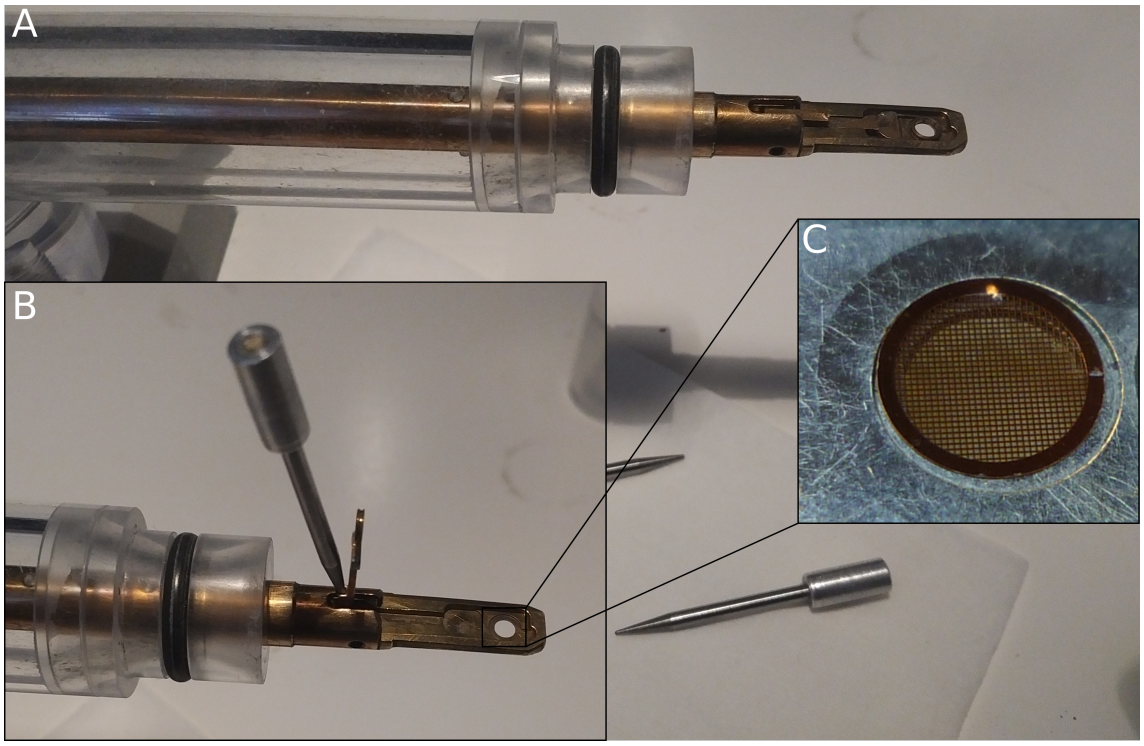


Figure 2.6: Single grid holder for the TEM used in this work. **A** The actual gridholder is fixed in a stage, ready to be loaded with an EM grid. **B** To place the grid, the flap is opened using a small pin. **C** Magnified image of an inserted grid visualizes the copper mesh of the grid.

struct high resolution images. However, phase shifts are also added to the electrons on their way from the sample to the imaging device by objective aberrations, which makes it difficult to interpret the image correctly. Therefore the minimization of aberrations such as spherical aberration, chromatic aberration and axial astigmatism (see Figure 2.7) is very important [24]:

- **Spherical aberration B** describes the way inciding beams are focused depending on their distance from the optical axis. Rays further away are refracted more strongly than rays closer to the optical axis. A point imaged by a lenses with spherical aberration will appear as a blurred disc.
- **Chromatic aberration C** affects the way electrons with different energies are focused by the lens. In a lens with large chromatic aberration, electrons with slightly different energies will be focused at significant different points. It is therefore important to have a well stabilized electron gun, that emits electrons only within a small range of energies.
- A lens with **axial astigmatism D** is not perfectly axial symmetric. Circles imaged with such a lens will be elliptic and the way objects are imaged would depend on their orientation to the plane of the specimen.

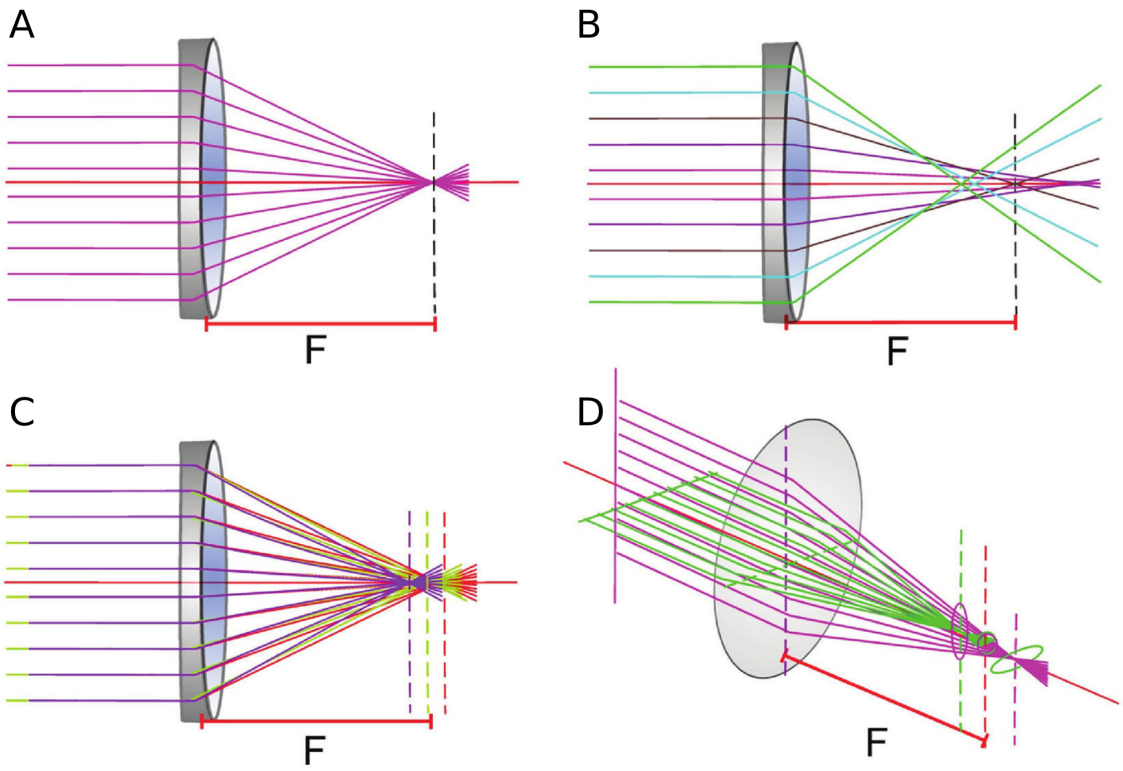


Figure 2.7: Origins of different lens aberrations. **A** An ideal lens, which focuses all parallel rays in one point. **B** A lens with spherical aberration will focus beams differently depending on their distance from the optical axis. **C** Chromatic aberration describes the effect of differently bended beams by the lens depending on their energy (equivalent to color). Only beams of electrons with the same energy are focused in the same point. **D** Lenses with axial astigmatism are not perfectly axial symmetric and circles will be imaged as ellipses. From [30]

2.2.5 Projector lens and imaging devices

Between the objective lens and the projector lens, several intermediate lenses are positioned. The strength of these lenses in combination with the projector lens define the final magnification, which can range from several hundreds up to several hundred thousands times. These lenses also control the imaging mode, since they can either focus the actual image or the diffraction pattern formed by the objective lens. The projector lens magnifies and focuses the final image onto one of the imaging devices:
A fluorescent screen is used for direct viewing by converting impinging electrons into visible light. When this screen is taken out of the electron beam, the image can be recorded either with a digital or an analogous device, such as a developable photographic film. A digital device is for example the standard charge-coupled device (CCD) and recent developments have made it possible to use direct electron cameras. While the field of view of a photographic film is larger than of a CCD, the digital recording has the advantages of immediate recording and viewing, a wider dynamic range, and the possibility to record incomparable larger amount of data in shorter time. [24]

2.3 Principles of image formation in the TEM

As already mentioned, electrons interact with the specimen in different ways. Electrons may be transmitted unscattered or scattered inelastically or elastically. Inelastically scattered electrons involve transfer of energy and have a narrow angular distribution. These electrons do not transmit any structural information but are the source of background noise in the image and also radiation damage to the specimen. However, due to the narrow angular distribution the resolution of this background is very low and thus usually tolerated. Information about the image are mainly obtained from elastically scattered electrons by two effects:
First, the amplitude of the transmitted electron beam is affected by the loss of electrons either due to electrons that are scattered outside the aperture or due to inelastic scattering. This contrast is therefore called *amplitude contrast*.
And second, contrast arises from shifts in the relative phases of the electron beam as it is transmitted through the specimen. Scattered electrons can be described by diffraction of electron waves. The interference of these diffracted waves produces intensity differences in the image, which leads to *phase contrast*. For thin specimens, this interference is directly related to the density variations in the sample and becomes dominant over the amplitude contrast. For thin samples, biological samples, the phase shift is relatively small due to weak scattering and the image formation is described by the *weak-phase-approximation*. From this theory the **(phase) contrast transfer function**, short CTF,

can be derived, which modulates the amplitude and phases of the electron diffraction pattern and thus describes the imaging characteristics.

2.3.1 Contrast Transfer Function CTF

Mathematically the CTF is defined as follows [24, 31]:

$$CTF\left(\mathbf{k}\right) = sin\left(\gamma\left(\mathbf{k}\right)\right) \tag{2.3}$$

with

$$\gamma\left(\mathbf{k}\right) = 2\pi\chi\left(\mathbf{k}\right) \tag{2.4}$$

and $\chi\left(\mathbf{k}\right)$ the *wave aberration function*

$$\chi\left(k,\phi\right) = -\frac{1}{2}\lambda\left[\Delta z + z_a sin2\left(\phi - \phi_0\right)\right]k^2 + \frac{1}{4}\lambda^3 C_s k^4. \tag{2.5}$$

Here $\boldsymbol{k}$ is the spatial frequency, λ the electron wavelength, Δz the defocus of the objective lens, z_a and $sin2\left(\phi - \phi_0\right)$ refer to axial astigmatism and C_s represents the quality of the objective lens in terms of spherical aberration.

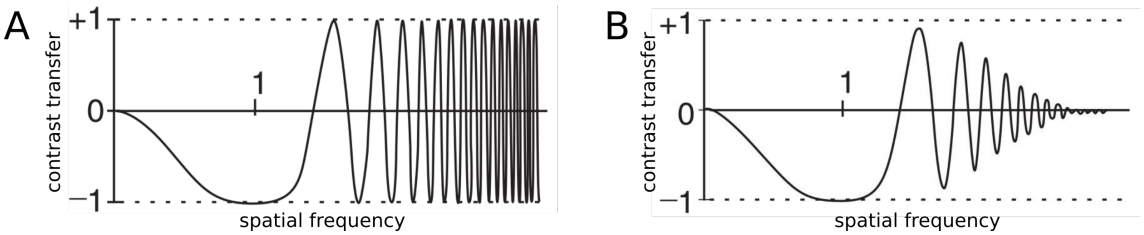


Figure 2.8: Undamped **A** and damped **B** contrast transfer functions (CTF). The contrast transfer is oscillating between negative and positive values with increasing spatial frequency. At frequencies where the CTF crosses the zero, no information is transferred, while negative values represent negative contrast and positive values positive contrast. The finite illumination as well as lens aberrations lead to a decreasing overall contrast transfer towards higher spatial frequencies, **B**. From [24].

This function indicates the amount of transmitted information in reference to the spatial frequency and is characterized by a sequence of alternating positive and negative zones (see Figure 2.8). The CTF oscillates between values of -1 and 1. Negative values describe a transfer of negative phase contrast (atoms appear bright on a dark background), while positive values represent positive contrast (dark atoms on bright background). As the CTF approaches the zero, less information are transmitted and at spatial frequencies where the CTF is exactly zero, no information is transferred. Theoretically, in the case of complete monochromatic electrons and in the absence of an aperture, the CTF would

extend towards high frequencies and thus transfer information up to almost unlimited resolution. This case is illustrated in Figure 2.8 **A**. However, due to a finite source size and varying electron energies in the beam, the CTF is dampened towards higher spatial frequencies which means a finite achievable resolution (see Figure 2.8 **B**). To summarize, the CTF describes:

- reversed information (negative values)
- diminished information towards higher frequencies and
- missing information at the zeros.

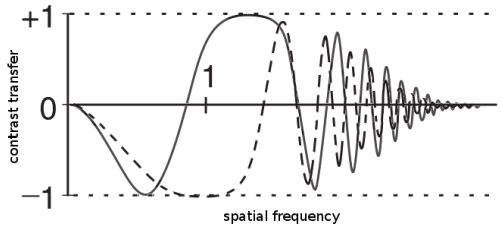


Figure 2.9: Two CTF's of different defocus values. Increasing the defocus shifts the first maxima and minima towards higher spatial frequencies. The combination of different defoci (solid and dashed line) compensates of the information loss due to zero crossings. Image modified from [24].

The problem of the reversed information can be compensated by simply flipping the negative values, and the degree of vanishing information can be controlled by the spot size (see section 2.2.2). The smaller the spot size, the less damping of the CTF and therefore higher resolution.
The third effect, missing information, can be also compensated: The CTF is dependent on the defocus value according to equation 2.5. The defocus defines the positions of the zeros. In order to restore all structural information, images can be recorded at a range of different defoci in order to cover all spatial frequencies. In Figure 2.9 two CTF curves for two different defoci are shown. The first zeros of each CTF can be compensated be each other, but there is still information missing at the second zero in both cases, which would make a third acquisition necessary. For image reconstruction, images are usually recorded under different defocus values, sorted into defocus-groups and analyzed separately by image processing methods. The different reconstructions are combined at the end to restore information of the whole spacial frequency range.
The corrections mentioned above are mostly important for cryo-EM, where the signal-to-noise-ratio SNR (see section 3) is very low but theoretical resolution high. In the case of negative stain on the other hand, the transmitted information after the first zero

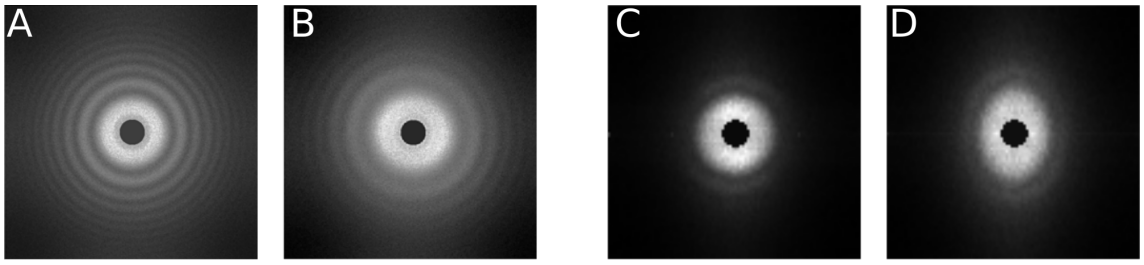


Figure 2.10: Representation of the CTF by diffraction pattern. The oscillations of the CTF are seen by Thon rings, where both, negative and positive values of the CTF, appear bright and the zero crossings dark. The extensions of the Thon rings are dependent on the spot size. **A** Small spot sizes result in little damping of the CTF and therefore in a large extent of Thon rings, representing high resolution. **B** Larger spot sizes and different defoci lead to a movement of the rings away from the center and to a stronger damping of the overall transmitted intensity. Thermal drifts and axial astigmatism can be recognized in the diffraction pattern **C** and **D** by not closed or elliptical Thon rings. Images **A** and **B** from [24].

crossing does not contribute significantly to the final resolution and hence corrections have little effect in practice. [24, 31]

The shape of the CTF can be also visualized in the form of a diffraction pattern (or power spectrum) of the electron micrographs as shown in Figure 2.10. From these images, more information on the quality of the images can be obtained. The power spectra represent a squaring of the CTF amplitudes hiding the fact of reversed information. The oscillating behavior is represented by so called Thon rings. The spectrum **A** in Figure 2.10 shows Thon rings extending to very high frequencies, representing high resolution, while the spectrum **B** is damped already at low frequencies and high resolution information is suppressed. Power spectra also reflect drift of the grid (image **C**), identified by not closed Thon rings and axial astigmatism (image **D**), which occurs for $z_a \neq 0$ in equation 2.5 and can be seen by an oval shape of the spectrum. A perfect image would have a power spectrum that has a round, uniform and large white disk.
Figure 2.11 demonstrates the effect of different defoci on the resolution by showing two micrographs and their corresponding diffraction pattern. As it is indicated, both images extend to similar high frequency but in the spectrum of micrograph **A** the first Thon ring can be seen, while it is not visible in micrograph **B**. Micrograph **B** was taken closer to focus, which causes the Thon rings to move outwards, thereby shifting the zeros and reversals to higher frequencies. Higher frequencies define small structural features, while lower frequencies represent the overall shape of objects. That effect is shown by the low contrast of particles in micrograph **B**, whereas good contrast in micrograph **A**. On the other hand, the shift of the zeros outwards (and therefore higher resolution information) is at the cost of the total intensity and thus information being transmitted at lower frequencies. Usually the right defocus is a compromise between loosing high frequency information and gaining particle visibility due to more power at lower frequencies.

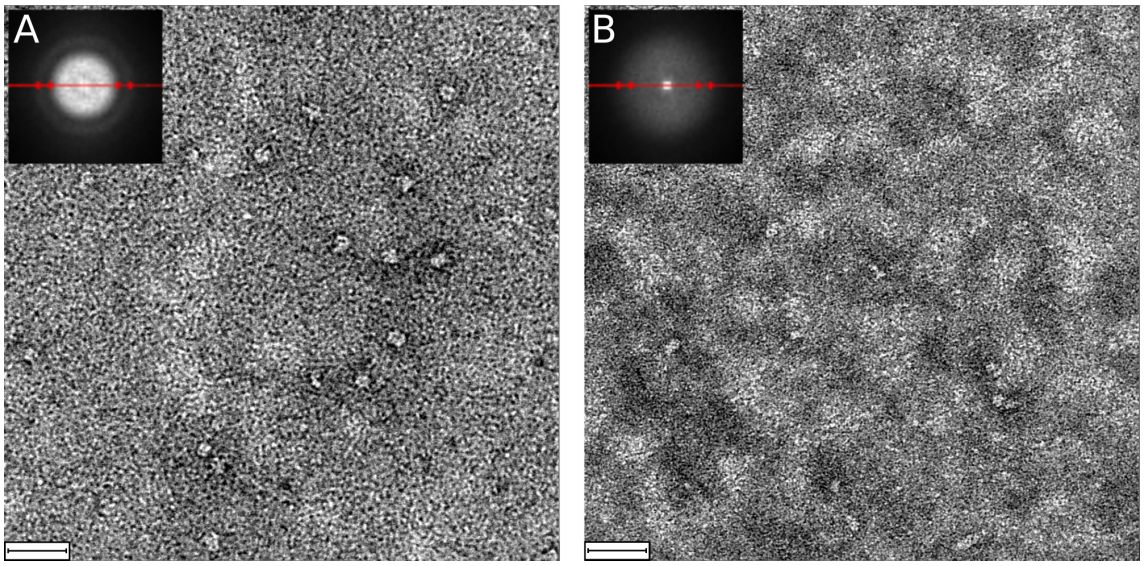


Figure 2.11: Two EM micrographs of nucleotide-free full length myosin XXI molecules recorded at different foci, with the corresponding diffraction pattern inserted. Micrograph **B** was taken closer to focus than **A**. Note the different pattern of the power spectra, both extent to similar high frequencies. Due to the overfocus of image **A** the Thon rings move towards lower frequencies and start to appear. While in micrograph **B** information of a larger frequency range is transmitted, the overall power of lower frequencies is higher in image **A**, resulting in better contrast and hence particle visibility. Scale bars: 40 nm

2.4 Negative staining: Principle and sample preparation

An established method for obtaining high contrast images of molecules is the method of ”negative staining”. Biological material applied to a TEM grid has usually very little electron absorbance and thus produces only very little contrast with the background. In order to improve this contrast, the specimen is stained with an electron dense material, usually compound of heavy metal salts with an high atomic number. After applying the sample to the grid, the stain in aqueous solution is added which spreads over the grid and embeds the particle. Stained areas scatter electrons strongly and are therefore almost opaque for electrons, while unstained or little stained particles are relatively transparent and therefore visible. One could also stain the particles, making them electron opaque against a light background. Since this ”positive staining” involves stain-protein interactions, however, negative staining is usually preferred. Depending on the sample, positive staining can also occur even if one aims for negative staining, as it is the case for e.g. DNA or ribosomes. Figure 2.12 illustrates both methods. As said, in negative staining the stain solution is mainly molded around the particle, but it also might penetrate into crevices and aqueous channels and is therefore able to reveal information about the internal structure [24]. The drawback of this method, apart from the possible distortions

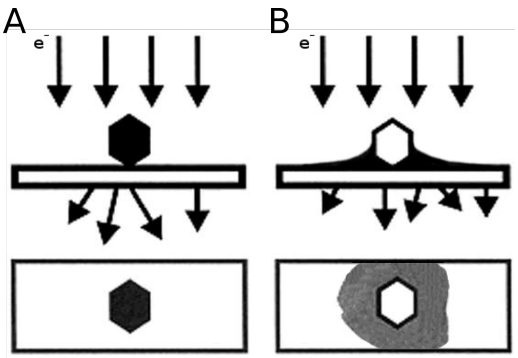


Figure 2.12: Comparison of positive and negative staining: **A** Positive staining includes the staining of the particles, resulting in a dark appearance on a bright background. **B** In the process of negative staining, the sample is applied directly to the TEM grid and stained with heavy metal ions, for example uranyl formate. The electron dense stain diffracts electrons more strongly compared to the weakly scattering biological specimen, which results in bright objects on a dark background. The stain also migrates into crevices of the objects, revealing also inner structural details. Image modified from [32].

of the molecules, is the ultimate resolution limit of about 10-20 Å caused by the characteristics of the staining solution itself. In order to apply negative staining correctly and to interpret the recorded electron micrographs appropriately, one has to consider various effects as described in the following:

- Negative staining is typically done by applying the specimen onto a glow-discharged of UV-treated carbon coated grid, followed by the application of stain, such as 2% uranyl formate in aqueous solution. After some time the excess liquid is blotted away by a filter paper and the grid is allowed to dry. The time in which the sample is exposed to the stain determines the contrast in the later EM imaging and staining procedures based on try and error are necessary to obtain the right contrast. However, due to the air drying of any remaining liquid, the molecules may be distorted, e.g. flattened to some extend [33]. At the end the forces during staining, blotting and air-drying are relatively large. Additionally, distortion may occur due to interactions between the molecules and the carbon film [34]. To minimize these interactions, the grid is glow-discharged prior to staining (i.e. making it hydrophilic by removing all organics causing the loss of hydrophilic characteristics). A comparison of negatively stained structures with structures obtained by other methods helps to quantify the degree of distortion due to staining [24, 34].

- As indicated in Figure 2.13, for the staining process itself it was important to use anti-capillary (**A**) instead of regular tweezers (**B**) in order to prevent stain migrating into the gap between the tweezers and under the grid. Due to the absence of sample on this side, it would create an equally distributed layer of stain, resulting in overall decrease of contrast.

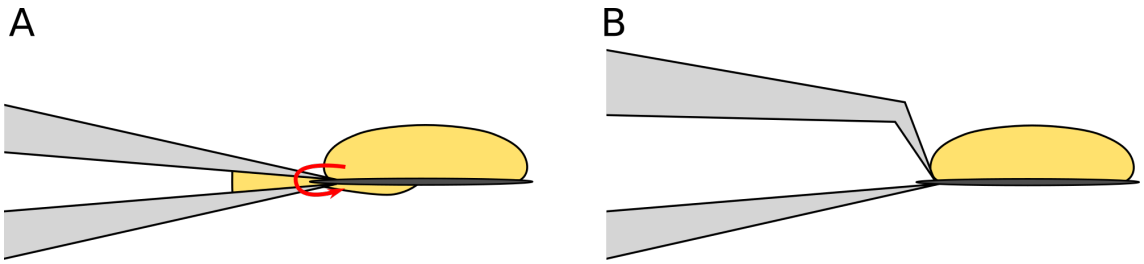


Figure 2.13: Stain migration into the gap of the tweezer and under the TEM grid (**A**) is avoided using a anti-capillary tweezer (**B**).

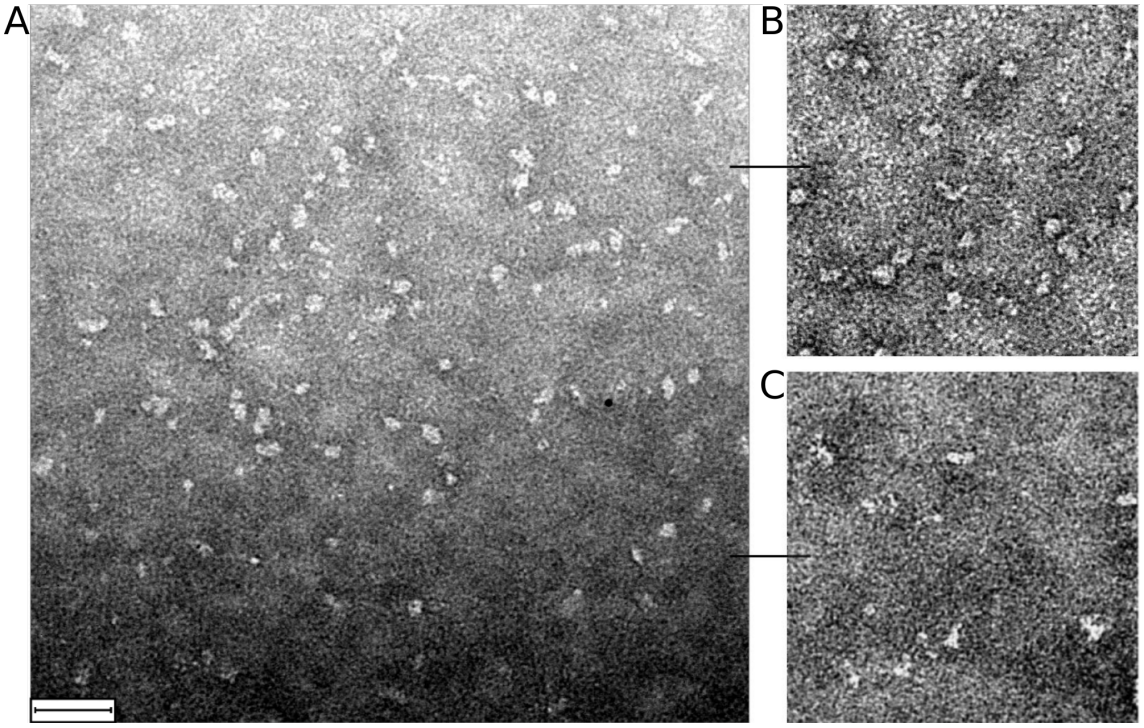


Figure 2.14: EM micrograph with higher stain depth at the lower end and more shallow stain at the top. Image **B** and **C** show these two areas after contrast balancing. Shallow stain **B** has the effect of clearly outlined particles, while deeper stain results in less overall particle contrast. Scale bar: 40 nm

- The quality of negatively stained electron micrographs and the issue of how representative the observed structures are, is also dependent on the right stain depth. An ideal stain would be just deep enough to embed the objects completely [24]. Apart from the effects of the staining procedure, the depth also varies across a single grid and it is important to find the areas with appropriate depth. When the grid is viewed under low magnification, darker and brighter areas can be identified (see Figure 2.14 **A**), representing areas of deeper and more shallow stain depths. A shallower stain depth has the effect that molecules appear outlined by an accumulation of stain, as one can see in a contrast balanced image of this area in Figure 2.14 **B**. This will enhance the image quality and particle visibility and is not seen in areas of deeper stain (lower half of Figure 2.14 **A**). Figure 2.14 **C** shows the corresponding contrast balanced image of that area.

- The concentration of particles on the grid should be between a few nM and a few hundred nM in order to obtain a reasonable distribution of particles on the surface. The exact values have to be determined experimentally and are dependent on the sample one is looking at. If for example single myosin molecules bound to actin are supposed to be studied, the concentration should not be too high since this may increase bundling of actin filaments or decrease the quality of filaments due to heavy decoration. On the other hand, if single molecules without actin are investigated, a high concentration has the advantage of recording several molecules within a single micrograph. The effect of the concentration on the staining quality is that too low particle densities will often lead to a too low depth of the stain as well.

- The particles should not be thicker than 30 nm, otherwise they would not be completely engulfed by the stain. Some part of the particle could stick out of the stain layer, which is consequently ”invisible” in an projection view from above. This ”partial projection” is not observable in single micrographs, but can be recognized after an image processing analysis. The images of small particles, lying in two orientations on the surface that are related by ”flipping”, can be usually brought into alignment by mirroring one of them. Images of bigger molecules on the other hand, such as the 40S ribosomal subunit, fail to obey each other by mirror relation [24].
  However, in this work myosin molecules and actin filaments are studied, which are small enough to legitimate mirroring for the alignment.

- There is also an effect of the electron dose on the behavior of the stain. Under low-dose conditions, usually higher resolution images can be obtained [35] and the

molecules are less radiation damaged. Nevertheless, high electron doses have the positive effect, that the stain crystallites may become fluid in the electron beam and begin to migrate. Therefore high-dose conditions can improve the contrast and the definition of objects against the background and certain features can be often seen already in raw micrographs[1]. In order to minimize radiation damage and maximize object identification, the right electron dose is a question of experience.

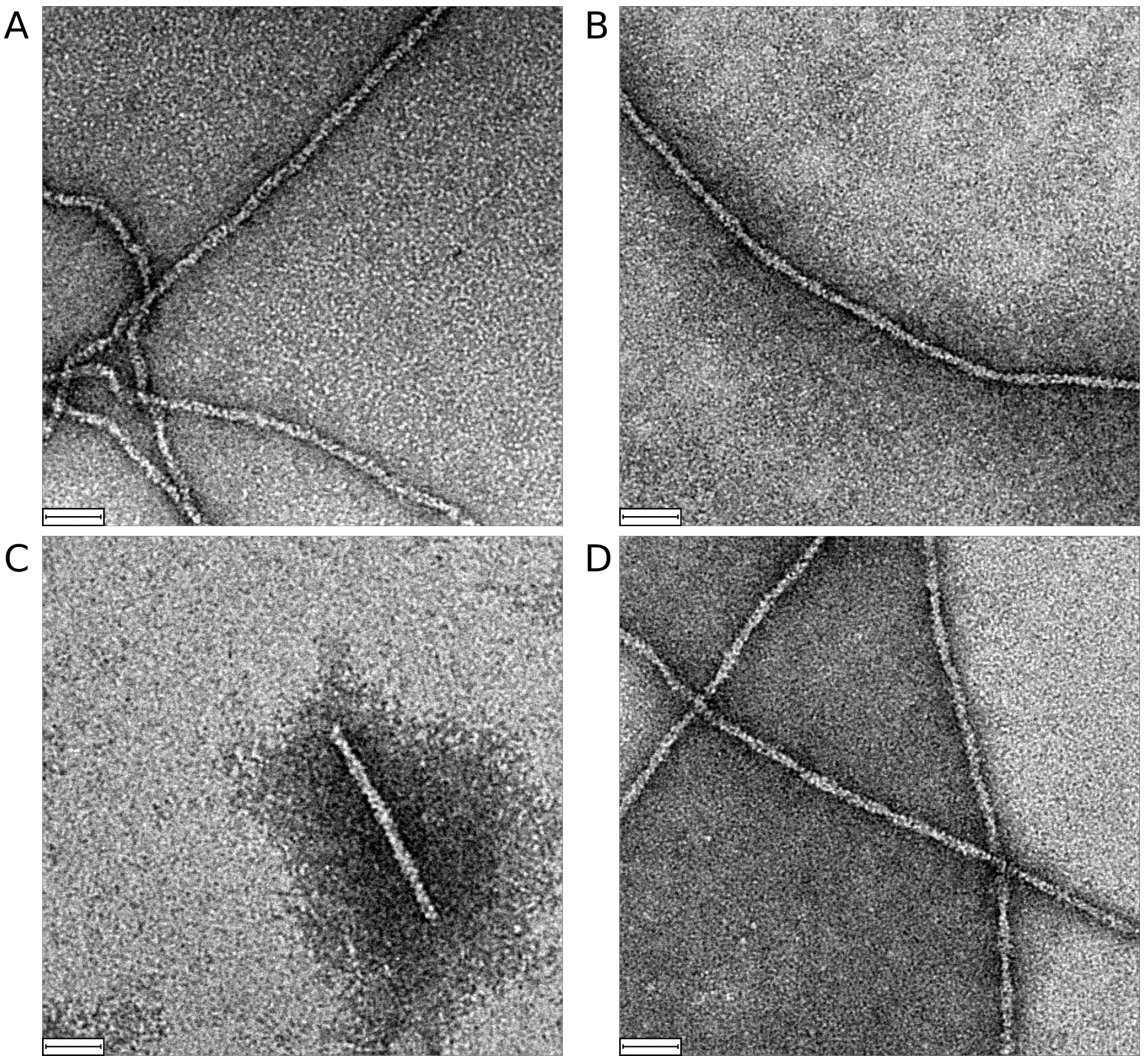


Figure 2.15: Exemplary micrographs of negatively stained actin filaments showing the effects of different stains. **A** The use of uranyl acetate leads to an sharp accumulation of the stain around the filaments, while uranyl formate **B**-**D** tends to form larger pools of stains. These pools are particularly visible in micrograph **C** and of advantage when small molecules attached to the filaments are supposed to be recorded. Scale bars: 40 nm

- Another factor affecting the quality of negative stain micrographs is the used staining material. Commonly used stains are uranyl acetate UA and uranyl formate UF. The difference is particularly visible when actin filaments are investi-

[1]Burgess S.A., personal communication, April, 2013

gated. While UA mainly forms contrast at the outlines of the filament (Figure 2.15 **A**), UF tends to form pools of stain, as shown in Figure 2.15 **B**-d. These pools often offer ideal and wider contrast, which is advantageous for small molecules attached to the filaments. The staining material has also an effect on the resolution. The heavy metals of the negative stain solution form microcrystals upon drying, whose size of about 20 nm are a natural limit to the resolution [36].

- And finally, as the electron microscopy is a much more sensitive compared to others such as spectrophotometry, the sample of interest has to have a high degree of purity. With the EM, even with negative stain EM, much more impurities can be detected than with any other biochemical approach. In addition to that, the visible parts of myosin molecules themselves are not large in size. The head and neck domain of myosin VI for example are approx.$15nm$ in length. The differentiation of the actual molecules of interest from impurities of similar size on the electron micrographs can be done by image processing, however the success of this analysis depends strongly on the sample purity.

Many of the drawbacks of negative staining mentioned above could be overcome by using another, stain-free techniques called cryo-EM, which was developed in 1984 by Dubochet and colleagues [37]. In cryo-EM, the sample is applied to a grid similar to the one used for negative staining, but without a continuous support-film. Instead, the e.g. carbon support comprises tiny holes in which the solutions is spanned. Instant plunge-freezing in liquid ethan creates a layer of thin ice which embeds the molecules in random orientations and conserves them in a near native state [38]. The fact, that the recorded micrographs contain projections of the particles from all possible angular views makes it very difficult to analyze the images by image processing methods. If the molecules themselves are additionally heterogeneous, analysis by cryo-EM becomes quickly challenging. Negatively stained particles on the other hand have the characteristic of being adsorbed to the surface in only a few preferred orientations, which increases the homogeneity remarkably (however on the cost of a decrease in the degree of freedom). Although cryo-EM is the ”state-of-art” due to much higher achievable resolution, negative staining is still a very powerful method, not least because it is a relatively quick and easy technique. More important, small particles below 200 kDa usually produce not enough contrast when observed by conventional cryo-EM (without staining) [39]. Therefore, negative staining is not only still preferred for first structural insights of unknown molecules, but it is also often the only method to analyze small objects, such as single myosin molecules (50-150 kDa).

3 Single particle image processing

Although negatively stained images have usually enough contrast to identify single molecules, the recorded signal does not only originates from the particle of interest, but also from unwanted background noise. This noise has various sources:
One is the irregular distribution of the stain crystals on the grid surface. The irregularity makes it impossible to eliminate this noise by simply subtracting a noise reference image. Furthermore, the carbon layer itself produces a structure, which appears superimposed on the image. Additionally, apart from noise that originates from the sample and the grid, the imaging devices also introduce noise. In the case of a CCD camera, noise is generated during the conversion of electrons into photons on the scintillator and during the following signal amplification [40].
One of the most limiting noise sources is the quantum nature of electric charge. It is produced by the statistical variations in the number of electrons that hit the recording device and its contribution to the noise increases with decreasing intensity of the electron beam (compare the statistical experiment of counting the heads and tails of a tossed coin: In the case of a large number of throwth, the difference is only a small percentage, while in the case of just a few throws, this percentage increases). Since the minimization of electron exposure is important, the contribution of this noise increases. [24, 30]
The amount of all the different noise sources is expressed by the **signal-to-noise ratio** (SNR), which relates the energy of the signal from the object P_{Signal} (described by the integral of the normalized power spectrum, see section 2.3) with the energy of the noise P_{Noise}:

$$SNR = P_{Signal}/P_{Noise}$$

The SNR of single micrographs is usually very low which makes an interpretation of the structural characteristics of small objects using only the raw micrographs challenging, if not impossible. The technological progress made it possible to increase the SNR by replacing the analogous way of recording electron micrographs by direct view-

ing using charge-coupled devices (CCD cameras). The most advanced form of recording the images produced by the EM is via direct detection, the first commercially-available direct detection camera for this purpose was introduced in 2008. However, no matter how revolutionary the imaging device itself is, it is crucial to analyze several hundreds to thousands images by superposition to filter out noise and allow for extracting the desired image information. Computerized and statistical analysis was introduced into electron microscopy as early as 1980 mainly by Joachim Frank and his colleagues [24, 41–43] and has been developed since. In addition, the fast growing computer industry speeds up the image processing and allows for processes that have not been possible before or were highly time extensive.

In order to improve the SNR by image processing, an aligned and homogeneous data set is required. While particles viewed in different orientations on the surface can be easily aligned by applying appropriate rotations and shifts in the image plane, particles absorbed at different binding sites to the surface and particles differing in their structure (e.g. due to flexibility of certain structural domains, different nucleotide, oligomeric and conformational states or the different appearance between intact and damaged molecules) make it necessary to classify the data set [24, 30] and thus compensate for heterogeneity. In the following section the basic principles behind the single particle analysis, i.e. particle picking, alignment, averaging and classification of negatively stained EM images, will be described. In the projects presented in this thesis, the single particle image processing was performed using mainly the software suite SPIDER [44] and EMAN2 [45] (see section 3.4 for more details).

3.1 Two-dimensional alignment and averaging methods

The averaging of particles in order to reduce the SNR requires precisely aligned particles. Aligning of two or more images is understood as the process of bringing a common motif contained in each image into register. Such common motifs are present when the set is homogeneous and all images of the particles are essentially the same. For a not completely homogeneous image set, a common, overall shape can be used as well for the alignment. The algorithm of finding the corresponding orientations is done using cross correlation: [24, 30]

3.1.1 The cross correlation function CCF

The cross correlation measures the consistency or dependency between two functions. In signal processing. the integral of the product of two functions that are constantly shifted against each other in time, is calculated as a function of time. The value of this **cross correlation function** (CCF) is called cross correlation coefficient (ccc) and has its maximum when the two functions are the most similar. For example, two identical signals, recorded in an interval of 2 ms, would result in a ccc-value of 1 for a shift of exactly 2 ms.

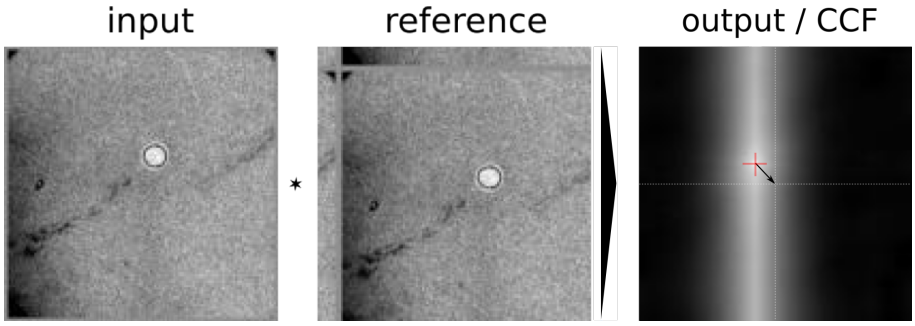


Figure 3.1: Demonstration of the cross correlation between to images that only differ in a translational shift. The resulting cross correlation function indicates the shift parameter that has to be applied to the input image to align it with the reference image. The shift is found by searching the coordinates of highest intensity within the CCF with respect to the center. Images modified from [44].

This concept can be directly applied to image analysis, if images are represented by their 2D functions, which should be normalized to exclude effects of varying exposure times. The two image functions are translationally shifted against each other and the CCF is calculated, which can be again represented as an image. The location of highest intensity within this image (with respect to the center of the image) corresponds to the 2D shift that has to be applied to one of the images for the best match, as demon-

strated in Figure 3.1. Rotational alignment can be obtained by representing the images in polar coordinates, where rotations are again described by translational shifts. Apart from calculating the appropriate rotations and shifts, the cross correlation coefficient is an important value to measure the similarity between two images. In this work, cross correlation is applied using the SPIDER commands *CC C* to calculate the CCF and *PK M* to search for the location of highest intensity [44]. These operations can be applied separately, but are also the basis for other aligning operations contained in SPIDER.
As it can be seen in Figure 3.1 (reference image), shifts are done by SPIDER in a circular enclosing manner, which may introduce artifacts. To avoid influences on the analysis, images can be padded into an larger image of double size with an average background value.

3.1.2 Reference-free and reference based alignment

In order to align a full set of images, there are two different techniques:
If the structure of the analyzed objects is know in advance, a model of its structure or a reference image can be used to align the data set. The SPIDER suite contains various operations for this **reference-based alignment**, for example *AP SHC*, which is mainly used in this work. This algorithm combines rotations and shifts within a specific range with the calculation of the cross correlation for each orientations. Moreover, the operation allows for restraining the translational and rotational search, which can be adjusted to the individual object sizes. Rotational restrictions are particularly useful when objects are known to have a polar structure, for example actin filaments. *AP SHC* also includes the option of multireference alignment, where the alignment is done for different reference images and the reference with the best match is listed in the resulting document file. This is for example needed to align an image set with different projections of an existing 3D model in order to identify the orientation of the recorded particles. The resulting translation and rotation can be applied using the command *RT SQ*.
The disadvantage of this approach however is that the alignment is biased by the choice of the reference image and the average is likely to appear similar to the reference even if the real structure looks very different. Figure 3.2 **A** illustrates this effect on an example data set of myosin VI. The unaligned data did not show any structural features but as a reference image of myosin VI was available (Reference 1), the data set was reference-based aligned with it. This resulted in an aligned global average, that could suggest great homogeneity within the data set with the majority of images looking similar to the reference. However, the alignment with an artificially created image (Reference 2) suggests that the objects have an ”A”-shaped structure. One way to overcome this problem is by applying several rounds of alignment in an iterative way, using the last aligned global

average as a new reference. This could wash out to fake information ("A"-shape), which however works only partly and the final result will always be somewhat biased towards the initially used reference. Therefore, the use of a reference-image at the beginning of the analysis should be avoided.

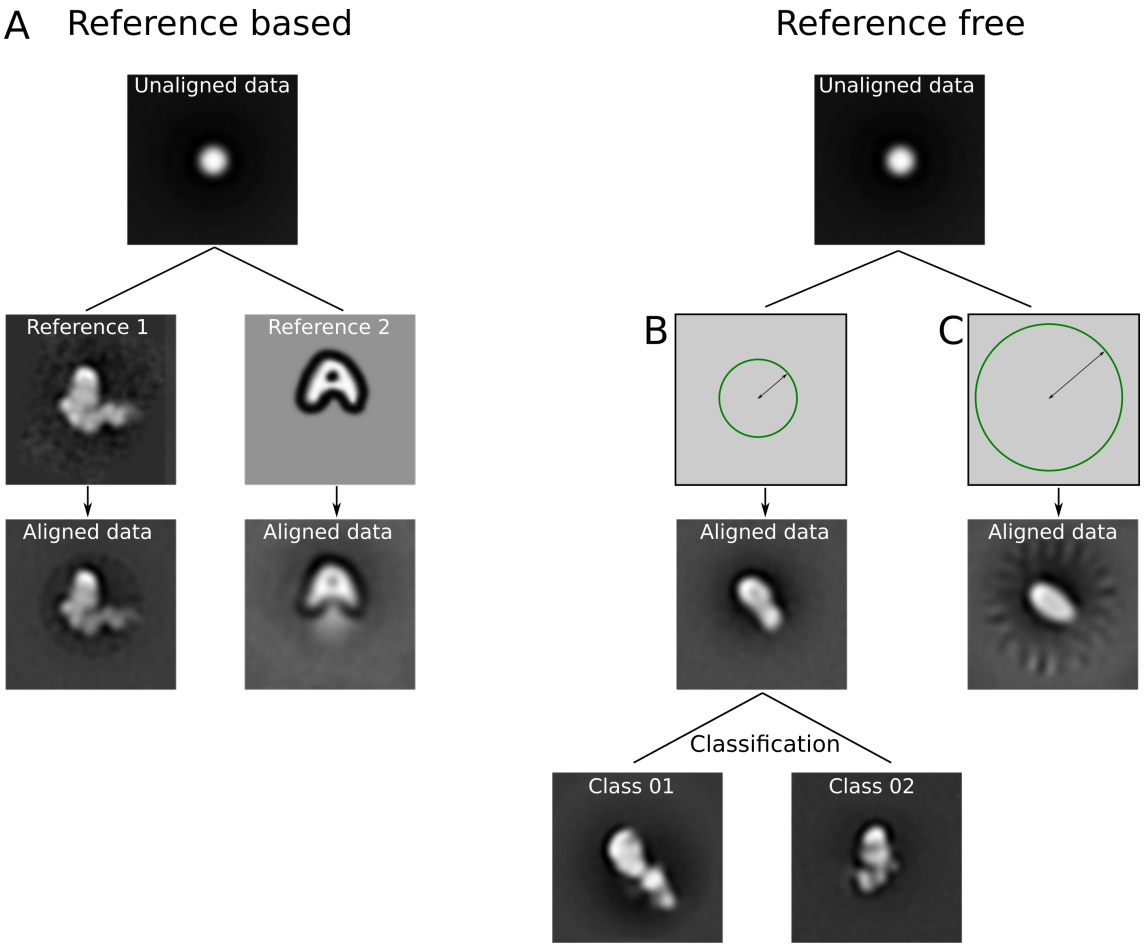


Figure 3.2: Reference free vs. reference based alignment: **A** The reference based alignment of the dataset with the two shown references indicates the strong bias during the alignment. Even though if the raw images vary significantly from the reference image (e.g. the A-shape), the aligned average suggests otherwise. **B,C** To eliminate this effect, the alignments should be done without any reference. In this case, the used mask and search radius (green rings) become important. If an overestimated mask is used, areas might be aligned that do not contribute to the actual object of interest. In case **C** e.g. background noise and stain crystals became aligned, resulting in the halo-like pattern.

The preferred method for initial alignments is completely reference-free, especially if the structure of the particles is completely unknown. One successful **reference-free alignment** approach by the SPIDER is the operation *AP SR* [44], which works in two steps. In the first step, two randomly chosen images are aligned by cross correlation and a very first average is created. This average is than aligned with a third randomly chosen image and a second average is calculated. This procedure is done until all images

are used, resulting in a first initial average. The second part is a refinement step, where iteratively a randomly chosen prealigned image (from the first step) is aligned with the average of the remaining images, until no further shifts of rotations are needed to align all images [46]. Figure 3.2 **B** shows the alignment on the same data set but without any reference image. The resulting average looks still relatively featureless, but classification reveals the underlying structures. Nevertheless, this method has also its drawback, namely the reproducibility. While two reference-based alignments with the same data set and parameters will always yield the exact same results, this is not necessarily the case for reference-free alignment. In particular if the data set is small (approx. 0-50 images, depending on the homogeneity), the random choice of the images will result in different averages for different alignment runs, even with the same input parameters. Therefore, one has to be carefully inspect the alignment and resulting aligned images and evaluate their homogeneity by classification.
However, if the reference-free alignment does not lead to satisfying results, there is a way to bias the alignment towards known features whilst keeping the alignment reference-free. As a standard, the SPIDER command *AP SR* has an algorithm implemented that tries to keep the images centered during the iterations using initially a Gaussian blob and the center of gravity for the following iterations. Instead of this blob one can use any arbitrary image, such as Reference 1 in Figure 3.2 **A**. This image will not be used to align the data set with, but to center all images and thus bias the alignment towards the overall structural properties of this image. This is especially useful if the objects differ significantly from blob-like objects such as actin filaments.

Obviously, both alignment algorithms do not differentiate between signals originating from the actual particle and signals originating from noise. For the algorithm an image is simply a matrix containing pixel coordinates and pixel intensity values. Therefore, the results of the alignment are very sensitive to the chosen parameters such as translational and rotational search range, image size, whether masks are applied or not, the shape if the masks and whether they have sharp or soft edges. The translation and rotation parameters should be set just large enough to include all pixels representing the particle, but small enough to exclude most pixels representing noise. Apart from this, a mask can be applied using the SPIDER operation *MM C* [44]. Areas, where the value of the mask falls below 0.5 the image will be replaced with a continuous background value. Both, translational and rotational parameter and the choice of mask have to be treated with care. If the parameters are wrongly set, background noise might become an aligning-factor itself. The same applies to the mask, where sharp edges (acting like strong image features) would become another factor for the alignment. The significance

of these effects is illustrated in Figure 3.2 **C**, where the search range / size of mask (indicated by the green circle) is chosen too big. The halo-like pattern in the aligned average has its origin in stain crystals being aligned. Smaller sizes of the mask result in better alignments (**B**).

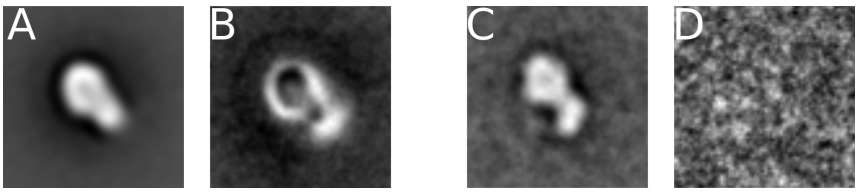


Figure 3.3: Global averages of an image data set (a and c) with their corresponding variance images (b and d). Regions of high intensity within the variance images indicate locations of high interimage variability. **B** The variance image of an aligned image data set of 2660 images of nucleotide-free myosin XXI monomers indicates high variability on the outlines of the molecules and particular at the lower right end, which is also seen by a blurred structure in the global average **A**. Image **C** shows a class average after classification of the data set. The homogeneity of this subclass is identified by the variance image **D**, which represents only variability due to background noise.

Whether an alignment is successful or not can be verified by inspecting the resulting average and variance image. With increasing similarity of the aligned objects, the variance decreases. The variance image is particularly helpful to spot regions of inconsistency within the images (interimage variances). On the one hand inconsistency is always given due to noise which is why the variance image of perfectly aligned images of the same object will be an image with equally distributed noise, representing that all images only differ in background noise. On the other hand, in case of misaligned objects or aligned objects that are not homogeneous, the variance image helps to identify areas where the images differ the most. The effect is exemplary shown in Figure 3.3. The average **A** of an aligned, but heterogeneous set of myosin XXI images appears blurred, but it is difficult to spot the areas where the images differ. The variance image **B** on the other hand reveals, that the biggest interimage variances are located at the lower right end of the molecules. Image **C** shows for comparison an average of only a few images from the data set. The variance **D** indicates, that this data set is very homogeneous.

3.2 Two-dimensional classification: Multivariate data analysis

While heterogeneity is usually inhibiting successful alignment, it is also the basis for the analysis of structural characteristics, such as flexible protuberances of a molecule, different conformational states or different orientational views of the same molecule. Various tools for statistical analysis exist that are capable of dividing a large image data

set into subsets of images with similar features. The following are the widest used and most efficient methods.
The basic idea is to represent each image of a data set by a vector in a multidimensional space R^J, where $J = LxM$ is the number of pixels in the images (it is required that all images are of the same dimension). For example a set of N images with only two pixels each, J=2, would be represented by points in the 2D space with coordinates according to the density value of each pixel. However, what is usually called an "image", consists of much more pixels and the space which represents the image set becomes quickly of very high dimensionality, beyond our ability to visualize. Exemplary, an image data set where the images only differ in the density of a single pixel, would form a straight line in the hyperspace, parallel to the axis describing this pixel. In a more realistic case, every pixel will vary across the data set and the representation would be a hyperdimensional cloud. Furthermore, moving along one axis only describes the change of one single pixel, while variances between structural subunits always involve the change of numerous pixels. This makes clear, that the initial choice of the axes is not appropriate and the aim is to find a new (orthogonal) coordinate system, whose axes describe these interimage variances rather than the densities of each pixel. [24, 30]

3.2.1 Principal component analysis PCA

As described above, the goal is to create a new coordinate systems with axes describing the interimage variances. These variances are described by the principal extensions of the data cloud, which form the basis of the new coordinate system and are obtained by first finding the maximum extension of the cloud followed by finding a vector perpendicular to the first one and pointing in the direction of the next-largest extension. All following vectors are found by the same principle, being perpendicular to the already found vectors. With increasing number of vectors, called *principal components*, the shape of the cloud is described with increasing accuracy.
Mathematically the search for the maximum extension $\mathbf{u}$ in of the space R^J of a data cloud can be solved by introducing a matrix $\mathbf{X}$ containing the image vectors as rows and performing an eigenanalysis (in statistics called principal component analysis PCA) of its covariance matrix $\mathbf{D}$:

$$\mathbf{Du} = \lambda\mathbf{u} \tag{3.1}$$

where $\mathbf{D}$ is the covariance matrix of $\mathbf{X}$

$$\mathbf{D} = \left(X\text{-}\overline{X}\text{'}\right)' \left(X\text{-}\overline{X}\right)$$

with

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1J} \\ x_{21} & x_{22} & \dots & x_{2J} \\ \vdots & \vdots & \dots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NJ} \end{pmatrix}$$

and $\overline{X}$ representing a matrix containing the average image in each row.
The solution of this eigenanalysis is the new basis in R^J, formed by the eigenvectors **u**, which represent certain interimage variances. Their contribution to the overall variance is given by the corresponding eigenvalue λ. [24, 30]

3.2.2 Correspondence analysis CA

A similar technique for the statistical analysis of data is the **correspondence analysis** (CA) which has in the case of image processing of EM data some advantages compared to PCA.
In statistics, CA is used to decompose a set of values within a table into a small number of "summary variables" and to quantify the correspondence between each of the values. The degree of correspondence is calculated by the χ^2 statistics[1], in contrast to PCA, where the distances are calculated in the euclidean metric. Applied to image analysis, the table mentioned is build up by N images (rows) with J pixels (columns), which can be represented by two diagonal matrices **M**(NxN) and **N**(JxJ), separating the "image space" R^N from the "pixel space" R^J. Additionally the diagonal entries of both matrices are expressed in "relative frequencies" [47]:

$$m_{jj} = \frac{1}{x_{.j}}, j = 1, ..., J \quad \text{and} \quad n_{ii} = \frac{1}{x_{.i}}, i = 1, ..., N$$

where $x_{.j}$ and $x_{.i}$ are defined by

$$x_{.i} = \sum_{j=1}^{J} x_{ij} \quad \text{and} \quad x_{.j} = \sum_{i=1}^{N} x_{ij}.$$

[1]The chi squared distance, in contrast to the euclidean distance, allows for calculating differences between values with the inclusion of weighting each value with reference to an expectation value

The advantage of CA in contrast to PCA at this point is the implemented rescaling / normalization of each value, which makes the outcome independent on images differing only by an overall multiplicative factor. Images of the same object recorded with different exposure times would be treated as identical in CA (given that the optical density is linearly related to the electron exposure), while PCA would treat them as different images.
With the definitions above, the eigenvalue equation (compare equation 3.1) becomes

$$\mathbf{X'NXMu} = \lambda \mathbf{u}. \tag{3.2}$$

Similar to PCA, the solution is a new set of orthogonal vectors (eigenvectors) representing directly interimage variances. The corresponding eigenvalues λ assign the share of the overall variance for each vector. Additionally to the analysis of image vectors in the "pixel space" R^J, CA simultaneously analyzes pixel vectors in the "image space" R^N with the result of another set of eigenvectors and eigenvalues, representing "interpixel" variances. [24, 30]
To summarize, PCA and CA are able reduce the complexity of multivariate data, represented as a cloud in a multidimensional space, by finding the principal extensions (eigenvectors, also called factors) of this cloud and by this describing the interimage variances. Furthermore, each eigenvector is associated with an eigenvalue which describes the "importance" of this factor.
In the new coordinate system, each image point can be reached by a linear combination of the eigenvectors. Using all vectors, the images would be completely reconstructed. However, often only the first few factors represent "real" differences between the objects, while most of the remaining factors (described by smaller eigenvalues) represent differences in background noise or differences in appearance due to uneven stain. Therefore, images can be reconstructed by a small number of vectors, which reduces the representation of the data set significantly without loss of information and speeds up the analysis. As an example [24] consider an aligned image set of 1000 images with a size of 64*64 pixel. A full pixel-to-pixel analysis would require the consideration of 1000*64*64 approx. 4 million values. On the other hand, the application of correspondence analysis with the inclusion of 10 eigenvectors requires only 10*1000 image coordinates plus 10*64*64 pixels of the 10 factors, in total approx. 50,000 numbers, which is a reduction by a factor of 80.
In this work correspondence analysis was applied using the SPIDER command *CA S*, which is also capable of performing PCA.

3.2.3 Factor maps

A useful way to visualize the outcome of the correspondence analysis is to display the image coordinates in a 2 dimensional map where each axis represents a specific eigenvector / factor. One can think of a projection of the cloud onto a layer spanned by two of the new coordinate axes (eigenvectors). The extension or a possible clustering of the cloud with respect to a specific factor can be easily inspected graphically with the use of such a "factor map", which can be created using the SPIDER operation *CA SM*. Images in this 2D map are represented usually by an image specific ID at the corresponding coordinates. However, the factor maps only give an overview on how the images are distributed along the principal variances. It does not show immediately, what changes are actually represented by the eigenvectors.

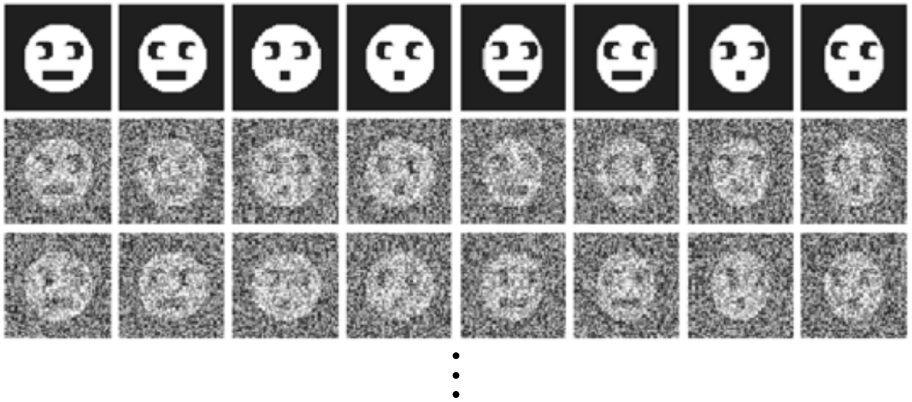


Figure 3.4: For the demonstration of correspondence analysis, a synthetic image data set was created with images of faces differing in three details, resulting in eight different classes: shape of head (round, oval), direction of eyes (left, right) and shape of mouth (big, small). Images are superimposed with random noise ten times, leading to a total image data set of 80 images. From [24].

One way to determine this is to use the fact that the eigenvectors are of the same dimensions as the images (dimension J, see section 3.2), which is why they can be interpreted as images again, called "eigenimages". These images show areas of major variances in the data set with respect to one or more factors. The construction of eigenimages can be done with the SPIDER operation *CA SRE* [44].
Another way is to calculate the averages of images that are lying either in distinguishable clusters or, if the cloud is rather continuously, to average images lying at the very extremes of the cloud. The calculated averages usually differ in one or two (depending if the extension of the cloud is parallel to one of the factors or not) distinct characteristics. The graphical user interface of SPIDER, called WEB, can be used to calculate the mentioned averages interactively by drawing a circle around image points in the factor map.
A histogram showing all eigenvectors ordered according to their share to the overall

variance, is furthermore a useful method to decide, which factor should be included into the analysis. This histogram can be created using the SPIDER operation *CA EIG-PCT* [44]. [24, 30]

The correspondence analysis and the use of the eigenimages and factor maps is best demonstrated with an example of a computer generated image set with clearly discernible features [24]:

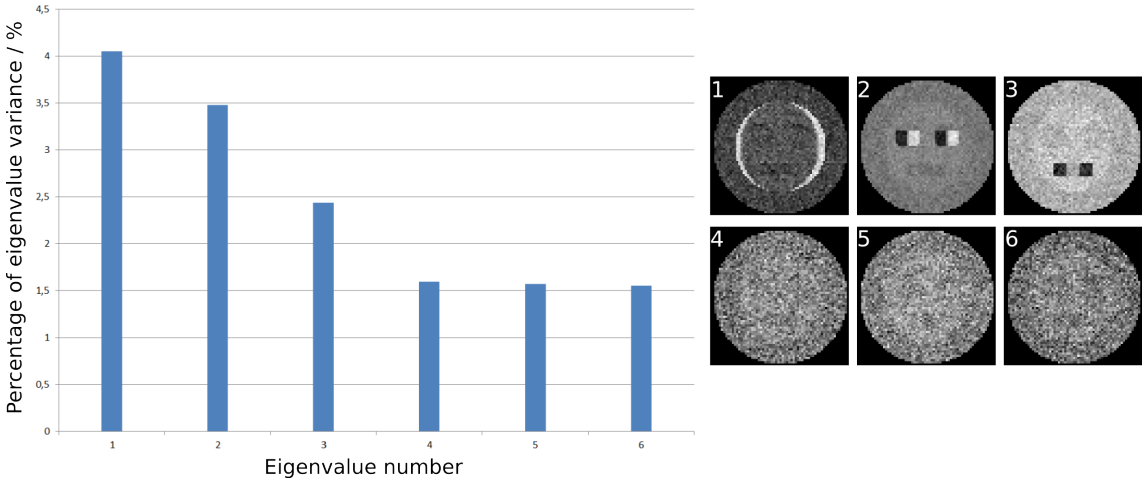


Figure 3.5: Histogram showing the percentage of interimage variance shared by each eigenvalue after applying correspondence analysis to the synthetic image data set in Figure 3.4. The first three eigenvalues are dominant and represent the three characteristics, shape of head and mouth and eyes, as shown in the eigenimages alongside. Other eigenvalues represent differences in noise.

The artificial data set contains images with faces differing in three details: Position of the eyes (left, right), shape of the mouth (small, big) and shape of the head (round, oval), as shown in the first line of Figure 3.4. Additionally, each image is superimposed with random noise in ten rounds to create a total set of 80 images, that is applied to correspondence analysis. The outcome of CA first investigated by plotting the share of each eigenvalue to the total interimage variance, shown in Figure 3.5. As it can be already seen, the total variance is mainly defined by the first three factors. Alongside the histogram are the corresponding eigenimages for the first 6 factors shown (the bigger circle visible in each image is due to the application of a mask during the procedure of CA). As expected, only the first three factors represent differences in the structure (image 1: head shape, image 2: eye position, image 3: mouth shape), whereas the remaining eigenimages indicate only differences in noise.

Figures 3.6 and 3.7 demonstrate the use of factormaps. Two maps are shown for the factors 1 & 2 and 3 & 4, respectively. Obviously, the images fall into two clusters for each of the first three, whereas along factor 4 the cloud is equally distributed. Using WEB, the averages of the images within the indicated circles are calculated, demonstrating clearly the interimage variances. As expected, images b and c in Figure 3.7 do not vary in any of

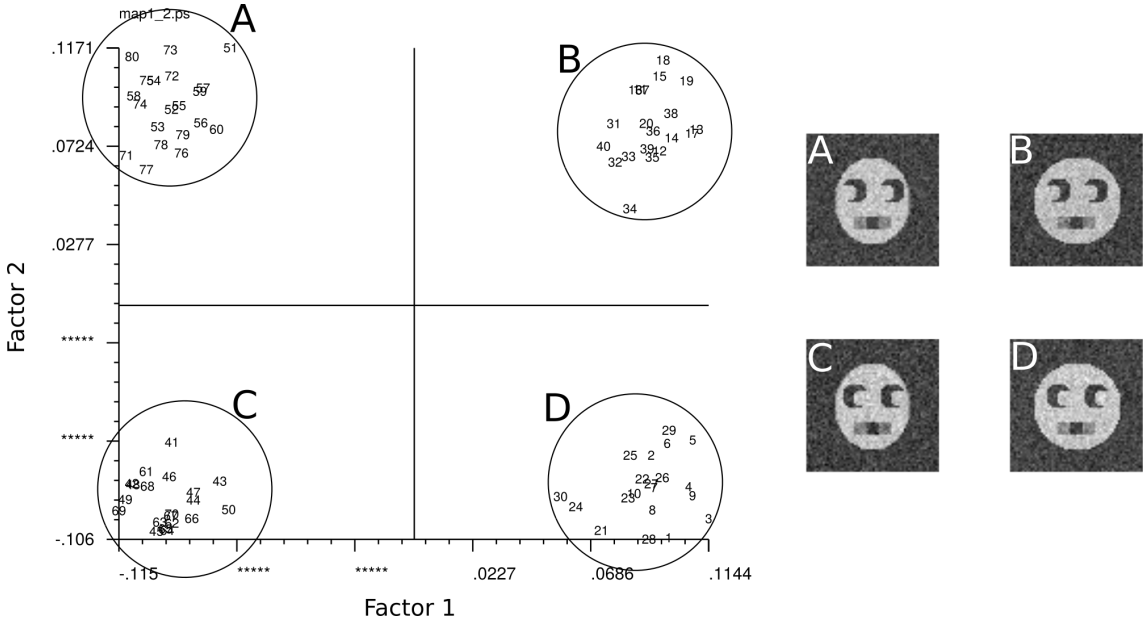


Figure 3.6: Representation of the hyperdimensional cloud of the face image data set in 2 dimensions by a factormap for the first two eigenvectors. Clearly visible are four clusters, identifying the classification into two classes for each factor. The right images show the averages of the four clusters, reproducing the shape of the head (images c/d and a/b) and the direction of the eyes (images a/c and b/d).

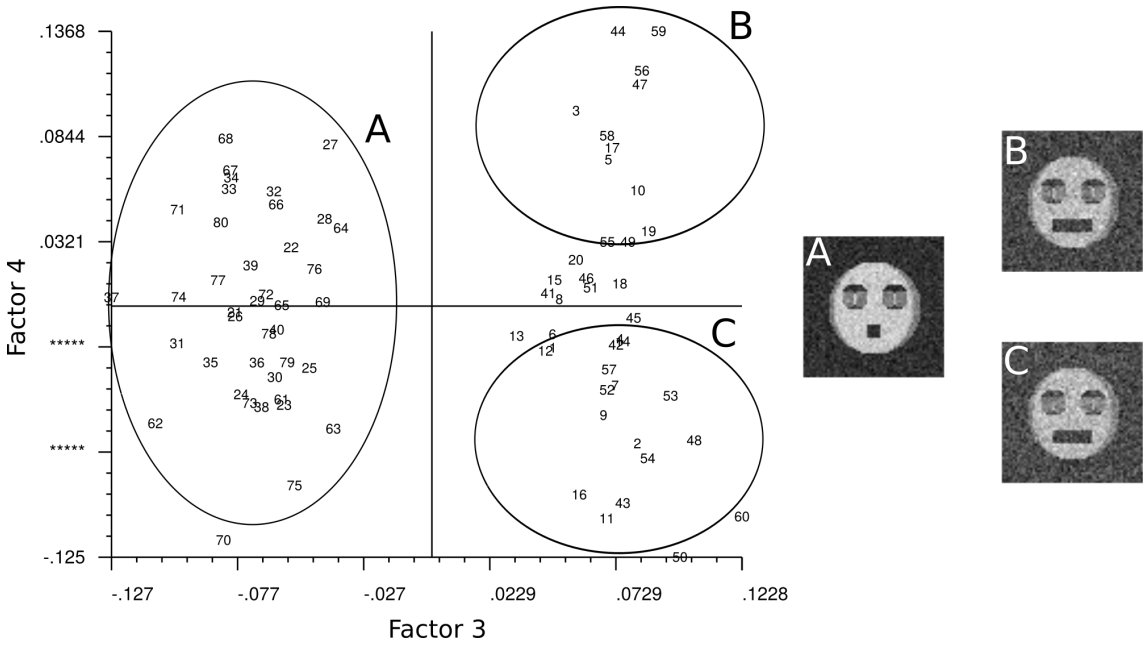


Figure 3.7: Representation of the hyperdimensional cloud of the face image data set in 2 dimensions by a factormap for the third and fourth eigenvector. Visible is only the clustering in two classes along factor 3, while both clouds are equally distributed along factor 4. The averages of the three circles indicate the classification according to factor 3 (shape of mouth), while no clustering is found for factor 4 (images b and c).

the face-characteristics but only in noise. Additionally,each average image has a blurred structure of the characteristic represented by the factor that was not taken into account, for example the shape of the mouth in Figure 3.6, which is described by factor 3.

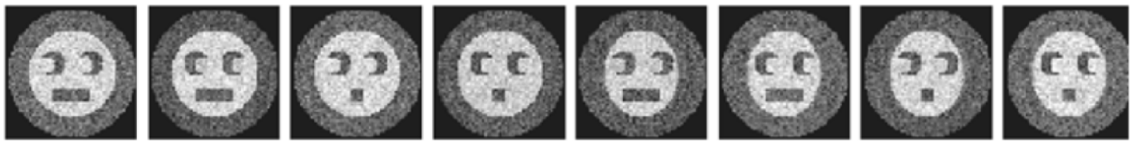


Figure 3.8: Reconstruction of the noisy image data set by a linear combination of the first three eigenvectors. Since factors, that represent differences in noise are not taken into account (factors ≥4), the SNR of the images is significantly improved. From [24]

Knowing that factors ≥4 represent differences in noise, the images can be reconstructed by a linear combination of the first three eigenvectors. As shown in the second row in Figure 3.8, this reconstruction leads to almost noise-free images (compare the initial noise-images in Figure 3.4). [24, 30]

The way of dividing the image set into subsets according to structural characteristics using the factor map is only used to get a first idea of what each factor represents and to see if the data set falls into clusters at all. However, a successful classification usually includes more than two factors, which is in the case of four factors already beyond our ability to visualize. Furthermore, the classification should be more automatic, controllable and reproducible by means of choosing the number of clusters and the weighting of certain factors. The basics of two different methods for classification, namely hierarchical ascendant clustering and K-means, are described in the following:

3.2.4 Classification by hierarchical ascendant clustering HAC

This classification method produces a hierarchy between all images of the data set by identifying similarities between images. Initially each image (or data point in the multidimensional space) is considered as an own cluster and the most similar images are merged into a new cluster. Following this, the newly formed clusters are again merged together with respect to their similarity. This procedure is done successively until one cluster is reached, containing all images. Images are considered as similar when their points in the multidimensional space are lying closely together, whereby this space has the dimension of the number of factors included in the classification algorithm. Moreover, it is useful to assign to each merging step a measure of similarity, by which all steps can be sorted. Various criteria have been invented, the mostly used one is the ”Ward’s criterion” [48]. This criterion is based on finding the minimum variance that is added to the new cluster compared to the variance within each previous group. From these measurements, a tree can be constructed that represents and visualizes the hierarchical

clustering, called ”dendogram”. To demonstrate the concept of hierarchical clustering using the Ward’s method it is applied to the previous exemplary face image set (Figure 3.4):

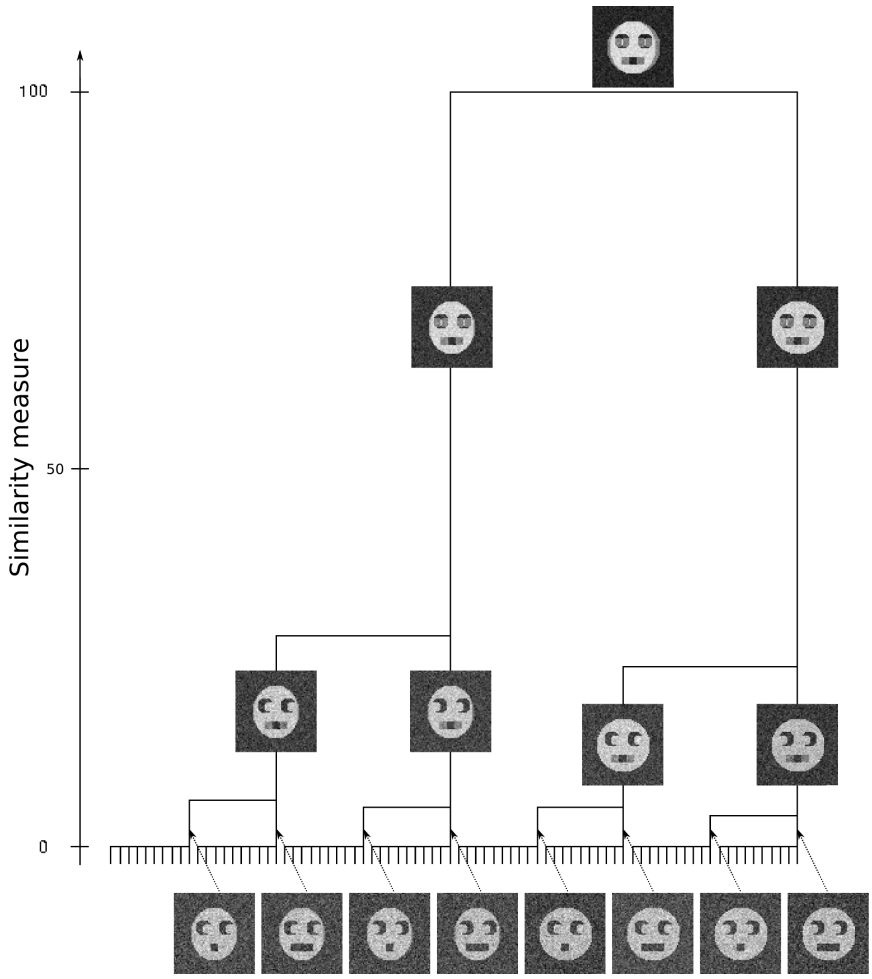


Figure 3.9: Dendogram from the classification of the face image data set by hierarchical ascendant clustering and the use of the Ward’s method (see text) with inserted classes at different heights. The lower vertical lines represent the individual raw images, which are successively merged into larger classes. The height of the vertical steps indicates the stability of the classes. A cut off at about 3% would successfully reproduce the initial eight classes, whereas a cut off at 50% leads to only two classes that differ in the most dominant characteristic, the shape of the head.

Figure 3.9 shows the dendogram of the data set with the first three factors included, created using the SPIDER operation *CL HC* [44] and visualized using WEB. Usually WEB only displays the resulting clusters at one specific height, but in the figure shown here the results for three different levels are inserted for clarification. The vertical lines at the very bottom represent the initial raw images, each forming its own cluster. These images are successively merged into larger clusters, indicated by the following vertical lines. The vertical axis on the left represents the measure of similarity in %, with higher values representing the merge of less similar clusters. If the dendogram is cut for ex-

ample at 30 %, two clusters would be created, representing the characteristic with the highest variance, the shape of the head (compare the histogram in Figure 3.5, where this characteristic is represented by the fist factor). A cut off at 3% would lead to the initial eight classes. The height of the merging steps can be also interpreted as a stability of the classification. Higher steps are more stable than smaller steps, since small variations in the cut off value do not affect the number of resulting clusters. [24, 30]

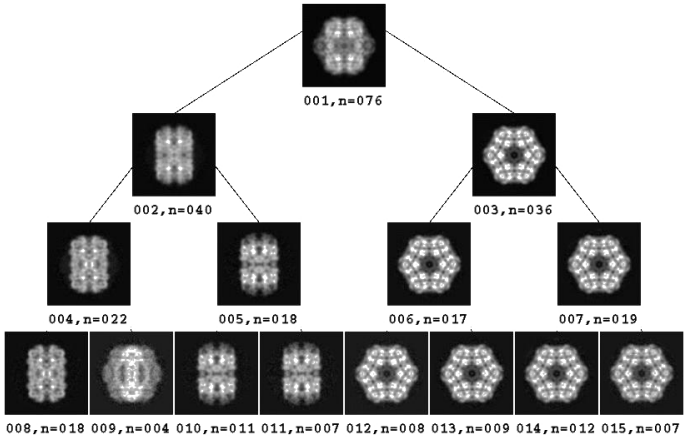


Figure 3.10: Classes produced from a dendogram after HAC at different cut off values can be represented within a single tree. This representation effectively identifies the necessity for further division of a cluster into subclusters. For example a classification of class 3 does not reveal additional information, in contrast to class 2. From [44]

As mentioned, the result of *CL HC* are only clusters at a specific cut off level. However, it is often difficult to decide, where to set this level. Therefore, it is useful to visualize the results for different levels together in one tree. Such a binary tree[2] is exemplary shown in Figure 3.10. In this representation it is easier to see at which point further division into smaller clusters becomes unnecessary (note that the clustering algorithm in fact goes the other way, merging smaller clusters into larger ones). In the example shown here, class 3 contains similar images, while class 2 should be still further divided.

3.2.5 Classification by K-means

The classification by K-means is different from the hierarchical clustering in the way that it does not use all images as initial classes but starts with a user defined number (K) of randomly chosen images as ”seeds” for a following iterative classification. To demonstrate this iterative process, an multidimensional image data set is represented in a reduced 2D space in Figure 3.11 **A**. If for example K=2, two points are randomly picked from the data set (filled circles), which define the first partition into two clusters, indicated by the dashed line. Subsequently (Figure 3.11 **B**), the center of gravities are

[2]The batch files to create a binary tree can be obtained from [44]

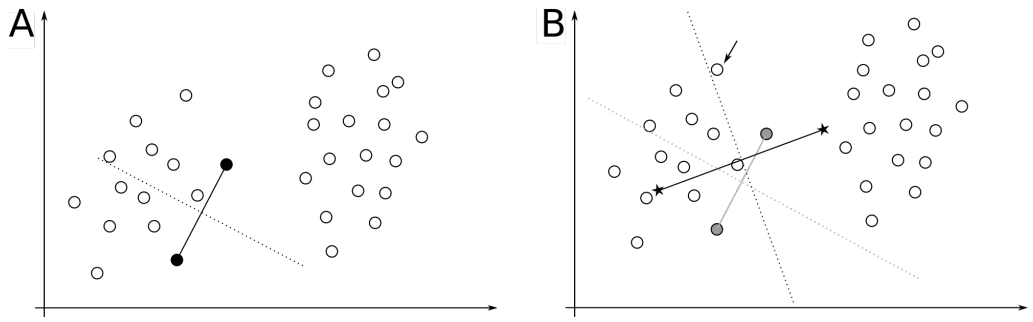


Figure 3.11: Illustration of the classification by K-means. **A** K points of an image data set, represented here by points in a 2D space, are selected randomly. These points divide the data set into K initial clusters (in this example K=2). **B** After the calculating the center of gravity for each cluster, the division of the data set is redefined. This process is done iteratively until convergence is reached. The arrow indicates an image point, that is still wrongly assigned after the second iteration.

calculated for both subsets (stars), to define the new points for the division into two, improved clusters. As it can be seen, the image, indicated by the arrow, is still assigned to the "wrong" cluster. A third iteration would be needed to successfully divide the image set into two clusters. In general, the procedure will continue until the division converges. Similar to the hierarchical classification above, the number of factors that should be taken into account has to be specified in advance.

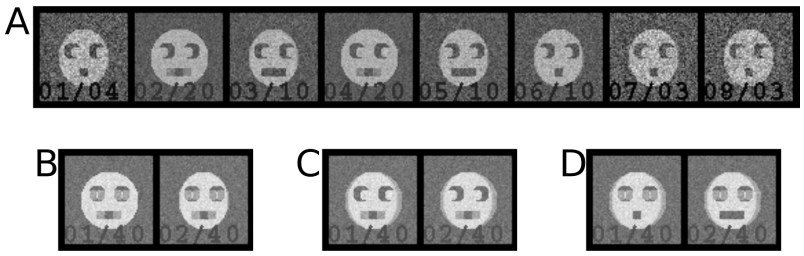


Figure 3.12: **A** Classification of the face data set by K-means into eight classes (K=8) under consideration of the first three factors. The initial characteristics of the faces are only restored partially, class 2 for example shows an undefined shape of the mouth. **B**-**D** Classification into two classes taking only one factor each into account. The three characteristics are clearly demonstrated.

As it becomes obvious, the initial choice of K defines the resulting clustering and also the random nature of this method has the effect that a repeated application with the same settings may lead to different results. Figure 3.12 demonstrates the application of classification by K-means to the example of the synthetic faces. Image **A** shows the classification with reference to the first three factors and K=8. This choice is obviously based on the knowledge, that eight the data set is constructed out of eight different classes. While some faces appear "noise-improved" and show clearly all three features (class 1, 3, 5, 6, 7 and 8), other classes such as 2 and 4 still have blurred characteristics with reference to the mouth.

Although the classification was less successful than with HAC, the K-means method can

be particularly useful to determine the effect of each factor by creating two classes (K=2) for each factor separately (images **B**-**D**).
A major disadvantage of this method is that the created clusters tend to be (hyper) spherically, which may lead to the splitting of elongated or more complex shaped clusters of the data set, which actually represent images that are closely related. [24, 30]

All in all both classification methods have their assets and drawbacks and it is useful to apply both methods and decide which one yields the best and most reasonable results. Furthermore one has to be always conscious about the way the classes are created to correctly interpret the resulting classes. And finally, in the case of clustering by K-means, various attempts should be made with varying starting K-values to find a compromise between a reasonable number of classes (to represent most of the variabilities within the objects) while keeping the number of images in each class high enough to maintain good contrast and SNR.

3.3 Resolution determination (2D) using FRC

An important factor for a correct interpretation of reconstructed EM images is the obtained resolution. Here a standard method is described for calculating the resolution, which is based on **fourier ring correlation** (FRC):

The idea is to split the image set of an EM reconstruction into two data sets and to compare the resulting sub-averages by calculating the spatial frequency at which both reconstructions can be considered to be consistent. The comparison is done by a Fourier transformation of both images, followed by cross correlating different corresponding rings in frequency space (therefore the name Fourier ring correlation)[3]. The result of this calculation can be visualized by plotting the cross correlation coefficients against the corresponding rings, which are equivalent to spatial frequencies. This curve is than evaluated by certain criteria to determine the resolution. The widest used criterion is to find the spatial frequency where the FRC falls below 0.5. Up to this frequency the initial reconstruction is said to be resolved. [49]

In Figure 3.13, a global class average of an example F-actin reconstruction is used to demonstrate the resolution determination, which was done in this work using the SPIDER operation *FRC*. The image set of 97 images was randomly divided into two equal sets and the corresponding subaverages were calculated (**A**). The FRC operation was then applied and the resulting curve plotted (**B**). In this example, the image had a resolution of 31.04 Å. However, the resulting value is dependent on the separation of the

[3] In three dimension, shells instead of rings are compared and the method is than called Fourier shell correlation, FSC

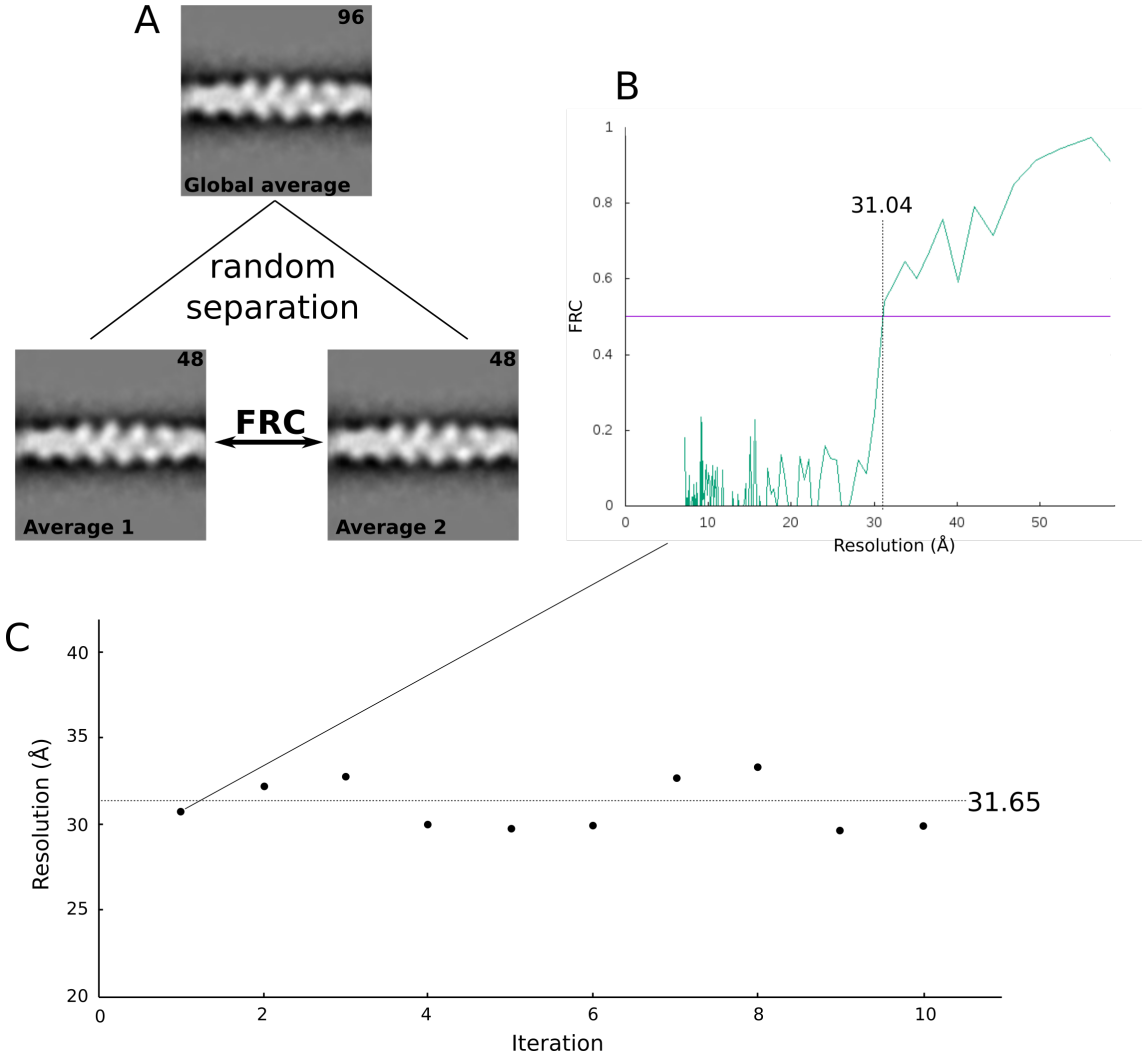


Figure 3.13: Principle of resolution determination of EM reconstructions using FRC on the example of a F-actin reconstruction: **A** The image data set of 97 images was split randomly into two subsets of equal size. These subsets were then compared in Fourier space by cross correlating corresponding rings in frequency space. **B** Up to the frequency where the correlation coefficient falls below 0.5, the images were considered to be consistent. **C** Since the splitting was randomly, different rounds of FRC calculation would lead do small variations. Therefore, 10 iterations were calculated and the average-value represented the final resolution. The average shown here had a resolution of 31.65 Å.

data set, especially when the data set is small. Therefore, the resolution was determined statistically by calculating the average of various FRC-applications (with randomly chosen images for the divided data sets). In the example here the average resolution after 10 iterations was 31.65 Å.

3.4 Software for image processing

There is wide variety of software to be used for single particle analysis of images produced by the TEM. One of the earliest approaches was made by Joachim Frank and his group in 1978 with the terminal based SPIDER suite (System for Processing Image Data from Electron microscopy and Related fields) [44] and is constantly developed and maintained by mainly ArDean Leith and Tanvir Shaikh. Around a decade later, the software EMAN (Electron Micrograph ANalysis) was developed by Steven J. Ludtke et al. [50], providing a more user friendly interface. With the vast improvements on cryo-EM and fast growing computer technologies, softwares concentrating on the analysis of cryo-EM data were developed, such as Relion by Sjors Scheres [51] and Xmipp by Jose Miguel de la Rosa Trevin [52], which also implemented new alignment and classification algorithms in order to speed up the process or improve the outcomes for specific applications (for example the analysis of very heterogeneous data sets). These programs are just a small part of all available softwares in this field and due to the fact that they were usually developed in small groups led to a large number of different file formats. An approach to combine the strength of different algorithms of these programs into one framework was made by Jose Miguel de la Rosa Trevin with the development of Scipion [53]. Scipion makes it possible to apply workflows from various programs on one data set without having the trouble of file conversion. In this thesis, mainly two softwares were used. For the particle picking, EMAN2 comprised an optimal mix of user friendliness, graphical interface and results. The following image processing was done with SPIDER, as it allows for a high degree of customization.

3.4.1 SPIDER / SPIRE

SPIDER is based on single operations that can be combined to create large workflows. Figure 3.14 shows the terminal window of SPIDER, where an operation or workflow can be entered. Apart from operations for the single particle analysis such as alignment or classification, a large variety of other operations are available to modify and analyze the images. With these additional operations, it is very convenient to create and apply masks, average images, rotate or shift them, interpolate images or apply various filters. Furthermore, operations that e.g. allow for the implementation of do- or if-loops and

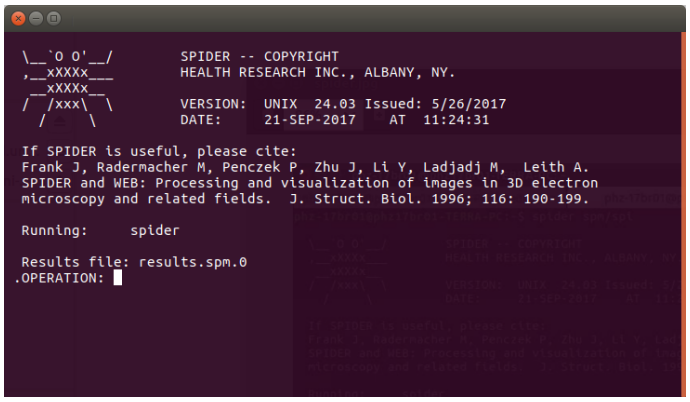


Figure 3.14: Terminal based SPIDER software for single particle analysis.

variable declaration as well as math-operations are available. In total, there are almost 500 different operations present that can be combined into a series of commands creating a custom procedure for the desired task.
Although the creation of such procedures and invoking them into SPIDER is convenient, a graphic user interface would make the analysis more comfortable and user-friendly. For this purpose SPIRE (SPIDER Reconstruction Engine), a graphical user interface GUI for SPIDER was developed by the same group in the early 2000's [54]. The main window of the interface is shown in Figure 3.15. As an example a simple cropping-operation is shown. First, the SPIDER script to crop images needed to be written by declaring variables and combining the corresponding operations. This script can then be implemented into a SPIRE dialogue / folder, in this case in "Micrographs", as shown in the Figure 3.15 and 3.16. All parameters can be adjusted ("Edit", Figure 3.17) before the script is executed or saved for later. Once a complete workflow with various different scripts is established the grouping of all scripts into different dialogs can be saved into one configuration file. This file, containing references to all contained scripts, can then be used to apply the procedure to different data sets.
The strength of SPIDER over other SPA softwares is the high degree of possible customization. While many programs provide only the user interface and already implemented "ready to apply", SPIDER procedures can be written from scratch[4] and therefore customized to match the purpose of the analysis.

[4]Of course, extensive work was and is done by the developers to write the underlying algorithms for each operation.

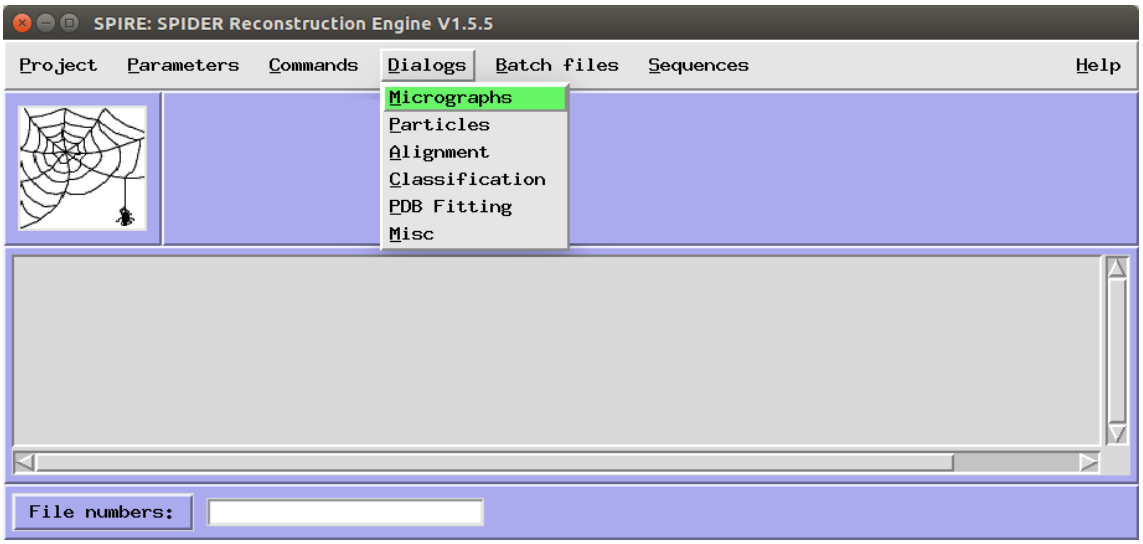


Figure 3.15: SPIRE, the GUI for SPIDER. Scripts based on SPIDER can be loaded into SPIRE and executed interactively. All scrips can be organized in dialogs.

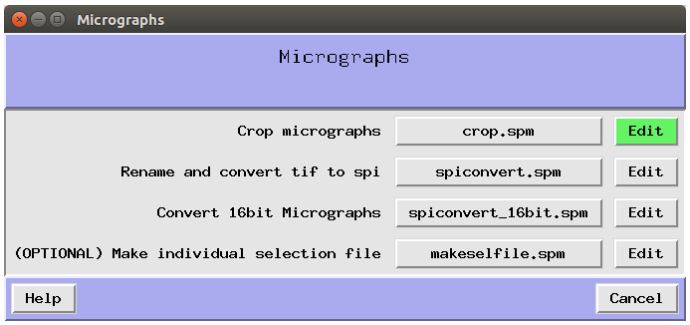


Figure 3.16: SPIRE, the GUI for SPIDER. Scripts in the dialogs can be either executed directly or edited for the setting of parameters.

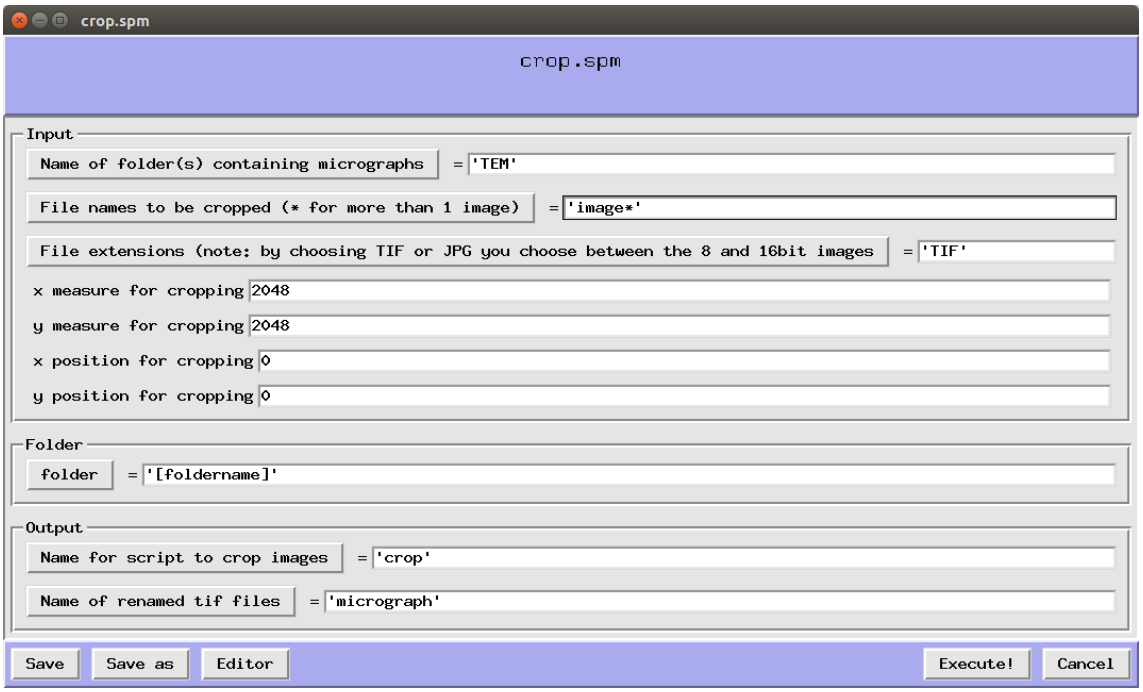


Figure 3.17: SPIRE, the GUI for SPIDER. The values for variables defined in the scripts can be entered here and the procedure executed. This way of entering the parameters is much more convenient than via the terminal-window.

4 Crystal structure: Modification and Projection matching

For the purpose of interpretation of EM class averages as a result of the single particle analysis, high resolution crystal structures can be used. However, not always corresponding crystal structures exist of the studied molecule and if they exist, they often comprise only parts of the molecule. Therefore, one has to choose carefully a model that can be used directly or with modifications adjusted to the requirements, including the combination of different crystal structures.
The following section describes shortly[1] the handling of crystal structure files, the possible ways to modify them and the actual process of using them to analyze electron microscopic data, called projection matching:

[1] More detailed descriptions will follow in the methods section of the corresponding chapters.

4.1 Modification

There are several ways to modify a crystal structure or to align it with other models. In this thesis, crystal structures were obtained in the PDB file format from the RCSB Protein Data Bank [55], a single worldwide archive of structural data of biological macromolecules, and imaged using Jmol [56]. A PDB-file contains information about the coordinates of each single atom as well as the assignment of atoms to various units such as amino acids and further to bigger complexes such as helical or sheet substructures. As a standard in the work presented here, the ribbon presentation of a structure was used which displays alpha-helices, beta-sheets and their connections. In addition to the spatial coordinates of atoms, different structural groups are separated in chains (for example the converter domain of myosin or a single G-actin unit of filamentous F-actin) and can be displayed in different colors. While certain chain-assignments are already contained in the PDB-file, they can be easily modified. With the background knowledge of which parts of the amino-acid sequence of a molecule represent structural units, new assignments can be implemented and thus various structural parts visualized in different colors. One way to do so is by using the software DeepView Swiss-PDBViewer [57].
In order to combine / align different crystal structures, the FATCAT (Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists) alignment method [58] can be used. This approach for protein structure comparison allows for the pairwise alignment of two complete structures or only parts of them. Furthermore, two alignment models are available, rigid and flexible, where the rigid comparison does not allow for any structural changes within the crystal models. The alignment with FATCAT results in a large number of information. One important piece of information is the P-value (Probability-value), quantifying the detected significance of structural similarity. The P-value is smaller, the more similar the two structures are. Two identical structures would have a P-value of 0.00 and up to a value of 0.05 the structures are considered to be significantly similar. The resulting aligned structures can be then combined and / or modified using SPDB Viewer or Jmol.

4.2 Projection matching

Once an appropriate crystal structure has been created based on the method described above, it can be used for EM interpretation by crystal structure projection matching, based on the idea published by Penczek PA et al. [59] and applied using developed script based on the SPIRE / SPIDER software [24, 44, 60]:

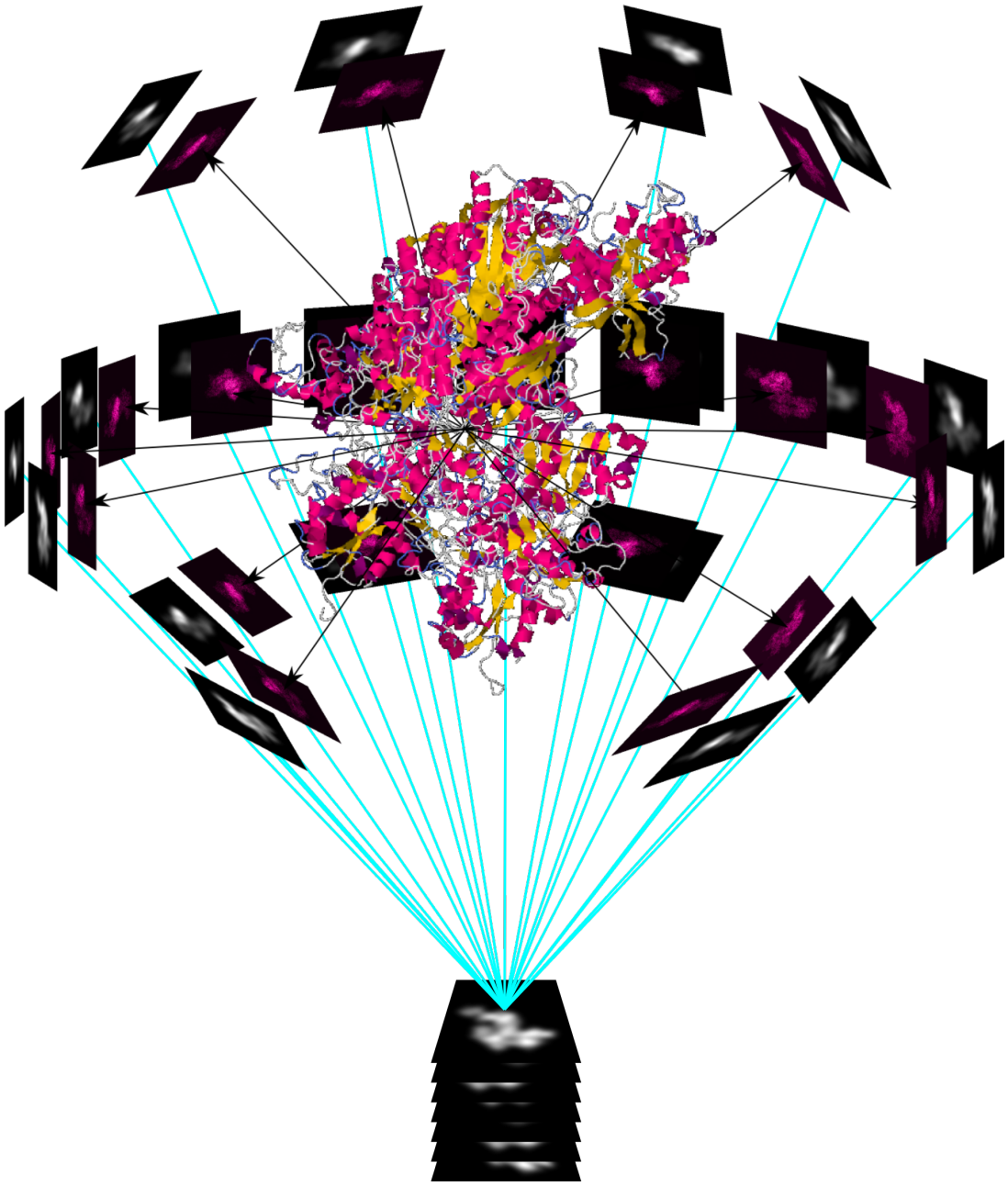


Figure 4.1: Crystal structure projection matching, Step 1: From a crystal structure, many 2D projections are calculated, low-pass filtered to simulate negative stain EM images and combined into one stack.

First, to recreate the projection views that TEM images yield, the PDB-file was converted into a file format that could be read by SPIDER using the operation *CP FROM PDB*. Additionally, a file was created that contained a list of combinations of the three Eulerian angles quasi-evenly spaced on the surface of a sphere (operation *VO EA*). Using this document-file, the operation *PJ 3Q* computes projection views of the crystal structure. Depending on the settings of the VO EA command, the density of angular distribution on the sphere could be controlled as well as the restrictions to for example only one changing Eulerian angle (i.e. rotation around one axis). Figure 4.1 illustrates the creation of 2D projections on a sphere. Each of these projection views was then converted into images simulating negatively stained EM averages by applying a low pass filter (operation *FQ*). The last step before the actual fitting process was to scale the simulated images to match the scale of the EM average and combine all projection views in one stack.

Each image of the created stack of simulated EM images was then used as a reference and aligned with the real EM image (operation *AP SHC*), generating sets of alignment parameter for each projection view (shift, rotation and mirroring). In Figure 4.2 **A** three example views from the stack are shown to illustrate the workflow. After applying the alignment parameter to the simulated EM images (operation *RT SF M*), the cross correlation value for each best fitting projection view with the EM average was calculated (operation *CC C*) and plotted in a graph (**B**). The model with the highest ccc-score was then considered to be the best fitting model and could be overlaid with the real EM average in order to on the one hand evaluate the fit and on the other hand identify intensities produced by the EM. In the example here the actin filament and parts of the myosin were covered by the crystal structure, leaving uncovered parts open for further interpretation (in particular of the myosin), e.g. by using a different crystal structure.

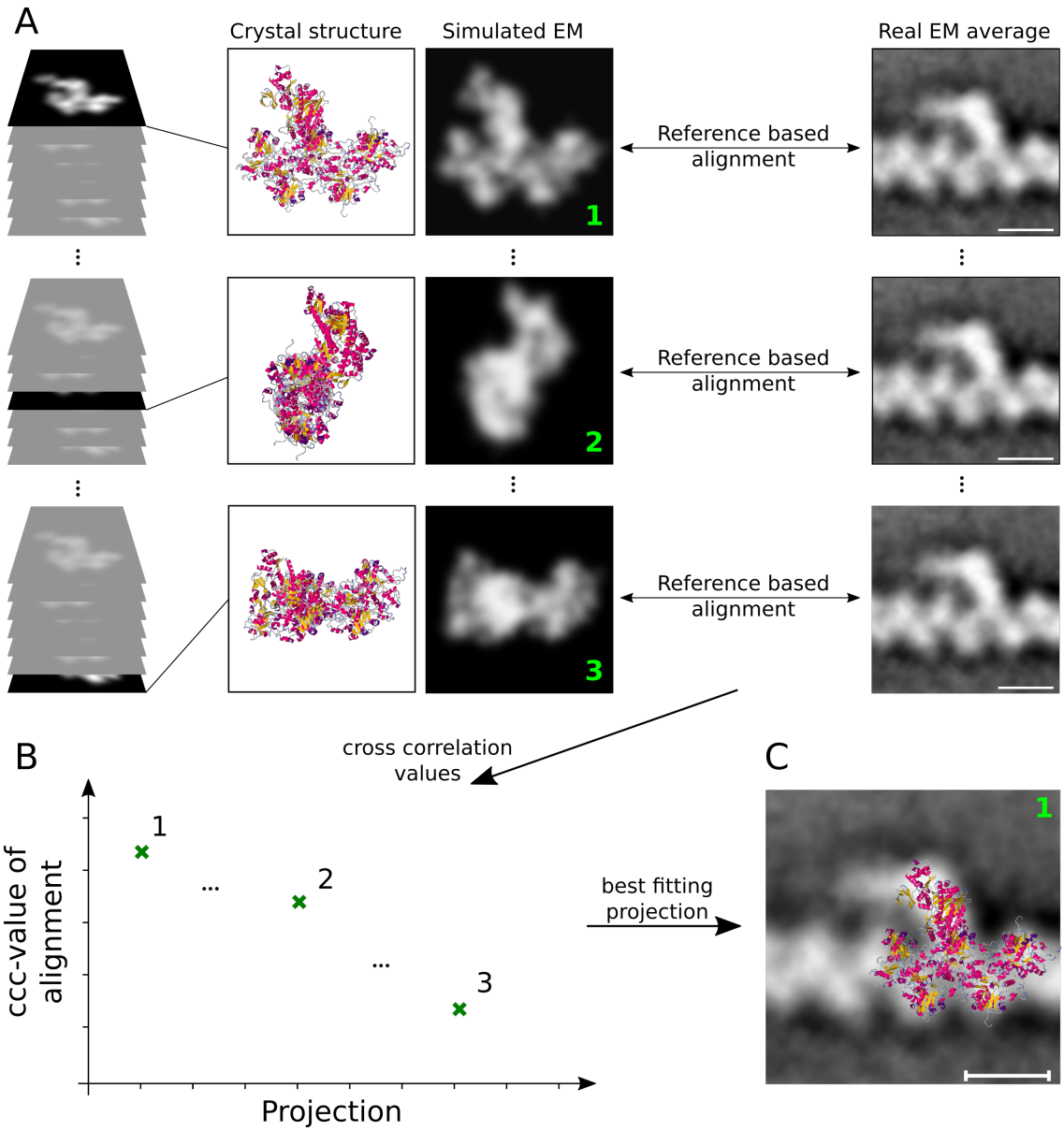


Figure 4.2: Crystal structure projection matching, Step 2: **A** Each image from the stack of low-pass filtered 2D projections of a crystal structure is aligned with the EM average and cross correlated to quantify the fit. **B** All cross correlation coefficients are then normalized and plotted. **C** The best fitting projection in its optimized orientation and with the scale adjusted can be overlaid onto the EM average to validate the fit and to identify the various structural parts seen in the EM. Scale bar: 5 nm

Part II

Results

5 Myosin VI

In human cells cytoskeletal motor proteins move along microtubules and actin filaments to generate complex cellular functions. The ability to coordinate the timing of motor protein activation lies at the center of a wide range of cellular processes, including endocytosis, cell division, and longer range transport of cargo. The cytoskeletal motor myosin VI is involved in a lot of these motile processes, e.g. cancer cell migration, and is the only myosin in the human genome shown to move towards the minus end of actin filaments [61]. This reverse directionality enables myosin VI to carry out highly specific tasks in the cell such as directed transport of cargo away from the cell surface. Myosin VI is also critically involved in endocytosis and functions that do not require processive motility, for instance maintenance of the cellular structure or the formation of stereocilia [62–64]. Apart from that, the expression of myosin VI is dramatically upregulated in carcinoma cells and is therefore used as an early marker of the cancer development and aggressiveness [65–69]. This aggressiveness is directly related to the movement behavior, cytoskeleton redistribution, and relocation of focal adhesions of migrating cells [70,71], which all have been shown to be influenced by localized calcium transients [72, 73]. On the basis that the mechanical function of myosin VI is directly linked to the properties of its calmodulin binding lever arm and that calmodulin is known to be a calcium sensor, we set out to study the effects of calcium and calmodulin on this myosin, in order to understand the underlying molecular mechanisms of how calcium affects the mechanical activation, cargo binding, and mobilization of myosin VI.

To elaborate these questions, we investigated full-length myosin VI monomers using negative stain electron microscopy combined with customized single particle image processing methods[1]. For the first time, intriguing calcium and calmodulin induced conformational changes of the myosin VI lever arm could be resolved[2]. The following chapter will present and discuss these rearrangements and what impact they have on the mechanical functionality of myosin VI, how they might affect its regulation and how they

[1] Unless stated otherwise, the use of "myosin VI" will from here on always refer to "FL myosin VI".

[2] The lever arm is crucial for the transduction of force generated in the motor domain.

could be able to activate and inactivate this unique molecular motor in the cell.

5.1 Introduction

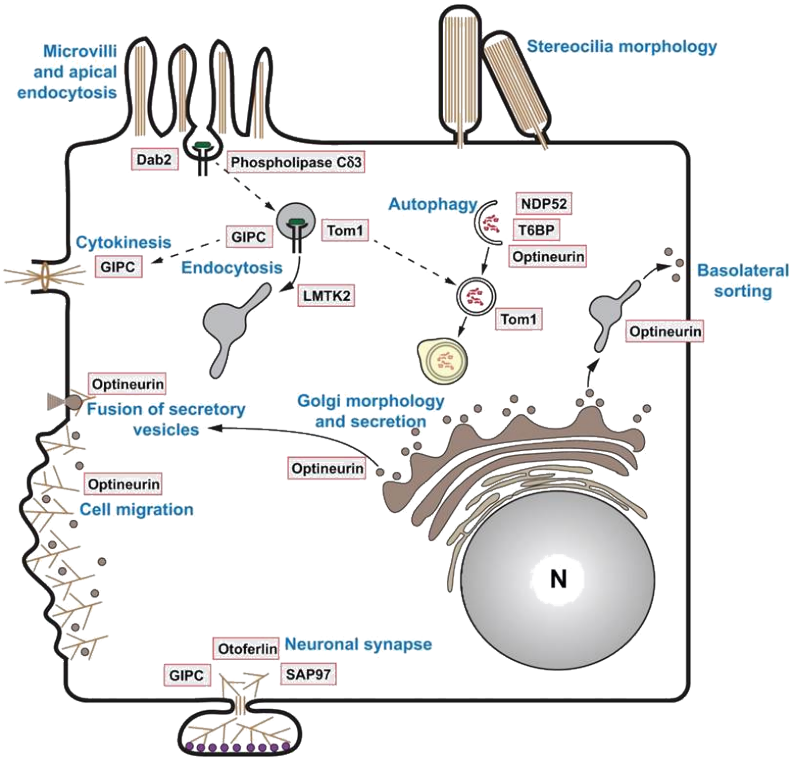


Figure 5.1: Illustration showing the wide range of cellular functions of myosin VI. It is involved in membrane trafficking pathways, secretory pathway and functions as an actin anchor. Furthermore, myosin VI is involved in cytokinesis and directed cell migration. Figure from [74]

Cytoskeletal motor proteins in human cells are essentially involved in a broad range of cellular functions such as muscle contraction, transport of cargo or endo- an exocytosis. In particular, the coordination and timing of the activation and inactivation of motor proteins is fundamental for many cellular processes. This chapter focuses on the unconventional motor protein myosin VI. This myosin class is like all myosin associated with actin filaments. Myosin VI however is different from all other classes as it is the only known myosin to date in the human genome that is able to perform a reversed, minus-end directed movement [61]. Since actin filaments in the cell are highly organized in a way that their plus ends are close to the plasma membrane, while the minus ends point towards the center of the cell [75], the minus-end directed motility of myosin VI is thought to provide unique properties. Among other functions (see Figure 5.1 [74]), this reversed transport enables myosin VI to be critically involved in endocytosis [64], (short-range) transport of cargo such as vesicles away from the cell surface and translocation of membrane to the leading edge in migrating cells [65]. Myosin VI as an actin anchor is also highly involved in a wide range of functions that do not require motility, such as maintenance of the cellular structure (i.e. Golgi morphology) [64], facilitation

of the secretory vesicle fusion at the plasma membrane [76, 77] and the formation of stereocilia in inner ear cells [62–64].

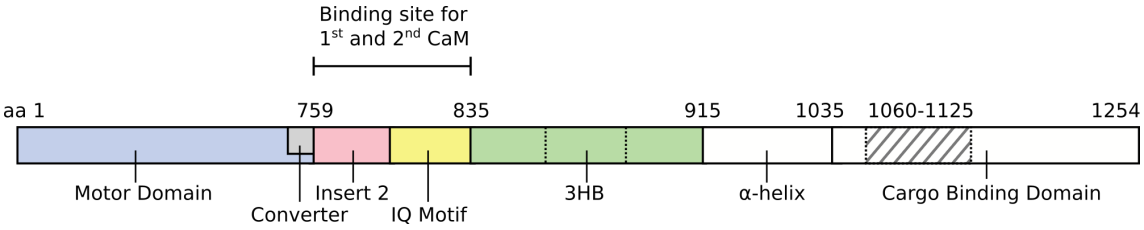


Figure 5.2: Sequence of the 1265 aa long FL myosin VI. Indicated are the motor domain (blue), converter (grey), unique insert 2 (red), IQ motif (yellow), three-helix-bundle (green) and a single α-helix (SAH) and cargo binding domain (white). The insert 2 and IQ motif are known binding sites for calmodulin.

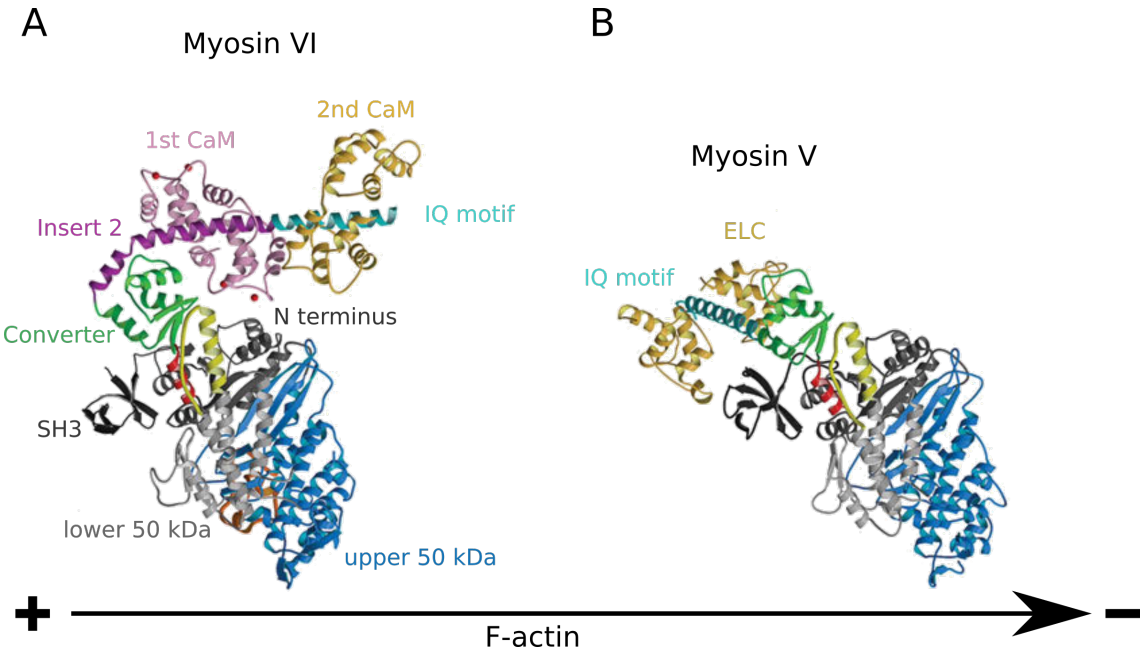


Figure 5.3: Structures of FL myosin VI and V together with a simplified actin filament indicating the actin-binding behavior of both myosin. **A** Crystal structure as ribbon diagram representation of nucleotide-free myosin VI. Apart from the main compartments of the motor domain (lower and upper 50 kDa domain, N terminus) and the converter domain, the unique insert 2 is highlighted. Furthermore, the IQ motif and two bound calmodulins (with the 1st CaM being in a calcium-bound and the second in the apo state) are shown. **B** Myosin V in a rigor-like state. Apart from the two unique inserts myosin V has a very similar overall structure. The orientation of the lever arm however is significantly different, pointing towards the opposite direction. The angle of its lever arm is different by approx. 150°. Modified from [78]

The general structure of the full-length, 1265 amino acid (aa) long myosin VI is similar to other myosin classes, being comprised of a highly conserved motor domain that identifies this molecule as myosin (see Figure 5.2). In this domain the nucleotide binding as well as the binding to actin filaments takes place. In contrast to plus-end directed myosin motors, myosin VI contains two unique inserts. Insert 1, being located close to

the nucleotide binding pocket, affects the nucleotide association rate. The second insert follows the converter domain and is stabilized by a calmodulin [78]. This insert 2 is thought to be crucial for the ability of myosin VI to move backwards, as it rotates into the opposite direction compared to similar calmodulin binding structural parts of any other myosin class (see comparison with myosin V in Figure 5.3 [61, 78, 79]. Adjacent to the insert 2 a second calmodulin can bind reversibly to a canonical IQ-motif [80]. Continuing the structure of myosin VI, the molecule is constructed of a semi flexible three-helix-bundle (3HB) that is able to unfold upon certain conditions such as dimerization as reported by [81] or upon changing calcium concentrations as proposed in this thesis. The remaining tail of the molecule comprises a single α-helix (SAH) [82] and a cargo binding domain. The SAH is flexible and contains a sequence (aa 1060-1125) that binds to the second calcium-free (apo) calmodulin attached to the IQ-motif [83]. Furthermore, the SAH domain is thought to be crucial for the formation of myosin VI dimers, although natural dimerization could not be shown yet. Apart from artificially fabricated dimers [84–88] it is not known whether this molecular motor works as monomer, dimer or oligomer or is able to switch between them. As a monomer, the conformation of the tail of myosin VI might be regulated by cargo binding and other unknown mechanisms and can exists in either a compact, folded shape or in an extended, unfolded state, depending on certain conditions in the cell [64]. This change of conformation is also discussed to be related to the activation or inactivation of myosin VI [89, 90]. However, the underlying mechanisms are not fully understood.
Given the importance of myosin VI in the functions mentioned above, it also has crucial responsibilities in human diseases such as deafness and cancer [64]. In particular, its role in the transport of membrane in migratory cells and the connected radical upregulation in migrating carcinoma cells (i.e. in breast, lung and prostate), make myosin VI an early marker for the aggressiveness of cancer and further an indicator for its development [65–69]. It has been shown that the redistribution of the cytoskeleton as well as the direction of the movement of migrating cells is connected to fluctuations in calcium concentrations in the cell [70, 72, 73]. Moreover, calcium is known to alter the kinetics and mechanical properties of myosin VI, shown by changing motility rates [91, 92]. As mentioned above, the lever arm of myosin VI is stabilized by calmodulin, which on the other hand is known to be calcium sensitive. And as the lever arm of myosin is crucial for the transduction of force generated in the motor domain, it seemed reasonable that calmodulin, influenced by calcium, plays an important role in the mechanical activation, regulation and mobilization of myosin VI.

This chapter will therefore examine and study the effects of calcium and calmodulin on the structural properties of myosin VI to better understand the underlying processes.

The focus will lie on the development of single particle image processing methods of negatively stained electron micrographs to analyze single nucleotide-free myosin VI molecules in the absence of actin and in four different conditions with respect to calmodulin and calcium concentration. These methods were mainly based on the SPIDER image processing system. In the discussion section, the results obtained by the application of these methods will be put into context with other, recently reported studies on myosin VI [83] and a model introduced that proposes a unique mechanism of how myosin VI might regulated by calcium.

5.2 Results

The following section will first present the effects of calcium and calmodulin on the general structure of myosin VI discovered by EM and SPA including a statistical evaluation of the found conformations. Following this, the focus will lie on a detailed examination of the calmodulin binding region of myosin VI.

5.2.1 Effect of calcium and calmodulin on myosin VI: Structure

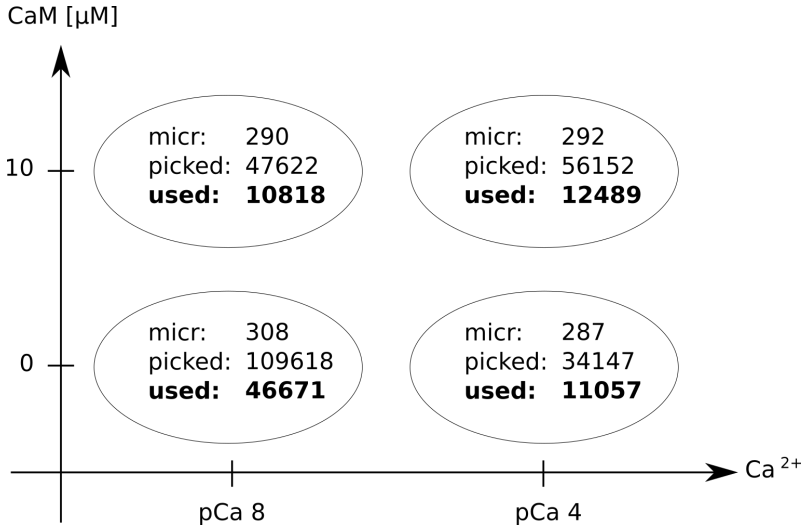


Figure 5.4: The number of recorded micrographs ("micr"), picked particles ("picked") and actually used images for the single particle analysis ("used") in all four conditions with respect to calcium and calmodulin concentrations.

Nucleotide-free myosin VI molecules were prepared as described in section 5.5.1 and a total of 247,539 images of myosin VI molecules in four different conditions with respect to calcium and calmodulin concentrations close to the physiological range (CaM: 0-10 µM, Ca^{2+}: pCa 8-pCa 4) were obtained. Figure 5.4 summarizes the number of recorded micrographs as well as the number of picked particles and actually used images for the single particle analysis. Each of the four data sets was separately analyzed by the single particle analysis methods explained in detail in section 5.5.2. One representative original micrograph for each conditions as well as summarized versions of the final classes are given in the supplementary materials (section 5.6): low calcium / low calmodulin: supplementary figures 5.1-5.7; high calcium / low calmodulin: supplementary figures 5.8-5.14; low calcium / high calmodulin: supplementary figures 5.15-5.17; high calcium / high calmodulin: supplementary figures 5.18-5.21.

As a first result of this analysis, in all four conditions myosin VI molecules were found to be absorbed to the carbon-coated EM grid in two main orientations. The so

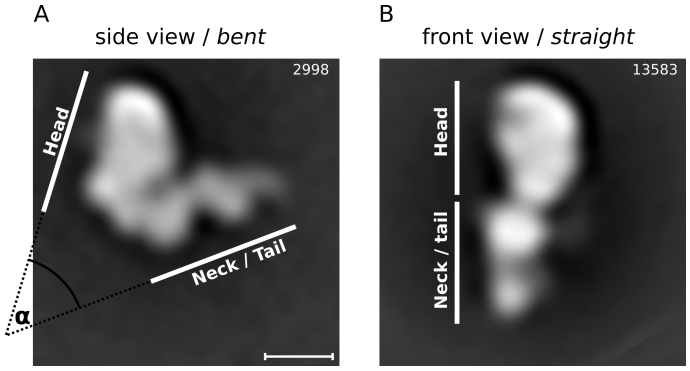


Figure 5.5: Full-length myosin VI molecules adhered to the EM grid in two preferred orientations, bent (A) and straight (B), also called side view and front view, respectively. In the side view, the direction of the lever arm is related to the motor domain by $\alpha \approx 53°$, while in the front view both are in line. The global averages shown here contain 2998 images for the side view and 13583 images for the front view. Scale bar: 5 nm

called *bent* conformation (Figure 5.5 **A**) shows the myosin VI molecules in a side view with the catalytic domain (head) and neck / tail domain at an angle of $\alpha \approx 53°$. And second, molecules were adhered to the grid in a *straight* conformation (Figure 5.5 **B**), most likely showing myosin VI in a front view with the head and neck domains in line.

The identification of the catalytic domain and neck / tail region was done by crystal structure projection matching as described in section 5.5.2.3. With the same method, further internal structural details of the EM averages could be determined with this method. Figure 5.6 shows the result of the analysis of the *bent* conformation using a X-ray model of nucleotide-free myosin VI composed of two crystal structures (PDB: 2BKI and 3GN4, for more details see section 5.5.2.3). The search of 799 low-pass filtered 2D projections of this model led to an optimized projection with an excellent fit to the EM average (ccc=0.95). The normalized and ranked cross correlation coefficients (ccc) of all projections can be seen in Figure 5.6 **A** together with three representative projections of the model (x1-x3; ccc(x1)=0.95, ccc(x2)=0.82, ccc(x3)=0.73). By overlaying the optimized orientation onto the EM average of the *bent* conformation (Figure 5.6 **B**), many structural parts could be identified: the catalytic domain with the SH3 domain (blue), converter (grey), 1st and 2nd CaM (red and yellow respectively), alpha-helix to which the CaM's bind (light green) and the first helix of the 3 helix bundle (3HB) (dark green). Following the orientation of the crystal structure, this view is also called *side view*.

The straight conformation in contrast could not be directly explained using the model. The best fitting of the 799 low-pass filtered projections of the 2BKI-3GN4 structure (Figure 5.7 **A**) was obtained by using only the catalytic domain for the alignment and showed great agreement with the EM average (ccc(x1)=0.95). Note the clear visibility of the cleft between the upper and lower 50 kDa domains (green arrow). In this projection on the other hand, the lever arm which lied underneath the catalytic domain (Figure 5.7 **A**,

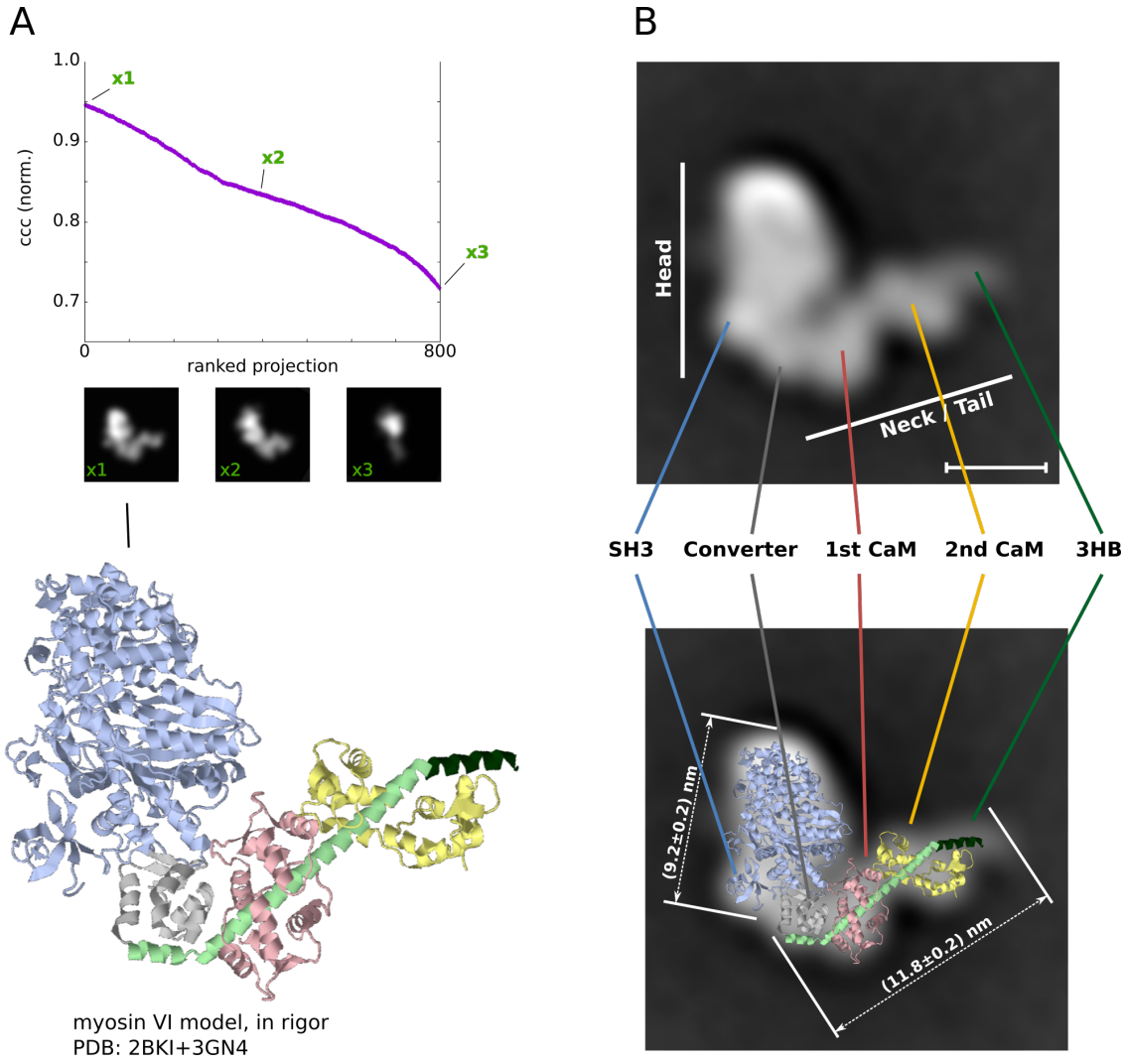


Figure 5.6: Analysis of the bent conformation of myosin VI by crystal structure projection matching: **A** Normalized and ranked ccc-values of 799 low pass filtered 2D projections of a myosin VI model (PDB: 2BKI + 3GN4) together with three representative projection views (best, medium and worst fit). The crystal structure is shown in its best fitting orientation. **B** By overlaying the model in its optimized projection, various structural parts of myosin VI can be identified in the EM average: Catalytic domain with SH3 and converter, 1st and 2nd CaM and the three-helix-bundle. Scale bar: 5 nm

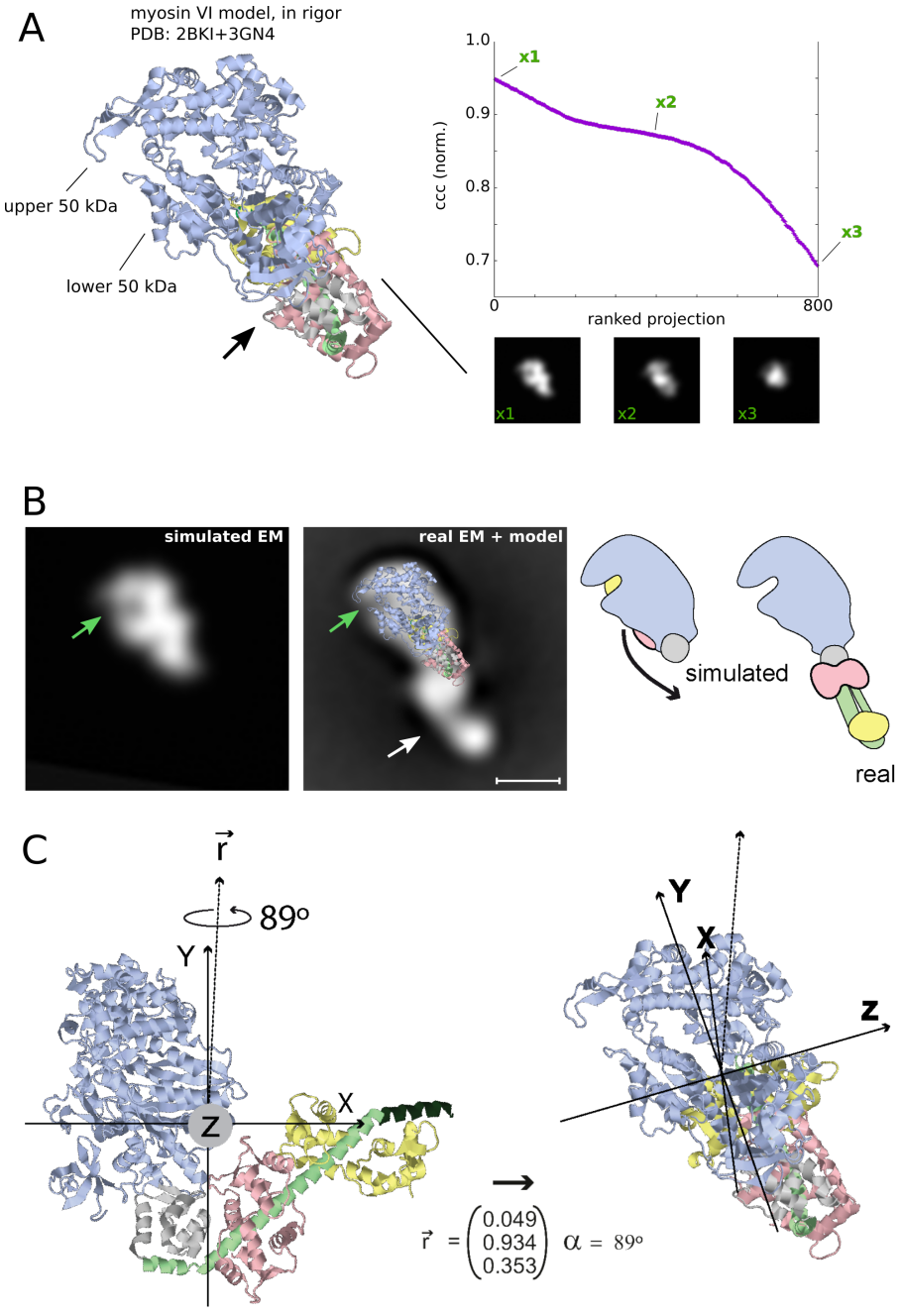


Figure 5.7: Analysis of the straight conformation of myosin VI by crystal structure projection matching using the same model as for the bent conformation (PDB: 2BKI+3GN4): **A** The best fitting orientation was found by using only the catalytic domain for the analysis. Shown here is the structure in its optimized orientation together with the normalized and ranked ccc-values for this analysis and three representative low-pass filtered 2D projections. The lever arm is in this view located behind the converter and motor domain (arrow). **B** Overlay of the model in its optimized orientation onto the EM average. The fit of the head domain identified e.g. the crevice between the upper and lower 50 kDa domains (green arrow), but left the lever arm unexplained (white arrow). The cartoon illustrates one possibility for the appearance of the lever arm in the EM data. **C** Relation between the two optimized orientations of the model for the bent and straight conformation by a rotation of $\alpha = 89°$ around an axis $\vec{r}$. Scale bar: 5 nm

black arrow), was inconsistent with the EM data. The average showed clearly an elongated neck, orientated in line with the catalytic domain (Figure 5.7 **B**, white arrow). The two optimized orientations for the *bent* and *straight* conformations were related by a rotation of $\alpha = 89°$ around an axis $\overrightarrow{r}$ (Figure 5.7 **C**), highlighting the position of the neck / tail and giving rise to the notation of *front view* for the straight conformation.

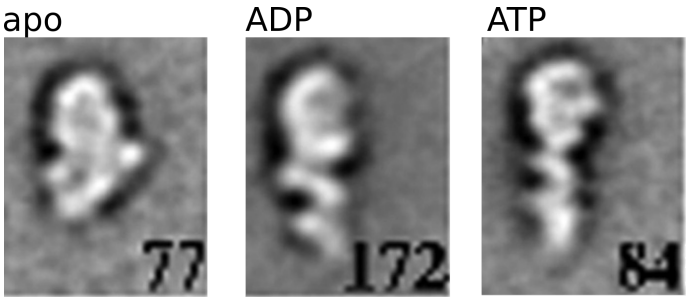


Figure 5.8: Representative negative stain EM class averages of studies on FL myosin VI in different nucleotide states by Song et. al [93]. Similar structures as the bent and straight conformation were found, however as an effect of nucleotide.

Electron microscopical studies on FL myosin VI by Song et. al [93] revealed shapes of myosin VI comparable to the *straight* conformation presented here. Figure 5.8 shows representative class averages from these studies on myosin VI without nucleotide (apo) as well as in the presence of ADP and ATP, indicating the unfolding / straightening of the calmodulin-binding lever arm upon increasing nucleotide concentrations.

To verify that the *straight* conformation discovered here was in contrast **not** an effect of nucleotide, a crystal structure of myosin VI in the presence of ADP (PDB: 4ANJ, see section 5.5.2.3 for more details) was used to analyze the data set as illustrated in Figure 5.9. **A** displays the normalized ccc-values for all 799 projections fitted to the catalytic domain, leading to the optimized fit of the ADP-myosin VI-model shown on the left side of the diagram (x1). At first sight the model in this orientation looked very similar to the EM average with a cross correlation coefficient of ccc(x1)=0.92 for the catalytic domain. The rest of the structure on the other hand comprises small, but crucial differences, shown by the overlay (Figure 5.9 **B**). Indicated with a white arrow, the point of emergence of the lever arm is inconsistent with the average. Furthermore, the orientations of the first calmodulins (black dotted lines: long axis in image-plane of 1st CaM of EM image, red dotted lines: long axis in image-plane of 1st CaM of model) are significantly different by $\alpha \approx 50°$ (in plane-angle). Mirroring the crystal structure on its long axis (image marked with an asterisk) would lead to the correct position of the lever arm and a better fit of the first calmodulin ($\beta \approx 20°$), the catalytic domain however did not agree with the EM-data anymore, in particular at the areas indicated by the red arrows.
As a negative control, the same analysis was applied to one of the class averages pub-

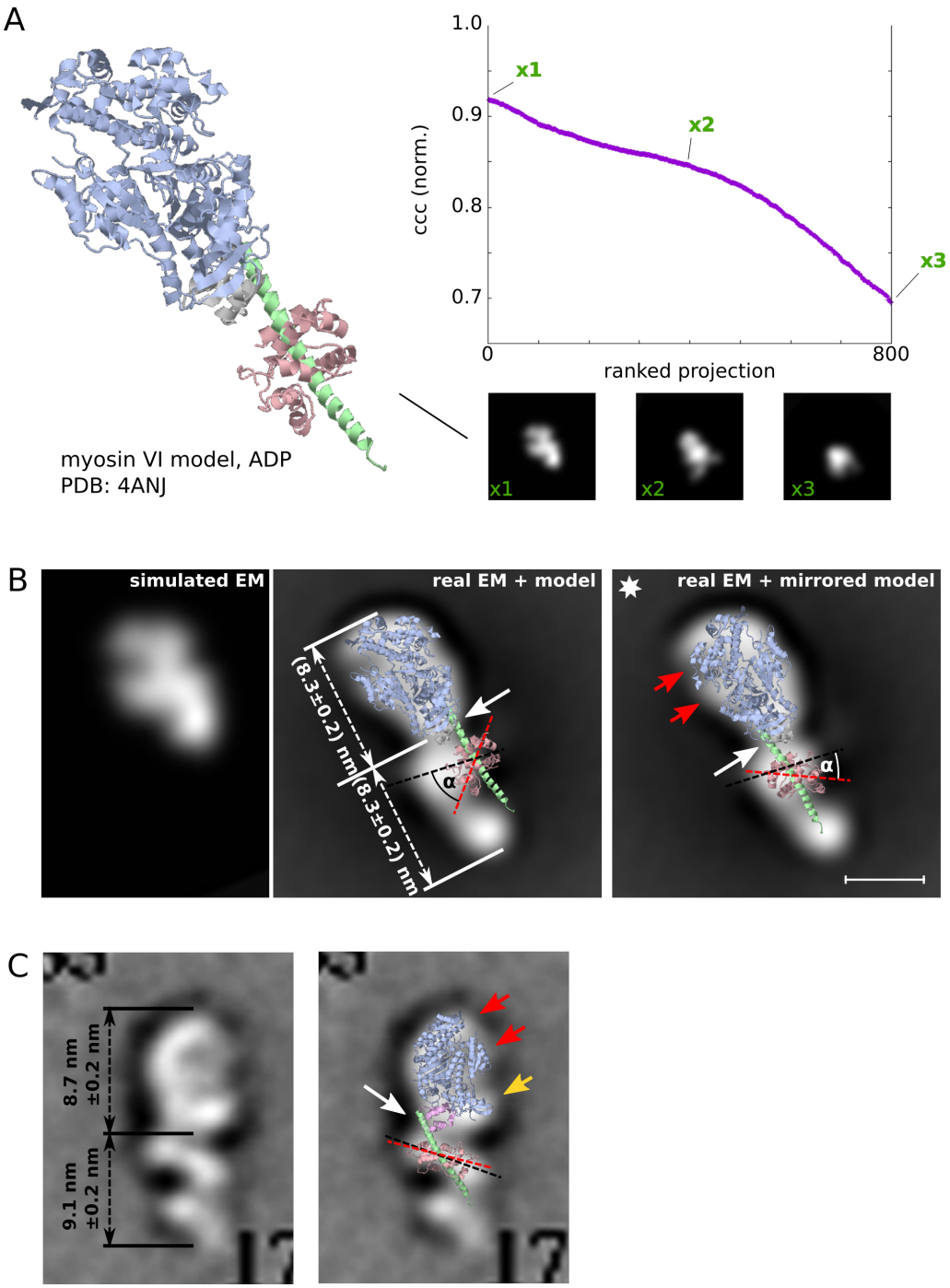


Figure 5.9: Analysis of the straight conformation of myosin VI by crystal structure projection matching using a model of myosin VI in the presence of ADP (PDB: 4ANJ): **A** The best fitting orientation was found by using only the catalytic domain for the analysis. Shown here is the structure in its optimized orientation together with the normalized and ranked ccc-values for this analysis and three representative low-pass filtered 2D projections. **B** The overlay of the optimized model onto the EM average revealed important differences. The angle of the first calmodulin was inconsistent with the EM data (by an angle alpha) as well as the point of emergence of the lever arm (arrow). Overlaying a mirrored version of the model improved the fit of the lever arm (incl. orientation of the CaM), but decreased the fit of the catalytic domain drastically (red arrows). **C** Using the EM data published by Song et. al for comparison revealed a great fit of the ADP-myosin VI structure as expected (50 kDa (red arrows), SH3 domain (yellow arrow), correct point of emergence of the lever arm and orientation of the calmodulin). While the general measures of head domain and lever arm were similar, the differences mentioned above confirmed that the straight conformation discovered here was not an effect of nucleotide. Scale bar: 5 nm

lished by Song et. al [93] (Figure 5.9 **C**). The measures for the catalytic domains of the molecule in both studies (**B** and **C**, (8.3±0.2) nm and (8.7±0.2) nm respectively) match when taking the errors into account. The lever arm in the other hand is approx. 1 nm longer in the studies by Song et. al. (**C**). Furthermore, the projection of the crystal structure in its optimized orientation showed a much better fit with the EM-intensity map. The catalytic domain (in particular the upper and lower 50 kDa domains (red arrow) and the SH3 domain (yellow arrow)), the point of emergence of the lever arm (white arrow) and the orientation of the first calmodulin (black dashed line) agree with the crystal structure, underlining that the elongated lever arm is in fact due to the presence of nucleotide as reported.
Based on this it can be excluded that the unfolded neck / tail domain seen in the *straight* conformation here was an effect of nucleotide or represents myosin VI in a prepower stroke. Instead, more likely the molecules were absorbed to the EM grid in a different binding mode and distorted by forcing the lever arm to unfold, caused by interactions with stain and surface charges.

This hypothesis could be further confirmed by a more detailed analysis of the lever arm itself. A closer look at the neck domain revealed additional masses on both sides, located on the right, left or both sides of the assumed position of the first calmodulin, as shown in Figure 5.10 and indicated by the white arrows. Furthermore, a second additional intensity was resolved close to the second calmodulin as marked by the white arrow-head, most visible in the low CaM / low Ca^{2+} condition. Interestingly, no such effect could be identified in the low CaM / high Ca^{2+} condition. One way to interpret these ”blobs” was by considering possible detachment of calmodulin lobes as it was reported previously by Terrak et al. [94] and Song et al. [93]. These studies however related the detachment to the presence of nucleotide, which was not the case in the samples analyzed in this thesis.
Another possible explanation of the additional mass is by exploring the flexibility of the neck domain. Images contained in the class average marked with the white star (this class contained enough images sufficient for the analysis) were analyzed by aligning the molecule with a mask covering only the catalytic domain. Shown in Figure 5.11 **A** are the first and last frames from the movie created using the method described in section 5.5.2.3 together with their outlines. This movie revealed that the appearance of the additional masses was in some cases correlated with the angle of the neck domain (with respect to the long axis of the head (blue)). The stronger the neck domain was bent, the more these masses became visible. As indicated by the cartoons below (**B**), these ”blobs” could be identified as the first and second calmodulin (red and yellow respectively), still being attached with both lobes. The flexibility not only arose from an in-plane rotation

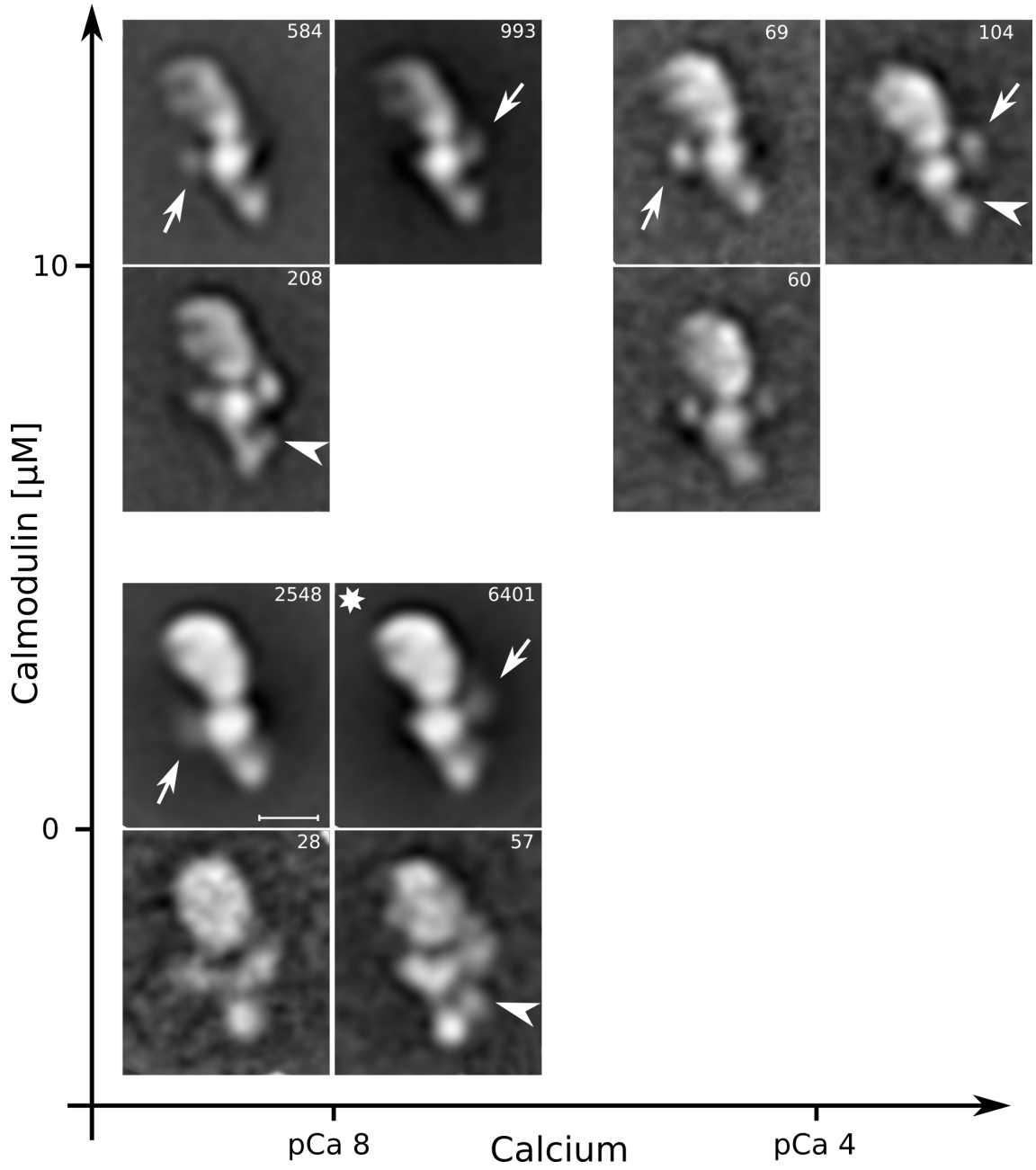


Figure 5.10: Analysis of the lever arm of the straight conformation revealed additional intensity blobs close to the first (arrow) and second calmodulin (arrowhead) in three conditions but not at low CaM / high calcium. The blobs might represent detached calmodulins as reported previously [94]. Upper right number: Number of images on average. Class average marked with a star used for further analysis (Figure 5.11). Scale bar: 5 nm

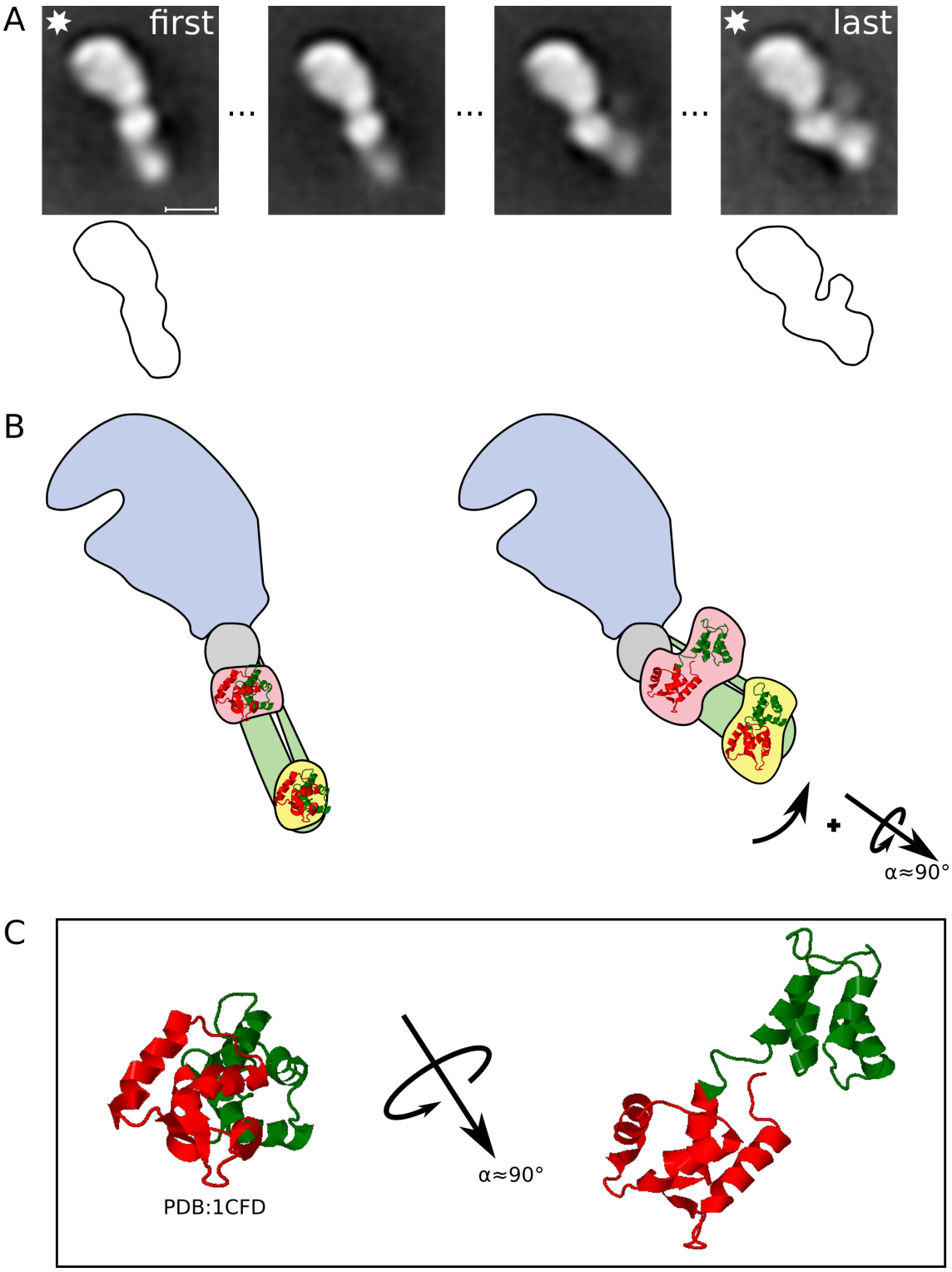


Figure 5.11: Analysis of the lever arm of the straight conformation by exploring its flexibility (**A**) revealed that the appearance of the additional masses was in some cases correlated with the angle between the neck and the long axis of the head. The stronger the neck was bent, the more the long axis of the calmodulins became visible, as illustrated in the cartoons (**B,C**). Scale bar: 5 nm

around a pivot point located in the area of the converter domain (grey) but also from a rotation around the long axis of the lever arm itself by approx. 90° as indicated. While in the first frame the two calmodulins are in a position where in the projection view their two lobes overlap, the rotation around the lever arm brings up the underlying lobe, making the calmodulins appear longer and explaining the additional intensity masses seen in the EM average. This effect is illustrated in Figure 5.11 **C**, using a calcium-free calmodulin crystal structure (PDB: 1CFD [16]). Although both interpretations seem to be quite reasonable, one still has to keep in mind that the appearance of the straight conformation in the first place was due to heavy deformations of the neck domain when myosin VI bound to the EM grid surface.

5.2.2 Effect of calcium and calmodulin on myosin VI: Statistics

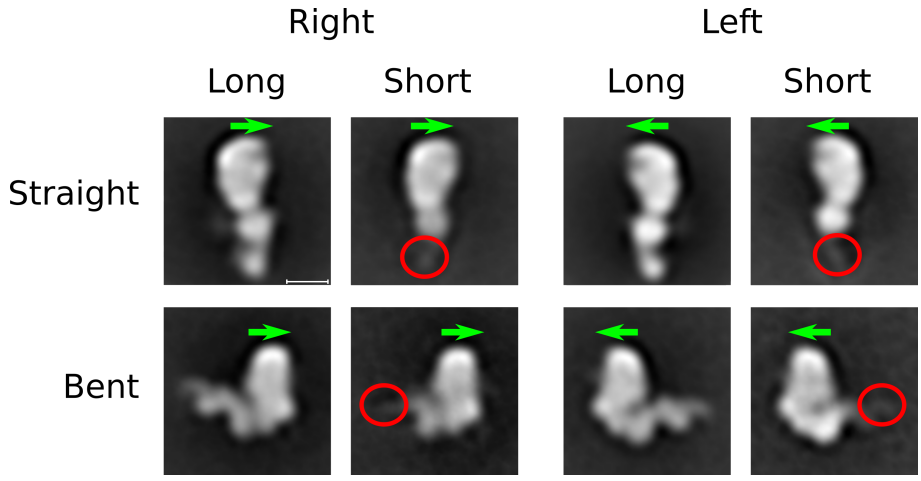


Figure 5.12: Myosin VI molecules were observed in eight different shapes, characterized by its sidedness (right,left; green arrow), conformation (straight,bent) and length of the neck domain (long, short; red circles indicate missing second calmodulin, resulting in short molecules). Scale bar: 5 nm

Apart from the main classification into *bent* and *straight*, myosin VI was present in a number of variations with respect to the length of the neck / tail domain (*long* / *short*) and mirror related orientations on the grid (mirror plane perpendicular to the EM grid / image plane), called *right* and *left*. This made a total number of eight different shapes, which are shown in Figure 5.12. The missing part of the *short* molecules in the lever arm could be identified as the second calmodulin, indicated here with red circles. Only the first calmodulin was resolved. The green arrows explain the "sidedness" of the molecules, which has been set arbitrary and orientated on the shape of the catalytic domain.

The figures 5.13 and 5.14 and Table 5.1 summarize the analysis on the statistics of these eight shapes in each of the four conditions with respect to calcium and calmodulin concentrations. Shown in Figure 5.13 are class averages together with the percentage

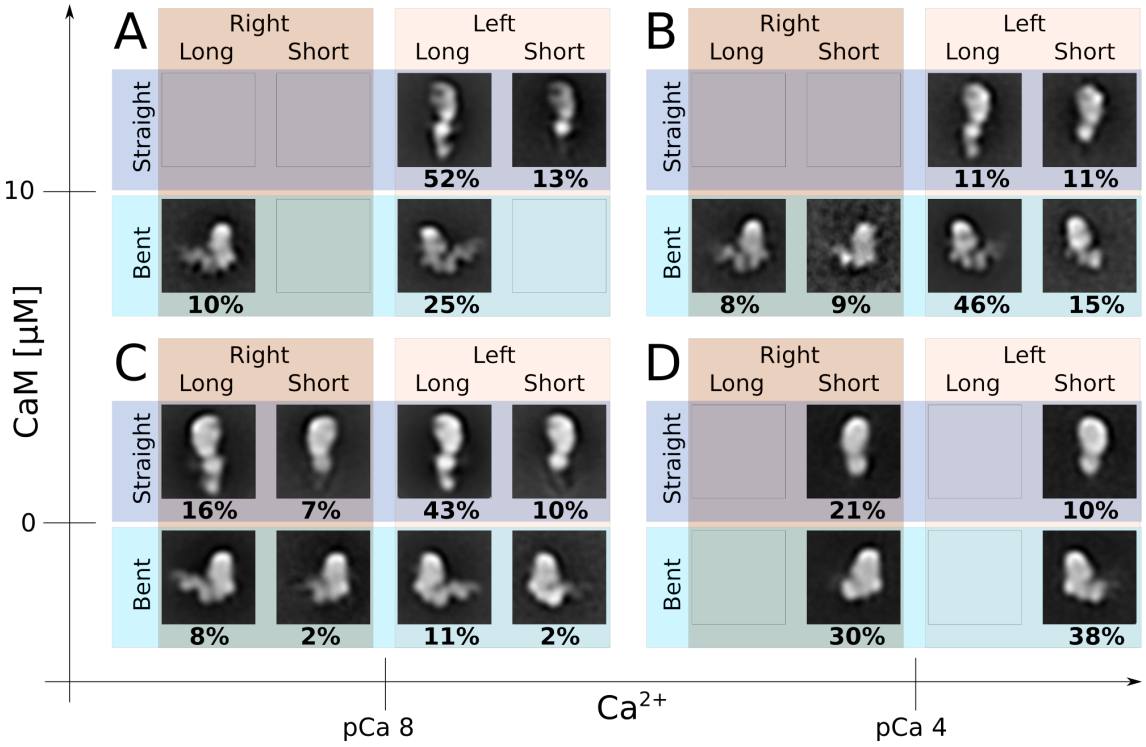


Figure 5.13: Statistics on the appearance of the eight shapes in dependence of calcium and calmodulin concentrations. Scale bar: Images are 20 nm wide.

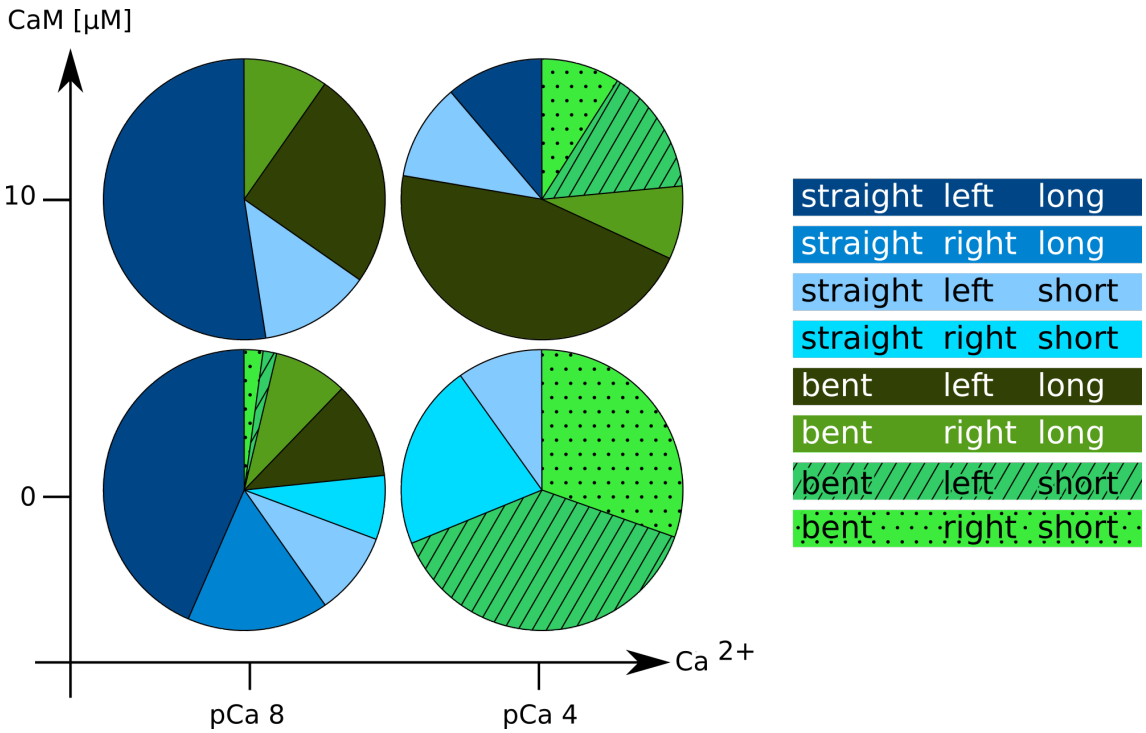


Figure 5.14: Absolute numbers of images found for each shape in all four conditions in the form of a piechart. Color code as indicated with straight molecules in blue shades and bent molecules in green shades.

CaM (µM) / Ca^{2+}	0 pCa 8	10 pCa 8	0 pCa 4	10 pCa 4
straight left long	20321	5677	0	1405
straight short long	75766	0	0	0
straight left short	4457	1380	1089	1374
straight right short	3424	0	2352	0
bent left long	5230	2720	0	5727
bent short long	3943	1041	0	1057
bent left short	749	0	4255	1813
bent right short	971	0	3361	1113

Table 5.1: Absolute numbers of images found for each shape in all four calcium / calmodulin conditions.

of their appearance in each case and Figure 5.14 presents this distribution in the form of a pie-chart (see Table 5.1 for corresponding image numbers). Interestingly, the number and ratio of in particular *long* and *short* molecules changed drastically. While at low calcium (pCa 8) and low calmodulin (Figure 5.13 **C**) all eight conformations were present, the increase of calcium to pCa 4 (but no increase of CaM) led to the total dissociation of the second calmodulin, resulting in significantly shorter molecules (**D**). No molecules with both calmodulins bound to the lever arm could be observed in this condition, while at the same time both the *straight* and *bent* as well as the *left* and *right* conformations were present. The adding of extra calmodulin however restored the binding of the second calmodulin to the lever arm and thus the appearance of *long* molecules in the high-calcium condition (**B**). At high calmodulin and low calcium (**A**) a decrease of *short* molecules could be observed, supporting the higher probability for the second CaM to be bound at this condition.

CaM (µM) / Ca^{2+}	0 pCa 8	10 pCa 8	0 pCa 4	10 pCa 4
straight	**77%**	65%	31%	**22%**
bent	**23%**	35%	69%	**78%**
long	79%	87%	-	66%
short	21%	13%	100%	34%
left	66%	90%	48%	83%
right	34%	10%	52%	17%

Table 5.2: Statistics on the shapes found in all four calcium / calmodulin conditions conditions. All eight shape characteristics were reduced to two parameters: straight / bent, long / short and left / right. The parameter long for example contains all images with two resolved calmodulins, independent of conformation or sidedness.

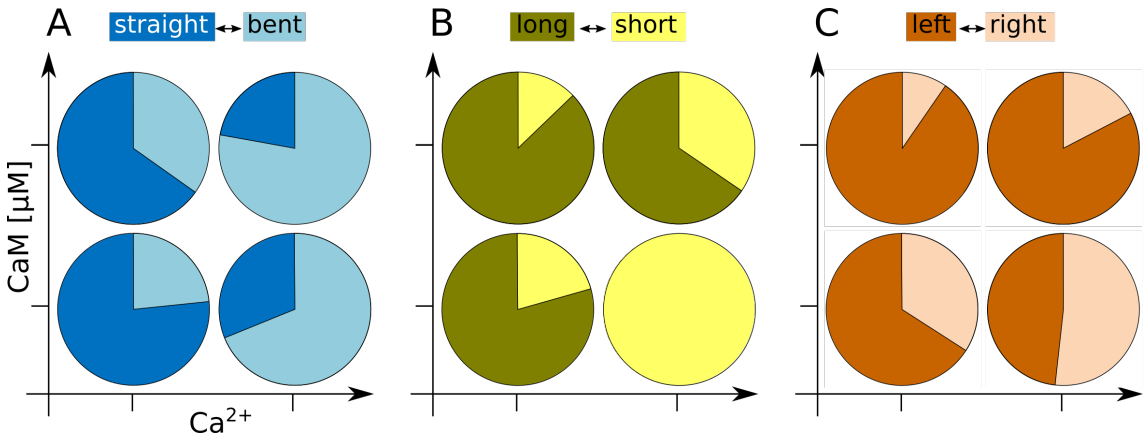


Figure 5.15: Statistics on the shapes found in all four calcium / CaM conditions conditions, reduced to two parameter, in the form of piecharts. Color code according to table 5.2.

Another way to analyze the statistical distribution of the eight different myosin VI-shapes in the four conditions is by decreasing the number of parameters from eight to 2. In Figure 5.15 **A-C** the features *straight* / *bent*, *long* / *short* and *left* / *right* are presented respectively, where for example the feature *long* included all long molecules, regardless whether they were straight, bent, left or right. Table 5.2 below lists the corresponding percentages.

One of the dominating result from this analysis was the change of ratio between *straight* and *bent* molecules due to an increase of calcium (Figure 5.15 **A**). At low calcium, the majority of myosin molecules (77%) adhered to the EM grid in the *straight* conformation compared to the *bent* conformation (23%). At high calcium on the other hand, the ratio reversed completely and the side view on the surface of the grid was preferred (78% for the *straight* conformation and (22%) for the *bent* conformation). The change of calmodulin concentrations had only a slight effect on the general shape of myosin VI. When the solution was substituted with extra CaM, the percentage of the *straight* conformation (front view) decreased at low calcium, while at high calcium the front view was somewhat less favored.

As mentioned earlier, the increase of calcium also had an effect on the length of the lever arm (Figure 5.15 **B**), letting the second calmodulin dissociate (or not bind in the first place) and therefore making the molecules appear shorter (from 17% on average at low calcium to 67% on average at high calcium). At high Ca^{2+} and low calmodulin, no long molecules could be observed at all. This effect could be compensated by adding extra calmodulin and thus increasing the binding probability of CaM and increasing the number of *long* molecules again (66%). At low calcium, the adding of calmodulin caused the same change, however with a change of 8% much less significant.

Moreover, the side of myosin VI, with which it bound to the EM-grid surface was influ-

enced by calcium and calmodulin (Figure 5.15 **C**). In almost all four conditions the *left view* was preferred and especially favored at high calmodulin conditions (87% on average). The increase of calcium had the largest effect at low calmodulin, making the *right view* becoming more favored (from 34% to 52%). As it was the case for the length of the lever arm, the calcium effect on the sidedness of myosin VI could be compensated by increasing the calmodulin concentration. Moreover, in both calcium conditions, low and high, the adding of extra CaM to the sample was correlated with an increasing probability of myosin VI to be absorbed to the grid in the right-facing orientation (from (right:left) 66%:34% to 90%:10% at low calcium and from 48%:52% to 83%:17% at high calcium).

These statistics indicated that the calmodulin binding lever arm and in particular the second calmodulin underwent the biggest structural changes.

5.2.3 Calcium induced conformational changes of myosin VI

Summarizing the results from the previous section, especially calcium had a significant effect on the calmodulin-binding lever arm. It influenced the shape and length of myosin VI and the way it was absorbed to the surface of the EM-grid during staining. The following part will therefore in more detail analyze the structural properties of myosin VI - in particular the lever arm - upon changing calcium concentrations. Furthermore, the focus will lie on the *bent* conformation, as it could be fully explained and described using the crystal structure model (see Figure 5.6).

Area	N	ccc high Ca	ccc low Ca	Δccc (low Ca-high Ca)
whole molecule	1	0.906203	0.929343	0.023140
catalytic domain	2	0.935078	0.920475	**-0.014603**
neck-tail	3	0.794921	0.945050	0.150129
1st CaM	4	0.927625	0.988483	0.060858
2nd CaM	5	0.763909	0.962096	**0.198188**
3HB	6	0.680171	0.897655	**0.217484**

Table 5.3: Cross correlation coefficients of areas N=1-6 of the EM averages at low and high calcium with the crystal structure model and their differences Δccc.

In order to evaluate the structural properties of the *bent* conformation in more detail, high resolution class averages were needed. Therefore, only images with sufficient quality and images showing all visible parts of the molecule (head, lever arm with two calmodulins and 3HB) were used to create one class average each for the low and high calcium condition. These class averages, shown in Figure 5.16 **A** (low calcium: 2998 images; high calcium: 2392 images, scale bar: 5nm) had a resolution of 18.06 and 17.76 Å

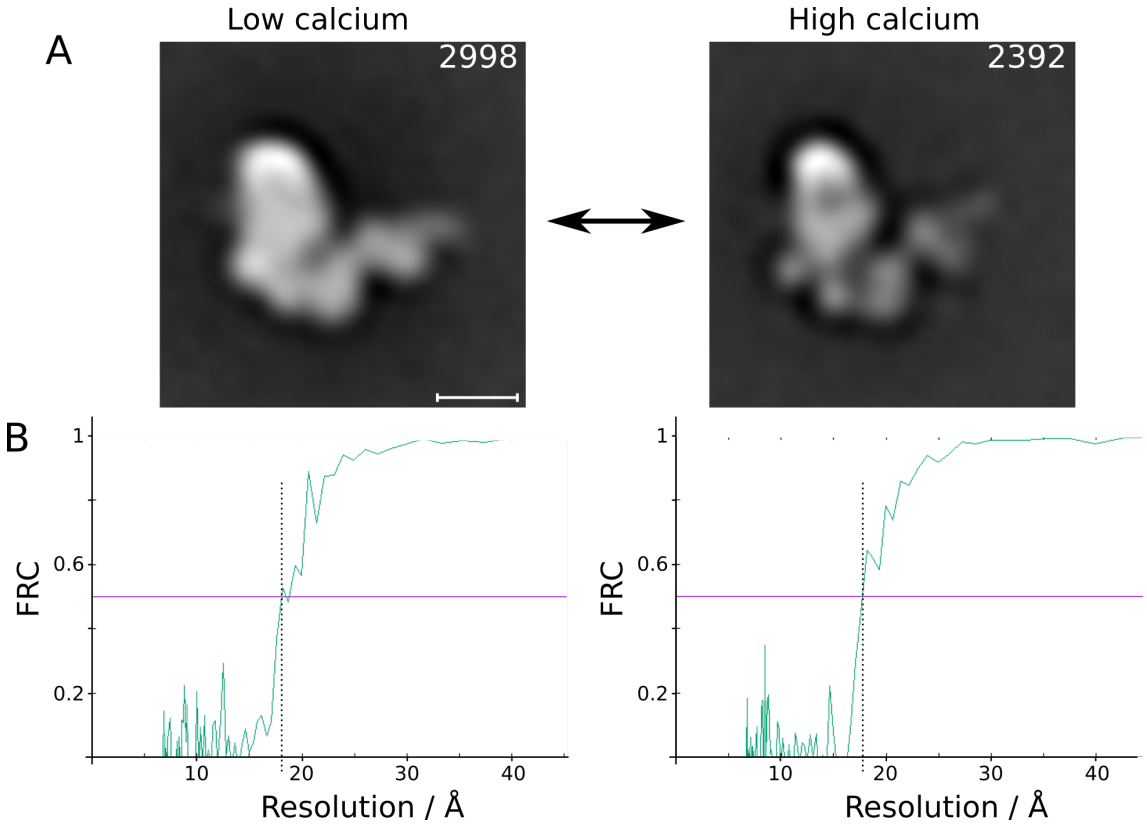


Figure 5.16: **A** Maximum resolution class averages of molecules in the bent conformation at low and high calcium. Only images with high quality were used to create the two averages. The upper right number denotes the number of images on average. **B** The resolutions of 18.06 Å and 17.76 Å (low and high calcium respectively) were measured using the FRC 0.5 criterion Scale bar: 5 nm

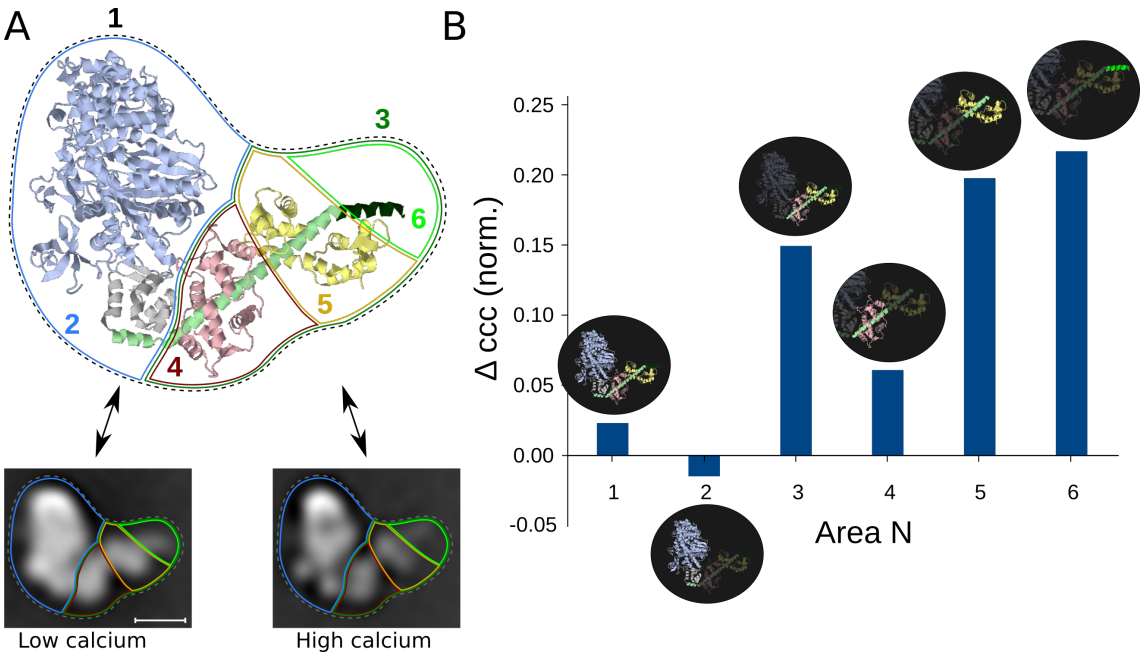


Figure 5.17: To compare the low and high calcium bent conformation of myosin VI, different areas *N=1-6* (**A**) of the corresponding EM averages were compared with a crystal structure model. **B** The normalized differences of the cross correlation coefficients for low and high calcium (positive values = better fit of low calcium EM average) reveal biggest differences in the structure of the second CaM (N=5) and three-helix-bundle (N=6). The negative value for the converter domain (N=2) indicates a better fit of the high calcium EM average to the model.

respectively, measured by calculating the FRC cutoff at a value of 0.5, as described in section 3.3 and shown in Figure 5.16 **B**. To localize differences between both EM averages, different parts of the images were compared using the myosin VI crystal structure model as described in section 5.5.2.3 (see Figure 5.37). Figure 5.17 **A** shows the analyzed parts, namely the whole molecule (dashed outline, N=1), catalytic domain with converter (blue outline, N=2), lever arm (dark green outline, N=3), first CaM (red outline, N=4), second CaM (yellow outline, N=5) and 3HB (bright green outline, N=6). The resulting (normalized) differences of the cross correlation coefficients between the low and high calcium class averages (Figure 5.17 **B**, $\Delta ccc = ccc(lowCa) - ccc(highCa)$) identified clearly the location of changes due to calcium (absolute numbers are given in table 5.3). Since Δccc was calculated by subtracting the values for the high Ca^{2+} average from the ones for the low Ca^{2+} average, positive values equaled a better fit of myosin VI at low calcium to the crystal structure, whereas negative values indicated a better fit of myosin VI in the presence of calcium.

Confirming the results from the statistics (section 5.2.2), the lever arm changed significantly under increase of calcium (Δccc[N=3]=+0.150129). A closer look into the lever arm revealed further that calcium had its biggest effect on the second calmodulin and three-helix-bundle (Δccc[N=5]=+0.198188 and Δccc[N=6]=+0.217484), while the struc-

tural changes of the first calmodulin were only slight (Δccc[N=4]=+0.060858). Interestingly, although the correlation between the catalytic domain of the model and the EM data was not as significantly affected by calcium (Δccc[N=2]=-0.014603). The negative sign represented a better fit of the head domain at high calcium. Looking closer, at low calcium the EM data comprised intensities in the area of the catalytic domain that could not be explained by the crystal structure model. These intensities were not present at high calcium.

The methods and results described above made use of an available crystal structure for comparison. To verify these results and examine the structural changes in more detail, the EM class averages also have been analyzed on their own without the need of a third parameter by plotting the intensity profiles of different parts of interest and comparing them as shown in Figure 5.18. In **A** displayed are both EM class averages for low and high calcium with the three lines for the profile plots overlaid (I: head domain, II: first and second calmodulin (only one loop each), III: first and second calmodulin (other loop) and 3HB). Each of these three intensity plots are shown in **B**, where the blue graphs correspond to the low calcium condition and the red ones to the high calcium condition. Each of the peaks was related to structural part as indicated by the labellings. Intensity minima, indicated by the black arrowheads, represented the presence of accumulated stain solution around the molecule, resulting in higher density and hence less transmitted electron, making these parts appear darker in the EM. The shape of the catalytic domain, (profile **I**), was similar at low and high calcium, as well as the location of the converter domain. Moreover, as expected the position of the first calmodulin remained stable under the influence of calcium, seen by similar positions for the intensity maxima for both of its lobes (**II** and **III** respectively). Supporting the previous findings that the second calmodulin was strongly influenced by calcium, the intensity distribution for the Ca^{2+}-CaM shifted significantly, however only for one of the lobes (graph **II**, red lines). The other loop on the other hand (**III**) remained mainly unaffected by calcium. This partly shifting suggested a rotational change of the second CaM. Furthermore, a significant change of the intensity profile representing the three-helix-bundle could be resolved in plot **III**. Although the peak for the low calcium condition was very weak, its position was clearly different from the high calcium peak.

Once the calcium-induced structural differences of myosin VI were localized and analyzed, the crystal structure model in its optimized projection was overlaid onto the EM data, as done in Figure 5.19 (color-code for the model: catalytic domain in blue, converter in grey, neck alpha helix in bright green, first calmodulin in red, second calmodulin in yellow and first helix of 3HB in dark green). For better contrast the underlying EM images were inverted and enhanced (**A** and **C**). Furthermore, as the biggest changes were

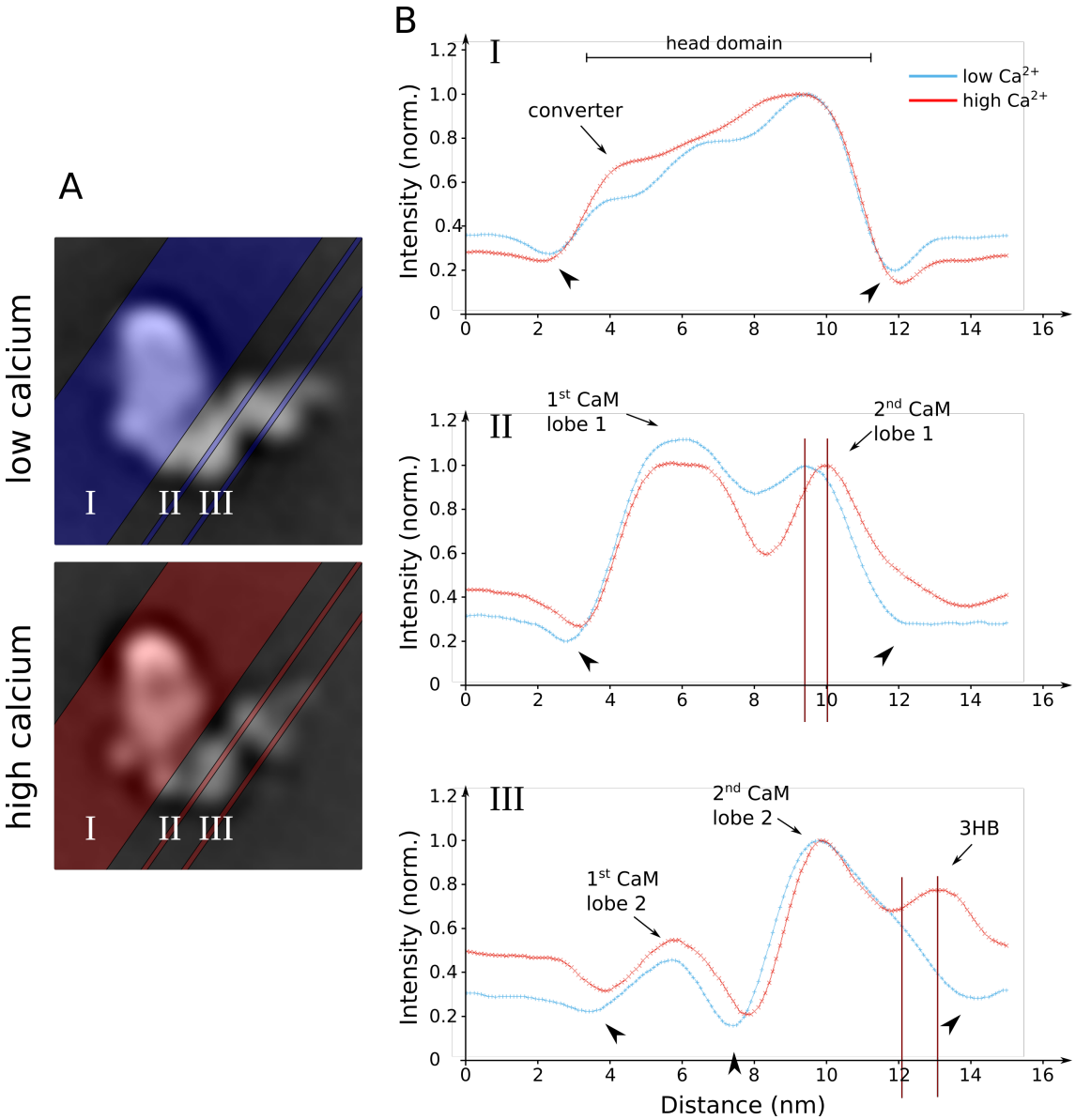


Figure 5.18: Further analysis of the myosin VI structure at low and hight calcium using intensity profile plots. **A** EM averages with indicated areas I-III for the calculation of intensity profiles in **B**. Minima in the plots indicate accumulated stain (arrowheads) and each peak can be identified as a structural part of myosin VI. While the converter domain as well as the first calmodulin (both lobes) had similar profiles in both calcium conditions, a significant shift of the peak representing the upper lobe of the second calmodulin could be identified (II). Intriguingly, the other lobe of this CaM did not change its position (III). Graph III also indicated a large shift of the 3HB.

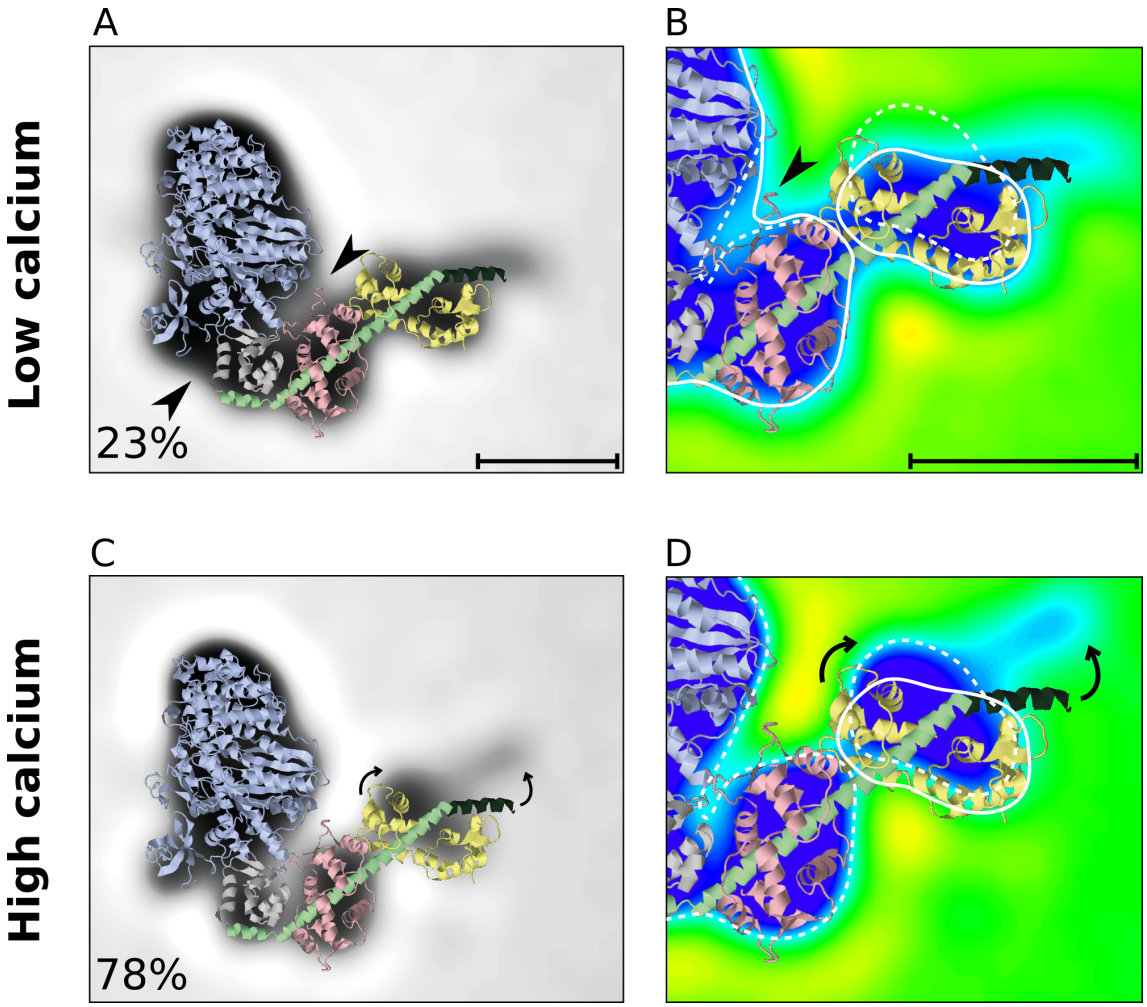


Figure 5.19: **A** Superposition of the crystal structure model in the optimized orientation and an inverted version of the EM averages of myosin VI in the bent conformation. Color code: catalytic domain in blue, converter in grey, neck alpha helix in bright green, first calmodulin in red, second calmodulin in yellow and first helix of 3HB in dark green. Apart from an overall excellent fit, some intriguing discrepancies were resolved. **B** Zoom into the lever arm. The EM average was color coded according to its intensities: low values in yellow, high values in blue. The superposition of the outlines from both conditions, low calcium (solid line) and high calcium (dashed line) indicated the calcium induced differences: structural change of second CaM and 3HB at high calcium (arrows) and an additional intensity in the area of the converter domain at ow calcium (arrowheads), consistent with a backfolded tail. Scale bars: 5 nm

discovered to be localized at the neck region, **B** and **D** show a zoom into the area of interest. Here, the EM data has been color-coded according to its intensities, with increasing intensity values from yellow via green to blue. Together with the contour lines that were created as described in section 5.5.2.3 (solid line: low calcium condition; dashed line: high calcium condition), the structural changes were better visualized:
At low calcium the modeled structure showed excellent agreement with the EM image (solid line), especially the first and second apo-calmodulins and 3HB (first helix), confirming the correct interpretation of the EM average. The cross-correlation-coefficients for these parts were ccc(1stCaM)=0.99, ccc(2ndCaM)=0.96 and ccc(3HB)=0.89. In the presence of calcium however, the orientation of the second CaM and 3HB (dashed line) did not match the crystal structure anymore (ccc(2ndCaM)=0.76 and ccc(3HB)=0.68). The 2nd calcium-calmodulin had rotated by approx. 30° in the projection plane (clockwise) and the 3HB approx. 30° anti-clockwise, as indicated by the black arrows. The first calmodulin remained relatively unaffected by calcium with a ccc-value of ccc(1stCaM)=0.93 at high calcium concentrations, as previously reported [80]. The black arrowheads show the structural parts in the area of the low calcium myosin VI catalytic domain that were not accounted for by the model. These intensities were not visible anymore at hight calcium, were the head domain showed a better fit to the model.

The structural and conformational changes of myosin VI due to calcium revealed by the electron microscopic studies could be confirmed and further interpreted in more detail using titration studies, intrinsic tryptophan fluorescence, pull-down and microscale thermophoresis, detecting specific calmodulin binding sites on on various peptides[3]. Figure 5.20 illustrates and summarizes the results from the experiments on 16 different peptides [83].

In Figure 5.20 **A-D** shown are the main results from these studies, which were summarized in **E**, a simplified sequence of myosin VI with highlighted positions of the main structural parts and peptides (P1-P4): Without calcium, the second calmodulin (apo-CaM) bound the aa 1060-1125 part of the tail (**B**) and further with both lobes to peptide P2 (**C**). With increasing calcium, the connection with the 1060 tail was lost (**B**) and only one calmodulin lobe could bind to the peptide P2, indicated by a stoichiometry of 1:2 for CaM:P2 at high calcium (**C**). Searching for the binding position of the second Ca^{2+}-lobe, the experiment with peptide P3 revealed no binding of the apo-CaM and a very weak binding of the calcium-calmodulin (**D**). The stoichiometry for the double peptide P2-3 however was 1:1 in both cases (**E**) with a higher affinity for the Ca^{2+}-CaM, indicating the presence of a bipartite binding site. The location of this second binding site for the calmodulin was identified to be located within the loop between the first and second he-

[3]These experiments were carried out by my colleges.

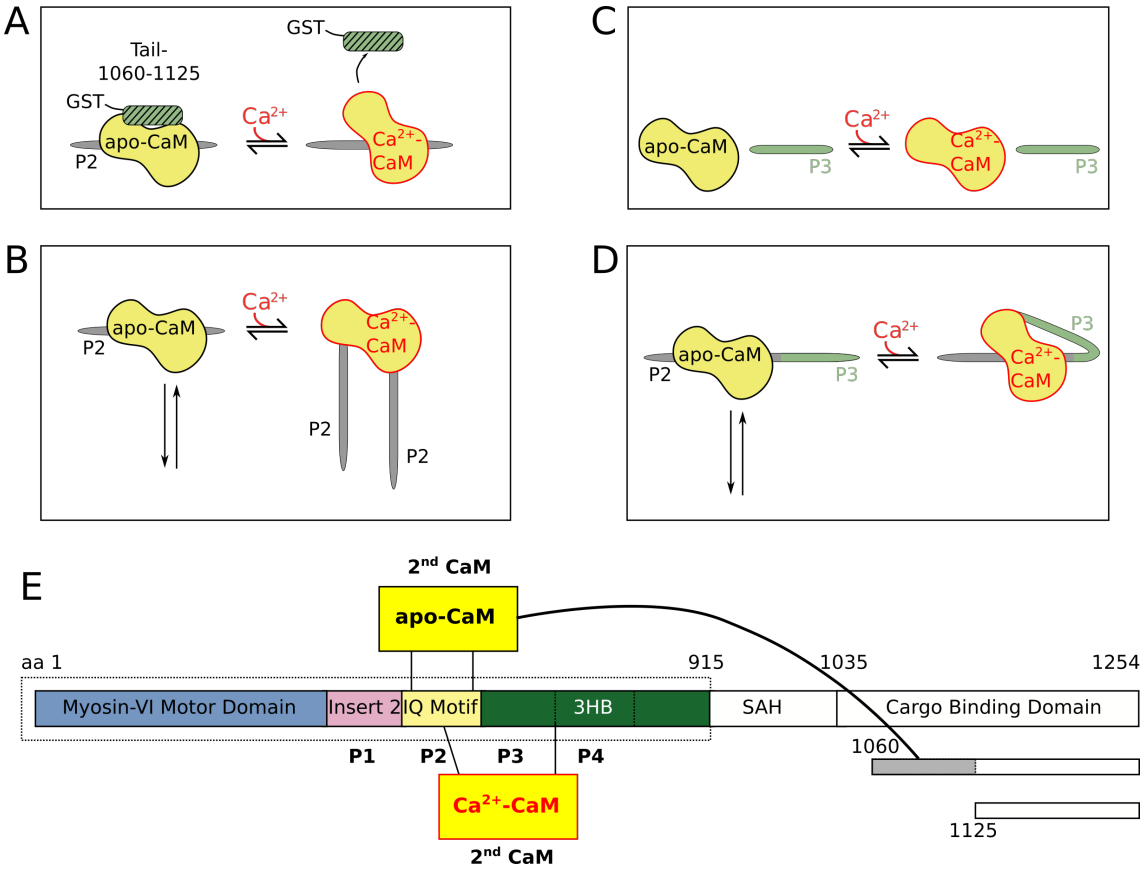


Figure 5.20: **A-D** Results from binding studies with 16 different peptides [83], which are summarized in **E**: Apo (second) CaM binds the tail segment aa 1060-1125 (**B**) and with both lobes to peptide P2 (**C**). Under binding of calcium the connection to the tail is lost (**B**) and the CaM bridges from peptide P2 to a non-adjacent part between P3 and P4 (**E**).

lix of the 3HB. Further studies showed that the detached tail from the second calmodulin in the calcium-state also provided a higher binding affinity to lipid cargo. [83].

The results from these experiments were in close agreement with the electron microscopic studies. The binding behavior of the second calmodulin due to calcium could be directly observed in the EM class averages. The partly detaching of the CaM and bridging to the 3HB (Figure 5.20) was consistent with the discovered clockwise rotation of the Ca^{2+}-CaM and anticlockwise rotation of the three-helix-bundle (Figure 5.19 **D**). Moreover, the previously mentioned additional mass in the EM average between the 1st CaM, converter and head at low calcium (Figure 5.19 **A** and **B**, black arrowheads) could not be explained as it was not accounted for by the crystal structure model. However, together with the results that the tail segment 1060-1125 bound to the second calmodulin at low calcium, the mentioned additional mass could be identified as part of the backfolded tail of myosin VI. As the tail was highly flexible, it could be only seen in areas with some kind of constraints, e.g. in the area of the second calmodulin and head domain. The detaching of the tail 1060 when calmodulin changed its structure due to calcium forced the whole tail to unfold, indicated by the disappearing of the mass close to the converter domain. Due to the even higher flexibility in this state it could not be observed any longer in the EM images (Figure 5.19 **C** and **D**).

5.2.4 Destabilization of the myosin VI lever arm induced by calcium

The previous results (section 5.2.3) suggested a regulation of myosin VI, where calcium induced structural changes of the calmodulin, which led to the unfolding of the tail at high calcium conditions, enabling myosin VI to bind lipid cargo. Furthermore, the fact that these structural changes were only observed for the second calmodulin while the first one seemed to be unaffected, and with the background knowledge that neck domains associated with calmodulin are known to be essential for the lever arm stability [95], the question arose whether the loss of contact of the two calmodulins had any effect on the mechanical properties of this myosin class. Naturally, the EM is not the ideal instrument to investigate mechanics and moving objects. Nevertheless it is possible to extract information on the flexibility of an object by SPA of EM images and indirectly relate it to its mechanical properties. With a method based on the idea of the "Muscle Group Leeds"[4] [96, 97] and described in section 5.5.2.3, the calcium-dependent stiffness of the neck domain, which plays a key role in the myosin's mechanical functionality, was investigated by creating an "EM-movie":

[4] Among others Matthew L. Walker, Stan A. Burgess

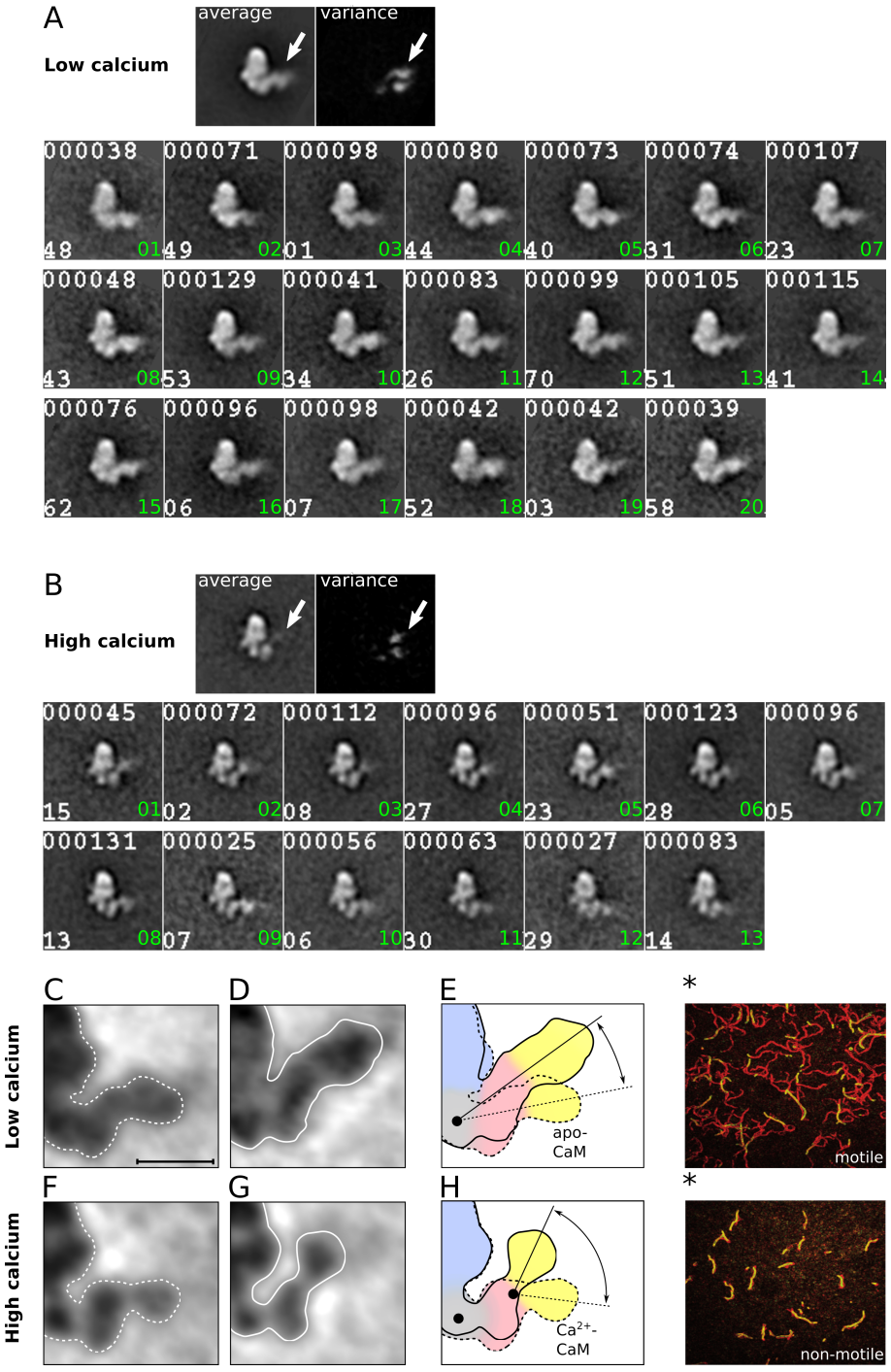


Figure 5.21: Lever arm stiffness of myosin VI investigated using EM-movies. **A**, **B** Classification of the low and high calcium data set with respect to the structure of the lever arm (arrow). Upper right number: number of images contained in average; lower left number: class number; Scale: images are 30 nm wide. **C-H** Outlining of the two extreme positions of the lever arm. A superposition of both outlines revealed the appearance of a second pivot point between the first and second calmodulin (red and yellow respectively) at high calcium, indicating a break in the lever arm structure. * Motility assays confirmed the relation of calcium induced loss of neck stability to the loss of mechanical functionality (motility) [83]. Scale bar: 5 nm

To visualize the flexibility of the lever arm, the catalytic domain was aligned only and the inhomogeneity of the lever arm was analyzed as described in section 5.5.2.3. Figure 5.21 shows the result for the two conditions low calcium / low calmodulin and high calcium / high calmodulin[5]. In **A** and **B** displayed are the global averages of this realignment together with their variance images. The catalytic domain was in both cases highly resolved as expected, while the lever arm appeared blurred in the averaged images, indicating the presence of inhomogeneity of this part within the data set (marked by the white arrow). This inhomogeneity, which was more clearly visible in the variance images, could be interpreted as some kind of flexibility. Shown underneath the average and variance images are the classified and sorted data sets (upper left number: images contained in the average, lower left number: original class number, lower right number: sorted class number). To further examine the flexibility of the neck region, Figure 5.21 **C**, **D**,**F** and **G** show magnified and inverted images of the extreme orientations at low and high calcium, overlaid with their outlines. When the two extreme outlines were overlaid (**E** and **H**, color code as previous), the structural behavior of this flexibility became visible:

At low calcium, there was only a single pivot point within the converter region around which the lever arm could rotate in a certain angular range (**C**-**E**). The neck domain itself however showed no signs of flexibility. At high calcium, a second pivot point arose, located between the 1st and 2nd calmodulin, making the lever arm flexible in its own (**F**-**H**). These results explain the further impacts of the structural changes of the second calmodulin when calcium was bound to it: Its rearrangement, rotation away from the first calmodulin (probably loosing contact to it) and binding to the bipartite binding site destabilized the lever arm, probably preventing the transduction of force generated in the converter domain, i.e. loosing mechanical functionality.

Further studies using gliding filament assays could confirm that this break in the structure of the neck domain caused the assumed loss of mechanical functionality [83]. The results are shown in Figure 5.21*, representing a superimposed 600 seconds movie, where the red lines are the tracks of actin-filaments moving over a layer myosin VI molecules. Shown in yellow are the initial positions of the filaments in the first frame. As an effect of the loss of stability of the lever arm at high calcium, the myosin VI motor was not able to establish any motility but remained bound to actin (indicated by the overlap and presence of the red and yellow lines). These experiments showed further, that the mechanical function (motility) could be completely restored by lowering the calcium concentration. The calcium-induced structural change of the lever am and in particu-

[5]Shown here are only two conditions of the left-facing bent conformation, as they represented the results the best. The analysis for the other six conditions (with respect to the calcium, calmodulin and sidedness) are discussed in section 5.3

lar the second calmodulin was therefore reversible, adopting its original conformation and restoring the stiffness of the lever arm, crucial for the generation of force and thus motility.

5.3 Discussion

This section will summarize and discuss the single particle analysis of myosin VI and in combination with published results and suggest a model of a novel two-step mechanism for the regulation of myosin VI by calcium and calmodulin.

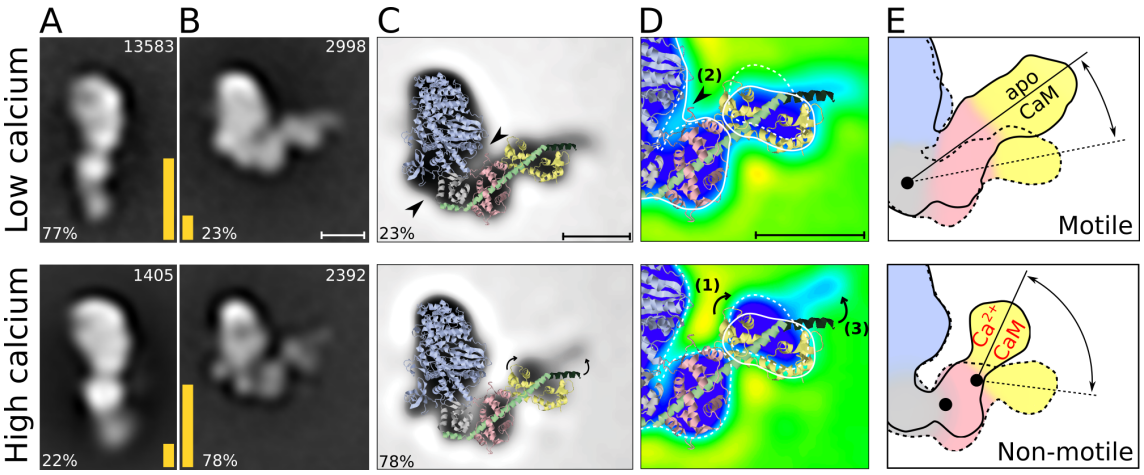


Figure 5.22: Summary of the single particle analysis on FL myosin VI in the absence and presence of calcium: **A**, **B** Myosin VI was found in mainly two conformations, bent and straight. At low calcium the straight conformation was more abundant (77:23%), whereas at high calcium this ratio switched to the opposite. **C,D** Superposition of a crystal structure model of myosin VI in its optimized orientation onto the EM averages of the bent conformation showed an overall excellent fit, however with some small but intriguing differences: Calcium induced structural change of second CaM and 3HB (arrows) and an additional intensity in the area of the converter domain at low calcium (arrowheads), consistent with a backfolded tail. **E** Superposition of two lever arm structures obtained from the creation of EM-movies reveal the appearance of a second pivot, indicating the loss of neck stability at high calcium and thus ability to transduce force. Scale bars: 5 nm

In the sections 5.2.1-5.2.4, intriguing discoveries of how calcium and calmodulin affected the structure and mechanics of myosin VI were made. Figure 5.22 summarizes the results of the single particle analysis. **A** and **B** show the two main conformations that myosin VI adopted. While both conformations appeared in the presence and absence of calcium, the ratio of how many molecules adopted the straight and how many the bent conformation was heavily influenced by calcium as indicated by the shown percentages. The low and high calcium concentrations (pCa 8 and pCa 4 respectively) were kept close to the physiological concentration range as they appear in the cell and at the same time were chosen to represent the extreme conditions in order to ensure saturated populations. At low calcium, 77% were found in the front view and the remaining 23% in the side view. This ratio reversed under calcium increase. The upper right number in the EM averages denotes the number of images contained in the averages. Note that these number do not coincidence directly with the given percentages, as for the statistics all different ”versions” of each conformation were taken into account, e.g. short molecules

with a missing second calmodulin. The averages on the other hand do not contain such images.

While the bent conformation could be explained in great detail by available crystal structures of myosin VI, the straight conformation most likely occurred due to deformations of the neck domain when myosin VI attached to the EM grid surface. The fact that the straight conformation was much more likely to be found at low calcium than high calcium indicated that changes of the structure as well as of the surface charges of myosin were induced by calcium. The closer investigation of these changes were done only on the bent conformation for the mentioned reason. Obtained by crystal structure projection matching, the best fitting model of nucleotide-free myosin VI in its optimized orientation was overlaid onto an inverted (**C**) and color-coded version (**D**, yellow→ blue = low intensities→ high intensities) of the EM averages. Apart from the excellent overall fit, three major differences were discovered:

- **(1)**: The position of the second calmodulin has changed upon increase of calcium (arrows). A closer inspection of this area of interest (**D**) compared the positions of the second calmodulin in the apo and Ca^{2+} state, revealing the structural changes in more detail. Here one has to keep in mind that EM images are initially always 2D projection views of a 3D object. Therefore, the rotation by approx. 30° as it was observed for the CaM and 3HB did not necessarily had to happen in plane, but more likely in three dimensions. What is seen in the EM averages is just the 2D projection of a 3D rotation. Furthermore, the large forces occurring during the binding to the EM grid, the negative staining and the blotting of excess liquid, are likely to flatten down the structure of the calmodulin and the helix bundle. To investigate the exact structural changes of the second calmodulin, 3D reconstruction of myosin VI using e.g. cryo-EM would be necessary. However, due to the size of myosin VI of approx. 150 kDa, this is very challenging. And although structures of less than 100 kDa have been reported to be resolved by cryo-EM in high resolution (for example the 64 kDa haemoglobin at 3.2 Å [98]) the negative stain approach was still the method of choice. In contrast to e.g. haemoglobin which consists of four symmetric subunits, myosin VI does not comprise such symmetric features, contains a highly flexible tail and partly flexible neck region, which all hinder the success of high resolution 3D EM reconstruction. Furthermore, the tail comprises such a high flexibility, that it could not even be seen in the negative stain images, which makes it even more unlikely to be discovered by cryo-EM[6].

- **(2)**: Indicated by the arrowheads, the EM intensity changed significantly in these

[6]The content of this statement will probably loose its value quite soon due to the current intense improvements in the field of EM.

areas when the calcium concentration increased. While at low calcium an EM mass in the area of the converter could be seen, which was not covered by the crystal structure, this mass disappeared at high calcium. As it was revealed by binding studies [83], the tail bound to calmodulin at low calcium but not to the Ca^{2+}-CaM, indicating an unfolding of the tail. Together with this piece of information, the additional mass at low calcium presumably represented the backfolded tail. However, as mentioned above, it could be only resolved in the area of the converter / head domain, the other tail segments stayed hidden due to the high flexibility. Another hint for the unfolding of the tail was the change of the 3HB structure.

- **(3)**: Also indicated by the arrow, a conformational change of the EM intensities representing the three-helix-bundle could be observed, characterized by an approx. 30° in-plane rotation. The overlaid crystal structure (Figure 5.22 **C** and **D**) was modified to only contain the first helix of the bundle to better highlight the difference between low and high calcium. The more profound reason behind that were the results from the binding studies summarized in Figure 5.20. In these titration studies the unique bipartite binding site of calmodulin was discovered. At low calcium the second calmodulin bound with both lobes to the regular, adjacent binding sites on the helix. At high calcium however, one of the lobes of this CaM was able to reach out and bind to a non-adjacent binding site, located within the loop between the first and second helix of the three-helix-bundle, possibly forcing the bundle itself to change conformation and even unbundle. That such an unfolding process of the 3HB in myosin VI could be possible has been already reported [81], however in this case it was related to dimerization. Whether calcium could induce a similar effect was unknown.
  A possible process behind the hypothesis of a calcium-induced unfolding of the 3HB by Ca^{2+}-CaM bridging to the loop between helix 1 and helix 2, based on both, the EM and titration studies, is illustrated in Figure 5.23, where the 3HB is represented with all three helices. Following this idea, a significant decrease of intensity should be observed in the EM average if this bundle unfolds at high calcium and the three helices do not align anymore.

	Max. intensity	Intensity$_{L1/H1}$	Intensity$_{L2/H2}$	Area under peak
Low Ca^{2+}	1.000	-0.174	-0.180	**23.98**
High Ca^{2+}	**0.494**	0.038	0.427	**08.40**

Table 5.4: Calcium dependent maximum intensities and area under the peaks in graph 5.24

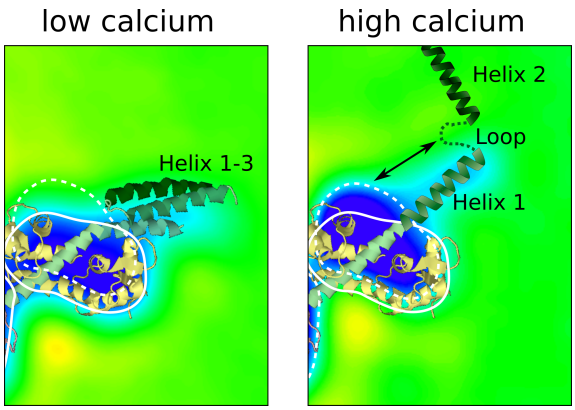


Figure 5.23: Possible calcium-induced unfolding of the 3HB to establish the binding of CaM to the loop between helix 1 and 2 at high calcium.

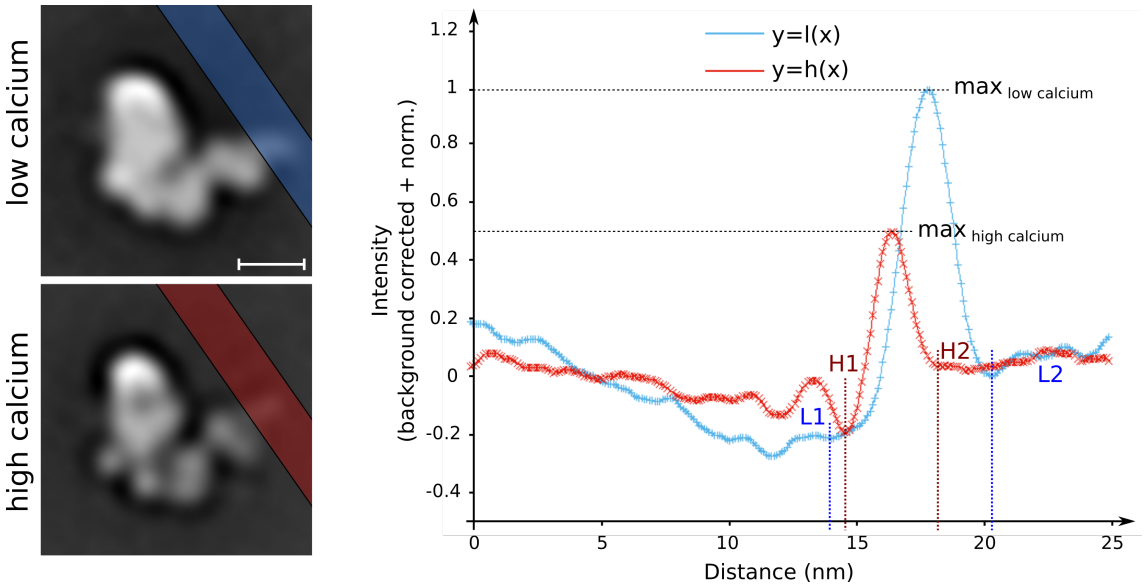


Figure 5.24: Background-corrected and normalized intensity profile plot of the 3HB at low (blue) and high-calcium (red). The important value were not the maximum peaks but the area under them, representing the total intensity of the 3HB and indicating a decrease of approx. 2/3 from low to high Ca^{2+}. Scale bar: 5 nm

To test this hypothesis, an intensity-profile of the corresponding EM masses was calculated as shown in Figure 5.24, where the blue graph $y = l(x)$ plots the profile of the low calcium condition and the red graph $y = h(x)$ plots the high calcium condition. The graphs shown here have been background-corrected by subtracting the average value of the curve before and after the peaks, and normalized to the maximum value of the low calcium peak. Apart from the expected shift of the peak-position (which represented the observed rotational change of the 3HB), the total intensity values showed also a calcium-dependency. While the maximum values already indicated a decrease of approx. 50% in intensity ($\max_{high\ calcium}$=0.49, arbitrary units), more of interest was the accumulated intensity of both peaks, i.e. the area under them. The first minima after each peak maxima were chosen to set the peak boundaries (L1 and L2: low calcium; H1 and H2: high calcium). The area under these defined peaks was calculated and compared (see Table 5.4). Strikingly, the change of total intensity representing the three-helix-bundle was in accordance with the expected decrease by 2/3 (64.89%):

$$\int_{L1}^{L2} |l(x)|\,dx = 23.98; \int_{H1}^{H2} |h(x)|\,dx = 8.42$$

The fact, that in the high calcium condition apparently only one helix was resolved, had its origin most likely in increasing flexibility of the two following helices.
Although this analysis confirmed the hypothesis of the 3HB unfolding process upon calcium, it has to be said that the loop between the first and second helix of the 3HB is relatively big on its own. Therefore, the connection between the second calmodulin and the 3HB as well as the change of the 3HB orientation might also be established without an actual unbundling.

Referring to the observed loss of contact between the two calmodulin due to the conformational change of the second calmodulin induced by calcium, the lever arm was observed to have lost its stability, as shown in the summary Figure 5.22 **E** and described in section 5.2.4. The analysis shown here, where the emergence of the second pivot point at high calcium indicated the loss of stability, was done on the low calcium / low calmodulin and high calcium / high calmodulin condition of the left-facing bent conformation. However, as shown in section 5.2.2, more conformations of myosin VI were discovered. Therefore, the same analysis was applied to the remaining conformations in order to certify the results.

Figure 5.25 shows the lever arm-flexibility-analysis of all bent conformations found in the four conditions. Above the cartoons, the two EM averages showing the extreme positions of the lever arm are shown and marked by dashed and solid outlines, which

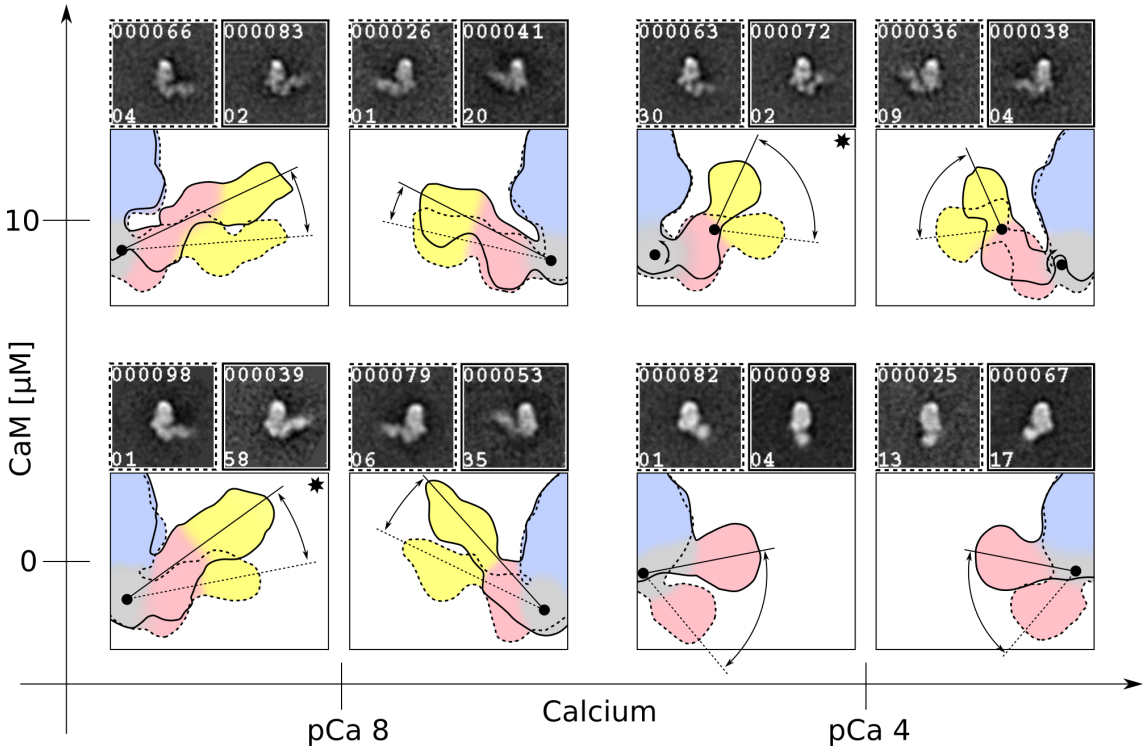


Figure 5.25: Analysis of flexibility of the neck and tail region of the bent conformation of myosin VI in dependence of calcium and calmodulin concentrations and orientation on the grid (mirror related). The alignment of the head domain and classification of the angular distribution of the neck region according to section 5.5.2.3 led to a series of images from which the to extreme orientations are shown here, together with an overlay of the two extreme outlines. The presence of one pivot point indicated neck-stability and the appearance of a second one results in the lost of this stability. It could be seen that the increase of calcium induces this break in the neck. At low calmodulin (and pCa 4) however the myosin had only the first calmodulin bound, resulting in maximum one pivot point. Scale: EM averages are 30 nm wide.

correspond to the outlines in the cartoon. In agreement with the two cartoons marked by the star (which were the ones shown previously, the other conditions yielded the expected behavior. Neither the sidedness of how the myosin landed on the surface (right- or left-facing), nor the calmodulin concentration had any influence on the flexibility. Only the calcium concentration was related to the emergence of a second pivot point. In the high calcium / low calmodulin condition only one pivot point in the converter region could be observed. This however was due to the fact that in this condition no second CaM was attached.

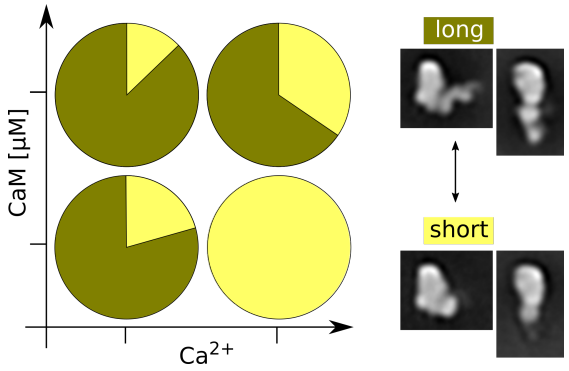


Figure 5.26: Distribution of long and short myosin VI molecules (with respect to the length of the lever arm due to the presence of absence of a second calmodulin) in dependency of calcium and calmodulin concentrations.

The total dissociation of the second CaM in this condition observed in the EM as well as the higher probability of shorter molecules even at high calmodulin (see pie-chart in Figure 5.26), led to a further interpretation of how the process of the second CaM bridging the two binding sites could proceed.

There are mainly two possibilities for the CaM to change its conformation, directly or indirectly: **A** Either it stays bound with one lobe to the peptide and the structural changes induced by calcium let the calmodulin rotate (1), i.e. detachment of the second lobe, followed by rebinding to the 3HB (2). Or **B**, the calmodulin detached from its initial binding position completely due to calcium binding (1). Subsequently, the same or more likely another Ca^{2+}-CaM attaches to the bipartite binding site of myosin VI (2), bridging to the 3HB (3). Both ways obviously lead to the same final conformation, the indirect change however includes steps with no second calmodulin bound, corresponding to the short conformation seen in the EM. Referring to the statistics on long and short molecules, a high concentration of calmodulin in the sample was necessary to compensate for the loss of CaM due to calcium. The significant increase of observed short molecules due to calcium suggested that the pathway shown in **B** might be the more likely one.

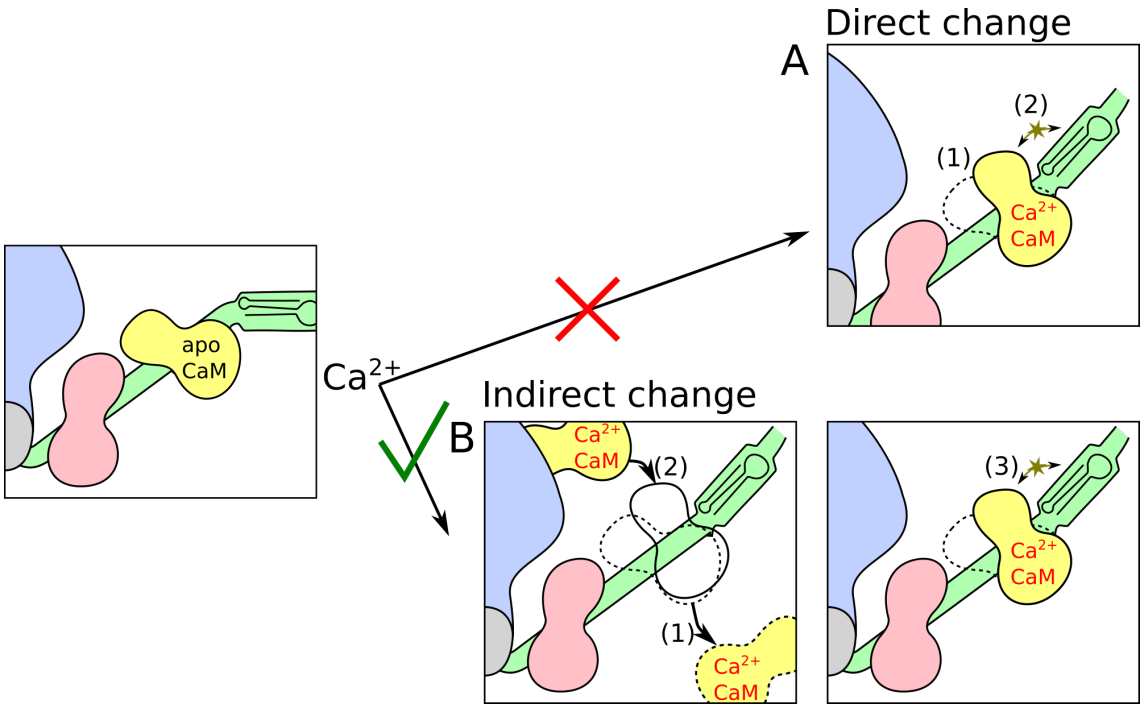


Figure 5.27: Possible pathways for CaM changing from the standard to the newly discovered bipartite binding site in myosin VI. **A** Direct change involves only a rotational rearrangement of CaM (1), which remains bound on the molecule, followed by bridging to the loop in the 3HB (2). **B** The more likely indirect pathway comprises total dissociation of CaM at high calcium (1) and rebinding to the bipartite binding site (2,3).

5.3.1 Model: Two-step mechanism for calcium regulation of myosin VI

The following list summarizes the results of the single particle analysis of myosin VI molecules in different calcium / calmodulin-conditions as well as results from different experiments that have carried out by my colleges simultaneously (such results [83] are marked with an asterisk in the following list).

At low calcium:

- The majority of myosin VI molecules adheres to the EM grid surface in the front view
- The second apo-calmodulin binds the tail segment 1060-1125 and forces the tail to backfold *
- The apo-calmodulin binds with both lobes to peptide P2 in the IQ motif *
- One single pivot point in the area of the converter domain allows the lever arm to rotate

- The close proximity of the two calmodulin in the lever arm provides the stiffness required for the transduction of force and ability of motility *

At high calcium:

- The majority of myosin VI molecules adheres to the surface of the EM grid in the side view
- The second Ca^{2+}-calmodulin undergoes major structural rearrangements, observed by a 30° clockwise in-plane rotation
- The three-helix-bundle rotates anti-clockwise by 30° (in-plane)
- The Ca^{2+}-calmodulin detaches with one lobe from the IQ motif (petite P2) and bridges to a nonadjacent binding site located in the loop connecting the first and second helix of the 3HB (peptide P2-3) *
- The tail unfolds as a consequence of the detachment of the tail segment 1060-1125 from the Ca^{2+}-calmodulin
- The release of the tail increases and enables the binding of lipid cargo to it *
- Apart from the pivot point at the converter region, a second one arises between the two calmodulin as an effect of the structural rearrangement of the second CaM (the first CaM remains mainly unchanged), indicating loss of stability within the lever arm
- Due to the instability of the lever arm, myosin VI is not able to provide functional motility *
- The calcium-induced structural changes are fully reversible once the calcium concentration decreases (back-folding of tail, restoring of motility) *

Based on these results, a model was suggested that explains and illustrates a novel two-step mechanism for calcium regulation, activation and mobilization of the myosin VI molecular motor, illustrated in Figure 5.28. The color-code is in correlation with the previous figures: catalytic domain blue, converter grey, 1st calmodulin red, 2nd calmodulin yellow, IQ motif and tail green.
A: When the calcium concentration is low, the calmodulin binds with both lobes to the lever arm. It further binds the tail segment 1060-1125, forcing the tail to backfold. In this condition, the myosin VI motor is in an inactive, dormant state.
B: As soon as the calcium concentration increases in the cell, calcium eventually binds to

calmodulin which undergoes major structural changes. In a new conformation it bridges from its initial position in the lever arm across to a binding site within the 3HB. In combination with the lowered affinity for the tail to bind to the 2nd Cam, the tail unfolds and is now ready to bind cargo.
C: As the structural change of the 2nd CaM results in the loss of rigidity of the lever arm, the myosin VI motor is not able to gain motility in this state. Instead it can now recruit lipid cargo, but is inhibited from translocating away.
D: The lowering of calcium restores the original conformation of the second calmodulin, making the lever arm stiff again. In the case that cargo is bound to the tail, the myosin VI motor is now mechanically activated and able to generate force and establish motility. For the case that no cargo is available, no cargo bound to the tail or the bound cargo is released, the myosin VI motor goes back to its original backfolded, inactivated state, ready for the next cycle.

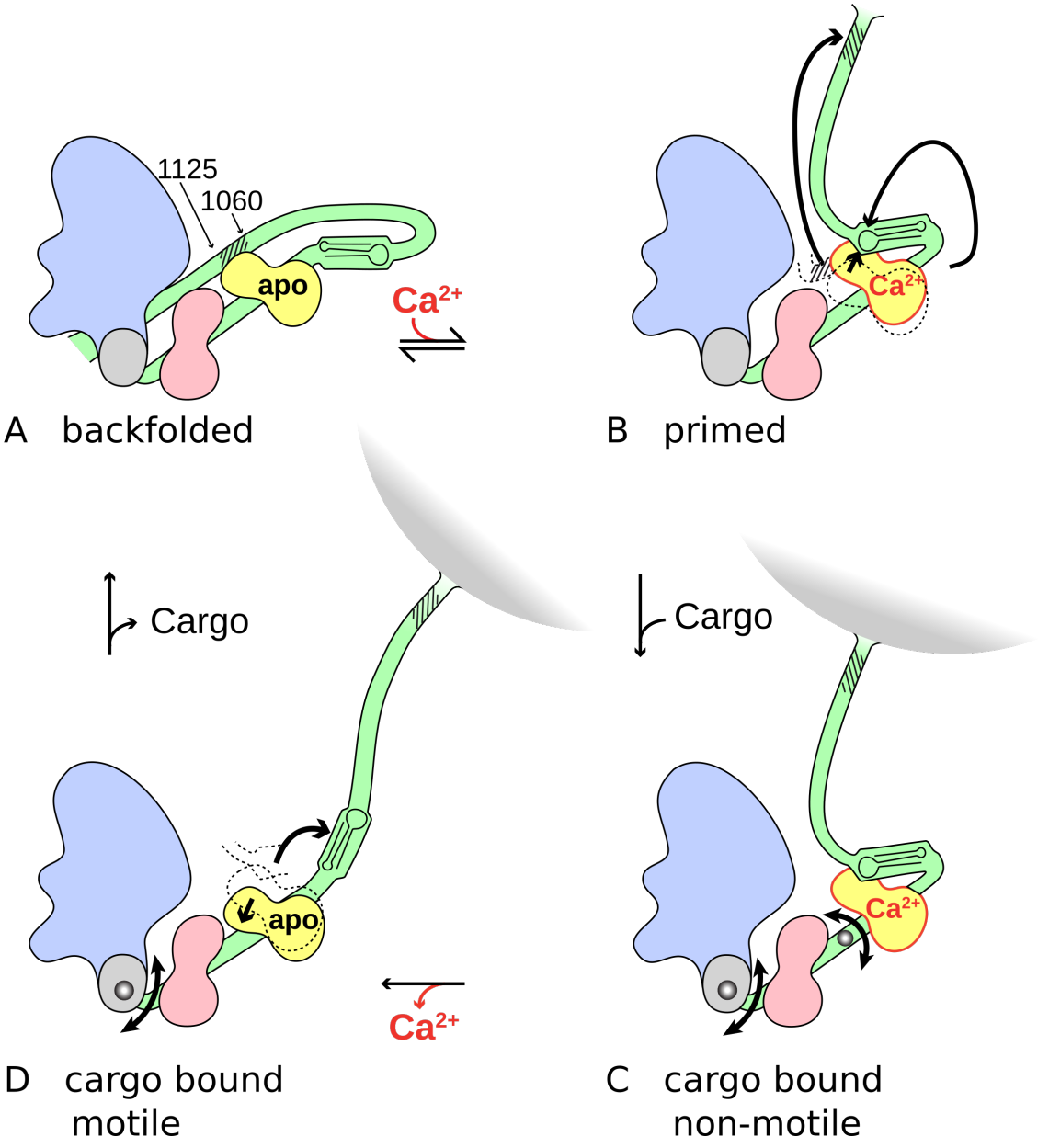


Figure 5.28: Model of two-step mechanism for calcium regulation, activation and mobilization of the myosin VI molecular motor. The backfolded, inactive myosin VI (**A**) can be primed for cargo recruiting by calcium-induced structural changes in the lever arm (**B,C**), still mechanically inactive. The return to low calcium restores the neck stability and generates either fully functional cargo bound molecular motors or, in the absence of cargo, lets myosin VI adopt its initial compact, backfolded conformation, ready for the next calcium-cycle.

5.4 Outlook

To conclude, the negative stain electron microscopy of the approximately 150 kDa myosin VI molecular motor and in particular the subsequent evaluation of these images using specialized single particle analysis methods resulted in intriguing findings. While the general shape of myosin VI was discovered to adopt mainly two conformations when absorbed to the EM grid, a more detailed investigation of the lever arm itself revealed for the first time calcium-induced structural changes of the 17 kDa small second calmodulin with a resolution of up to 18 Å. Furthermore, major structural rearrangements of the three-helix-bundle as well as the tail domain could be resolved. Although the electron microscopy itself is a static technique of measuring, which does not allow for the acquisition of moving objects directly, special image processing tools applied on the EM data can unveil underlying dynamic processes. With these methods and in combination with other techniques, including fluorescence spectroscopy and motility assays, the discovered conformational changes could be directly related to functional and mechanical properties of myosin VI. As published in Batters and Brack et al. [83], our group showed that calcium is the cellular switch that directs the rearrangement of the motor domain from a dormant, inactive state at low calcium to a cargo binding non-motile state at high calcium. The return to low calcium generates either cargo bound active motors, which relocate towards the center of the cell, or refolded inactive motors ready for the next calcium influx.

The results worked out in this chapter not only demonstrate the power of single particle analysis and negative stain EM, but also provide the basis for future high resolution studies using cryo-EM and introduce further questions to be looked into. For instance, as elaborated in section 5.2.3 (see Figure 5.19), the tail domain of myosin VI was too flexible to be observed directly. Instead, only unstructured additional EM intensities in the area of the catalytic domain indicated the presence of parts of the tail there. One of the future studies will thus concentrate on methods to resolve the tail in more detail. These include on the one hand techniques to restrict the tail's flexibility and thus increase the signal-to-noise ratio of the image-processed data, and on the other hand additional approaches, for instance small-angle X-ray scattering (SAXS) to calculate envelopes of the myosin VI tail in the absence and presence of calcium. Apart from that, it will be highly interesting to introduce nucleotide to the system in order to examine the (calcium dependent) structure of myosin VI in different stages of its power-stroke cycle and relate these to results from other methods.

5.5 Methods

The following methods part will first summarize how myosin VI was prepared for negative staining. Following a description of the staining itself and the TEM image acquisition, the focus will lie on the single particle image processing methods used to analyze single myosin VI molecules in four different conditions.

5.5.1 Preparation, staining and image acquisition of nucleotide-free full-length myosin VI

As mentioned in section 2.4, the first requirement for a successful single particle analysis using the EM was a high purity of the sample. In order to achieve that, special care was taken when myosin VI from chicken was expressed and purified with methods based on published protocols[7] [83, 89, 99]. The sample was then checked for purity by electron microscopy and stored at -80 °C by shock freezing in liquid nitrogen at a stock concentration of 2 μM. In this conditions, the sample could be stored for several months without loosing quality.
To investigate the effect of calcium and calmodulin, freshly defrosted myosin VI samples were diluted to 200 nM in different buffers containing varying concentrations of Ca^{2+} and (human) CaM (table 5.5), leading to four combinations:

- low CaM / low Ca^{2+}
- high CaM / low Ca^{2+}, with 10 μM CaM
- low CaM / high Ca^{2+}, with 100 μM Ca^{2+} (free)
- high CaM / high Ca^{2+}, with 10 μM CaM and 100 μM Ca^{2+} (free)

The concentration range of 10-100 μM ($\widehat{=}$ pCa 8 and pCa 4) of free calcium was chosen to be slightly larger than the actual physiological concentration range, pCa 7 - pCa 5. This ensured the binding of calcium to calmodulin to be saturated and high calcium and also more homogeneous populations / conformations of myosin VI molecules, which was crucial for the success of the subsequent single particle image processing. The same applied to the calmodulin concentrations, where the sample was supplied with 10 μM CaM to create high CaM-conditions.

[7] Highly purified myosin VI samples were provided by my colleges

CaM (µM)	0	10	0	10
Ca^{2+}	pCa 8	pCa 8	pCa 4	pCa 4
NaCl	25	25	25	25
TrisHCl, pH 7.5	20	20	20	20
Imidazole, pH 7.5	20	20	20	20
$MgCl_2$	5	5	5	5
DTT	10	10	10	10
EGTA	1	1	-	-
$CaCl_2$	-	-	0.1	0.1
CaM	-	0.01	-	0.01

Table 5.5: Myosin VI buffer for negative stain EM to investigate the effect of calcium and calmodulin

After an incubation time of approx. 10 minutes after mixing, the samples in each of the four conditions were applied to carbon-coated copper grids[8] that have been hydrophilized immediately prior to staining. Hydrophilization was done by glow-discharging the grids in a plasma cleaner for 30 seconds using air[9]. Several negative staining methods have been tested and as mentioned in section 2.4 the quality of the micrographs depended heavily on the staining-method, which has to be adjusted for each sample. For the study of single myosin VI molecules, the optimal staining method was the following:
A 5 μl drop sample was applied directly to the coated side of the grid and allowed to rest for 2 minutes. Meanwhile, as the following washing and staining steps have to be carried out quickly and to prevent the sample from drying out too quickly, two drops, 5 μl and 20 μl, of prepared 2% Uranylformat (UF) were placed on a piece of parafilm. Once the resting time was over, the solution was blotted away by a filter paper and the first, 5 μl drop of UF was picked up from the parafilm by touching the drop carefully with the coated side of the grid. As this step is for the purpose of washing, the drop is being blotted away instantly and the second, 20 μl drop of UF picked up in the same way. The sample was then stained for 10 seconds before any excess solution on the grid was removed. In this last step is was important to not move the filter paper once it touched the grid. It had to be held still until no significant solution-flow from the grid to the paper could be observed. Finally, the finished grid has to completely dry in air for approx. 30 minutes to prevent any liquid being transferred to the high vacuum in the TEM.
The prepared myosin VI grids were then subsequently examined with a Philips CM 100 TEM[10] at 100 kV. Micrographs were manually recorded using a CCD camera at a mag-

[8]ECF400-Cu-50-UL: TEM Grids, Carbon Film coated, 3-4 nm (Ultrathin), 400 Mesh, Cu, 50 pieces; Science Services
[9]Hendrick Dietz, Laboratory for Biomolecular Nanotechnology, Technische Universität Munich
[10]Hendrick Dietz, Laboratory for Biomolecular Nanotechnology, Technische Universität Munich

nification of 42k (which on this microscope corresponded to a resolution of 0.332886 nm/px).

5.5.2 SPA workflow for full-length myosin VI

The following part will describe the workflow for the single particle analysis of nucleotide-free FL myosin VI molecules. As it was analogous for all of the four conditions it will be described only once on the example of the low CaM / low Ca^{2+}-condition. The particle picking was done using EMAN2 [45] and the following alignment, classification and crystal structure projection matching steps were done by applying my own specifically adapted and customized scripts based on and used with the SPIRE / SPIDER software suite [24, 44, 60]. The complete set of scripts for the whole workflow (from micrograph converting to crystal structure matching (PDB Fitting)) together with various different small scripts such as "renaming" were combined in one single conformation file to be used with SPIRE. This file made it possible to repeat the whole workflow with different data sets (see figures 6.1-6.7 in the supplementary IV for the SPIRE-GUI with all needed dialogs and their contained scripts). The following description will refer to the names of the corresponding scripts (extension ".spm") shown in these figures.
Despite recording the micrographs in a manual manner, all images were one more time carefully inspected in order to discard any with insufficient quality (such as unequal staining-depth, too big / small defocus or too few / many particles). All remaining micrographs were then further processed using the two-step (manual and automated) particle picking protocol "e2boxer" implemented in EMAN2 [45] and the Spider script *Particles/particlepicking_eman_spider.spm*, in order to create a stack containing 90x90px (or 30x30nm) images presumably showing single myosin VI molecules. The following structural analysis was divided into three steps:
The first step was aimed at the detection of general shapes present in the data set. These could then be used in step 2 to more specifically and accurately search the data set for these particular shapes. And finally in step 3, the resulting global averages underwent a detailed structural analysis with various rounds of alignment and classification. If not stated otherwise, alignments were performed reference-free (see section 3.1.2) and classified using the K-means algorithm (see section 3.2.5).

5.5.2.1 Step 1: General shape finding

As a first step, in order to detect and extract general shapes present in the large data set, all images were reference free aligned (script: *Alignment/align_apsr.spm*). To reduce the influence of noise in the background and also the influence of neighboring particles that

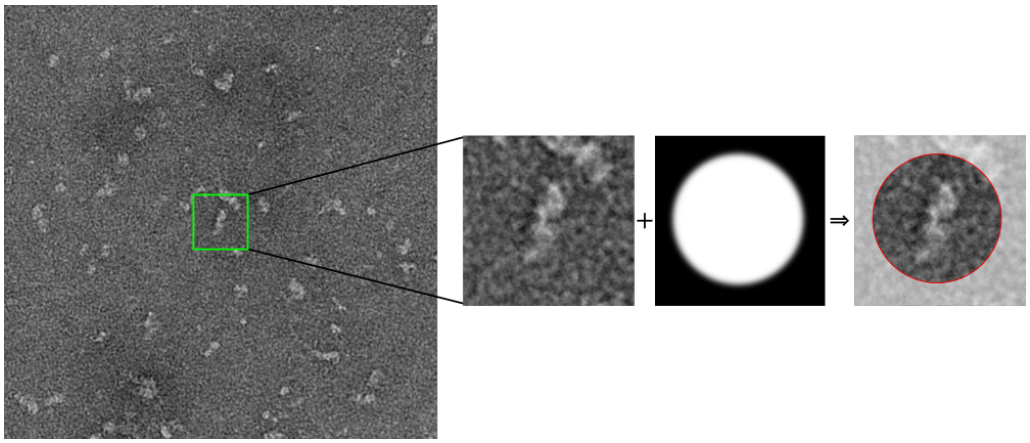


Figure 5.29: Application of a round mask with blurry edges to each individual picked particle to avoid the alignment being influenced by background and possible neighboring particles. Scale: small images are 30 nm wide.

might be partly visible in the image, a mask was applied to the data set prior to the first alignment (see Figure 5.29). As mentioned in section 3.1.2, the influence of the mask itself had to be minimal and therefore was chosen to be round and blurred. Figure 5.30 **A-D** illustrates step 1 of the SPA on the example of the low CaM-low Ca^{2+}-condition. Numbers in the upper left corner indicate the number of images contained in the average and the class number is shown in the lower left corner.

Following this first alignment (**A**), the dataset was classified by K-means (script: *Classification/clkm.spm, clkm_manyclasses.spm, selectclasses.spm*) with only the first few factors taken into account (which represented the main differences) and using a mask that roughly followed the shape of the aligned average (**B**). This classification led to four classes, whose global averages represented main differences: elongated objects, small round objects and uniformly elongated objects in different sizes. Even though the number of classes was still small, it made sense to combine classes with similar features, like *class02* and *class03*, where the lower left ends of the objects were blurred, in contrast to *class01* and *class04*. Each class was then again reference free aligned, using a mask that was more specific: In this case the shape of the mask followed the elongated shape of the first global averages, which provided a more accurate cutoff of background noise. The success of using this kind of mask could be seen in the three new global averages after the alignment, where the ray-like halo pattern around the objects (coming from noise being aligned) disappeared (**C**).

Several more rounds of alignment and classification had to be performed until sets of classes were finally obtained that were appropriate to distinguish good from bad classes (**C**). ”Good” and ”bad” referred to the presence of any kind of structure of the right size. Small round blobs or large objects filling the whole image for example were considered as ”bad”. The number of alignments and classifications depended on the homogeneity of

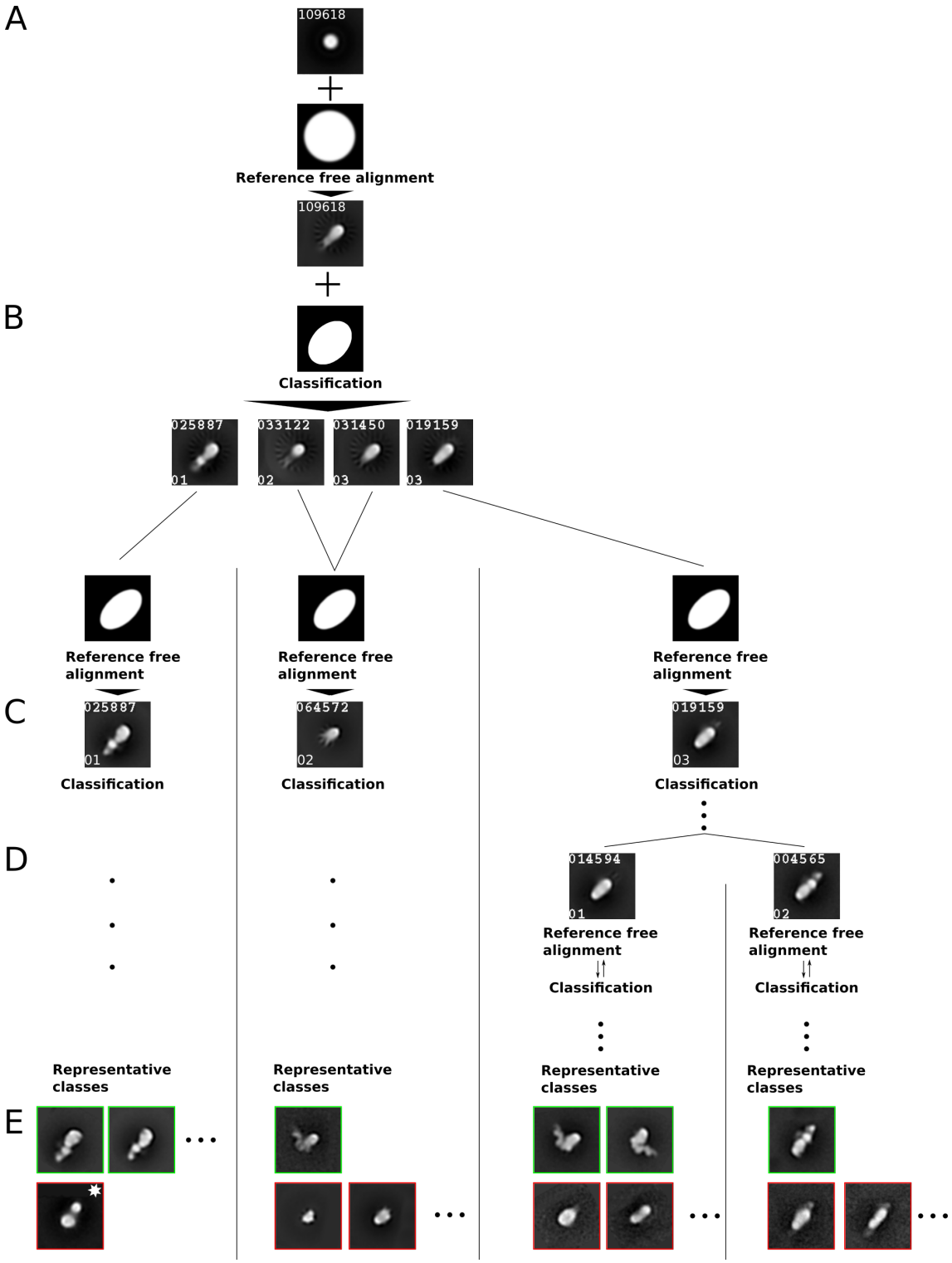


Figure 5.30: Step 1 of the SPA on myosin VI on the example of the low CaM-low Ca^{2+}-condition. **A** Reference free alignment and **C** classification into a small number of classes. **C,D** Various rounds of reference free alignment and classification. **E** Sorting of final set of classes into good (green) and bad (red). Some bad classes should be investigated further (marked with a star, see next figure). Scale: images are 30 nm wide

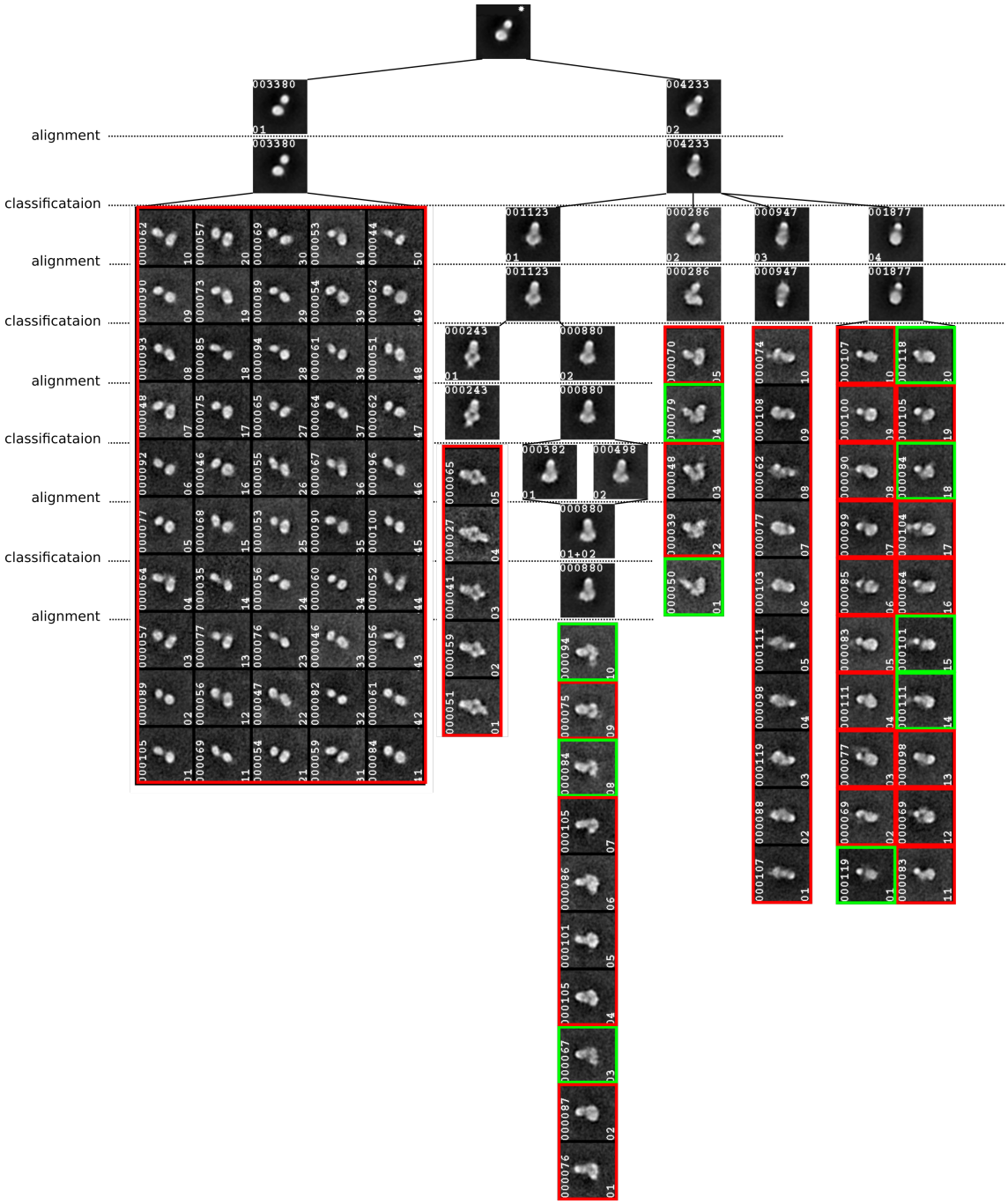


Figure 5.31: Initially bad class from previous figure (marked with a star) contained still good images. This example highlights the importance of detailed examination of presumably final classes. Scale: images are 30 nm wide

each class. *Class 03* for example showed great heterogeneity and was therefore divided into two more classes (**D**). In step (**E**) representative final class averages are shown, and highlighted to indicate whether they were used for further processing (good / green) or discarded (bad / red). Naturally, the decision whether a class is good or bad was subjective. Therefore, to be as objective as possible and to avoid good particles being discarded, or bad ones being further used, all final class averages (the good and bad ones) underwent various more, separate rounds of alignment and classification. This is illustrated in Figure 5.31 on the example of the class average marked with an asterisk in Figure 5.30 **E**. It showed how an originally bad class (red) could still contain good, usable images (green) with structures presumably representing myosin VI.
In this case, existing studies on myosin VI using the EM [93] as well as existing crystal structures of at least parts of myosin VI (PDB: 2BKI [78], PDB: 3GN4 [81]) helped to identify good and bad images.

5.5.2.2 Step 2: Multireference alignment using shapes found in step 1

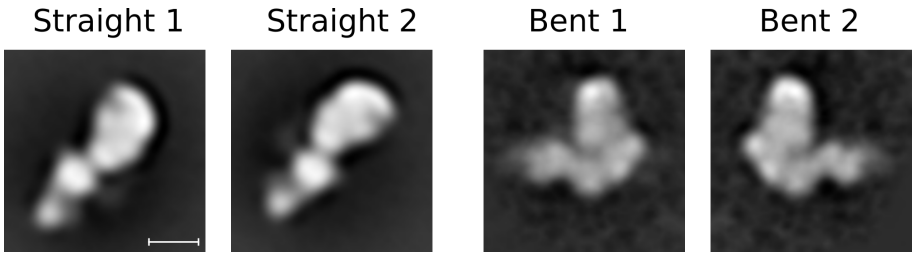


Figure 5.32: Step 2 of the SPA on myosin VI on the example of the low CaM-low Ca^{2+}-condition. The preceding classification revealed two main conformations, straight and bent, each present in mirror related orientations on the grid. Scale bar: 5 nm

Once the data set was analyzed, the usable class averages were manually compared in order to find strongly represented shapes. There were mainly four shapes detectable called *straight* and *bent*, both present in two, mirror related orientations. In Figure 5.32 the four corresponding averages are displayed, each created by averaging and aligning all images belonging to this class.
As shown in Figure 5.33 **A-G**, these averages were than used as references (**A**) to realign and reclassify the original data set by a multireference-based alignment algorithm (script: *Alignment/align_apshc_moreref.spm*). Additionally, a blob-like image was also used as reference to filter out images with missing structural information. The multireference-based alignment (with the application of a round, blurred-edge mask) led to the creation of new classes based on the best fit to the corresponding reference (**B**). As mentioned in section 3.1.2, reference-based alignment can mislead and should be treated with care. Therefore, each class was once more reference-free aligned, this time using a mask that

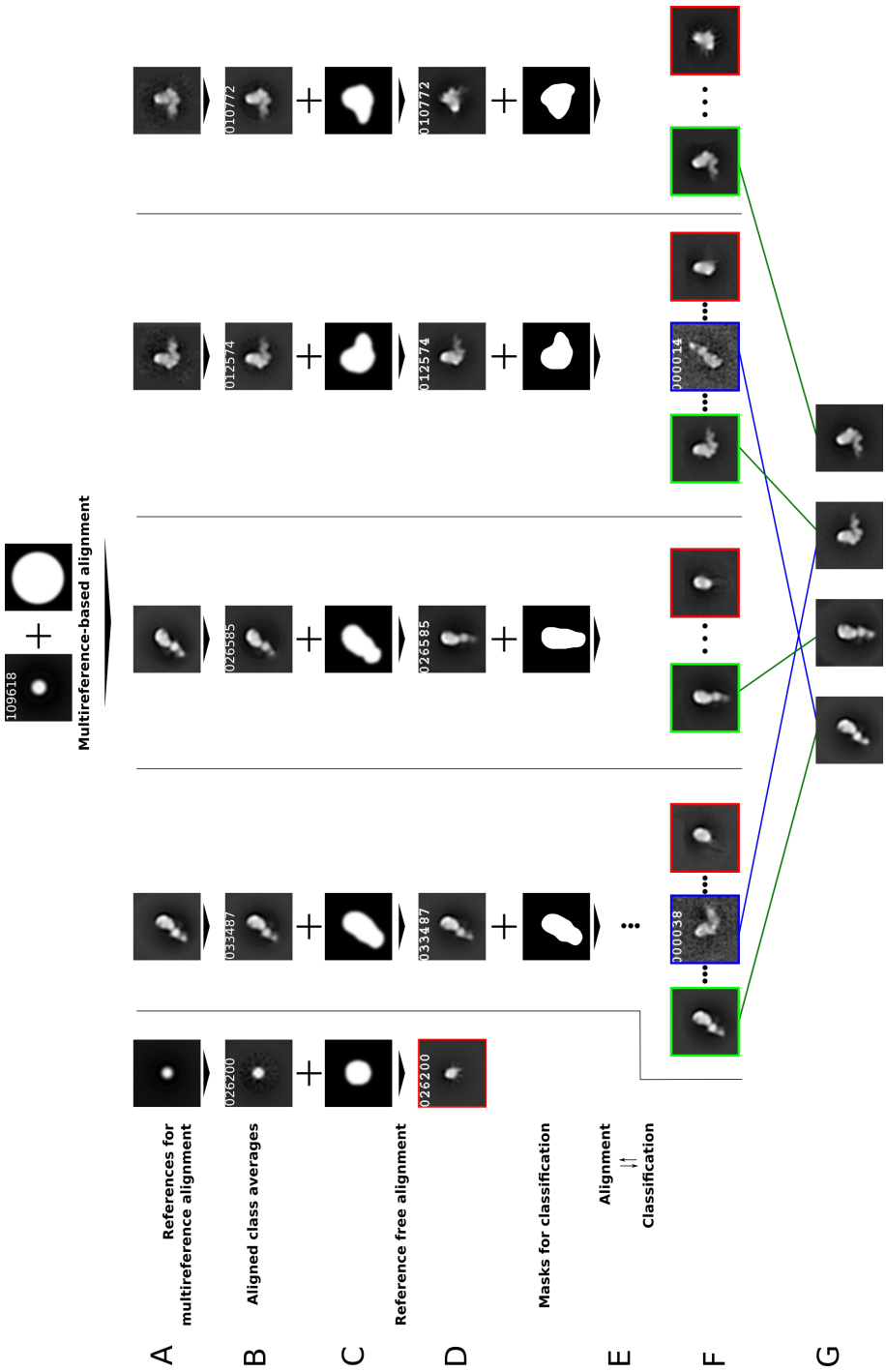


Figure 5.33: Realignment of the initial data set by multireference-based alignment (**A,B**) and using the previously found shapes and one bloblike average to filter out void as references. Repeated rounds of classification and alignments (**E**) lead to a prefinal set of classes (**F**). Sorting of misaligned images creates the final classes.

followed the shape of the myosin more accurately (**C,D**). These averages were then analyzed further by various rounds of alignment and classification (**E**) in order to detect good (green), bad (red) and mis-assigned (blue) images (**F**). Finally, after the proper sorting and reassigning of all images, four final, homogeneous classes were created that contain only images showing the discovered shapes: bent and straight, each in two mirror-related orientations (**G**).

5.5.2.3 Step 3: Detailed analysis

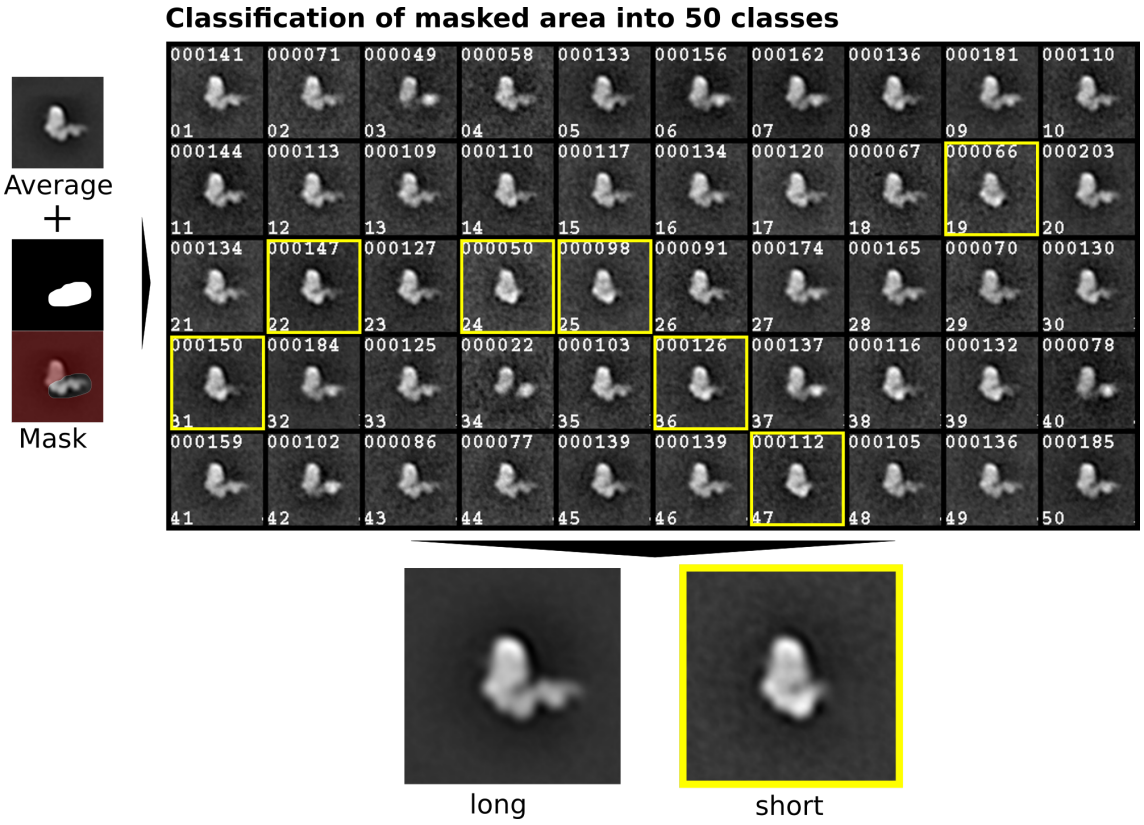


Figure 5.34: Step 3 of the SPA on myosin VI on the example of the low CaM-low Ca^{2+}-condition. Each class was further analyzed in more detailed by applying classifications on specific parts of the image. In the example shown here, the classification with respect to the tail revealed the presence of long and short molecules

Statistics:

On the basis of the shapes found in step 2, a more detailed analysis of the images was done in the final step 3. Each class was further manually classified and aligned in several iterations with the application of different masks in order to find further differences. As illustrated in Figure 5.34 on the example of the bent conformation, the classification with use of a mask that covered only the neck / tail domain revealed the presence of *long* and *short* molecules (with respect to the length of the neck / tail domain). Investigation, analysis and counting of all images / classes led to the statistics shown in Figure 5.13 in

the results-section 5.2.2.

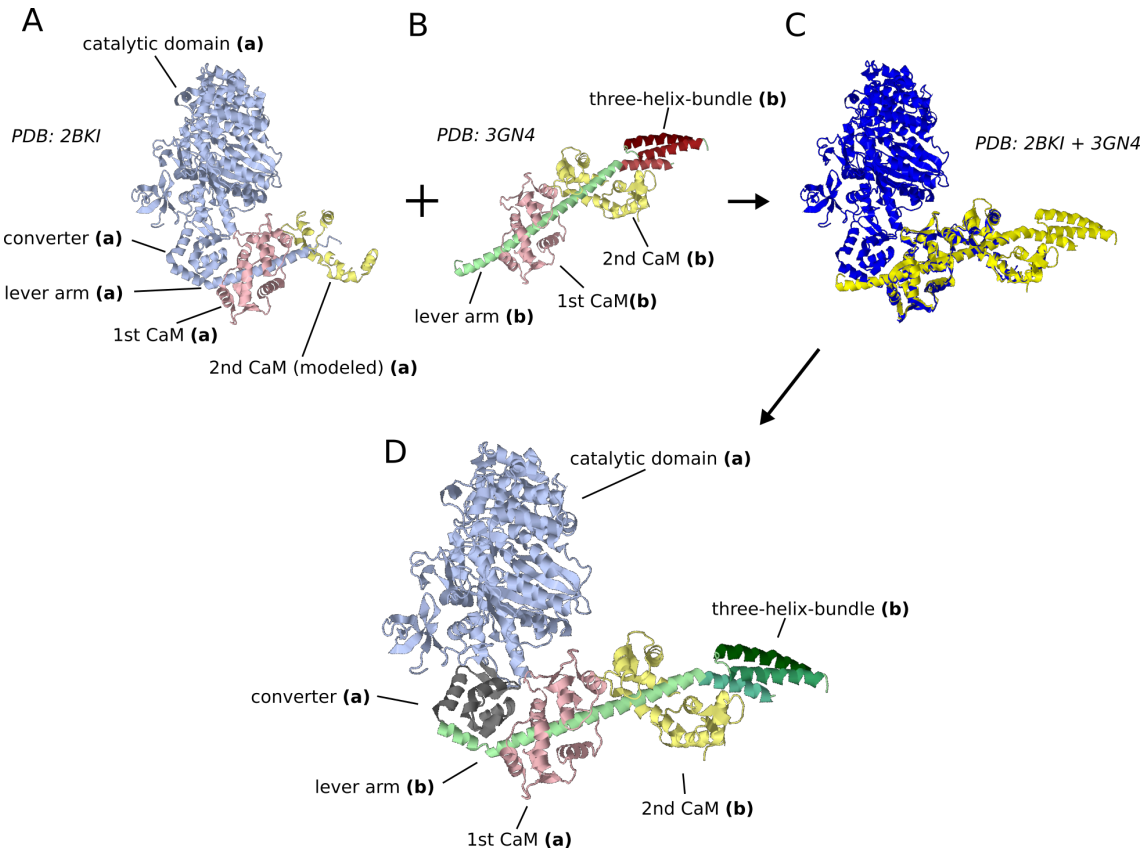


Figure 5.35: Creation of a nucleotide-free myosin VI crystal structure model by combining to two existing structures: The FATCAT rigid pairwise alignment method aligned a structure of the myosin VI lever arm (**B**), including the 1st and 2nd calmodulin and the 3HB (PDB: 3GN4), with a structure of myosin VI containing the catalytic domain, converter region and first calmodulin (**A**, PDB: 2BKI). **C** Resulting superposition of bot structures in orientation that yielded the best fit. **D** Color coded final crystal structure model.

Crystal structure projection matching: The interpretation of the final nucleotide-free myosin VI EM averages was done by crystal structure projection matching based on the description in section 4.2 and using the scripts *PDB Fitting/pdbtospi.spm, findorientation.spm, writejmolmacros.spm.* For the creation of the model, the generating of the 799 projections, the cross correlation and the editing and viewing of the structures the software SPIDER [44], Jmol [56] and the DeepView Swiss-PDBViewer [57] were used. The crystal structure used for the bent conformation was achieved by the following process (Figure 5.35 **A-D**): The most complete X-ray structure of myosin VI (in-rigor) available at the time was a structure comprising the catalytic domain and the neck region with two calmodulins (**A**), PDB: 2BKI [78], of which however the second calmodulin was not resolved and modeled on. Furthermore, a published structure of only the lever arm with

two calmodulins and the following three-helix-bundle was available (**B**), PDB: 3GN4 [81]. As depicted in (**C**), these two structures were aligned using the FATCAT rigid pairwise alignment method [58]. The P-value (see section 4.1), that describes the significance of similarity between two structures, was 1.79E-06 in this case, which indicates great similarity. By combining the catalytic domain, converter and first calmodulin of 2BKI and the lever arm, second calmodulin and three-helix-bundle of 3GN4, the final structure of myosin VI (aa 1-847) was modeled (**D**). This model was than used for the crystal structure projection matching: A total number of 799 low-pass-filtered 2D projections was created to simulate negative stain TEM images. After adjusting the size of the image projections to the scale of the real EM images (0.332886 nm/px), each projection was cross-correlated with the real EM averages and the normalized CCC values were sorted in a descending order to identify the 2D projection that matched the best. The projection in the optimized orientation was then overlaid onto the EM average in order to identify the different parts of the real EM image and relate them to the different parts of the modeled structure.

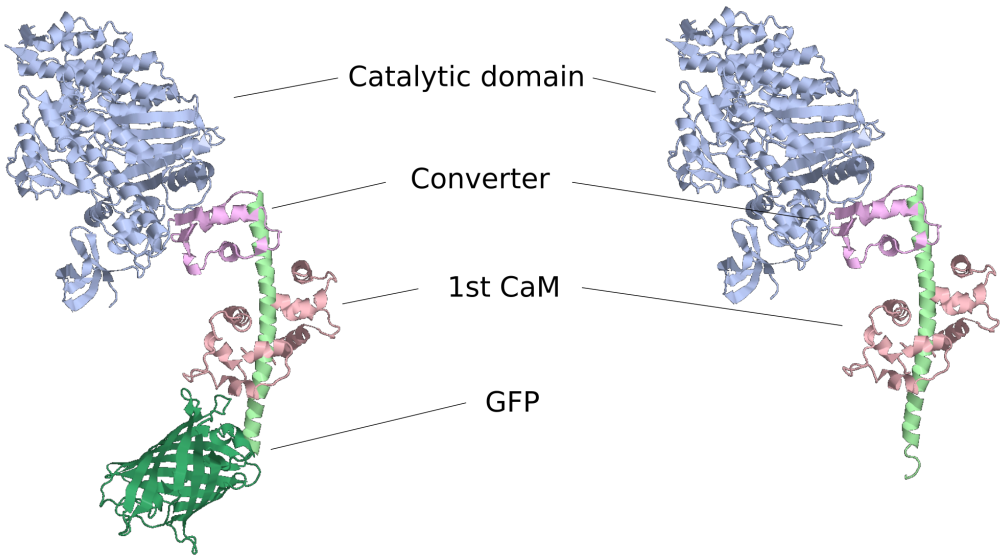


Figure 5.36: Crystal structure of myosin VI in the presence of nucleotide ADP and a GFP-molecule on its C-terminus (PDB: 4ANJ). For the purpose of crystal structure projection matching, the GFP molecule was deleted.

Since the straight conformation of nucleotide-free myosin VI described in the results section (see section 5.2 and Figure 5.34) showed similarities with published EM images of myosin VI in the presence of nucleotide [93], a crystal structure of ADP-myosin VI was also used for crystal structure matching: PDB: 4ANJ [100]. This structure comprised - apart from the catalytic domain, converter, neck and first CaM - a GFP-molecule. For

the purpose of matching to the EM averages, the GFP-part was deleted (aa 1012-1228) as shown in Figure 5.36.

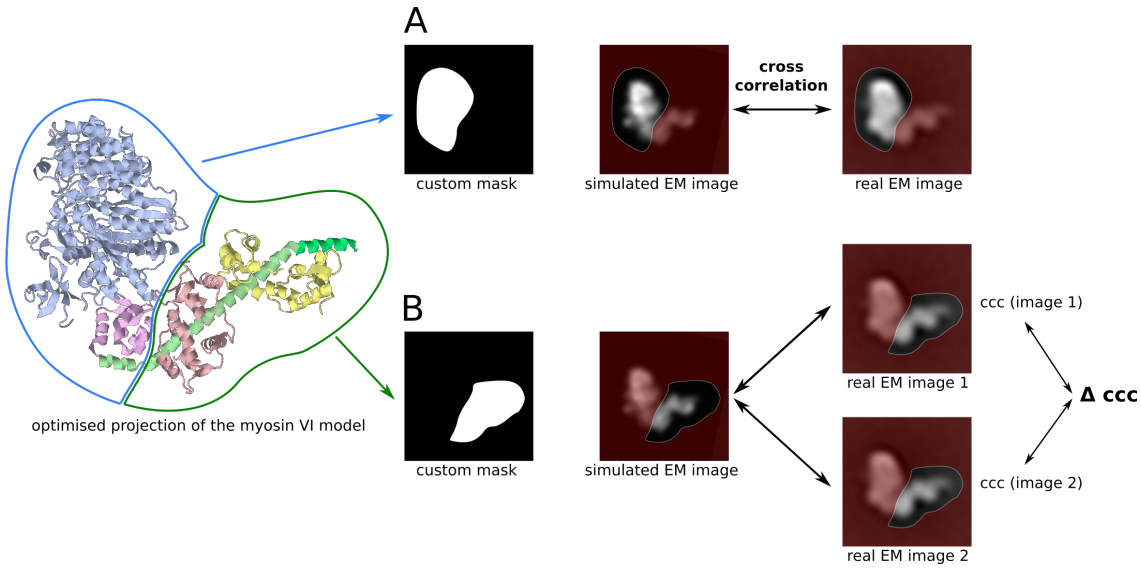


Figure 5.37: Detailed analysis of EM averages using the crystal structure model in its optimized orientation. The cross correlation between model and EM average can be calculated separately for different areas using custom masks (**A**,**B**).This comparison localizes good and bad fitting areas (**A**) and, in the case of more than one EM average, can identify structural differences including the degree of these variations (Δccc).

Once the optimized projection for the modeled nucleotide-free myosin VI structure was found, the usage of masks allowed for a more detailed analysis, as illustrated in Figure 5.37. A custom mask was applied to cross correlate only parts of the molecule with the simulated 2D projection and thus quantify the fit of only specific parts. The mask was created by drawing an outline of the desired area on top of the crystal structure and convert it into a binary image. In the figure shown here, one mask analyzed the head+converter domain only (blue outline), the other analyzed the lever arm with two calmodulin and the 3HB (green outline). With this method, the fit of specific areas could be identified (Figure 5.37 **A**) and more importantly the differences between two or more EM class averages localized by calculating the differences of ccc-values (Figure 5.37 **B)**.

Image modification for better analysis: For better analysis of the bent conformation of myosin VI in low and high Ca^{2+}, the EM averages were processed to improve contrast as well as to accurately outline the shape of the molecule. Figure 5.38 illustrates the workflow. Using ImageJ [101], the lookup-table (LUT) was changed to *spectrum* with the following custom entries: *176-255* to *36/0/255*. Furthermore, the levels were adjusted to *100-288* and *68-186* for the low Ca^{2+} and high Ca^{2+} condition respectively. To outline the shape, first the image was binarized by thresholding using Gimp and the resulting

outline (= edges) were vectorized using Inkscape.

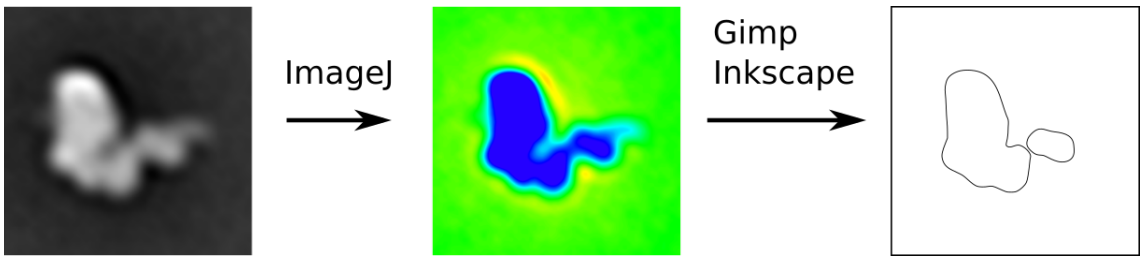


Figure 5.38: Workflow for creating outlines of EM averages and color coded versions of the EM average. The color coded image was obtained by the application of a different LUT (lookup-table) in ImageJ. The outlines were created by applying a threshold to the average and converting it into a contour using Gimp and Inkscape.

Creation of movies to visualize flexibility: To identify and, if present, visualize flexibility of different parts of myosin VI, masks that cover only specific parts of the molecule were applied to the images prior to the alignment. As mentioned before, these masks needed to have blurred edges to minimize influences on the alignment. As the example in Figure 5.39 shows, the alignment of the low Ca^{2+} bent conformation of myosin VI with a mask covering only the catalytic domain (**B**) revealed great detail of this part, while the area that has been excluded from the alignment appeared blurred (marked with a white arrow in **C**), indicating the presence of heterogeneity of this part (here neck / tail domain). A subsequent classification of this area made it possible to separate the images into classes that were homogeneous in themselves. By inspecting all class averages and selecting (**E**, marked in green) only those with sufficient quality (meaning resolving the structural parts of the neck / tail domain), the averages could be sorted into a sequence of images (**F**) and played as a movie (script: *Classification/sort_movie.spm*). This movie enabled to explore of the presence, extend, location and pathway of flexibility of certain parts of the molecule. In order to express the movie in one picture, one can outline and overlay the extreme orientations.

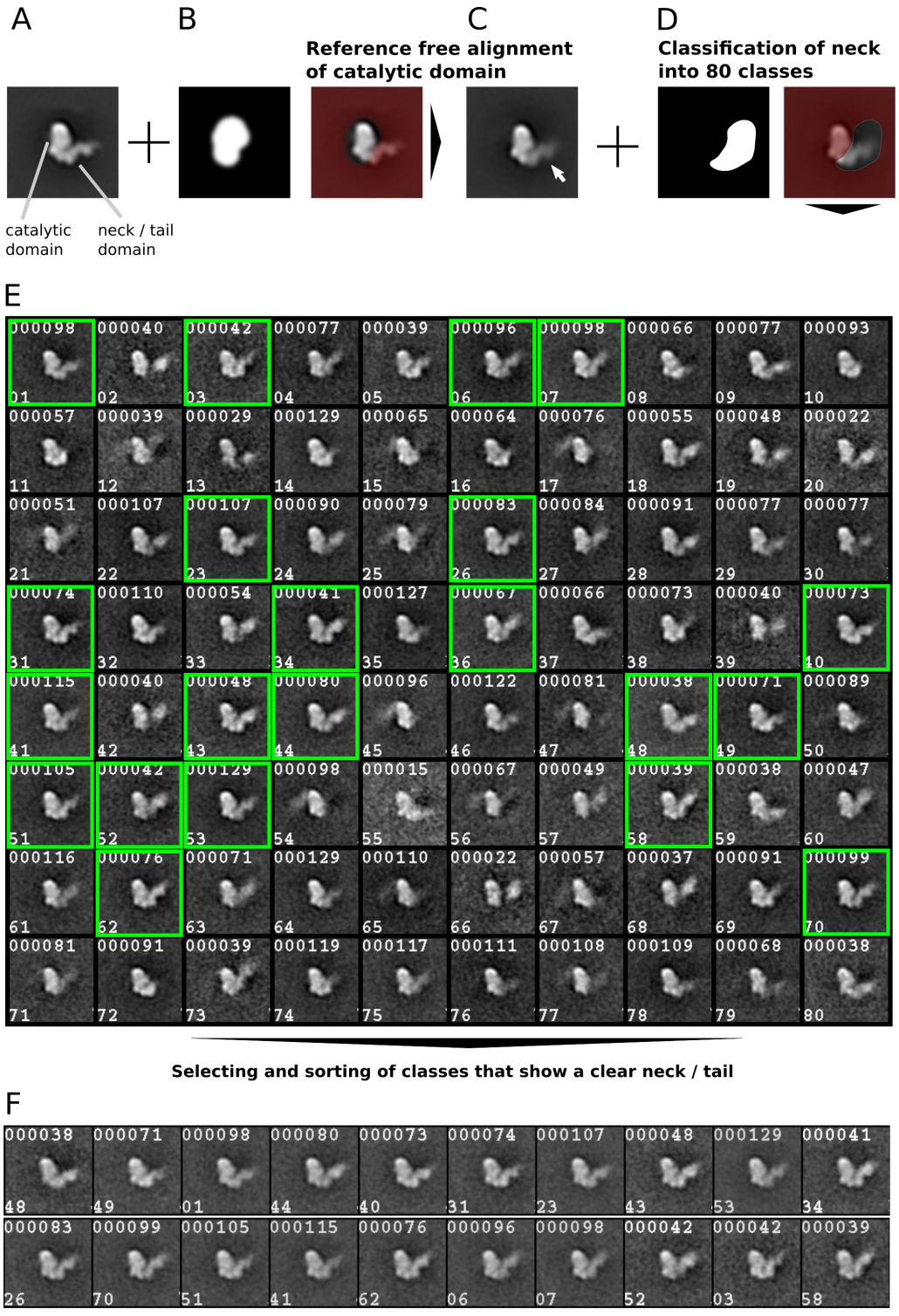
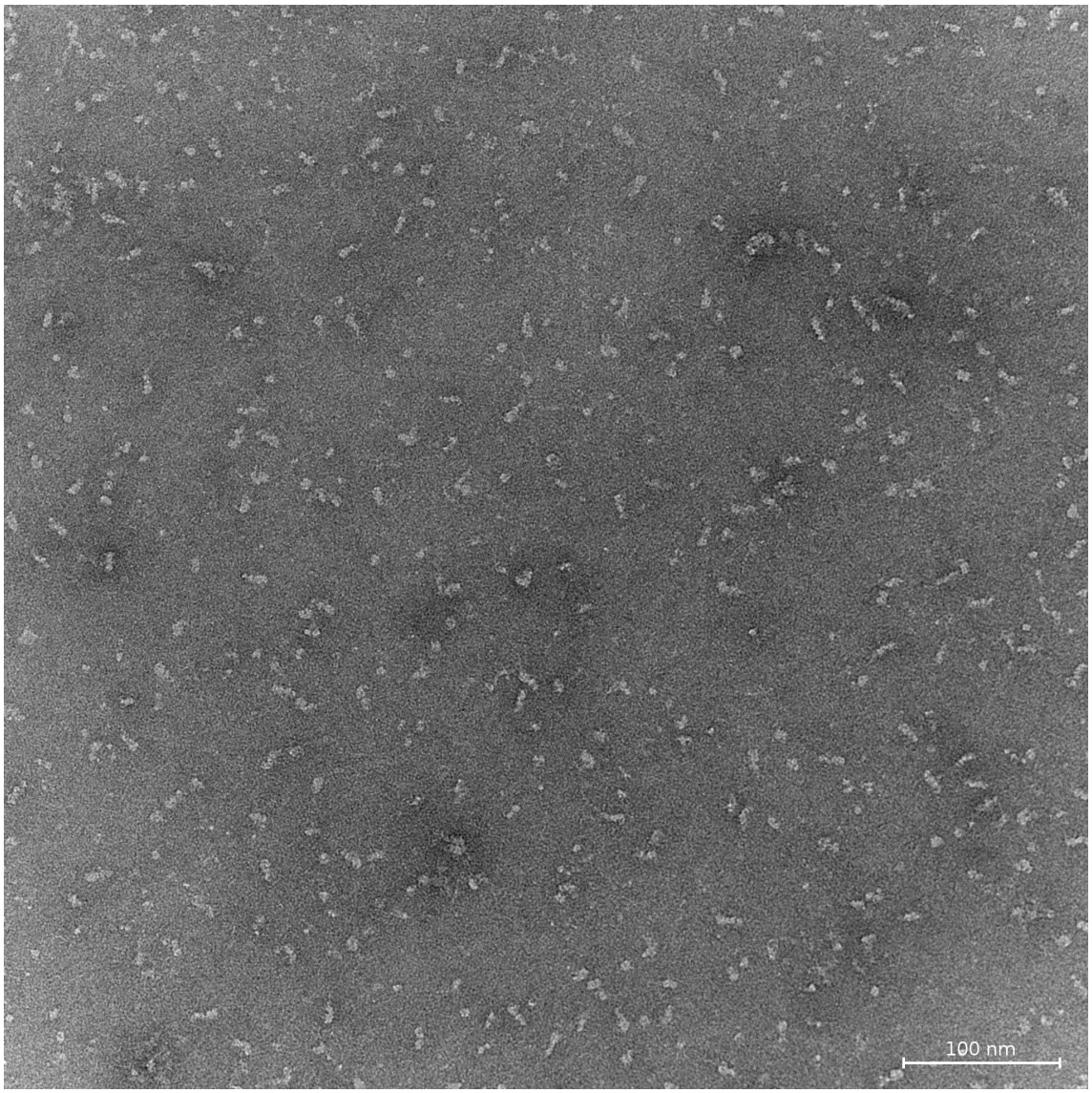


Figure 5.39: Analysis of flexible parts within EM classes by alignment of assumed fixed parts and classification of the remaining intensities, followed by the creation of a movie. Here the workflow is demonstrated on the bent conformation of myosin VI: **A-B** Application of a mask during the reference-free alignment of the data set to align only structural parts that were predicted to be fixed. **C** Unaligned areas appear blurred (arrow) depending on the heterogeneity of the data set. **D-E** Classification of images with respect to the parts excluded from the mask and sorting classes with sufficient resolution (green) into a series that can be played as a movie (**F**).

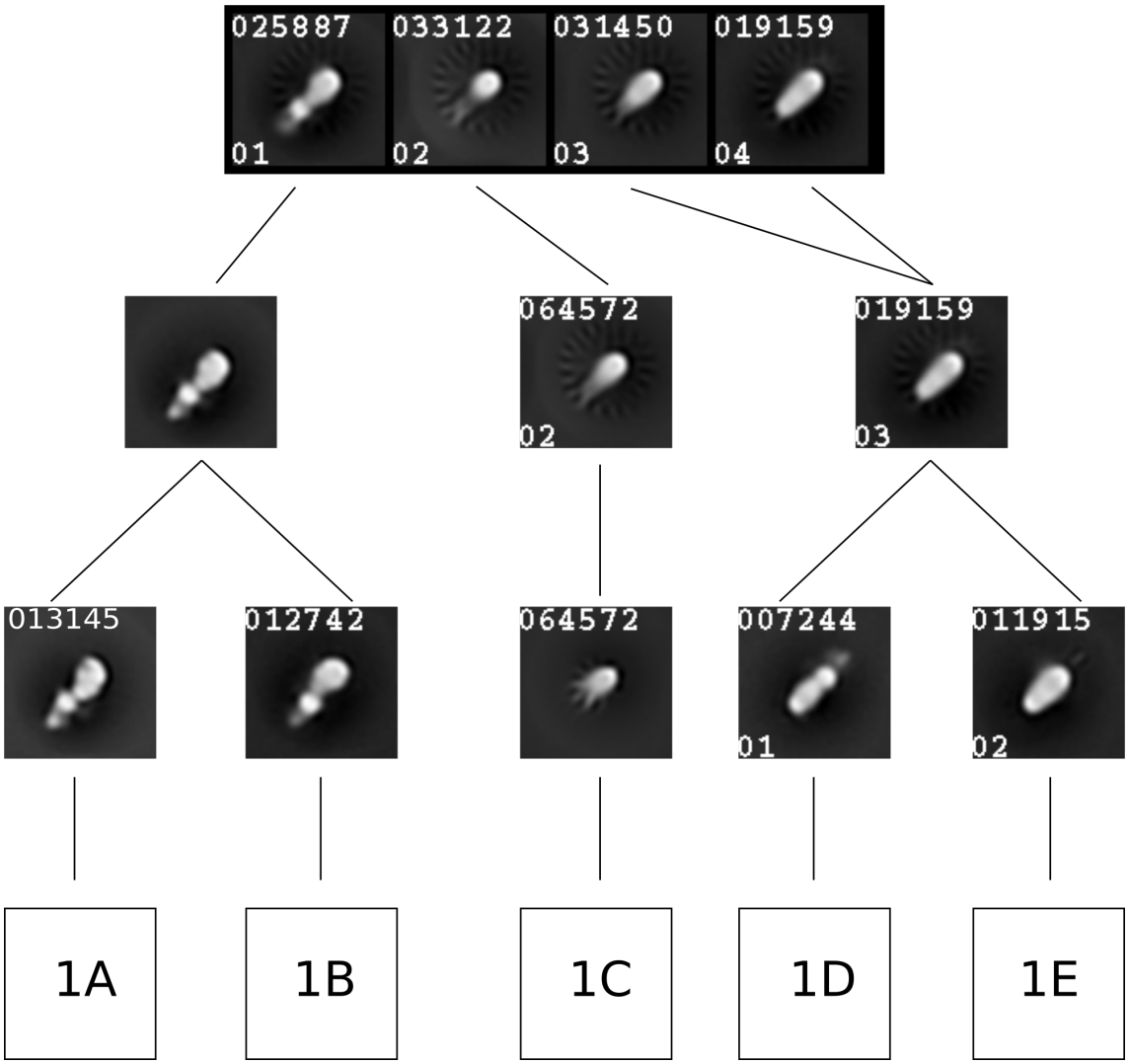
5.6 Supplementary material

The following section contains the supplementary material that has been referred to in this chapter. For each of the four conditions of myosin VI (with respect to calcium and calmodulin concentrations), one representative negative stain electron micrograph is given, followed by the complete alignment and classification chart. These charts are divided into an overview showing the general classification and alignment and also representative class averages of the final classification steps:

1) low calmodulin / low calcium

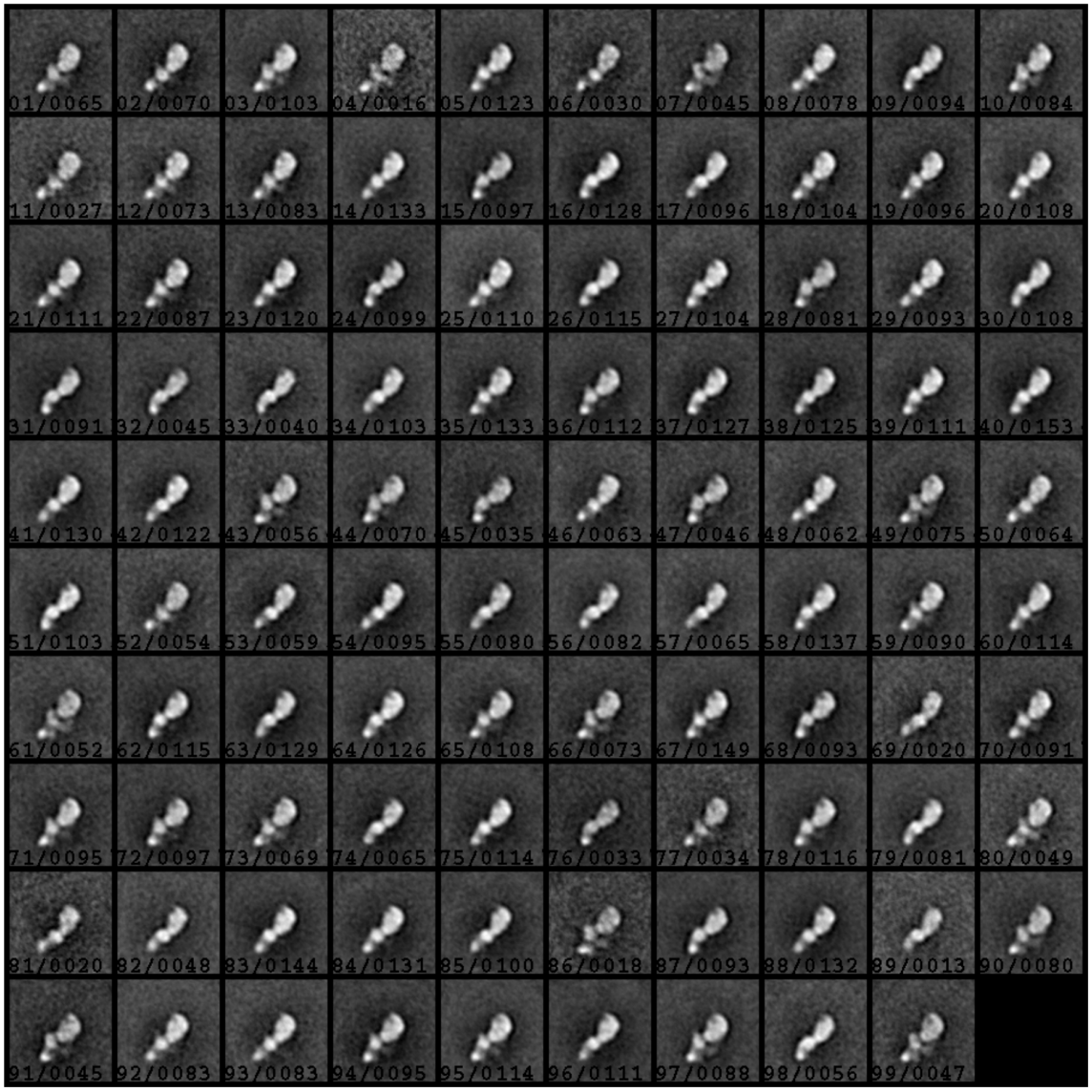


Supplementary Figure 5.1: Example negative stain electron micrograph of myosin VI at low calcium and low calmodulin. Scale bar: 100 nm

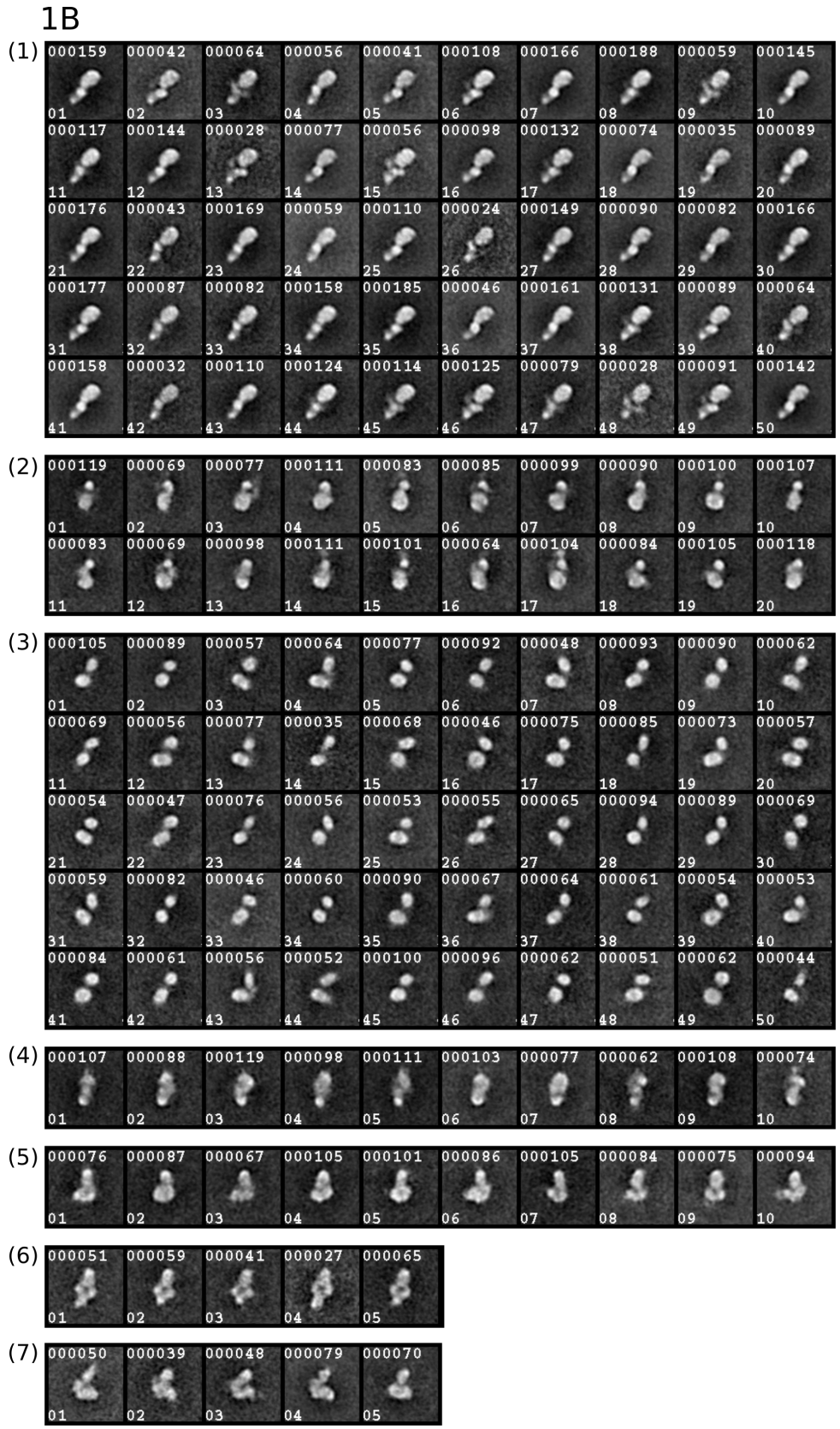


Supplementary Figure 5.2: Alignment and classification chart of 109618 images of myosin VI molecules at low calcium and low calmodulin. The boxed letters represent further charts seen in the figures 5.3-5.7. Lower left number: class number. Upper left number: number of images in average. Scale: EM averages are 30 nm wide.

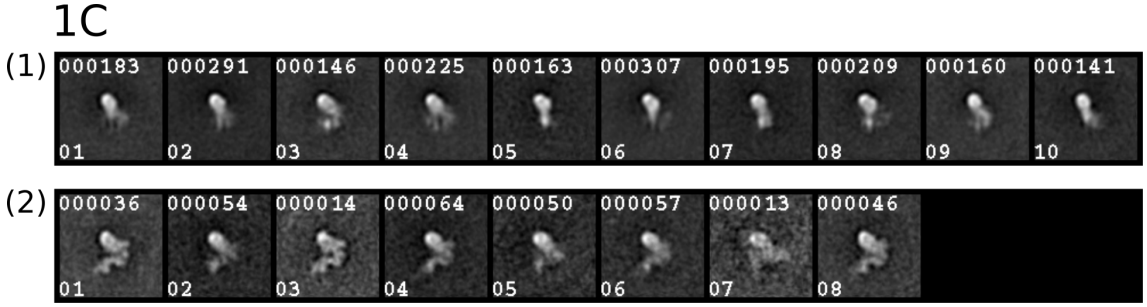
1A



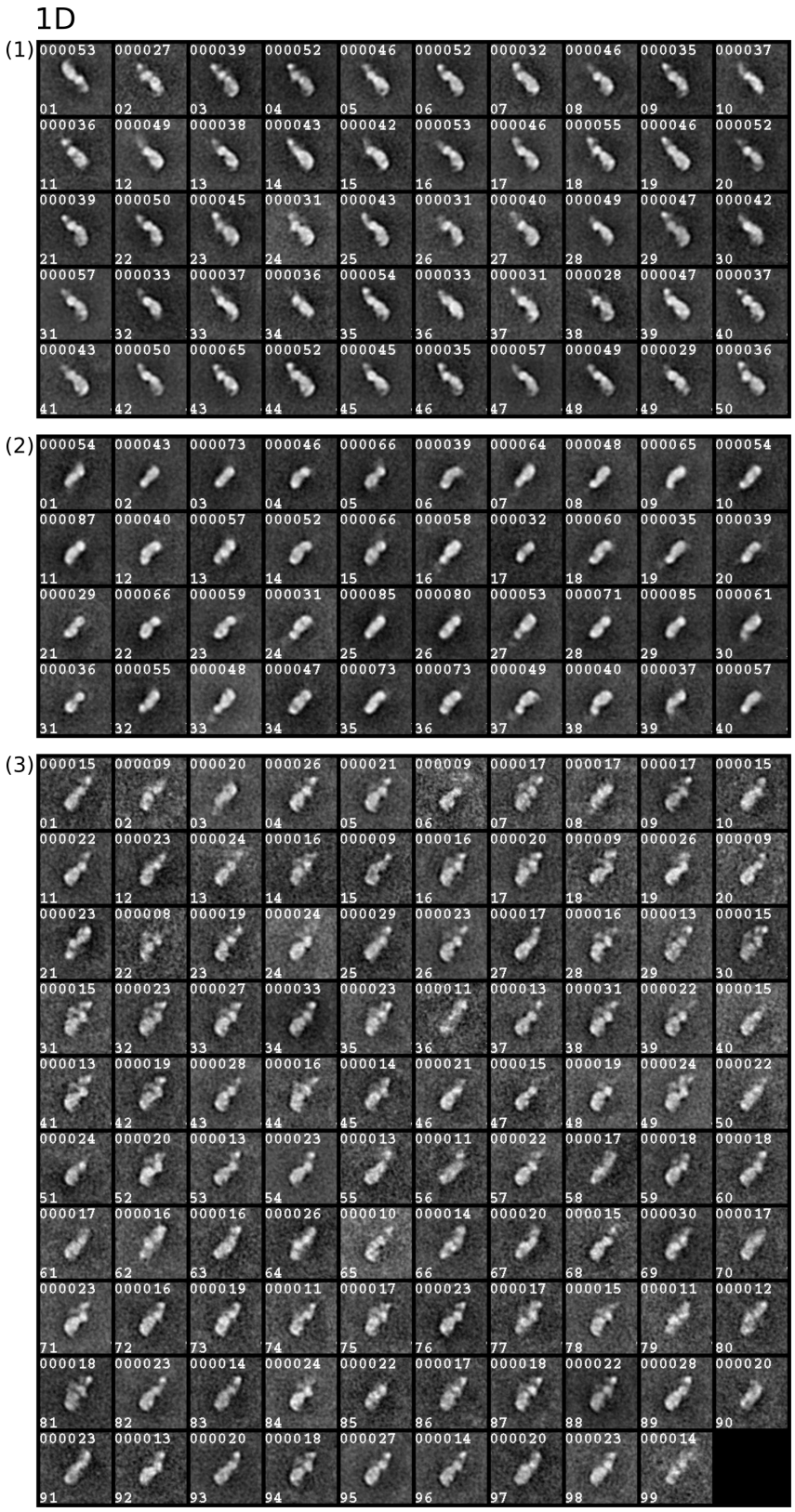
Supplementary Figure 5.3: Classification of 13145 images of myosin VI molecules at low calmodulin / low calcium in class 1A (see Figure 5.2) into 99 subclasses. All averages show myosin VI in a left facing straight conformation. Lower left number: class number. Lower right number: number of images contained in the class average. Scale: EM averages are 30 nm wide.



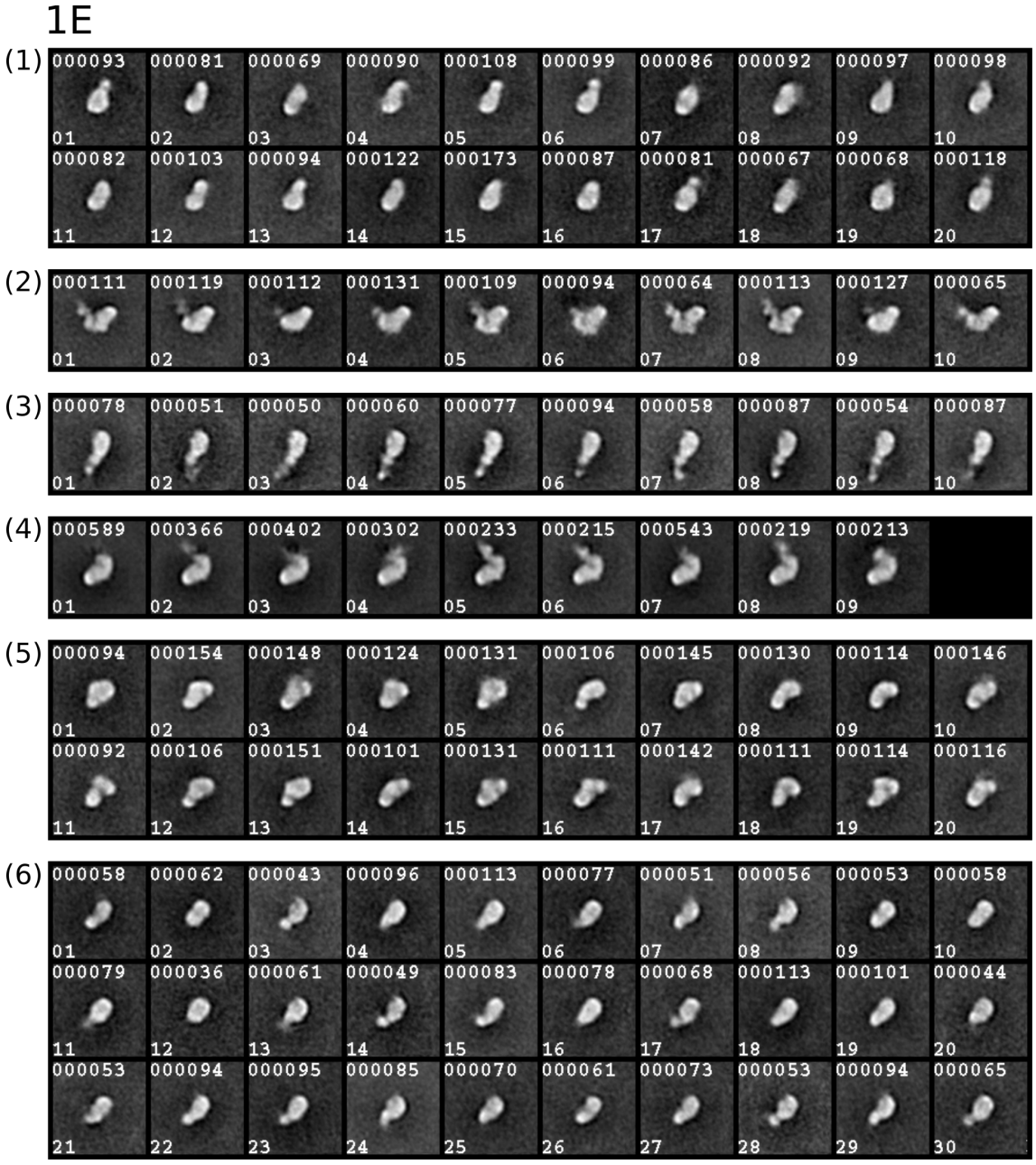
Supplementary Figure 5.4: Classification of 12742 images of myosin VI molecules at low calmodulin / low calcium in class 1B (see Figure 5.2) into 7 classes, each further classified into 5-50 subclasses. This classification revealed different shapes of myosin VI in class 1B: (1) long right facing straight conformation, (2-4) undefined structures of two connected and separate blobs, (5) long right facing bent conformation, (6) undefined structure, (7) long left facing bent conformation. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.



Supplementary Figure 5.5: Classification of 64572 images of myosin VI molecules at low calmodulin / low calcium in class 1C (see Figure 5.2) into 2 classes, each further classified into 8-10 subclasses. This classification revealed different shapes of myosin VI in class 1C: (1) short right facing bent conformation, (2) long right facing bent conformation. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

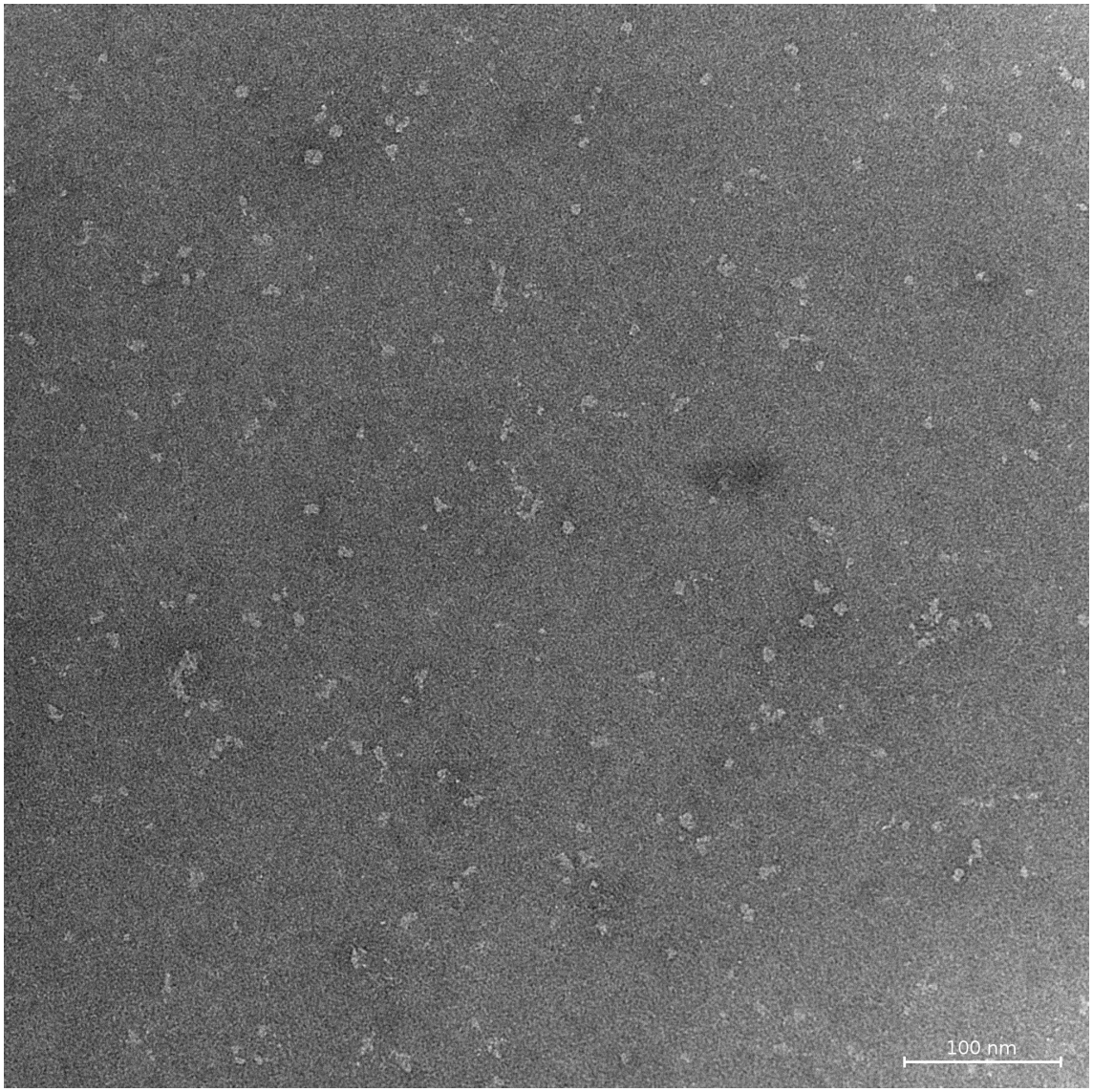


Supplementary Figure 5.6: Classification of 7244 images of myosin VI molecules at low calmodulin / low calcium in class 1D (see Figure 5.2) into 3 classes, each further classified into 40-99 subclasses. This classification revealed different shapes of myosin VI in class 1D: (1) long right facing straight conformation, (2) short right and left facing straight conformation, (3) long left facing straight conformation. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

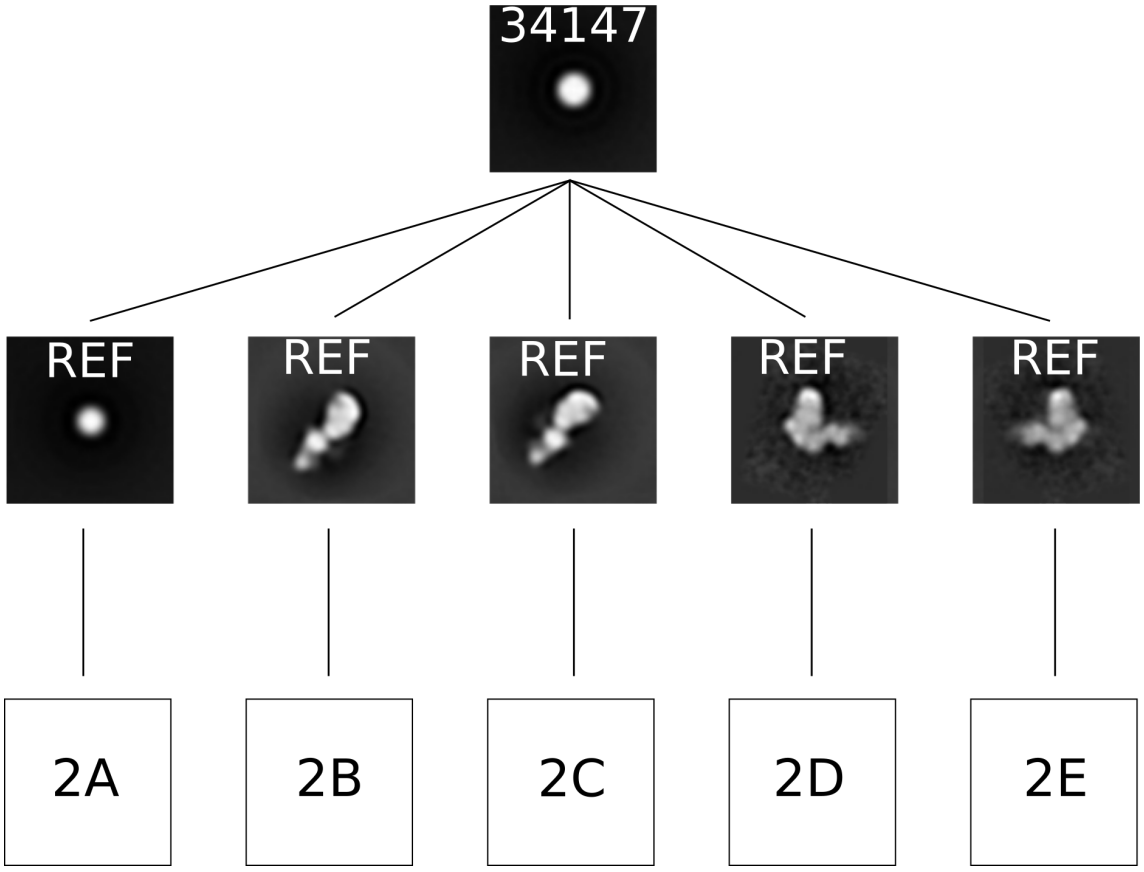


Supplementary Figure 5.7: Classification of 11915 images of myosin VI molecules at low calmodulin / low calcium in class 1E (see Figure 5.2) into 6 classes, each further classified into 9-30 subclasses. This classification revealed different shapes of myosin VI in class 1D: (1) undefined structure, (2) long right facing bent conformation, (3) long left facing straight conformation, (4) short and long left facing bent conformation, (5) short left and right facing bent conformation, (6) short left and right facing straight conformation. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

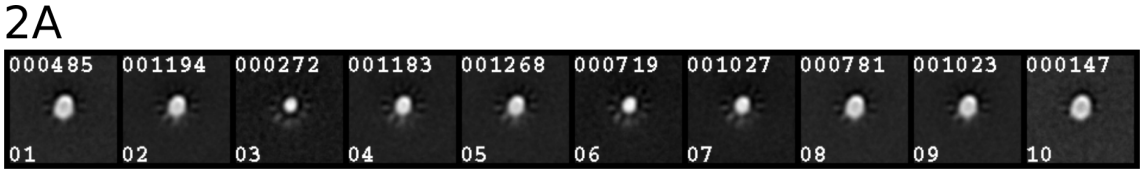
2) low calmodulin / high calcium



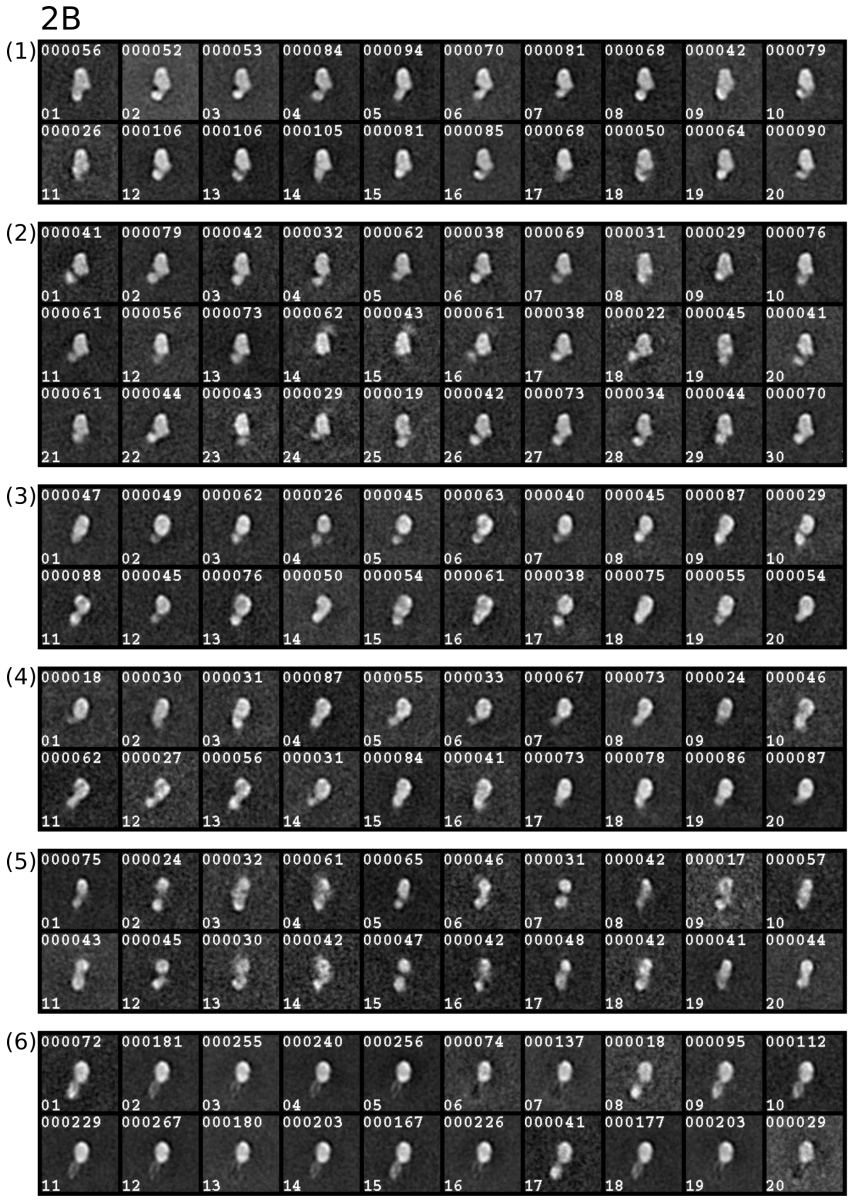
Supplementary Figure 5.8: Example micrograph of myosin VI at high calcium and low calmodulin. Scale bar: 100 nm



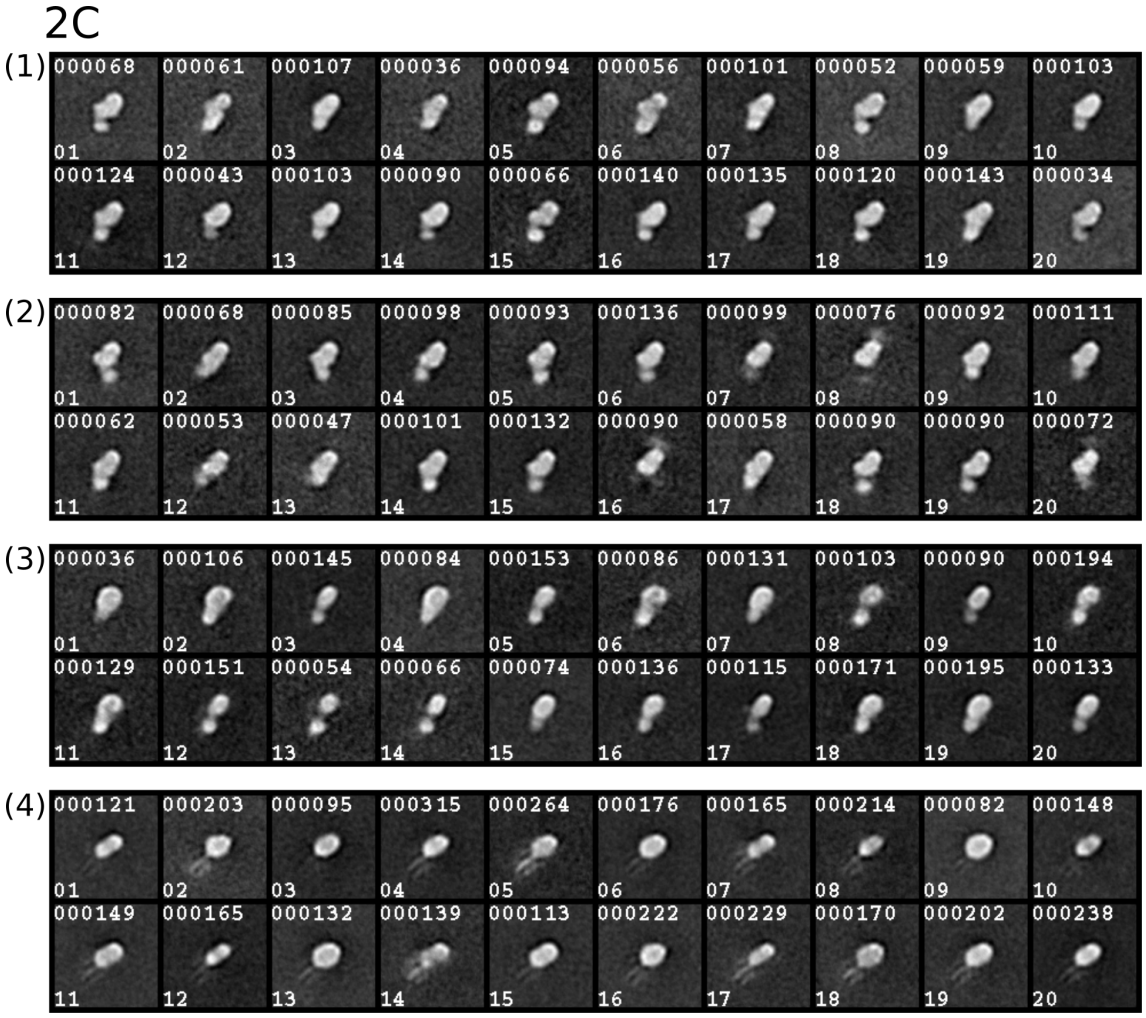
Supplementary Figure 5.9: Multireference-based alignment and classification chart of 34147 images of myosin VI molecules at high calcium and low calmodulin using five different reference images. The boxed letters represent further charts seen in the figures 5.10-5.14. Scale: EM averages are 30 nm wide.



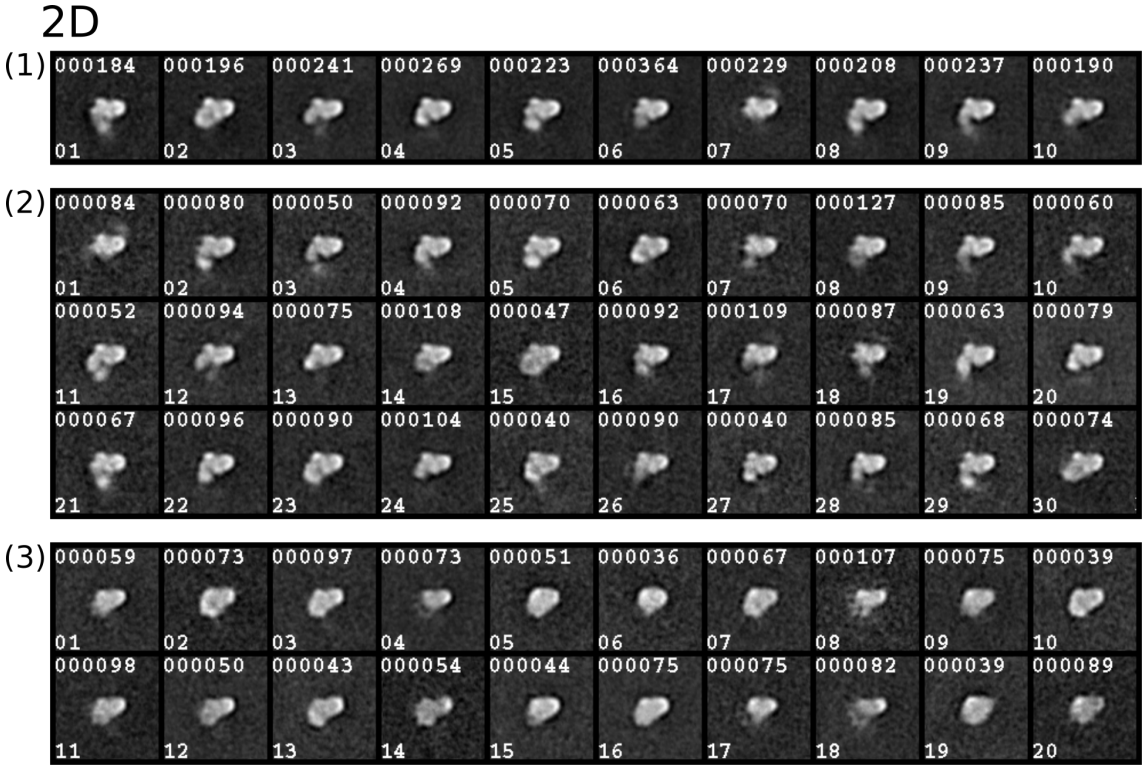
Supplementary Figure 5.10: Classification of representative aligned images (blob-like image as reference) of myosin VI molecules at low calmodulin / high calcium in class 2A (see Figure 5.9) into 10 subclasses. All averages show myosin VI in a short right facing bent conformation. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.



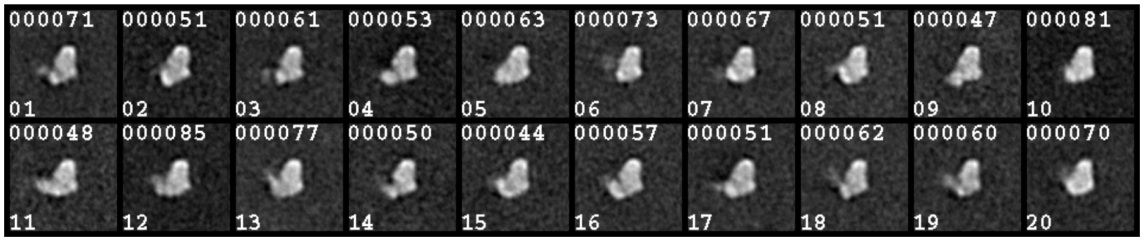
Supplementary Figure 5.11: Classification of representative aligned images (long left facing straight conformation as reference) of myosin VI molecules at low calmodulin / high calcium in class 2B (see Figure 5.9) into 6 classes, each further classified into 20-30 subclasses. This classification shows myosin VI in mainly short right and left facing bent conformations. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.



Supplementary Figure 5.12: Classification of representative aligned images (long right facing straight conformation as reference) of myosin VI molecules at low calmodulin / high calcium in class 2C (see Figure 5.9) into 4 classes, each further classified into 20 subclasses. This classification shows myosin VI in mainly short right and left facing bent conformations. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

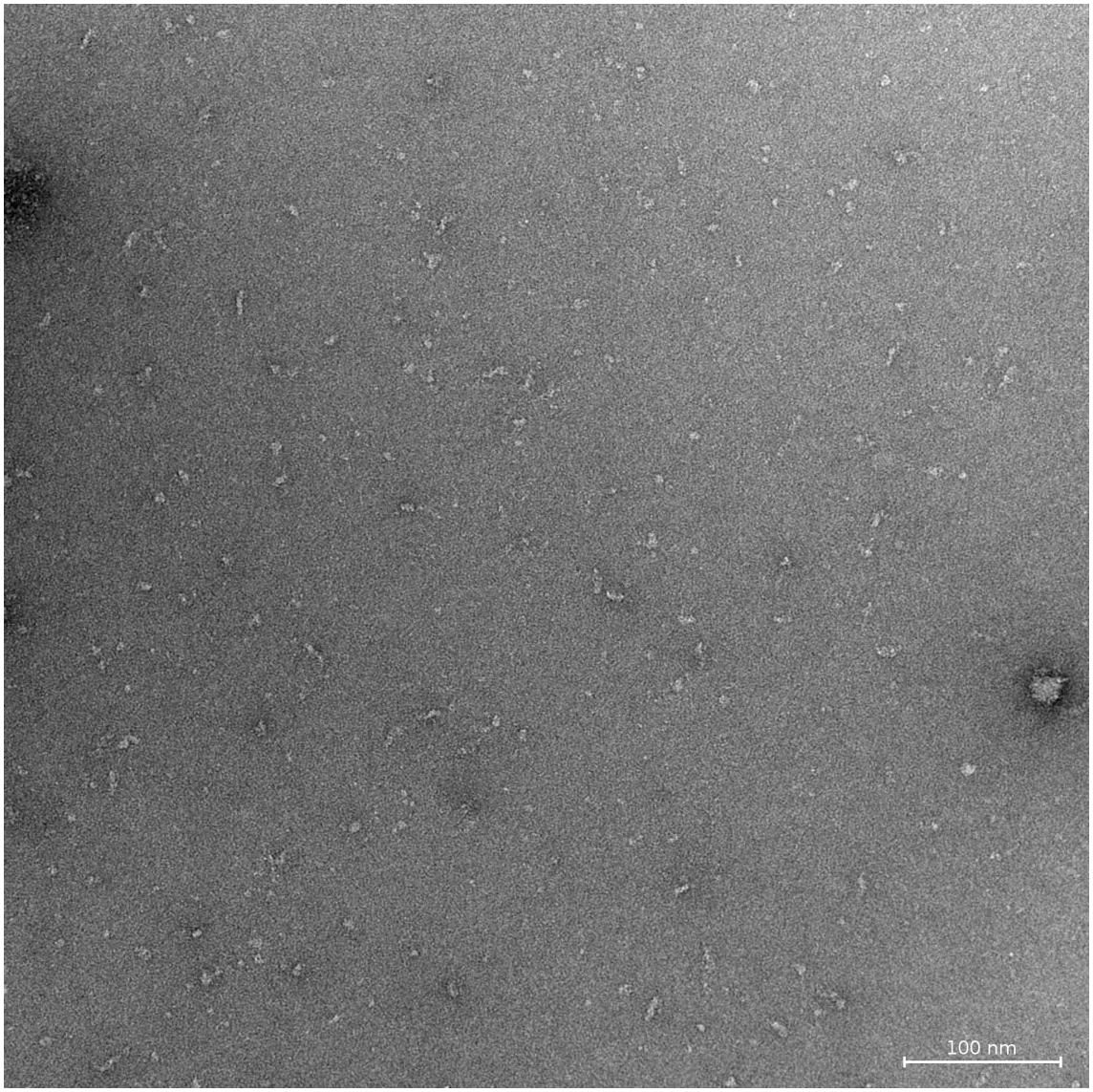


Supplementary Figure 5.13: Classification of representative aligned images (long left facing bent conformation as reference) of myosin VI molecules at low calmodulin / high calcium in class 2D (see Figure 5.9) into 3 classes, each further classified into 10-20 subclasses. This classification shows myosin VI in mainly long and short right and left facing bent conformations. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

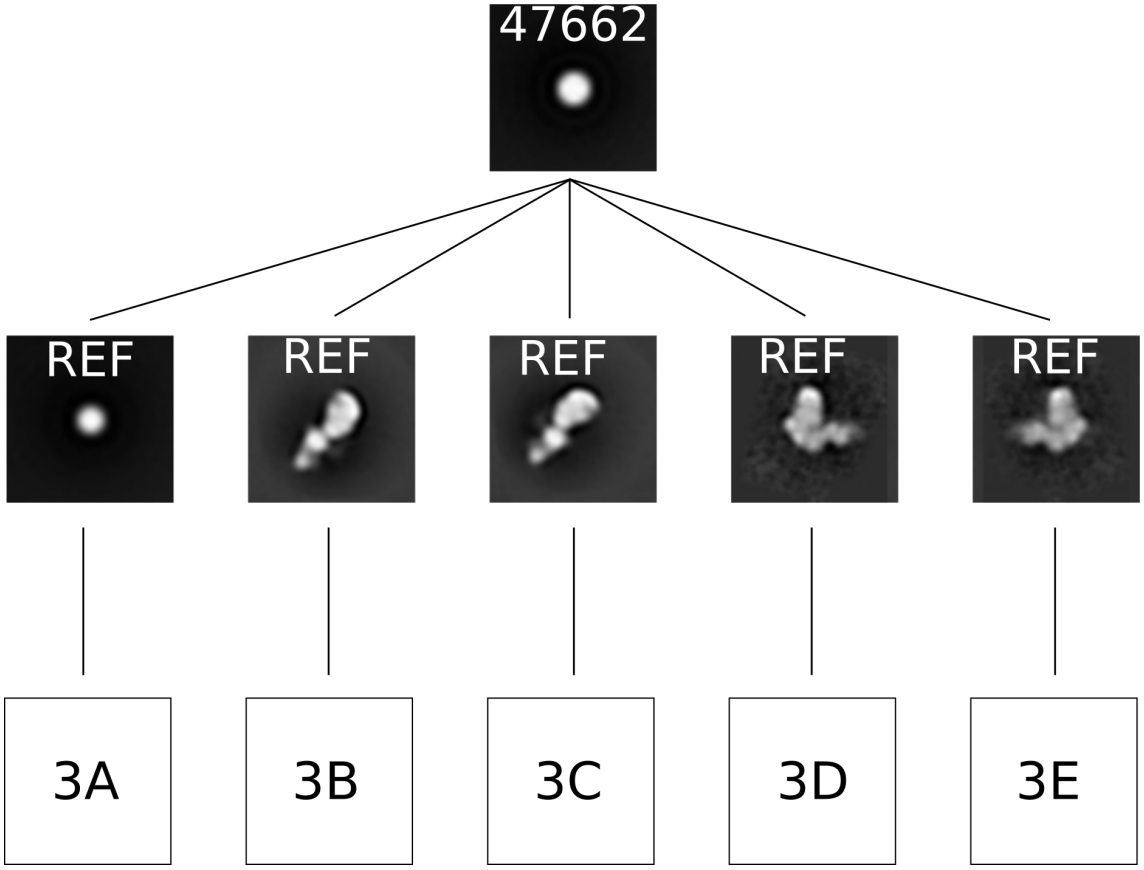


Supplementary Figure 5.14: Classification of representative aligned images (long right facing bent conformation as reference) of myosin VI molecules at low calmodulin / high calcium in class 2E (see Figure 5.9) into 20 subclasses. This classification shows myosin VI in long right facing bent conformations. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

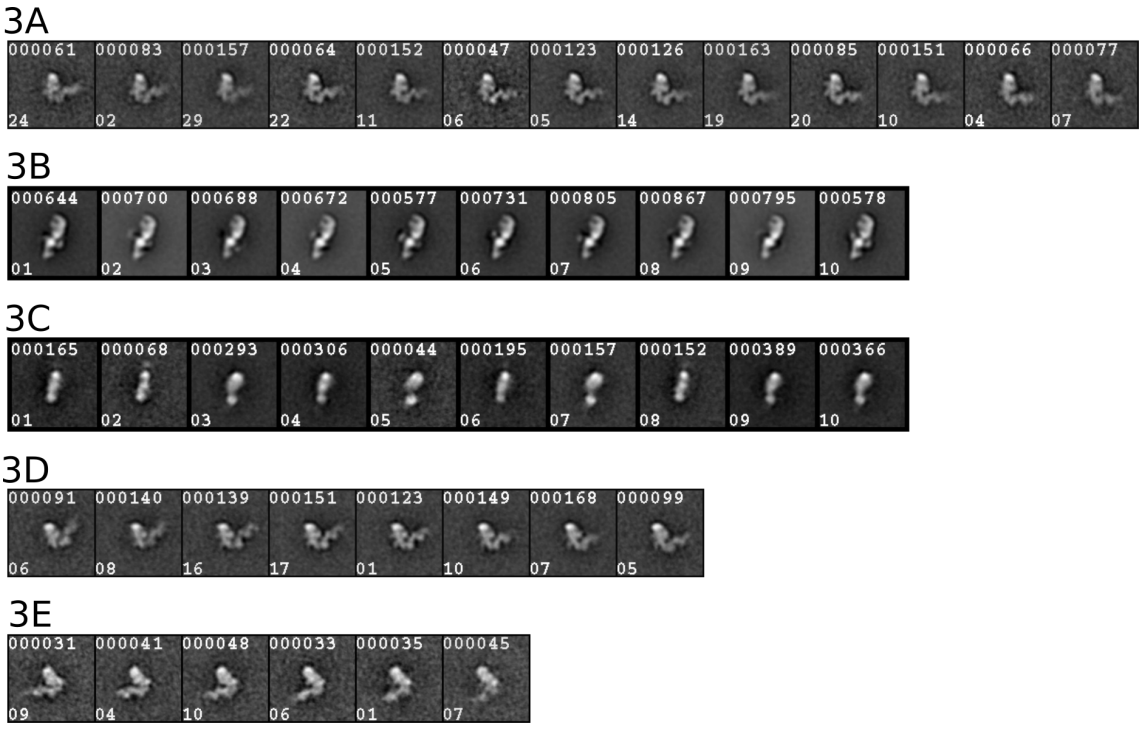
3) high calmodulin / low calcium



Supplementary Figure 5.15: Example micrograph of myosin VI at low calcium and high calmodulin. Scale bar: 100 nm

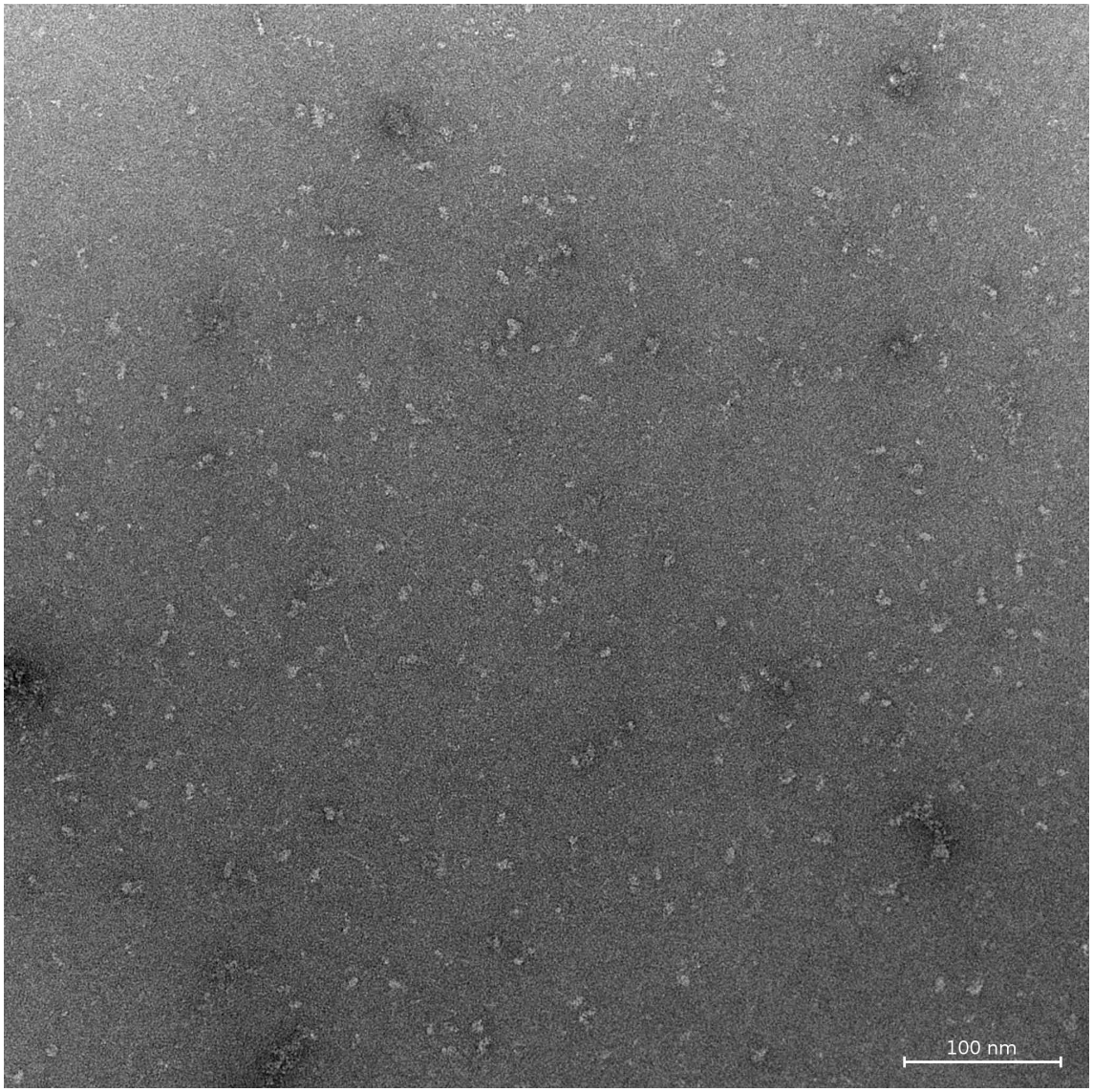


Supplementary Figure 5.16: Multireference-based alignment and classification chart of 47662 images of myosin VI molecules at low calcium and high calmodulin using five different reference images. The boxed letters represent further charts seen in Figure 5.17. Scale: EM averages are 30 nm wide.

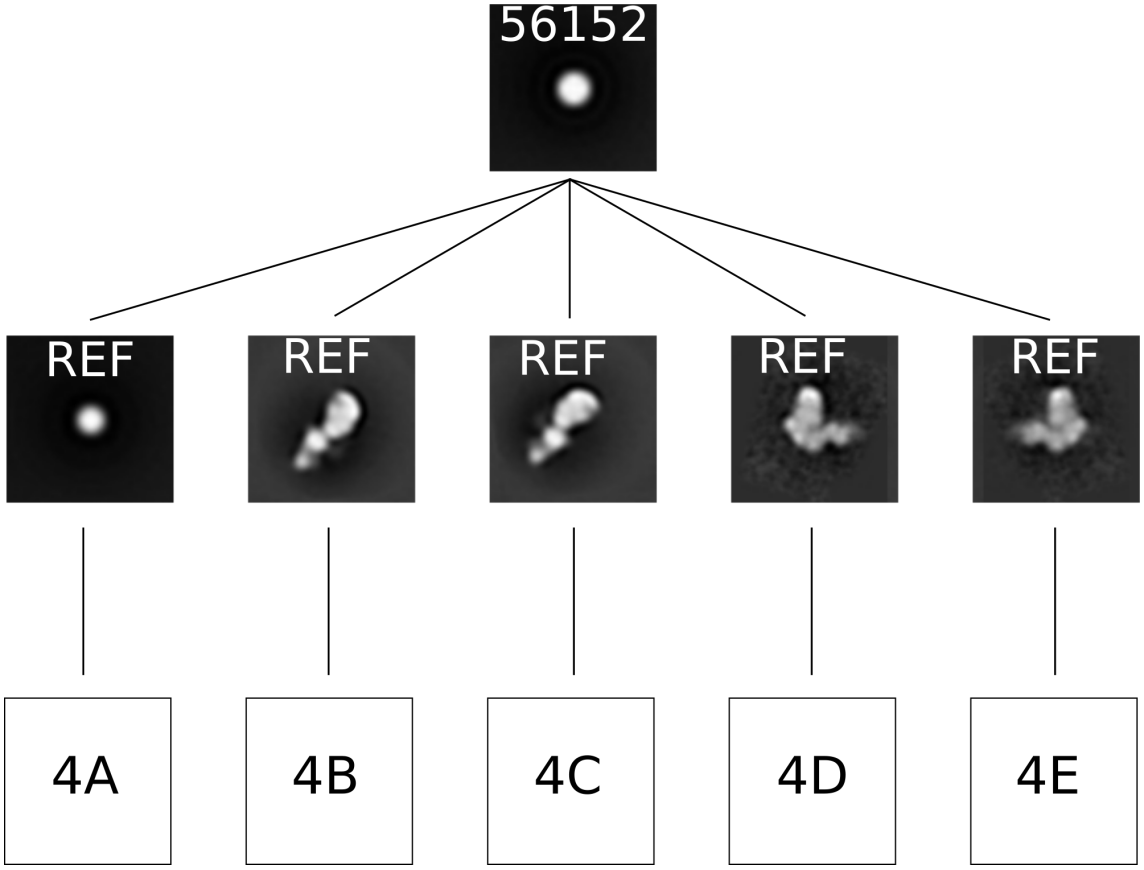


Supplementary Figure 5.17: Classification of representative aligned images (with five different reference images, see 3A-3E in Figure 5.9) of myosin VI molecules at high calmodulin / low calcium into 6-13 subclasses. These classifications show myosin VI in all of the four conformations. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

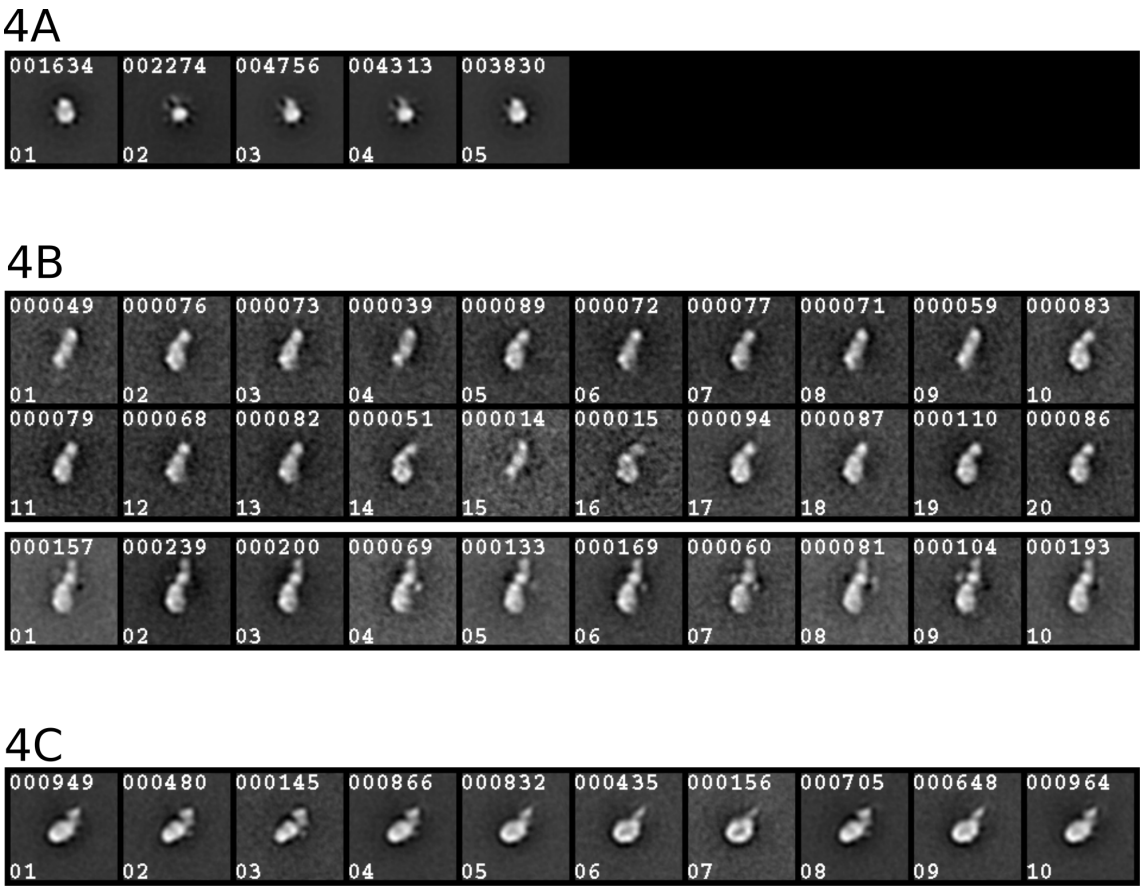
4) high calmodulin / high calcium



Supplementary Figure 5.18: Example micrograph of myosin VI at high calcium and high calmodulin. Scale bar: 100 nm



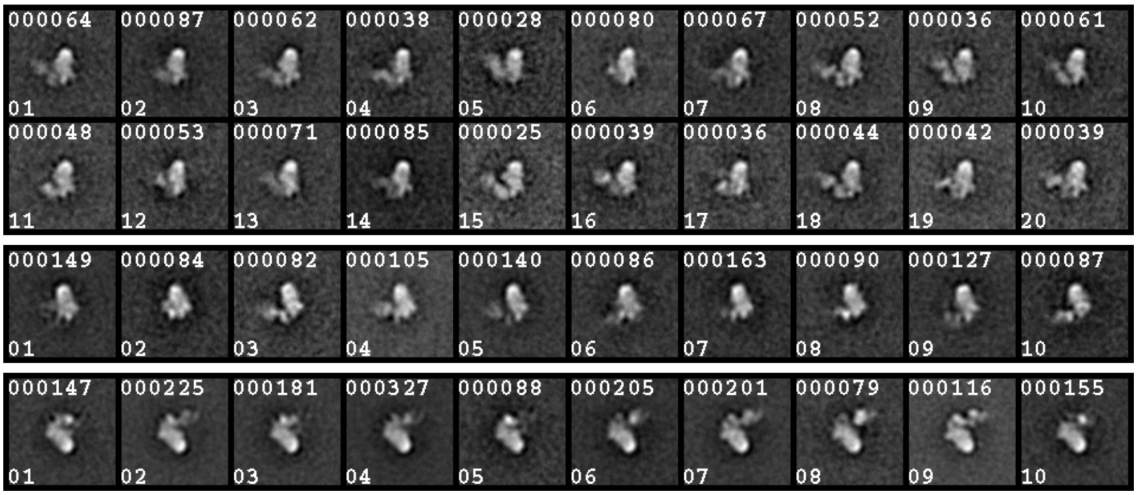
Supplementary Figure 5.19: Multireference-based alignment and classification chart of 56152 images of myosin VI molecules at high calcium and high calmodulin using five different reference images. The boxed letters represent further charts seen in Figure 5.20 and 5.21. Scale: EM averages are 30 nm wide.



Supplementary Figure 5.20: Classification of representative aligned images (with three different reference images, see 4A-4C in Figure 5.19) of myosin VI molecules at high calmodulin / high calcium into 5-20 subclasses. These classifications show myosin VI in long straight conformations (4B) or short bent conformations (4C). Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

4D

4E



Supplementary Figure 5.21: Classification of representative aligned images (with two different reference images, see 4D-4E in Figure 5.19) of myosin VI molecules at high calmodulin / high calcium into 10-30 subclasses. These classifications show myosin VI in long and short, left and right facing (4D and 4E respectively) bent conformations. Lower left number: class number. Upper left number: number of images contained in the class average. Scale: EM averages are 30 nm wide.

6 Myosin IX

A wide range of cellular processes, including cell polarization and collective cell migration, are regulated by the organization of cytoskeletal network formation. One of the proteins largely found in the cytoskeleton is the actin associated monomeric molecular motor myosin class IX. Myosin IX is expressed in two isoforms, myosin IXa and IXb, each located at different sites in mammals. While myosin IXa is mainly found in the brain and testis, the isoform myosin IXb is an important molecule in migrating cells of the immune system. Structurally, the catalytic domain of myosin IX is unique due to a large insertion of approx. 200 amino acids into loop2 of the catalytic domain, not present in any other known myosin class. This insert provides myosin IX with unique properties. It was reported that, although being monomeric, myosin IX is able to move processively along actin filaments [102–107]. While the directionality of most myosin classes is well studied, there is a lot of controversies about the direction of the myosin IX processivity. Although Inoue et al. have shown that myosin IX is minus-end directed [104], recent studies by Liao et al. concluded that myosin IX in fact moves towards the plus-end of actin filaments [106]. Additionally, it is not completely understood for which functions the processivity of myosin IX is of importance. On the other hand, myosin IX might be involved in dynamic actin-myosin networks, as it was reported that myosin IXa plays a crucial role in the collective migration of epithelial cells, established by actin bundles associated with cell-cell adhesions [108, 109]. However, the underlying processes of myosin IX organizing the actin cytoskeleton is unknown as well as how myosin IX might be involved in the assembly of actin-myosin networks is unknown.

To address these questions, we investigated the interaction of myosin IX with actin. A completely unexpected discovery was made that the catalytic domain of a single myosin IXa molecule is able to uniquely incorporate the functions of both, a molecular motor and a stabilizing molecule via the formation and regulation of actin networks in the cell. The basis for the organization of such networks lies in the presence of a second actin binding site in the above mentioned unique loop2-insert of myosin IX [110]. To further investigate myosin IX on a structural level, negative stain electron microscopy

followed by specialized single particle analysis methods were applied to answer the following questions: What are the structural properties of single actin crosslinking myosin molecules? What is the structure of larger scale actin bundles? How are these actin-myosin networks formed? Do the linked actin filaments have distinct characteristics in terms of polarity and phase? And what are possible regulation mechanisms of these networks? Does calcium play a role as we have for instance discovered for myosin VI (see chapter 5)?
The results of this study, which revealed intriguing first time structural information about myosin IX, are presented in the coming chapter.

6.1 Introduction

Present in all cells, the cytoskeleton is a network of complex form and composed of mainly three protein types, microtubuli, actin and intermediate filaments. Each of these filaments are associated with specific binding partners. The cytoskeleton can have highly complex forms and at the same time is able to adjust rapidly. These two factors are crucial to adapt to the wide range of function that the cytoskeleton is responsible for. Most importantly, it provides the cell with a shape and mechanical resistance against deformation [111].

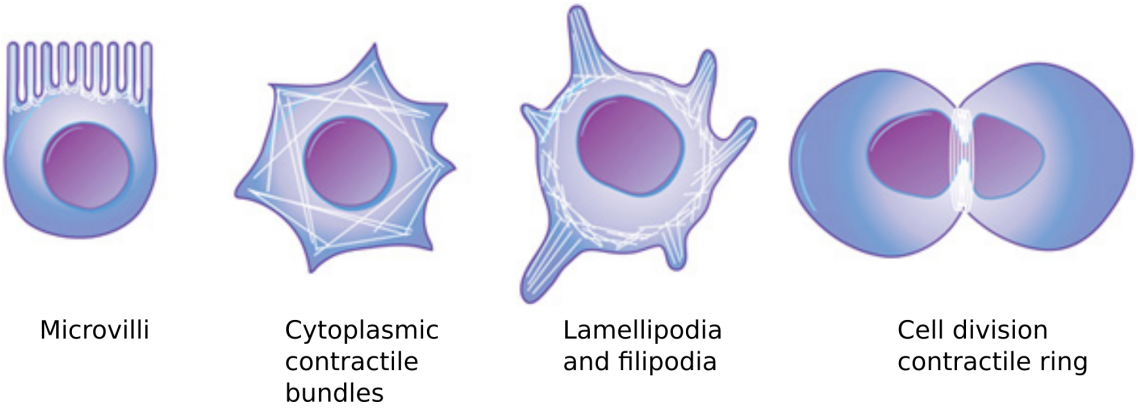


Figure 6.1: Actin filaments are involved in a variety of structures in the cell. From: [112]

While the microtubuli based cytoskeleton is primarily involved in the organization of the cytoplasm, such as the positioning of organelles, and is located in proximity of the nucleus, the networks based on actin filaments can be largely found towards the plasma membrane (Figure 6.1). In combination with its associated proteins, such as various forms of myosin, the actin cytoskeleton is largely found in muscle cells and involved in cell movement (e.g. through structural reorganization in lamellipodium and filipodia), cytokinesis, cell polarization and morphogenesis [113–117]. Despite intense research on these highly interesting networks, many underlying mechanisms are still unknown.
One of the myosin forms, class IX, has been shown to be highly involved in the processes of actin-myosin networks. Myosin class IX is a molecular motor that is present in the form of a monomer and is primarily found at the location of cell-cell contacts [108] which are known to be essential for directional migration [118, 119]. These areas in the cell are exposed to high tensions and it could be shown that the lack of myosin IX had a destructive effect on the linearity of cell boundaries [120], resulting in defective cell-cell contacts and hindered collective cell migration [108].
Myosin IX is expressed in two isoforms, myosin IXa and myosin IXb, both being found in different cell types. The isoform IXa is primarily and largely expressed in the brain and testis and crucial for the epithelial differentiation and morphology of cells [120, 121].

Myosin IXb on the other hand can be found in cells of the immune system and the intestinal epithelium [15].

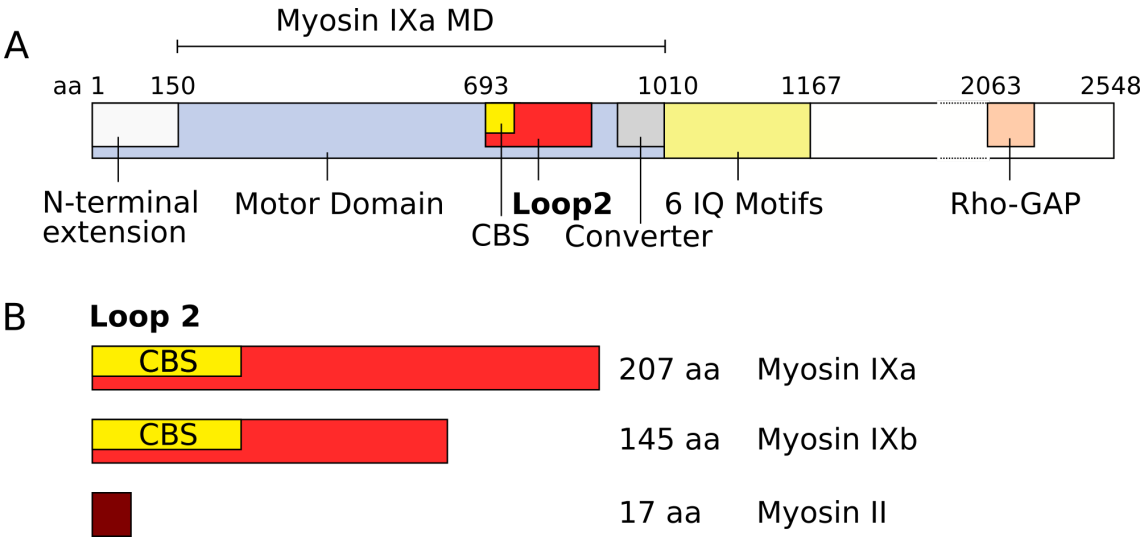


Figure 6.2: **A** Sequence of the 2548 aa long myosin IX. Indicated are the motor domain with a N-terminal extension, the unique long loop2 with a calmodulin-binding-site (CBS) and the converter. This domain is followed by 6 IQ motifs and tail containing a Rho-GAP domain. **B** A large insert extends the size of loop2 to 207 aa (myosin IXa) or 145 aa (myosin IXb), in contrast to the regular size of approx. 20 aa (e.g. 17 aa for myosin II).

Structurally, the 2548 amino acid (aa) large myosin IX comprises the conserved catalytic domain, the converter, a lever arm with six IQ motifs and a tail (Figure 6.2 **A**). Indicated in the figure is as well a special part of the molecule named myosin IXa MD, which includes the entire motor domain (MD) but without the N-terminal extension. As only this construct will be used in the following structural investigation (starting section 6.2), it will be named ”myosin IX” only for simplicity.

Apart from these sequences, the myosin class IX has a unique characteristic in its loop2 that has not been found in any other myosin. The loop2, bridging the upper and lower 50 kDa domain, is present in all characterized myosin classes and is known to be critically involved in actin binding (in combination with other parts of the upper and lower 50 kDa domain) [122, 123]. Its length usually comprises less than 20 amino acids.

In myosin IX however, an unstructured, exceptionally large insert extends this loop to up to 207 aa (myosin IXb has a slightly shorter insert of 145 aa), as shown in Figure 6.2 **B** [102, 103]. The function of this special insert was long unknown and is still not fully understood but given the fact that the catalytic domain of myosin IX was sufficient for its processivity on single actin filaments, the unique insert, which has been reported to bind calmodulin [106] (indicated in the figure by CBS, Calmodulin Binding Site), had to play a critical role in these processes, presumably by introducing a second actin binding

site [102, 103, 105, 107, 124]. In order to establish processive motility for a single-headed myosin motor[1], it was proposed that the insert in loop2 might act as an flexible actin tether and thus allowing the head to bind to adjacent binding sites on the filament and move in an inchworm kind of fashion along the filament track [102, 103, 126]. The ATPase activity of myosin IX, which is related to the speed of its processive movement, was measured to be substantially lower than other molecular motors [127]. Furthermore, recent studies have suggested that myosin IX as well as actin filaments might be involved in cell-cell-adhesions to regulate collective cell movements [108, 109]. Both results indicated, that the role of myosin IX is more on a structural level, possibly organizing the actin cytoskeleton instead of transporting cargo.
In its full length, myosin IX comprises six IQ motifs following the motor domain and a tail which includes a Rho-GTPase-activating protein domain (RhoGAP). This domain induces GTP hydrolysis of RhoGTPases, which are important for the transduction of signals in the cell, and by this can inhibit their signaling function. The activity of such GTPases is linked to structural rearrangements of the cytoskeleton [128]. Therefore, the presence of a RhoGAP domain in the tail of myosin IX suggests that this myosin might be able to combine its mechanical functionality with signaling activity in the context of actomyosin structures [124, 129].
Another unique feature of myosin IX is the large N-terminal extension in the motor domain (aa 100-150) whose function however remains to be discovered [130].

In order to better understand the underlying processes of collective motion and organization of the cytoskeleton, in vitro model systems that include actin filaments, class II myosin motors and molecules known to be able to form and stabilize actin-crosslinks have been developed as reported recently [131, 132]. Several actin-crosslinker are available that mainly differ in the distance of crosslinked filaments (Figure 6.3):

Shortest crosslinks can be formed by fascin which connects parallel orientated actin filaments and with a distance of approx. 8 nm [134]. Slightly larger distances of 11-12 nm are obtained by the crosslinker fimbrin [135]. Another well-studied linker is alpha-actinin, forming ordered actin networks with an interfilament distances of approx. 34 nm [136]. Other molecules able to link actin filaments are spectrin or filamin which, however, are much larger molecules, hence able to span several µm between two filaments, which are not necessarily in a parallel orientation.
All of the above mentioned molecules are pure actin-crosslinker. Intriguingly, our group could show that myosin IXa too was able to induce the formation of actin crosslinks. Whereas previous studies have already reported crosslinking features of other myosin

[1]Processivity of myosin motors usually requires a dimeric motor molecule with two heads interacting with actin in a hand over hand fashion, for instance myosin V [125] or myosin VI [84]).

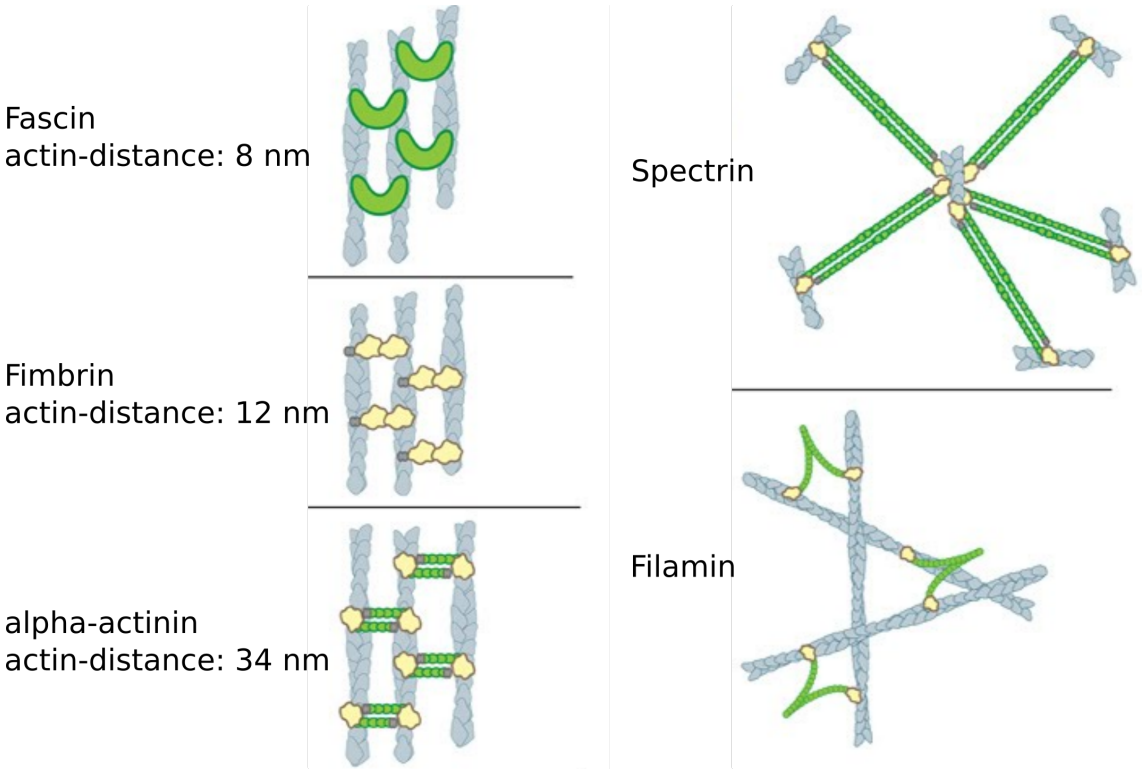


Figure 6.3: Several actin crosslinker are identified: Fascin, Fimbrin, alpha-actinin, Spectrin and Filamin. Modified from [133]

classes, such as myosin XXI [137], these usually either involved additional crosslinking molecules or the presence of dimers in order to generate two actin binding sites. In the case of myosin IX however, one single catalytic domain comprised all features needed (most importantly the unique insert in loop2) to incorporation both functions, a molecular motor and actin crosslinker. [110].

Figure 6.4 shows fluorescently labeled actin filaments mixed with the motor domain (MD, aa 149-1019) of either myosin IXa wild type (**B**) or a mutant where the myosin IXa-specific loop2 was replaced by the loop of skeletal muscle, missing the special insert (**A**). The color code indicates the fluorescence intensity as also shown in **C**, which plots the intensity along the indicated path. In the case of the IXa mutant, exclusively single actin filaments were observed, whether the wild type IXa, containing the insert, induced the formation of structures involving more than one actin filament, indicated by the change in fluorescence. [110]

These results, which suggested that the insert in loop2 not only enabled myosin IX to tether itself to one actin filament as reported [102–104], but also is able to reach out and bind to other filaments in order to form crosslinks, drove my studies towards the structural investigation of actin crosslinks formed by myosin IX.
The structural analysis of myosin IX - actin - crosslinks presented in this thesis was

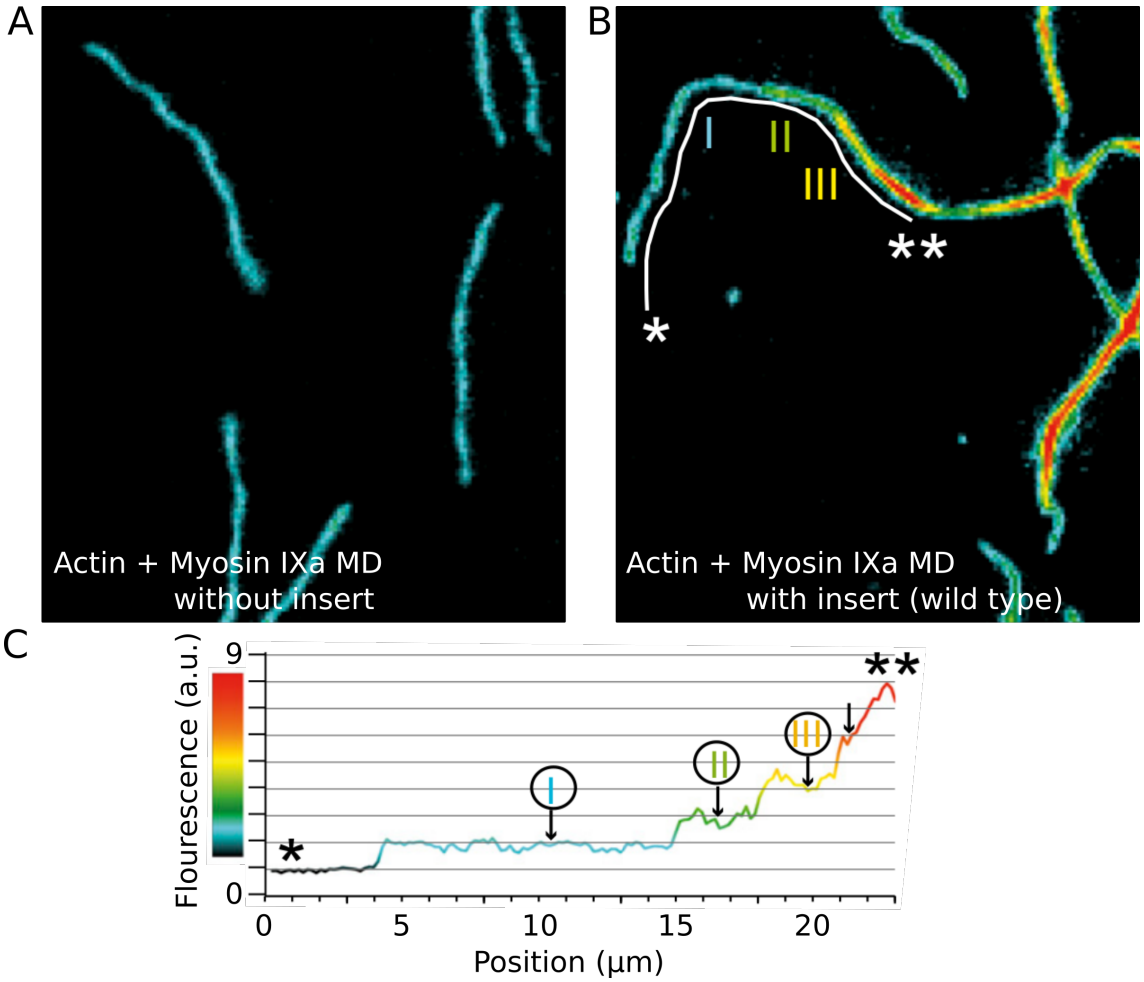


Figure 6.4: Formation of actin crosslinks induced by myosin IX, modified from [110]. Fluorescently labeled actin filaments in the presence of the mutant myosin IX (**A**) or wild type (**B**), which has its unique loop2 replaces by a loop from skeletal muscle myosin and therefore lacks the special insert. The color code indicates fluorescence intensity which is related to the number of actin filaments in a bundle with a diameter below the resolution limit if the fluorescence microscope. **C** From one up to more than three filaments in involved in one crosslink could be resolved.

approached by negative stain electron microscopy, followed by single particle analysis. The special properties of myosin crosslinking actin, the different shapes it might adopt and the examination of the polarity of these actin filaments within the networks required the development of specialized SPA-methods. The elaborated scripts described in the methods section were primarily based on the SPIDER / SPIRE software. The application of these scripts led to intriguing results on the structure of myosin - actin - crosslinks. They are explained and discussed in the following chapter. Based on the finding that myosin can be regulated by calcium as e.g. discovered for myosin VI (see chapter 5), the elaborated single particle analysis workflow could be also used to analyze new data sets of samples in the presence of calcium and in combination with what was known about myosin IX, these results were put into a broader (cellular) context.

6.2 Results

6.2.1 SPA of myosin IX - actin - crosslinks

Samples of F-actin and myosin IXa MD were prepared, diluted, negatively stained and imaged using negative stain electron microscopy according to section 6.4.1. Following the process of particle picking described in section 6.4.2.1, a stack of 981, 300x300px images of actin-crosslinks was created. These images were then used to analyze the general physical dimensions of the crosslinks, the periodicity and structural properties of the crosslinking molecules and the polarity of the crosslinked actin filaments. The results of this analysis were then interpreted using crystal structures and put into context with previous studies on myosin IX:

6.2.1.1 Myosin IX crosslinks actin filament in highly ordered lattices

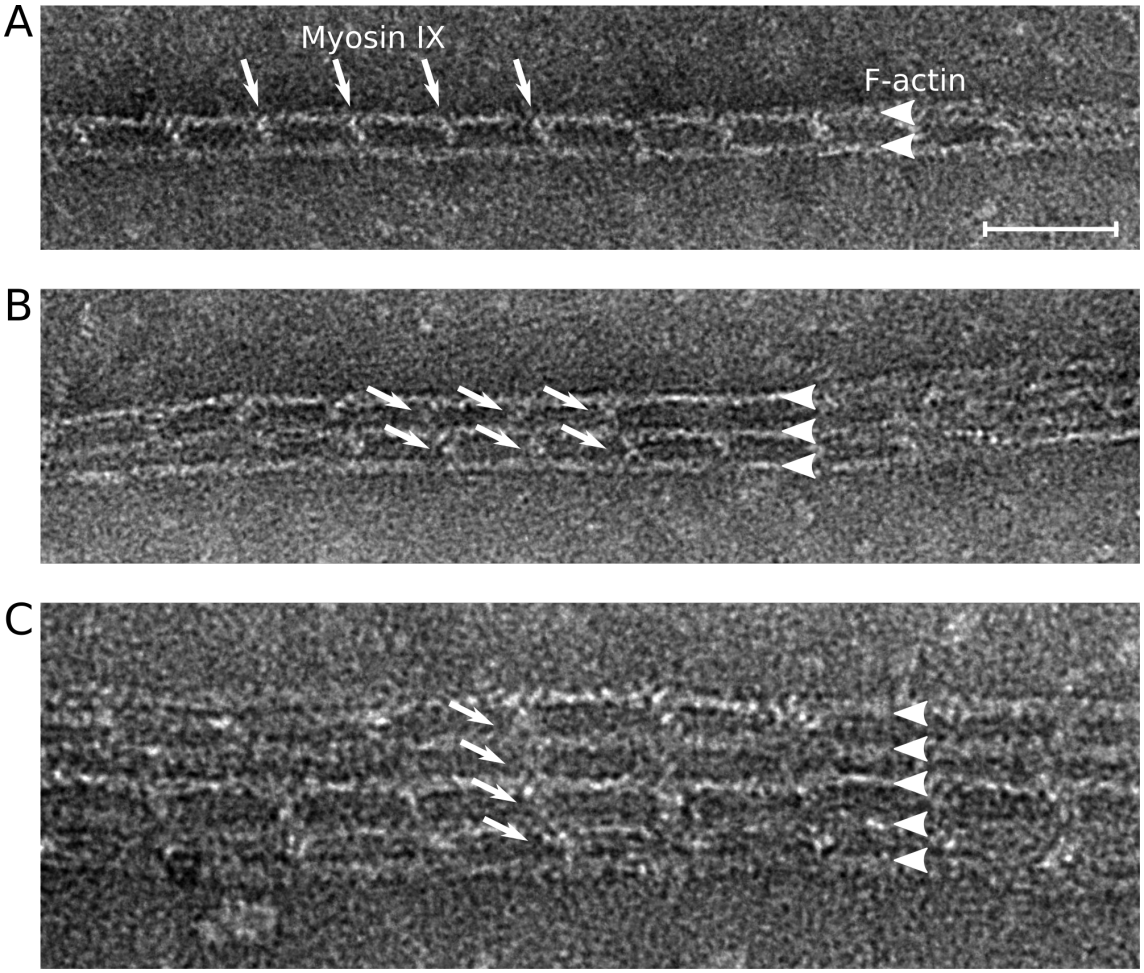


Figure 6.5: Negative stain electron micrographs of single myosin IX molecules (arrows) crosslinking either two (**A**), three (**B**) or more (**C**) actin filaments (arrowheads). Scale bar: 50 nm

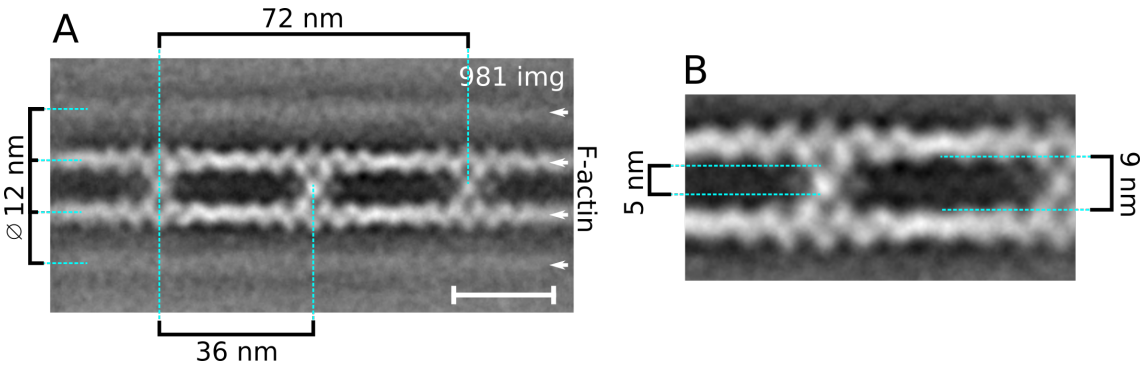


Figure 6.6: **A** Global average of 981 images of myosin IX - actin - crosslinks (polarity of actin filaments not taken into account) revealed a crosslink periodicity matching the actin helical repeat (36 nm) and an interfilament distance of approx. 12 nm. **B** Magnified area of the average to measure the minimal and maximal distance between the actin filaments. Scale bar: 20 nm

Figure 6.5 shows example raw micrographs of single myosin IX molecules (arrows) crosslinking either two (**A**), three (**B**) or more (**C**) actin filaments (arrowheads). Already in the raw micrographs it could be seen that the periodicity of the myosin molecules was highly ordered within one bundle. Moreover, the crosslink events were highly cooperative over the range of different crosslinks, as it can be seen in **C**. To explore the properties in more detail, the stack of 981 images was aligned and averaged according to the methods described in section 6.4.2.1. The global average can be seen in Figure 6.6 **A** and **B**. The average image revealed a distance between the myosin IX monomers of precisely 36 nm, matching the helical (pseudo)repeat of F-actin. Furthermore, apart from the two actin filaments in the center of the average, more filaments could be seen above and below (indicated by the arrows). These filaments and the crosslinking myosin however were very poorly resolved and highly blurred due to the fact that not all images contained more then two filaments. Furthermore, as the picking of images was concentrated on good quality crosslinks, the myosin molecules above and below might have had poor quality. Another reason was that although the actin - myosin - crosslinks were relatively stable constructs, small flexibilities would affect the resolution of the average. Even the resolution of the two filaments in the middle drops towards the edges on the left and right sides for the same reason. Nevertheless, the average distance between two crosslinked actin filaments could be measured relatively accurately. Using the (horizontal) center of each filament, the mean distance was 12 nm. Figure 6.6 **B** shows a section of the crosslink together with the minimal and maximal distances between the actin filaments, 5 nm and 9 nm respectively.

Interestingly, in almost all (raw) crosslink images, all available positions were filled by myosin IX monomers and no molecules were observed to be situated in between two crosslinks. Moreover, not even one single myosin molecule bound to a single actin

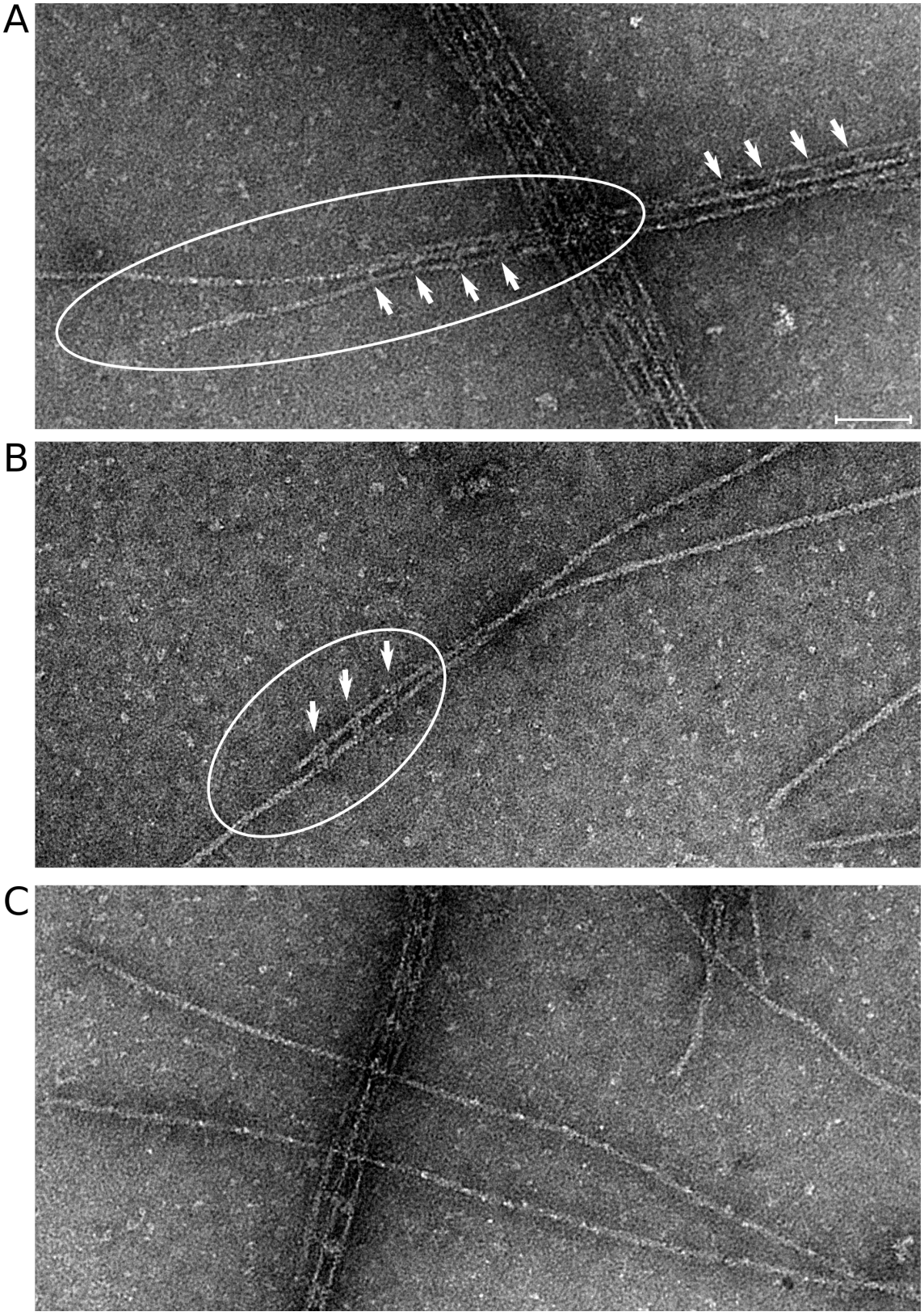


Figure 6.7: Negative stain electron micrographs of single myosin IX molecules (arrows) crosslinking either actin filaments demonstrates the cooperative binding of myosin. Myosin binding to actin was only observed in combination with formed crosslinks (circled areas). Single myosin bound to single filaments or irregularly crosslinked filaments seemed not to occur. Scale bar: 50 nm

filament was observed. This suggested that the formation of crosslinks was highly cooperative. Figure 6.7 shows three examples of this cooperative binding. Whenever myosin IX bound to actin filaments (white arrows), crosslinks were formed. Any other single actin filaments not involved in crosslinks did not have any myosin molecule bound. Furthermore, the micrographs showed nicely the stability of crosslinks that could be established already by a small number of myosin crosslinkers. The circled area in **A** for example highlights two actin filaments that were partly crosslinked. Only where myosin IX was involved, the filaments adopted a straight conformation. Without the crosslinker, this parallel conformation of two filaments was very unlikely to occur. Another example in micrograph (**B**) shows two relatively long filaments, crosslinked by only three myosin monomers. Given the large sterical forces, this suggested a high rigidity of such crosslinks. The two filaments in the example image **C** are not crosslinked at all. One obvious reason was the absence of available myosin molecules in this area. Another explanation however could be the wrong arrangement of the two filaments to each other. As it will be described in the next section, myosin IX crosslinked only actin filaments that were aligned in parallel to each other (in the absence of calcium). The filaments shown here might be simply antiparallel.

6.2.1.2 Actin filaments crosslinked by myosin IX have parallel polarity

The following section will explore how two actin filaments, crosslinked by myosin IX monomers, were related to each other with respect to their polarity and phase. The workflow for this analysis will be shortly summarized at this point[2]. Figure 6.8 illustrates the four possibilities of how two actin filaments can be related to each other, resulting from combinations of parallel, antiparallel, in phase and out of phase filaments[3]. Before the actual polarity analysis was carried out, all 981 images were windowed to 90x90px images and classified according to the binding positions of the myosin. This classification resulted in 451 images with identifiable binding positions, whereas the remaining 530 images were not of sufficient quality to be used further. Three distinct binding conformations, each present in mirror related orientations on the grid, could be resolved as shown in Figure 6.9 **B**. As described in the methods section, the information about these conformations was used to create a stack of 451 "prealigned" images[4] (**C**), which was then used for the actual polarity determination:

[2] A detailed description is given in section 6.4.2.3

[3] These are the general possibilities, small relative shifts of the filaments against each other are further variations of each possibility

[4] "Prealigned" in this case meant to create a stack where all images show the myosin binding to the lower left binding position, achieved by mirroring accordingly

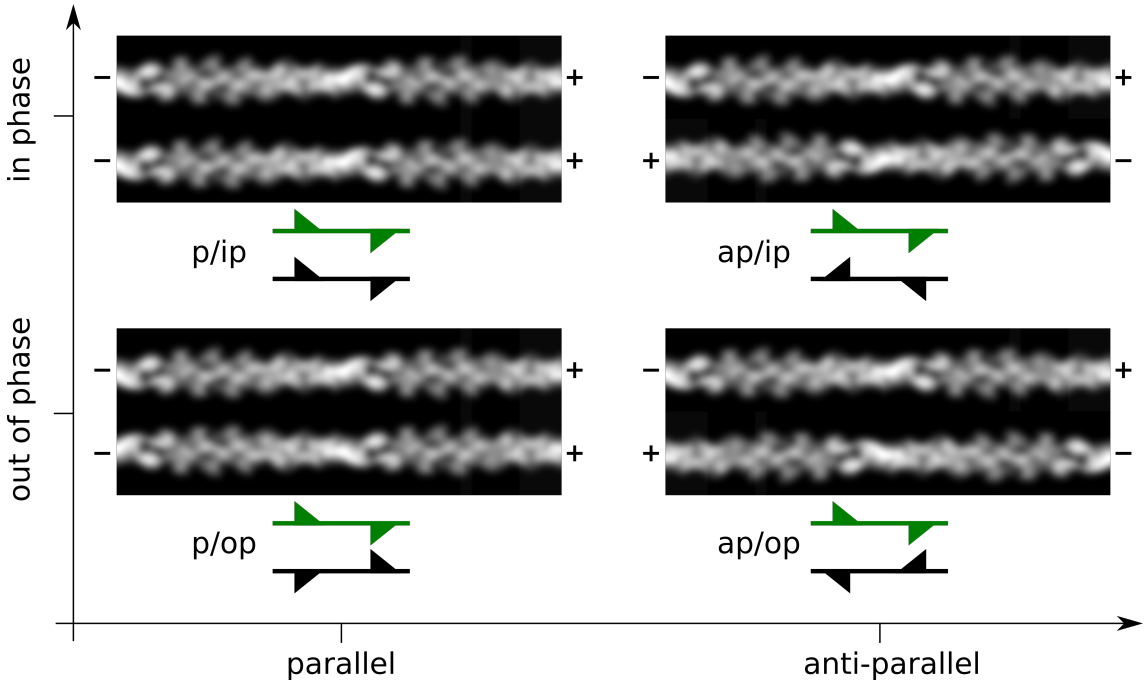


Figure 6.8: Illustration of the four possibilities of how two actin filaments can be related to each other: parallel, antiparallel, in phase and out of phase. The filaments shown here are actin models according to section 6.4.2.2.

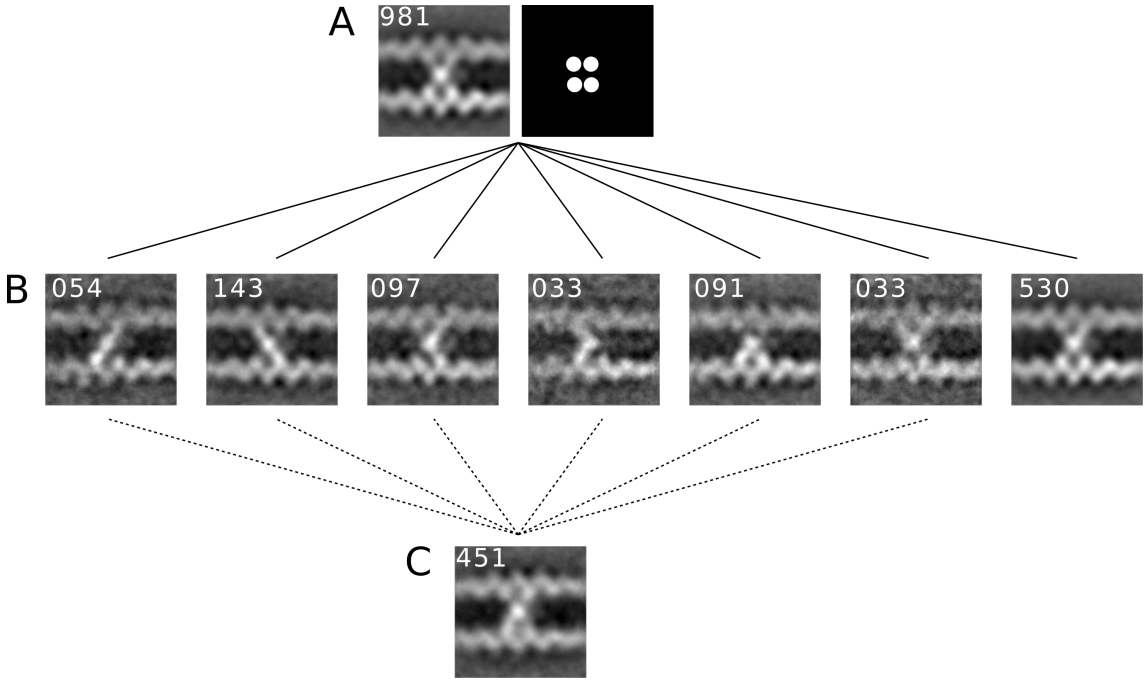


Figure 6.9: **A** Windowed global average of 981 crosslink images. **B** Classification of the data set with respect to the binding positions, see mask in **A**. 451 images had sufficient resolution to identify one of the three discovered conformations. **C** Creation of a prealigned data set where all images contain myosin bound to the lower left position, obtained by mirroring each image accordingly. Upper left number: Number of images on averages. Scale: images are 30 nm wide.

In analogy to the established workflow for the polarity determination of actin-crosslinks described in section 6.4.2.3, the upper and lower filaments of the ”prealigned” data set were first processed separately and checked for consistency. Figure 6.10 summarizes this first step, where shifting up / down (**B**) and aligning with two different references (**C**) led to the graphs **D,E** showing the rotation and mirroring parameters. The combination of these graphs is shown in the next Figure 6.11 **B** and the number of images for each of the cases and the alignment parameter are given in Table 6.1.
The comparison of two graphs within each filament, revealed information about the consistency for each filaments separately (upper and lower), shown in Figure 6.11 **D** and summarized in Table 6.2. For the lower filament, 350 images (77.61%) were consistent, while for the upper filament, 346 images (76.72%) showed clear polarity features. Although the analysis was done for both filament separately, one had to always remember that two images (one of the upper and one of the lower filament) belonged to one and the same crosslink image. Because of that, although the total number of consistent upper and lower filaments was similar, this did not mean that they belonged to the same crosslink image. The next step was therefore to identify only those crosslink images with both filaments consistent (Figure 6.11 **E**). The corresponding values are given in Table 6.3. Although 49.4% of the data set were inconsistent, the remaining 228 images revealed the preferred polarity and phase of the crosslinks.

	lower filament polarity 1	lower filament polarity 2	upper filament polarity 1	upper filament polarity 2
both	166 img	164 img	157 img	158 img
mirror at y axis	59 img	90 img	56 img	101 img
180° rotation	139 img	128 img	145 img	120 img
nothing	87 img	69 img	93 img	72 img

Table 6.1: Absolute values for graphs **C** in Figure 6.11

		lower filament	**upper filament**
1) **consistent**	different polarity different phase	254 img / 56.32%	256 img / 56.76%
2) **polarity consistent**	different polarity same phase	096 img / 21.29%	090 img / 19.96%
3) **inconsistent**	same polarity same phase	101 img / 22.39%	105 img / 23.28%
1)+2)	different polarity diff.+same phase	350 img / 77.61%	346 img / 76.72%

Table 6.2: Absolute values for graphs **D** in Figure 6.11

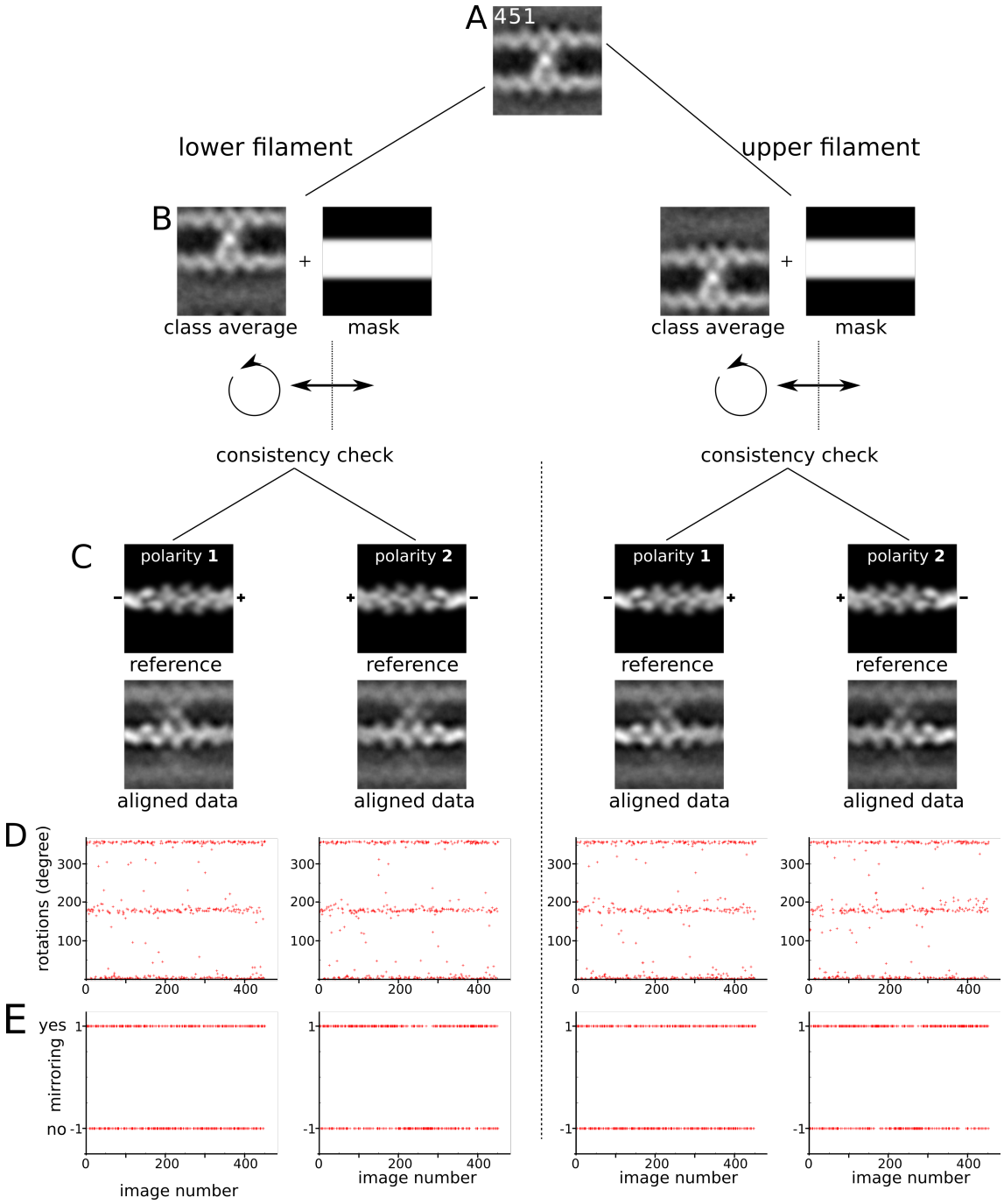


Figure 6.10: **A,B** Separate analysis of lower and upper filaments in the crosslink images. **C** Each image is aligned with a simulated actin filament in two polarities as a reference to examine the quality of this filament in terms of consistency. **D,E** Optimal rotation and mirror parameter applied during the alignments. Scale: Images are 30 nm wide.

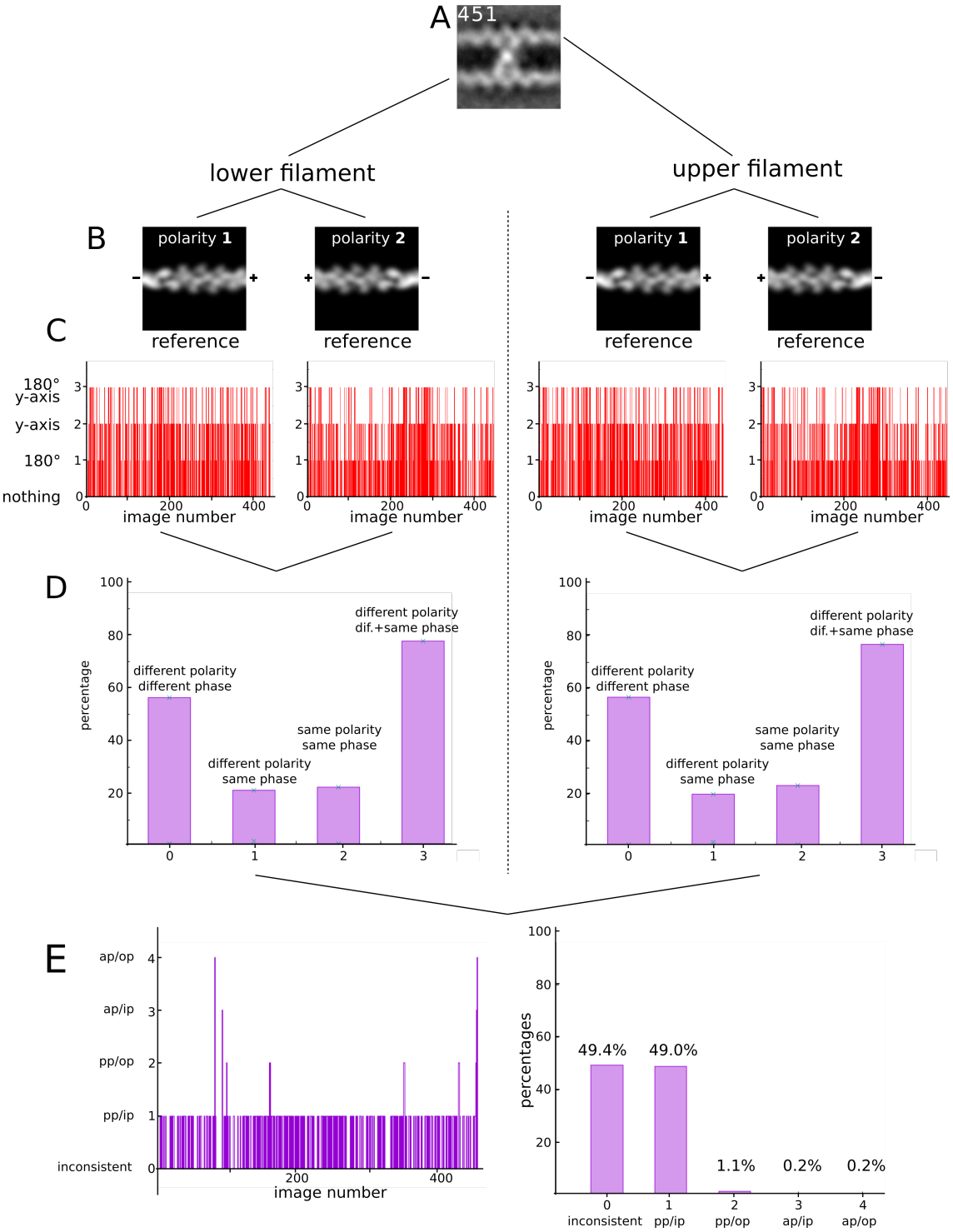


Figure 6.11: **C** Combined and simplified alignment parameter from the alignment of the lower and upper filaments to the reference images **B** with polarity 1 and 2. **D** Comparison of alignments with polarity 1 and 2 of lower and upper filaments to identify consistent filaments. **E** Comparison of lower and upper filaments. 223 images (49.45%) were inconsistent, the remaining images clearly identified the crosslinked actin filaments to be parallel and in phase.

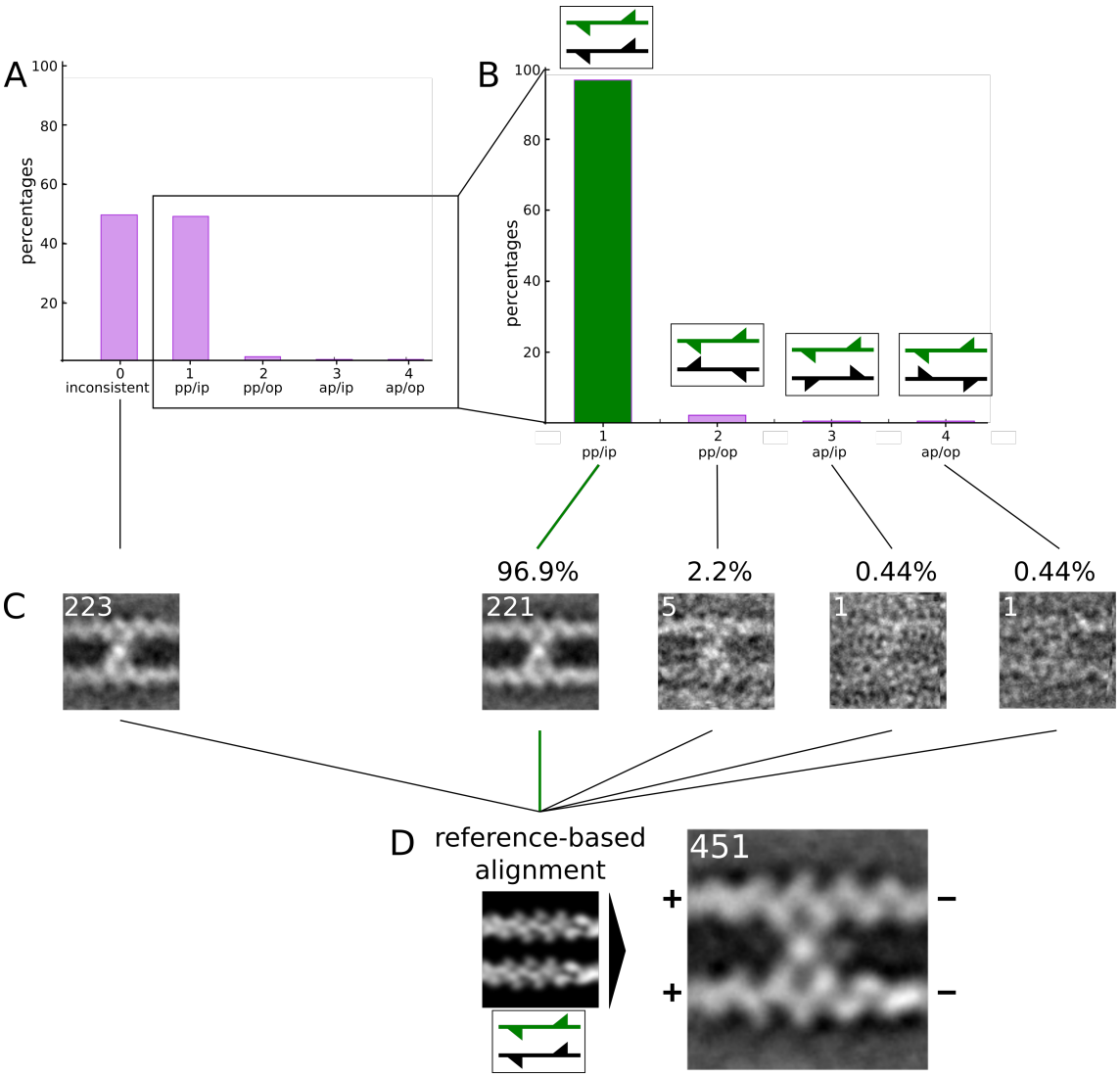


Figure 6.12: Polarity and phase assignment to all (**A**) 451 and only consistent (**B**) crosslink images. **C** Corresponding class averages for each case. **D** Global average of all images aligned with a pp/ip reference. Upper left number: Number of images on averages. Scale: images are 30 nm wide.

		all images (451 img)	only consistent images (228 img)
inconsistent	223 img	49.45%	
parallel polarity / in phase	221 img	49.00%	96.93%
parallel polarity / out of phase	005 img	01.11%	02.19%
anti-parallel polarity / in phase	001 img	00.22%	00.44%
anti-parallel polarity / out of phase	001 img	00.22%	00.44%

Table 6.3: Absolute values for graphs **E** in Figure 6.11

In the next Figure 6.12, the graph **B** shows the statistics for only consistent images (see also Table 6.3). A striking ratio of 96.93% of these 228 crosslink images were formed by actin filaments with parallel polarity and in phase. Only a small number of 7 images (3.07%) made up for the other three conformations, from which 5 had also parallel polarity but with the two filaments out of phase. The corresponding images for each condition were averaged and are shown in **C**. In these averages the polarity of the actin filaments was hidden due to averaging over different orientations on the grid. The myosin on the other hand could still be identified as bound at the lower left binding position. This was due to the prealignment of the data set (Figure 6.50). In order to improve the resolution of the actin filaments, the images were reference-based aligned, with the corresponding simulated crosslink image as a reference. Because of the fact that from the approx. 50% of the data set almost all crosslinks had one distinct polarity / phase-relation, it could be assumed that the inconsistent images had the same properties. This however could not be determined due to low quality of one or both filaments. However it was legit to not only align the 221 images of pp/ip crosslinks with the corresponding reference, but in fact the whole data set of 451 images. The resulting average of this alignment is shown in **D**. It showed asymmetric structures of both, the actin filaments and the crosslinking myosin. Figure 6.13 highlights these properties by showing the average one more time together with its variance: The arrow-like, pointing structure of F-actin (indicated by the lines) was resolved in both filaments and pointed towards the same direction, equivalent to parallel polarity. The arrangement of the circles, indicating the phase of the two filaments, could also be resolved and was similar for both filaments, which equaled the property ”in phase”. The overall good resolution of the actin filaments could be seen in the low variance in these areas. Focusing on the crosslinking myosin (indicated by the arrowhead), the four binding positions were not discernible anymore. The additional mass (arrow) however appeared on only one side of the myosin. This indicated that the orientation of the myosin itself was consistent in the polarity-aligned data set with different binding patterns possible. The resolution[5] of the whole crosslink average (see Figure 6.14) was 33.09 Å and of the actin filaments alone (using a mask that cuts out the myosin) was, as expected, better: 25.76 Å.
Once the polarity and phase of the crosslinks could be revealed, the structure of the myosin IX was explored in more detail.

[5]measured by calculating the FRC cutoff at a value of 0.5 as described in section 3.3 (script: *Misc/frc_resolution*)

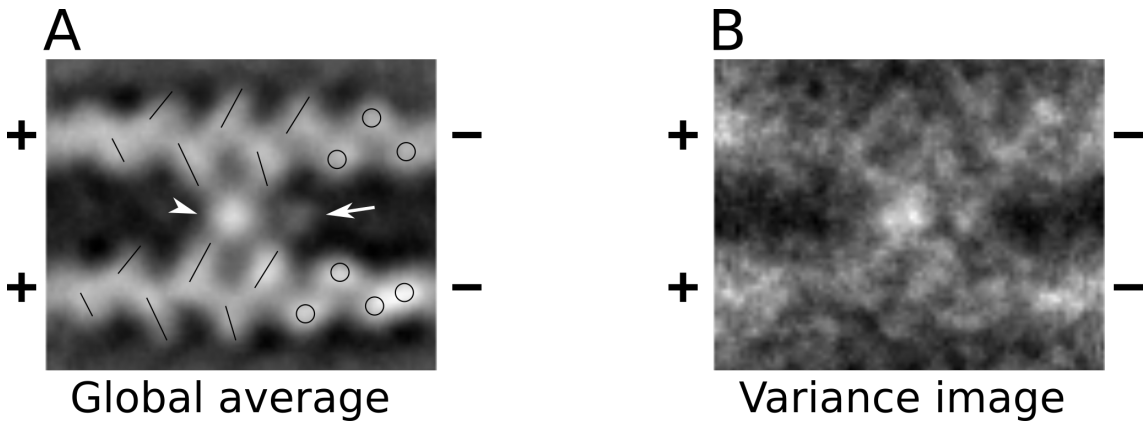


Figure 6.13: **A** Global average of polarity and phase aligned EM crosslink images showed clear polar features of the actin as well as the myosin (additional mass mainly on one side, arrow). **B** The variance images shows great homogeneity in most of the image parts but still interimage variances for the crosslinking myosin itself (arrowhead in **A**), arising from averaging over all three conformations. Scale: images are 30 nm wide.

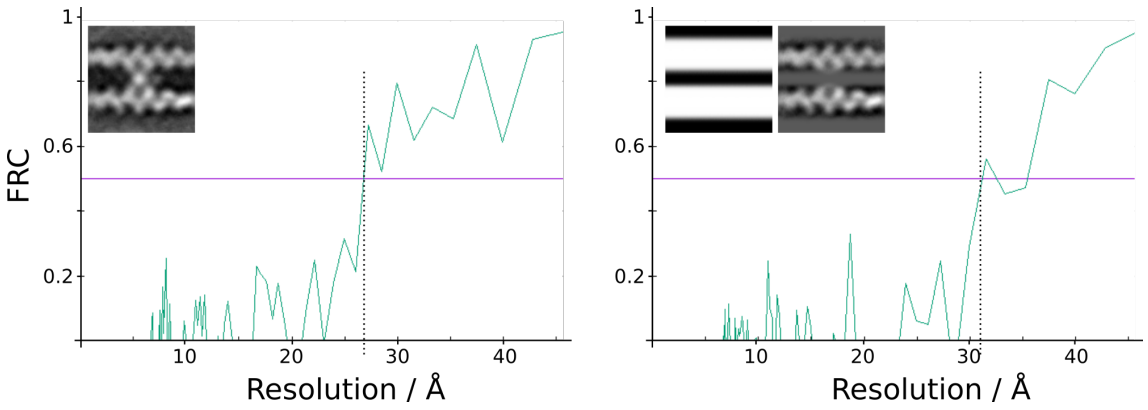


Figure 6.14: Resolution determination of the whole crosslink average (**A**) and only the actin filaments (**B**) using the FRC 0.5 cutoff criteria and averaging over 10 iterations. The resolution of actin filaments alone (25.76 Å) was better than of the whole image (33.09 Å). Scale: EM images are 30 nm wide.

6.2.1.3 Myosin IX adopts three different conformations

General shape:

From 981 images of Myosin IX - actin - crosslinks, 451 had sufficient resolution to determine the binding behavior of myosin to actin. Classification of these 451 images according to section 6.4.2.2 revealed three distinct binding patterns, where the two actin binding sites of myosin IX bound either the same or different actin filaments:

- Conformation **(I)**: Binding to one actin monomer of each filament in a diagonal way (binding position 1&4 or 2&3), termed "**diagonal**"
- Conformation **(II)**: Binding to one actin monomer of each filament, vertically arranged (binding position 1&3 or 2&4), termed "**bent**"
- Conformation **(II)**: Binding to two actin monomers of the same filament (binding position 1&2 or 3&4), termed "**inchworm**"

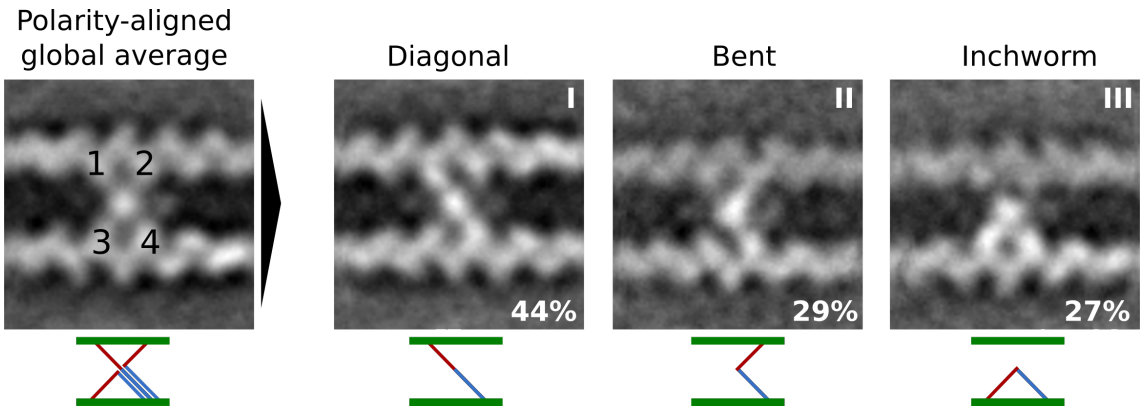


Figure 6.15: Reclassified polarity aligned data set with respect to the three binding conformations diagonal (44%), bent (29%) and inchworm (27%). The diagonal and bent conformation crosslink two parallel actin filaments, whereas the inchworm conformation bridges two adjacent actin monomers of the same filament. Lower right number: Percentage of occurrence. Scale: images are 30 nm wide.

As discovered in the previous section, 97% of the crosslinks were formed between two actin filaments that had parallel polarity and were aligned in phase. Using this information, the polarity-aligned data set was re-classified, which resulted in the three final class averages shown in Figure 6.15. The lower numbers in the class averages denote the percentages with which each shape was present in the data set. The diagonally shaped crosslink was with 44% the most abundant, followed by the bent conformation with 29% and inchworm conformation (27%). The absolute numbers are given in Table 6.4. Interestingly, only two of the conformations (diagonal and bent) comprised actual crosslinks, connecting two different filaments. In the inchworm conformation, the myosin did not crosslink but formed a bridge between two adjacent actin monomers on the same actin

filament. Myosin adopting only this conformation could not establish crosslinks and also each inchworm shaped myosin IX molecule in crosslinked filaments destabilizes the crosslink. Only a dynamic equilibrium of all three conformations in already established bundles can allow for the existence of the inchworm conformation. This however only applied, if the occurrence of the inchworm conformation was not dominating and therefore it was to be expected that this conformation was the least occurring one.

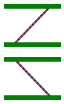
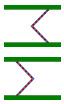
Conformation	Number of images	Percentage	Combined percentage of usable images
	54	05.50%	**43.68%**
	143	14.58%	
	97	09.89%	**28.82%**
	33	03.36%	
	91	09.28%	**27.50%**
	33	03.36%	
	530	45.06%	-

Table 6.4: Absolute numbers of the statistical distribution of the diagonal, bent and inchworm conformation in the polarity aligned data set (451 images).

Detailed analysis

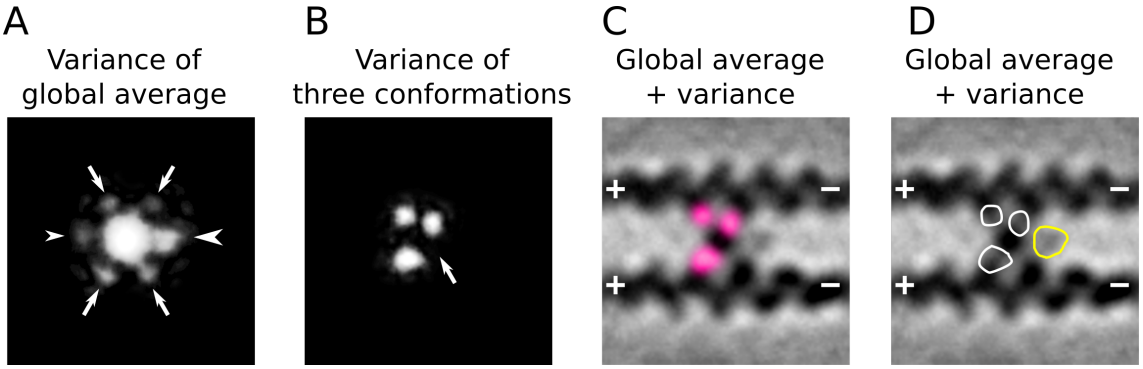


Figure 6.16: **A** Variance calculation of the global average with the simulated actin crosslink images indicated several spots, representing the central myosin mass, actin binding positions (arrows) and the observed additional masses (arrowheads). **B** Variance of the variances of the three conformation with the simulated crosslink image. The three spots indicated the positions of the second actin binding site of myosin IXa, whereas the missing spot (arrow) represented the myosin primarily binding.

In order to analyze the structure of the actin-crosslinker myosin IX in more detail, the variances of the class averages shown above were compared as described in section

6.4.2.5. This comparison revealed which parts of the intensities belonged to the myosin motor domain, which to the actin binding sites and which to the actin filaments. The variance of the global average with the simulated actin crosslink images, shown in Figure 6.16 **A**, identified several spots: One central mass, the four connections to actin (arrows) and two additional masses of significant different intensities at the left and right side of the central mass (indicated by the arrowheads) later identified as presumably representing a calmodulin bound to the insert in loop2. Furthermore, the variance of the variances of the three conformations (**B**) showed three spots, representing the varying second binding position of myosin IX. The missing (variance) spot close on the lower right position, see arrow, indicated that this binding site was present in all three conformations and therefore represented the primarily actin-binding-site as it is conserved in all myosin classes. Both, the varying and conserved binding positions reflected the diagonal, bent and inchworm myosin conformation. To highlight the identified binding behavior, and to more specifically localize the binding sites, the global class average (inverted) was overlaid with a colored version (pink) and the outlines of the variance image (white) including the yellow outline of the (dominating) additional mass on the right hand side (**C** and **D** respectively).

To analyze the myosin structure even further, the average was interpreted by crystal structure projection matching using modeled structures of myosin IE bound to actin in different variations (PDB: 4A7F±3B63) as described in section 6.4.2.5. The cross correlation of 360 low-pass filtered 2D projections from the structures with the global average (see Figure 6.60 in section 6.4.2.5) resulted in the best fitting structure shown in Figure 6.17 **B** in its optimized orientation (ccc_{max}=0.61[6]). This best fitting model was comprised of one myosin IE molecule bound to the middle of an actin filament with five actin monomers. The normalized and ranked ccc-values for all 360 views are shown in **A** together with three representative 2D projections (best, medium and worst fit, ccc-values 0.61, 0.54 and 0.42 respectively). The overlay of the optimized model onto the EM average (**C**) visualized the great fit of both, the actin filament (green) and the myosin motor domain (blue). A zoom into the outlined area however showed still discrepancies between the EM data and the crystal structure (**D**). The upper two indicated intensities (arrows) were not covered by the model but could be explained by two actin monomers from the upper actin filament. The third not covered intensity, outlined in yellow (using the calculated outline from Figure 6.56) was caused by structural parts of myosin IXa not present in the myosin IE model. The most unique property of myosin IXa was

[6]This ccc-value differs from the one shown in the methods section, Figure 6.60, because for the final ccc-calculation here no mask was applied

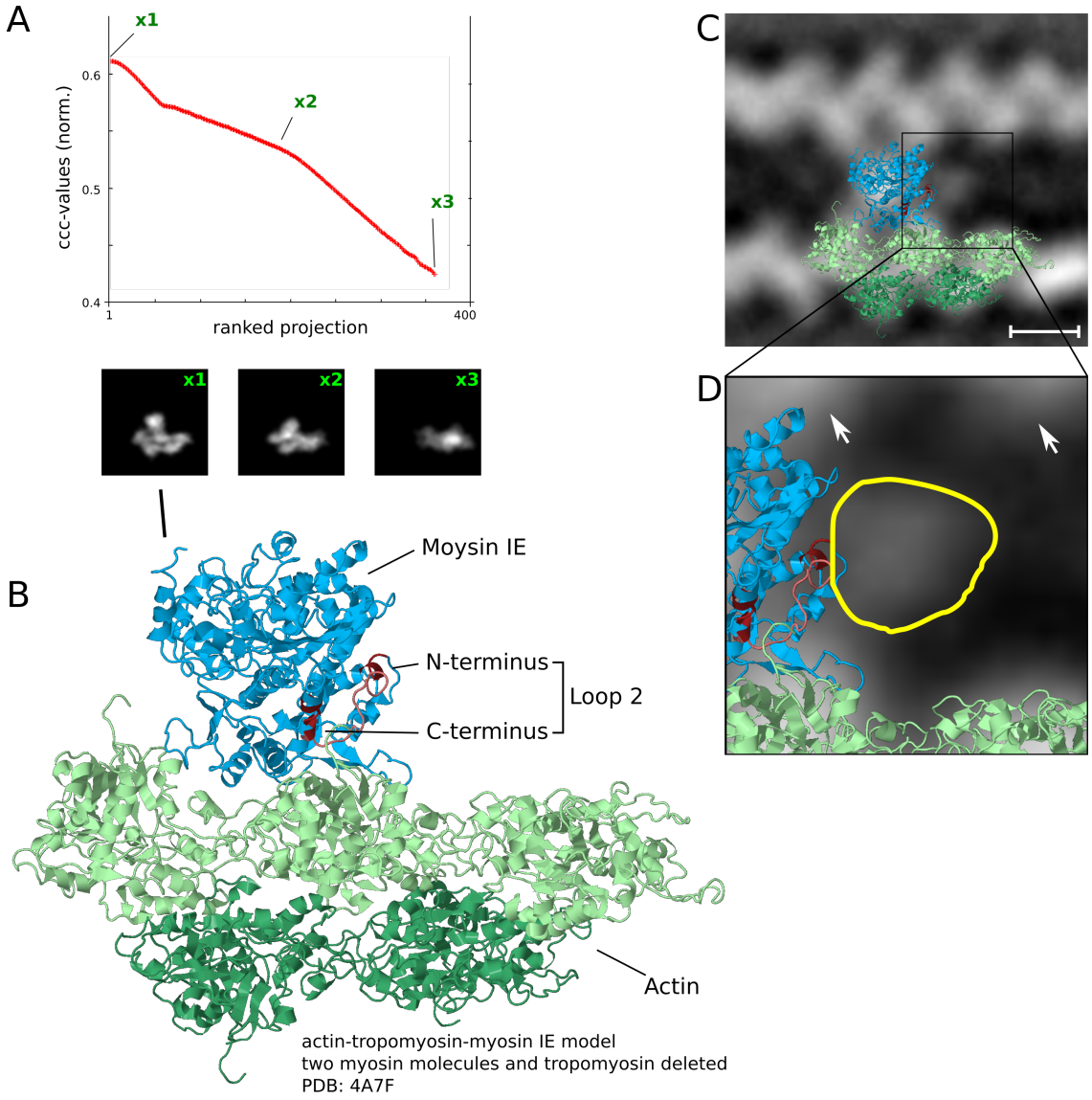


Figure 6.17: Crystal structure projection matching of the polarity aligned EM average using a model of myosin IE bound to actin (modified PDB:4A7F). **A** Normalized and ranked cross correlation coefficients with three representative simulated EM images.The crystal structure model in its optimized orientation alone (**B**) and overlaid onto the EM average (**C**). The N- and C-termini of loop2 are indicated in red. An additional mass on the right side was not accounted for by the model, as seen in the magnified section in **D**. The two arrows indicate the actin monomers from the upper actin filament. Scale bar: 5 nm

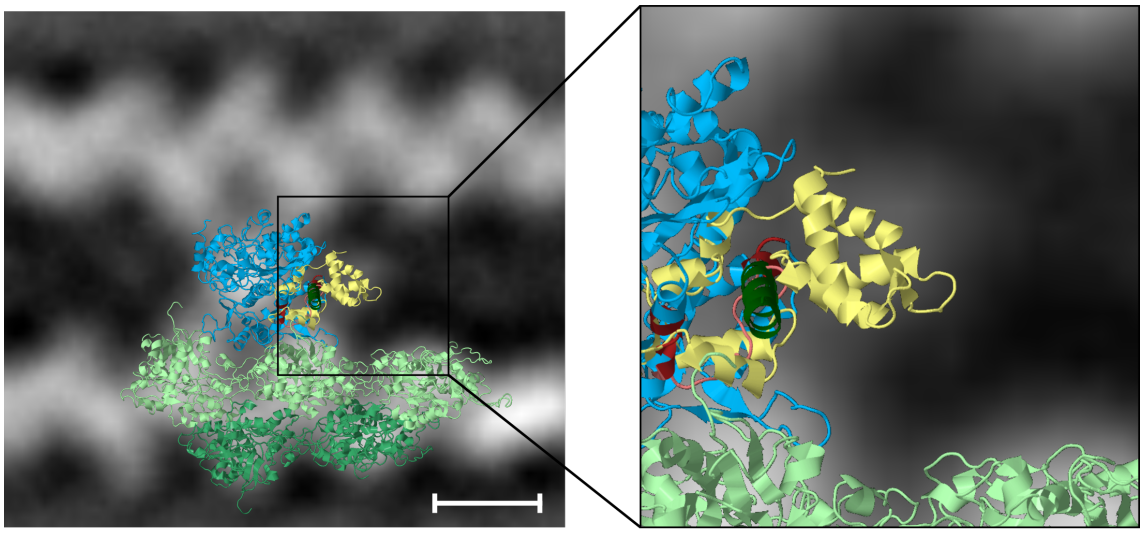


Figure 6.18: Superposition with a myosin IE structure with a calmodulin modeled onto the N-terminus of loop2 explains the observed mass. Scale bar: 5 nm

the approx. 207 amino acid long insert in loop2 (the N- and C-terminus of the loop2 in myosin IE are colored in red). As shown by previous studies [110] and mentioned in section 6.1, this special insert was able to bind calmodulin at a position close to the N-terminus. Therefore, a structure of calmodulin was modeled onto the myosin IE structure as described in the methods section (6.4.2.5). Figure 6.18 shows the same overlay as the previous figure together with the magnified area but using the calmodulin-modified-structure instead. Here, the calmodulin of the model could account for the intensity seen in the EM average. This on the one hand confirmed the discovered ability of the insert to bind calmodulin and furthermore revealed structural insights of how this calmodulin might be arranged.

6.2.2 Myosin IX - actin - crosslinks in the presence of calcium

The results presented above were obtained with actin and myosin in the absence of any nucleotide or other myosin-regulators. Since we found that myosin IXa MD binds calmodulin and since calmodulin is widely known as a calcium sensor, it was investigated whether calcium has an effect on myosin IX - actin - crosslinks. The SPA-methods established for the crosslinks in the absence of calcium could be directly applied to the new data sets. To explore calcium-induced effects, the myosin IX sample was supplied with a calcium concentration close to the physiological range (≈ 10 µM, pCa 4), mixed with F-actin, negatively stained, imaged and analyzed analogous to the methods described in section 6.4.1 and 6.4.2.

6.2.2.1 Calcium does not influence the order, periodicity and properties of the myosin - actin - crosslinks

Figure 6.19 shows three raw micrographs of myosin IX - actin - crosslinks in the presence of Ca^{2+}. The first difference to the calcium-free condition was the high tendency of myosin to form aggregates. The preparation and staining of the sample, as well as the acquisition of suitable areas was challenging due to the high amount of background myosin. Unfortunately the concentration had to be kept high since most of the molecules rather formed aggregates than crosslinks. The circled areas in micrograph **A** shows such aggregates, which would make the analysis of underlying crosslinks impossible. However, the overall formation of crosslinks was similar to the case without Ca^{2+}. Areas free of aggregates, such as the one between the circled areas in **A**, indicated the formation of crosslinks with several actin filaments included. Apart from that, crosslinks with only two or three filaments did also occur as seen in the micrographs **B** and **C**. In **B** the cooperative and highly ordered crosslinking of myosin IXa (arrows) could be observed similar to the Ca^{2+}-free condition. On the other hand, as indicated by the turquoise arrows in micrograph **C**, few events of myosin IXa bound to an actin filament and not being involved in the formation of crosslinks could be seen only in the presence of calcium.

For a more detailed analysis of these crosslinks, whether the general shape, periodicity, polarity and myosin structure itself was influenced by calcium, 201 images of crosslinks were picked and analyses as described in the methods section[7].

In contrast to the no-calcium data set, the images picked here had lower quality (mainly due to aggregates of myosin) and were fewer in total. 201 images could be used to create the first global average. The alignment however was not done completely reference free, but using the aligned global average of the Ca^{2+}-free condition (Figure 6.6) to keep the images centered during the alignment (see section 3.1.2 for more details). The aligned average of myosin IXa MD crosslinking actin filaments in the presence of calcium is shown in Figure 6.20. Although the quality was lower, the average was still resolved well enough to measure the inter-crosslink distances as well as the distances between the actin filaments. These measurements were in exact agreement with the non-calcium condition, indicating that calcium did not have any effect on the periodicity of the crosslinks. The magnified image in **B** shows, apart from the minimal and maximal distance between the two actin filaments (5 nm and 9 nm respectively), similar binding patterns of the crosslinking myosin. This was consistent with myosin IXa MD binding to F-actin at four actin-monomers.

[7] Although the basic SPA methods were the same as for the Ca^{2+}-free condition, small modifications had to be done, which will be indicated whenever necessary.

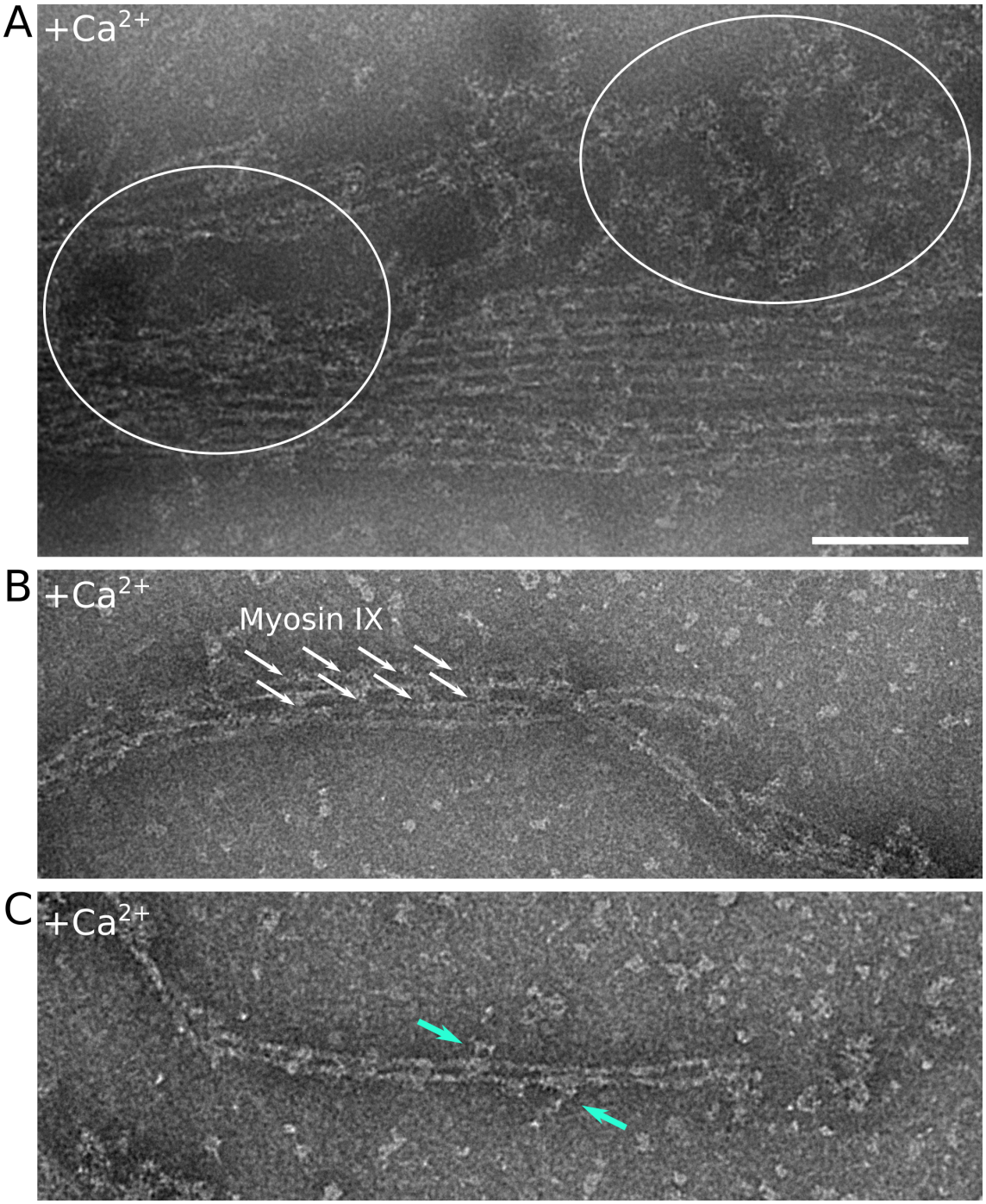


Figure 6.19: Raw micrographs of myosin IXa and actin in the presence of calcium (pCa 8). **A** Samples with calcium produced more myosin-aggregates (circled areas). **B** Crosslinks formed in a similar cooperative and highly ordered pattern as without calcium (white arrows indicate myosin IX molecules). **C** In the presence of calcium, singe binding events to actin of myosin could be observed (turquoise arrows). Scale bar: 100 nm

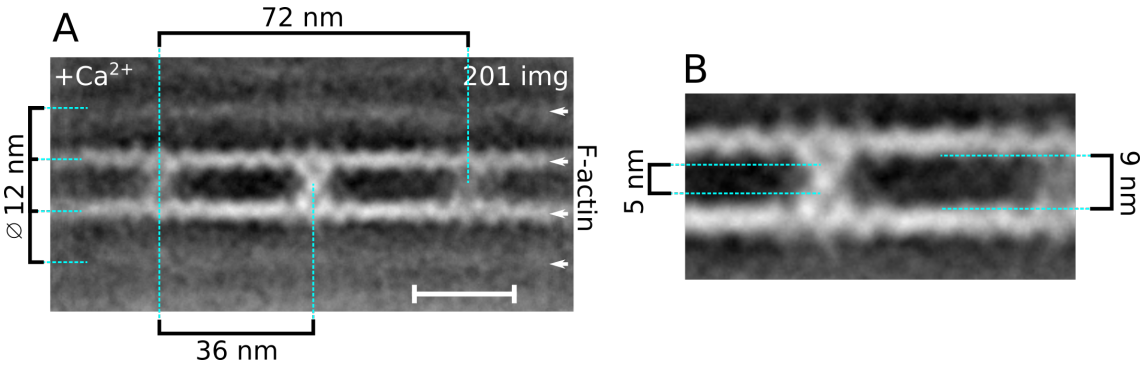


Figure 6.20: Global aligned average revealed the general shape of crosslinks. **A** The distance of 36 nm between to myosin IXa molecules matches the actin helical repeat. Actin filaments have a mean distance of 12 nm, with 5 nm and 9 nm as extreme values for the distance between two filaments (**B**). Upper right number: Number of images in the average. Scale bar: 20 nm

To further investigate this property, images were windowed to the size of 90x90px and further processed as described in the following.

6.2.2.2 Calcium influences the polarity of actin filaments crosslinked by myosin IX

In analogy to the non-Ca^{2+} condition, all 201 images were first classified with respect to the four binding positions using the mask shown in Figure 6.21 **A**. Interestingly, the same three binding patterns were discovered, namely the diagonal, bent and inchworm conformation, each present in mirror-related orientations on the grid (**C**). 47 images did not have enough resolution to determine the binding pattern and thus were not used for further processing. The remaining 154 images were combined to create a prealigned data set, shown in **C**. This image stack then underwent the established analysis of polarity and phase:

Figure 6.22 **B** shows the averages of the centered images, which were aligned with two references (**C**). The two graphs summarize the necessary rotations for the reference-based alignment (**D**) as well as the information whether mirroring was necessary (**E**). These two parameters for each of the four data sets (lower / upper filament, polarity 1 / 2) were combined and shown in Figure 6.23 **C** (see also Table 6.5 for absolute image numbers)). The comparison of the resulting statistics for each of the two filaments with respect to the alignment with both polarities (consistency check), is shown in **D** and the corresponding percentages listed in Table 6.6. For both filaments, the majority of images (48.70% and 50.65% for the lower and upper filament respectively) were consistent with respect to polarity and phase. Only approx. 30% were inconsistent. These were similar values to the non-calcium condition, although with approx. 22% there were less

inconsistent images. This reflected the overall better quality of the images of myosin IXa crosslinks in the absence of Ca^{2+}.

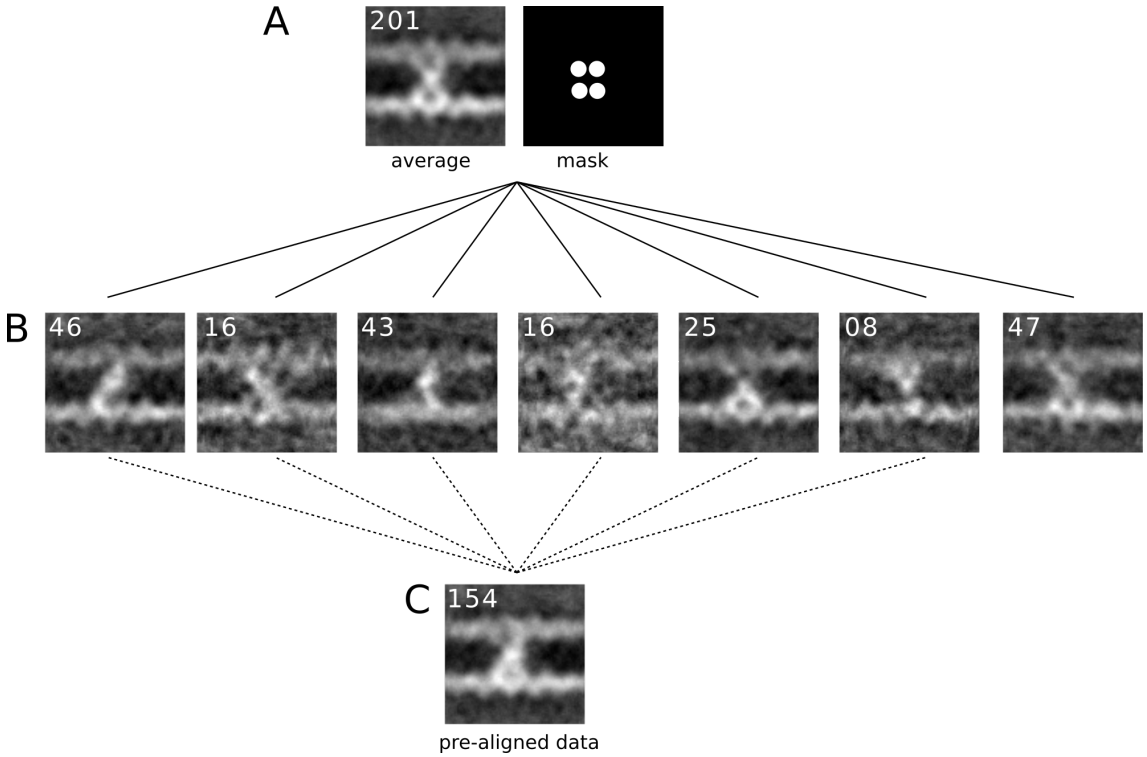


Figure 6.21: **A** The windowed data set of 201 crosslink images in the presence of calcium was classified with respect to the binding sites of myosin using the shown mask. **B** Three binding patterns were observed on 154 images, each present in mirror related orientations. 47 images had not sufficient resolution. **C** Creation of a prealigned data set where all images contain myosin bound to the lower left position, obtained by mirroring each image accordingly. Upper left number: Number of images on averages. Scale: images are 30 nm wide.

	lower filament polarity 1	lower filament polarity 2	upper filament polarity 1	upper filament polarity 2
both	31 img	27 img	45 img	47 img
mirror at y axis	43 img	48 img	30 img	31 img
180° rotation	30 img	25 img	35 img	43 img
nothing	51 img	55 img	45 img	34 img

Table 6.5: Absolute values for graphs **C** in Figure 6.23

The results from the next step, comparing the images of the lower with the upper filament, is shown in Figure 6.23 **E** and the corresponding percentages can be seen in Table 6.7. Approx. 80% of images could not be used for the polarity / phase assignment, as either one or both filaments were inconsistent. Only 31 images had sufficient resolution in both filaments to do so. Figure 6.24 **B** illustrates the distribution of consistent images with the result that the majority of crosslinks (45.16%) comprised the two actin filaments

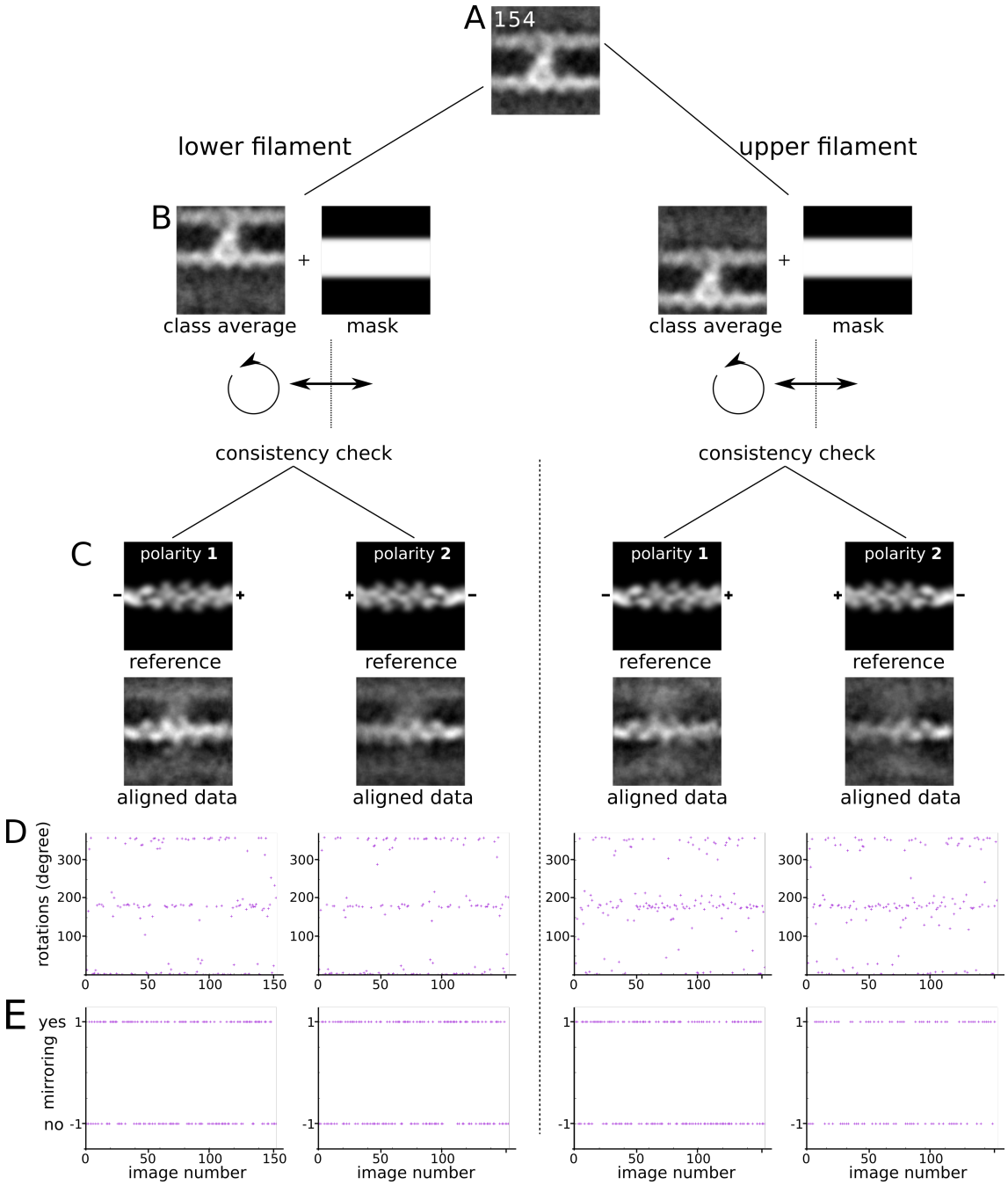


Figure 6.22: **A,B** Separate analysis of lower and upper filaments in the crosslink images in the presence of calcium. **C** Each image was aligned with a simulated actin filament in two polarities as a reference to examine the quality of this filament. **D,E** Optimal rotation and mirror parameter applied during the alignments. Scale: Images are 30 nm wide.

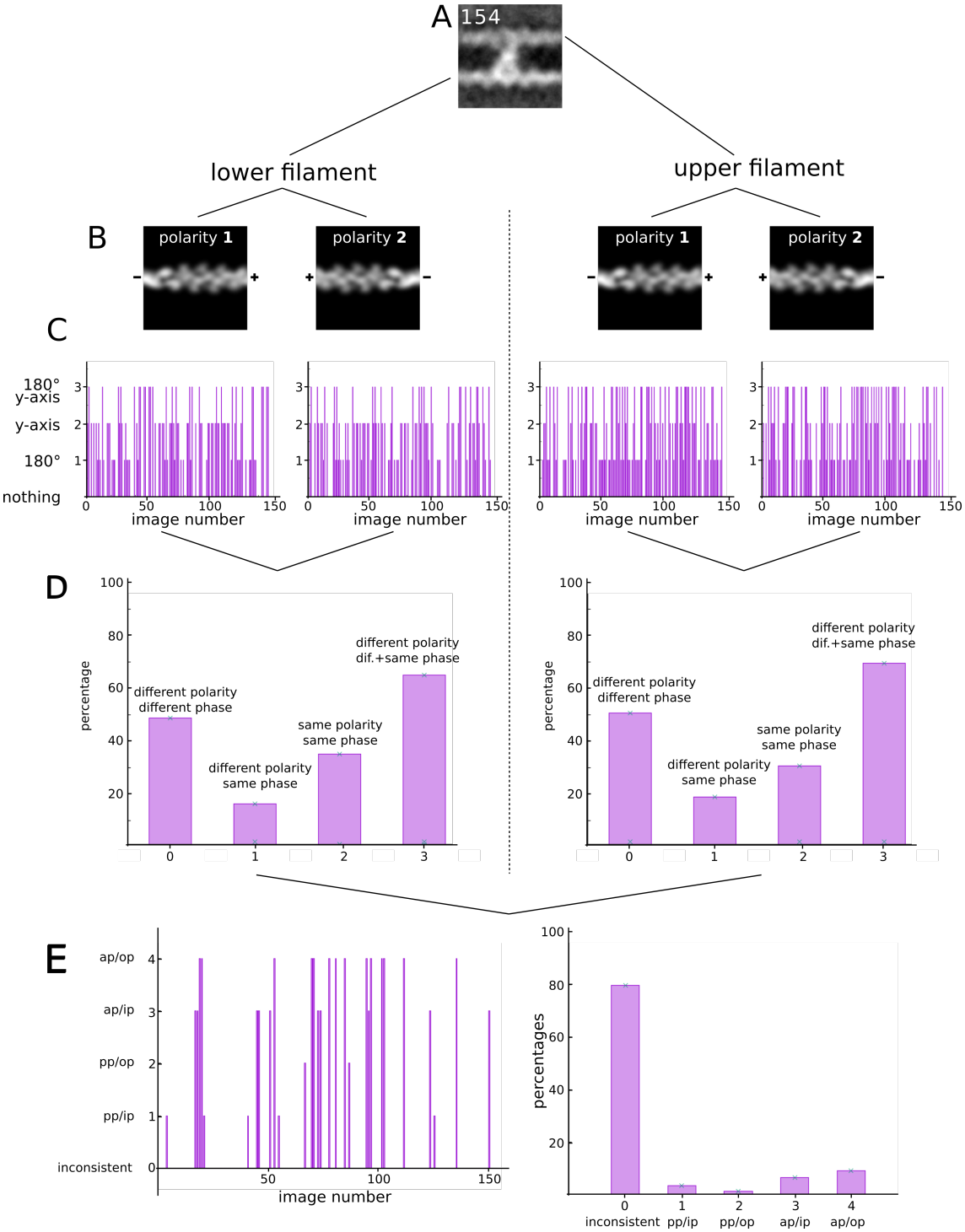


Figure 6.23: **C** Combined and simplified alignment parameter from the alignment of the lower and upper filaments to the reference images **B** with polarity 1 and 2. **D** Comparison of alignments with polarity 1 and 2 of lower and upper filaments to identify consistent filaments. **E** Comparison of lower and upper filaments. 123 images (79.87%) were inconsistent, the remaining images identified a trend towards anti-parallel actin filaments and out of phase (14 images / 9.09%).

		lower filament	upper filament
1) **consistent**	different polarity different phase	075 img / 48.70%	078 img / 50.65%
2) **polarity consistent**	different polarity same phase	025 img / 16.23%	029 img / 18.83%
3) **inconsistent**	same polarity same phase	054 img / 35.06%	047 img / 30.52%
1)+2)	different polarity (diff.+same phase	100 img / 64.94%	107 img / 69.48%

Table 6.6: Absolute values for graphs **D** in Figure 6.23

in anti-parallel polarity and out of phase (see also Table 6.7). This was in complete reversal to the non-calcium condition where approx. 97% of the crosslinks had actin filaments in with parallel polarity and in phase. Furthermore, the result for the $+Ca^{2+}$-condition was less significant with 32.26% being anti-parallel but in phase, and 22.58% with parallel polarity actin filaments (16.13% in phase and 06.45% out of phase). Therefore, the images were separated according to their polarity / phase relation and separately aligned with the corresponding references, as shown in Figure 6.24. The resolution of these images however was with 38.02 Å (10 iterations of FRC 0.5 cutoff criterion) very low due to the small data set of max. 14 images. If it can be assumed that the polarity of actin filaments crosslinked by myosin IX was as strongly restricted as it was in the absence of calcium, but the resolution and size of the EM data was simply too low to reveal this here, the whole data set could be aligned with the reference of two actin filaments with anti-parallel polarity and out of phase **D**. However, even in the polarity-aligned average, the resolution was still relatively low, especially of the actin filaments. The crosslinking myosin on the other hand was not much lower resolved than in the non-calcium condition and the closer investigation of its conformations will be subject of the next section.

		all images (154 img)	only consistent images (31 img)
inconsistent	123 img	79.87%	
parallel polarity / in phase	005 img	03.25%	16.13%
anti-parallel polarity / out of phase	002 img	01.30%	06.45%
anti-parallel polarity / in phase	010 img	06.49%	32.26%
anti-parallel polarity / out of phase	014 img	09.09%	45.16%

Table 6.7: Absolute values for graphs **E** in Figure 6.23

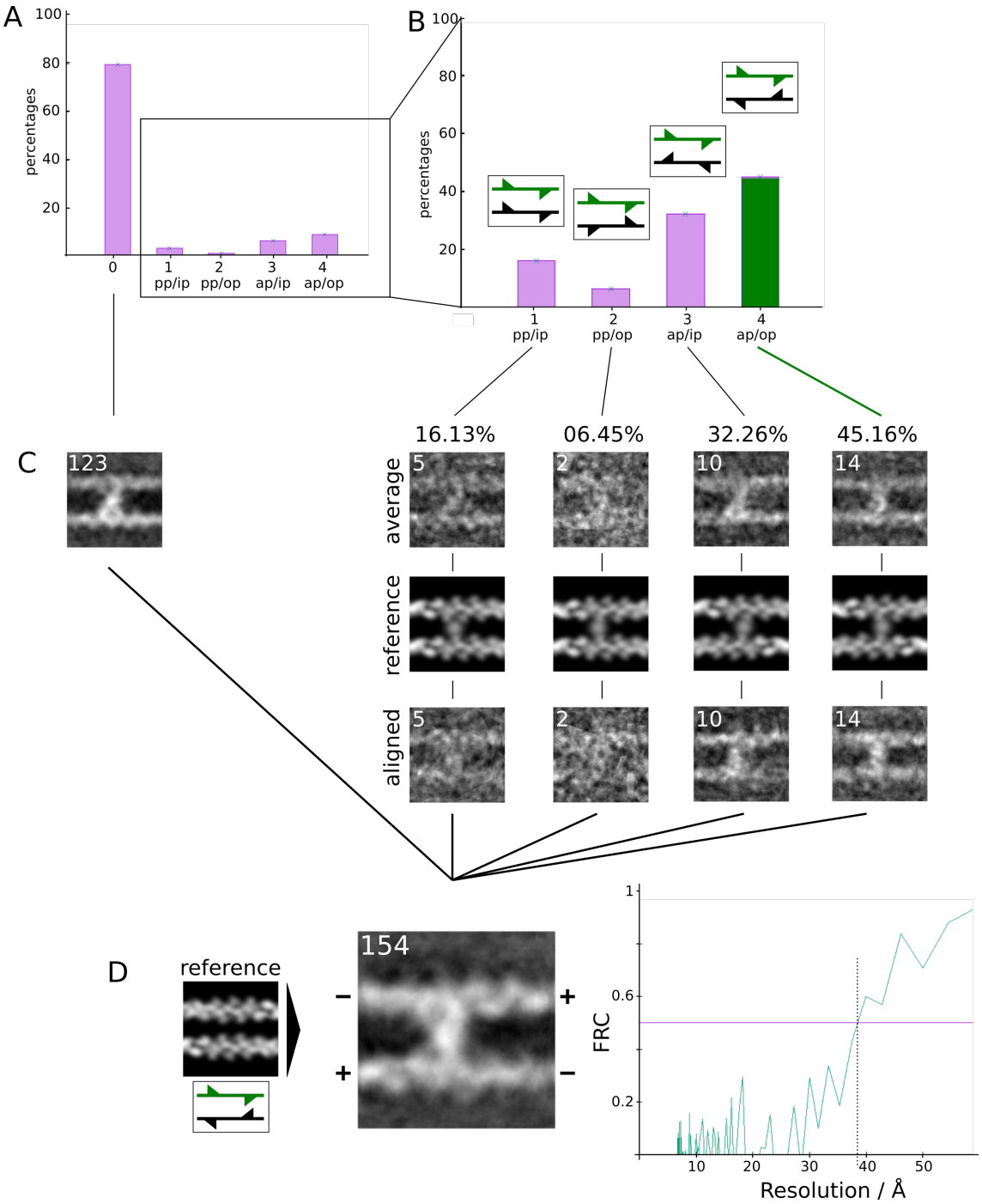


Figure 6.24: Polarity and phase assignment to all (**A**) 451 and only consistent (**B**) crosslink images in the presence of calcium. **C** Corresponding class averages for each case. **D** Global average of all images aligned with the dominating pattern, anti-parallel and out of phase ap/op. The global average of the aligned data comprised a resolution of 38.02 Å (FRC 0.5-cutoff criterion). Upper left number: Number of images on averages. Scale: EM images are 30 nm wide.

6.2.2.3 Myosin IX conformations in the presence of calcium

Using the polarity information of myosin IX - actin - crosslinks in the presence of calcium revealed in the previous section, the polarity-aligned images data set (it was assumed that all crosslinks have anti-parallel and out of phase actin filaments) was reclassified with respect to the myosin binding positions. Figure 6.25 shows the three conformations which were present in similar ratios compared to the calcium-free data (see Table 6.8). The majority of myosin adopted the crosslinking diagonal or bent conformations (40% and 38% respectively) and only 22% were found to bridge to actin monomers of the same actin filaments and therefore not contributing to the crosslink directly.

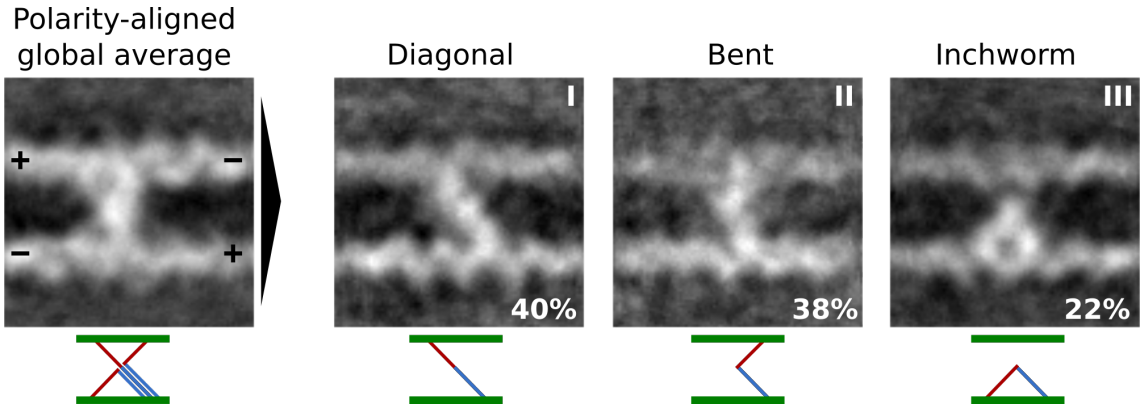


Figure 6.25: Myosin IXa adopts three conformations in the process of actin-crosslinking: As in the calcium-free condition, two crosslinking conformations (diagonal and bent) and one non-crosslinking, inchworm conformation were found.

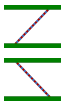
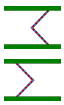
Conformation	Number of images	Percentage	Combined percentage of usable images
	46	22.89%	**40.26%**
	16	07.96%	
	43	21.39%	**38.31%**
	16	07.96%	
	25	12.44%	**21.42%**
	08	03.98%	
	47	23.38%	-

Table 6.8: Statistics on shape-appearance. Given are the absolut image numbers, percentages and combined percentages where the last group (unidentifiable conformation) was excluded.

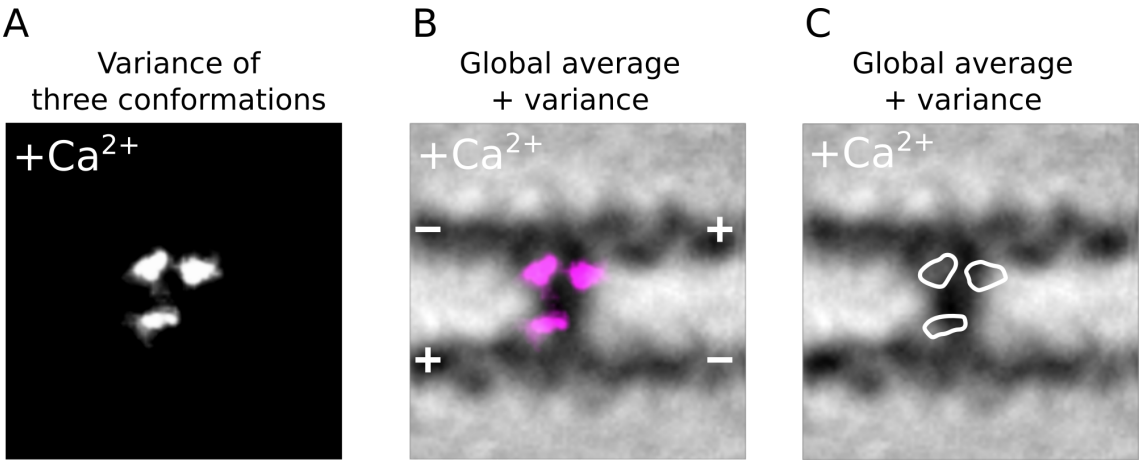


Figure 6.26: Variance analysis of myosin IX crosslinks in the presence of calcium. **A** Variance of the three conformations identifies loci of the second actin-binding. **B,C** Variance overlay onto an inverted version of the class average

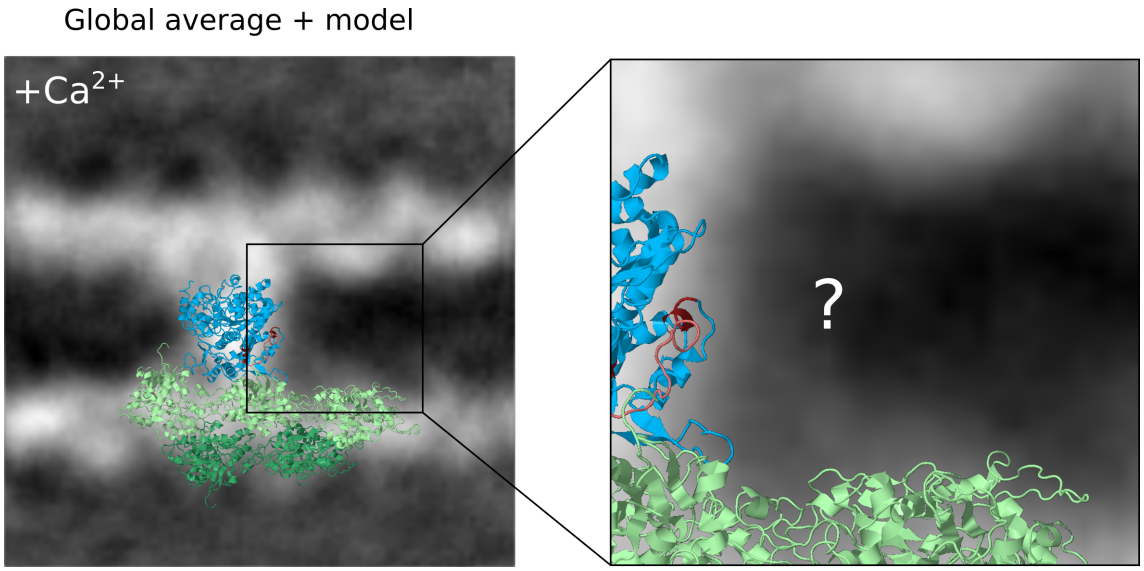


Figure 6.27: Myosin IE - actin crystal structure in the minus-calcium optimized orientation superpositioned with the EM average. An additional mass presumably representing the calmodulin could not be observed in this global average.

To confirm and identify the structure of the crosslinking myosin in more detail, the variance image of the global average was investigated as shown in Figure 6.26 **A**. The superposition of the variance indicated the position of the varying second binding site of myosin (**B** and **C**). These findings were in full agreement with the non-calcium condition.

Due to the fact that the overall resolution of the data was much lower, a proper crystal structure projection matching could not be applied. Instead, the same actin-myosin-model (one myosin IE molecule attached to the middle of five actin monomers) in the "non-Ca^{2+}-optimized" orientation was overlaid with the EM average, see Figure 6.27. Both, the actin monomers and the mass in the center could be explained by the crystal structure. However, an additional mass (at the position of the question mark) was present in the absence of calcium. This mass was identified to presumably represent a calmodulin bound to the insert in loop2 (see Figure 6.18) and was not seen here in the global average of myosin IX in the presence of calcium. Previous studies on the other hand have shown that calcium concentrations did not influence the binding of calmodulin to the insert [110], and therefore a CaM should be seen here as well.

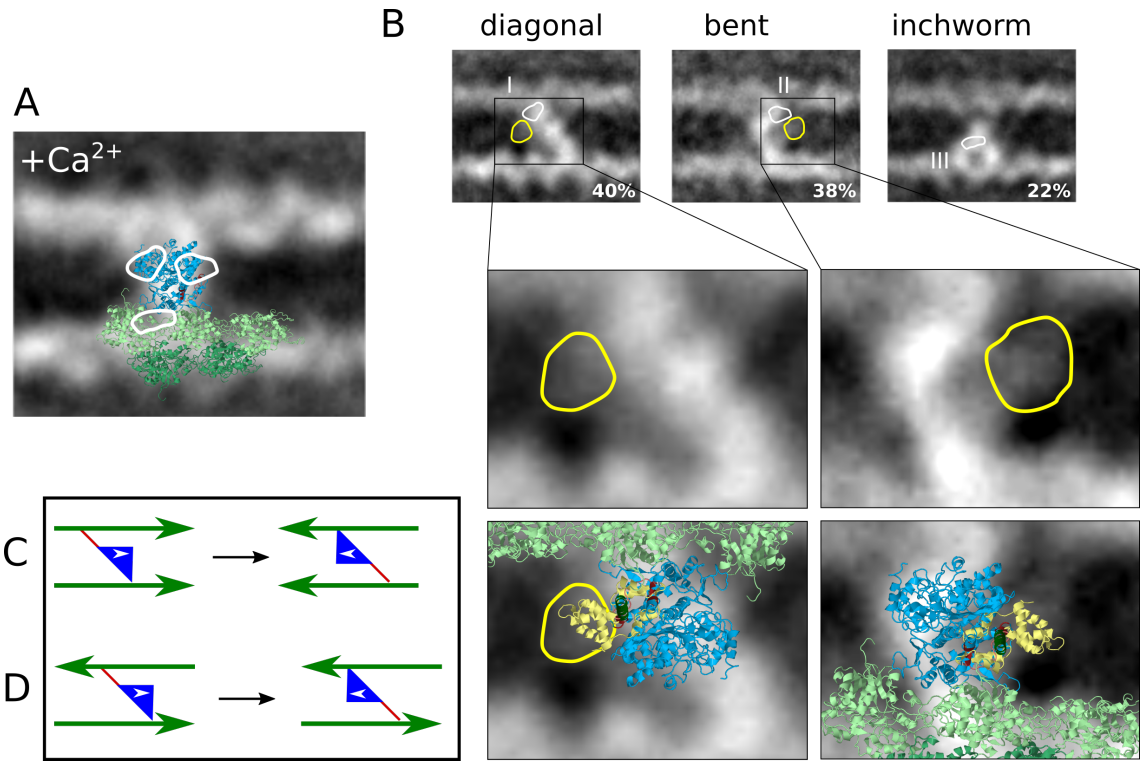


Figure 6.28: Separation of EM data (**A**) into the three conformations (**B**) revealed in conformation I and II the presence of additional mass on the left and right side of the myosin mass. The magnified areas explain these additional masses by overlaying the myosin IE crystal structure.

If the image data set was separated into its three conformations, however, the diagonal and bent conformations (I and II) showed weak intensities (note the size of the data set) of an additional mass (Figure 6.28 **B**, yellow outline, overlaid with the calmodulin-modified myosin VI model in the low-Ca^{2+}-optimized orientation, see section 6.4.2.5). While the high calcium bent conformation showed similarities with the low calcium condition (calmodulin on the same side and position of the central mass), the calmodulin in the diagonal conformation seemed to be on the other side of the central mass. Given the fact that calmodulin bound to the N-terminal of the insert and keeping the size of calmodulin in mind, it was very unlikely that calcium induced such strong structural changes that the CaM changed its position to the other side of the head domain. More likely, this average shows a myosin bound to the upper filament as indicated in the figure. Such an effect is only possible in the case of crosslinks with filaments of anti-parallel polarity, as indicated in the cartoons **C,D** (actin in green, myosin head in blue, insert in red, white arrowhead indicate polar binding of myosin to actin): If the crosslinks have parallel actin filaments, a rotated structure of myosin resulting in the same diagonal orientation can only be achieved if the whole polarity changes (**C**). In the case of antiparallel crosslinks on the other hand, the polarities stay the same, even if the myosin rotates by 180°. This explains, why only here a mass representing the calmodulin was observed on the left side of the aligned data set. One possible explanation for the missing mass in the inchworm conformation could be that the calcium-CaM underwent structural changes, which positioned it in-line with the motor domain itself. In the 2D projections of the negative stain images its structure would be hidden by the much larger intensities of the rest of the catalytic domain.

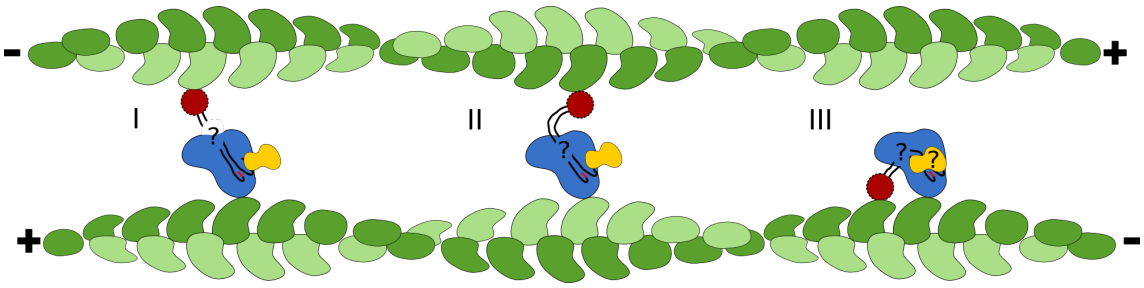


Figure 6.29: Summary of myosin IX - actin - crosslinks in the presence of calcium. The crosslinked actin filaments were mainly antiparallel and out of phase. Myosin IX adopted three conformations where in two a possible attached calmodulin could be resolved (I and II). The polarity switch in comparison to the non-calcium conditions suggested conformational but still unresolved changes in the loop2.

These results are summarized in the cartoon in Figure 6.29, where an overall anti-parallel polarity of the actin filaments (and being out of phase) was assumed. The presumably hidden position of the calmodulin is indicated by the question mark in the inchworm conformation. Furthermore, as the polarity of the actin filaments was in complete

reversal in the presence of calcium, the calmodulin in its calcium-state might have had a structure-altering effect on the loop2, which however could not be resolved yet.

6.3 Discussion and Outlook

In the following section the results of the single particle analysis of negatively stained electron micrographs of myosin IX - actin - crosslinks in the absence of calcium will be summarized, put into context with already published characterizations of myosin IX. Furthermore a model of how a monomeric myosin is able to regulate the formation of actin-crosslinks is suggested. Eventually, these findings will be then compared with the results on the same crosslinks in the presence of calcium.

6.3.1 Model: How a monomeric myosin self-organizes actin networks

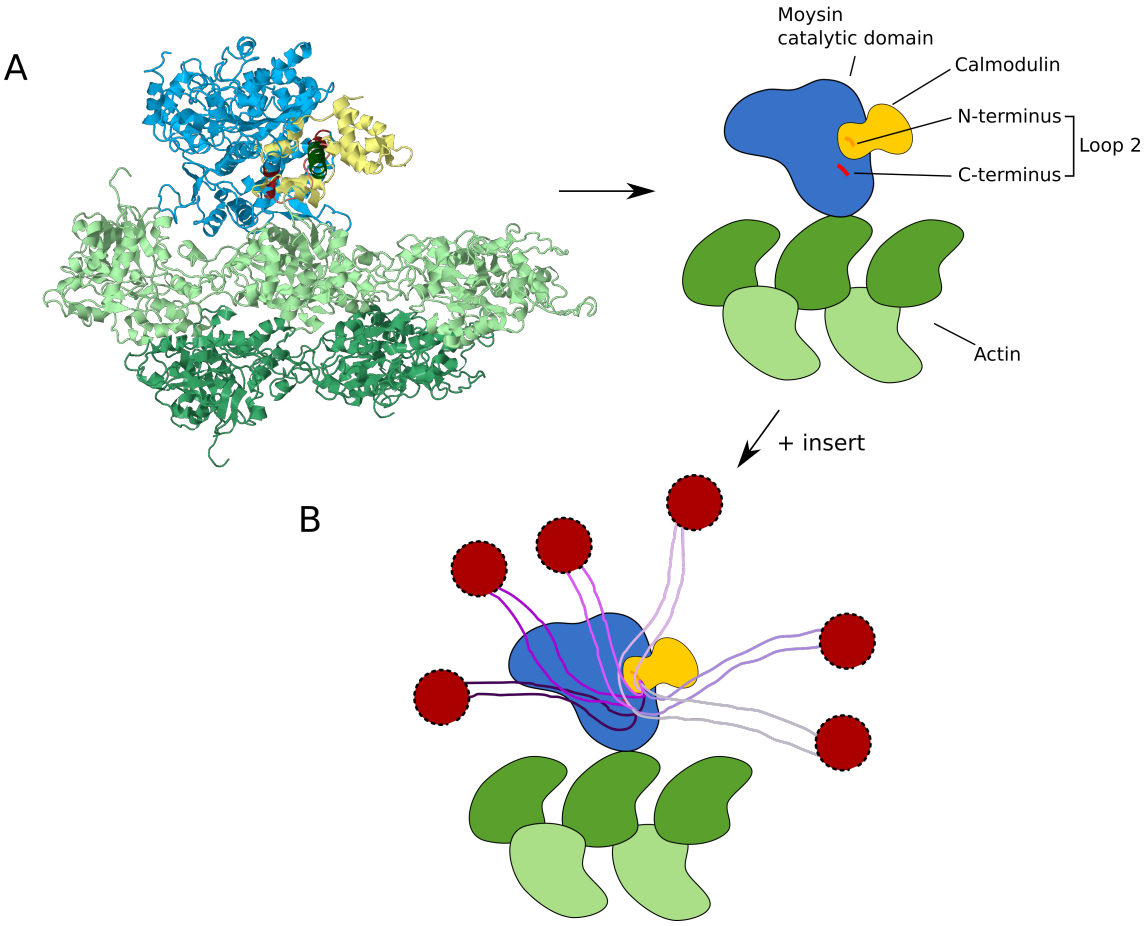


Figure 6.30: **A** The crystal structure of myosin IE bound to actin in an orientation optimized to match the EM data was converted into a cartoon. **B** The myosin IXa specific, 207 aa long insert in loop2 has a yet unresolved structure and is therefore shown in various possible orientations. Note that these represent just a small number of possibilities. The known second actin binding site in this insert is illustrated as a red blob.

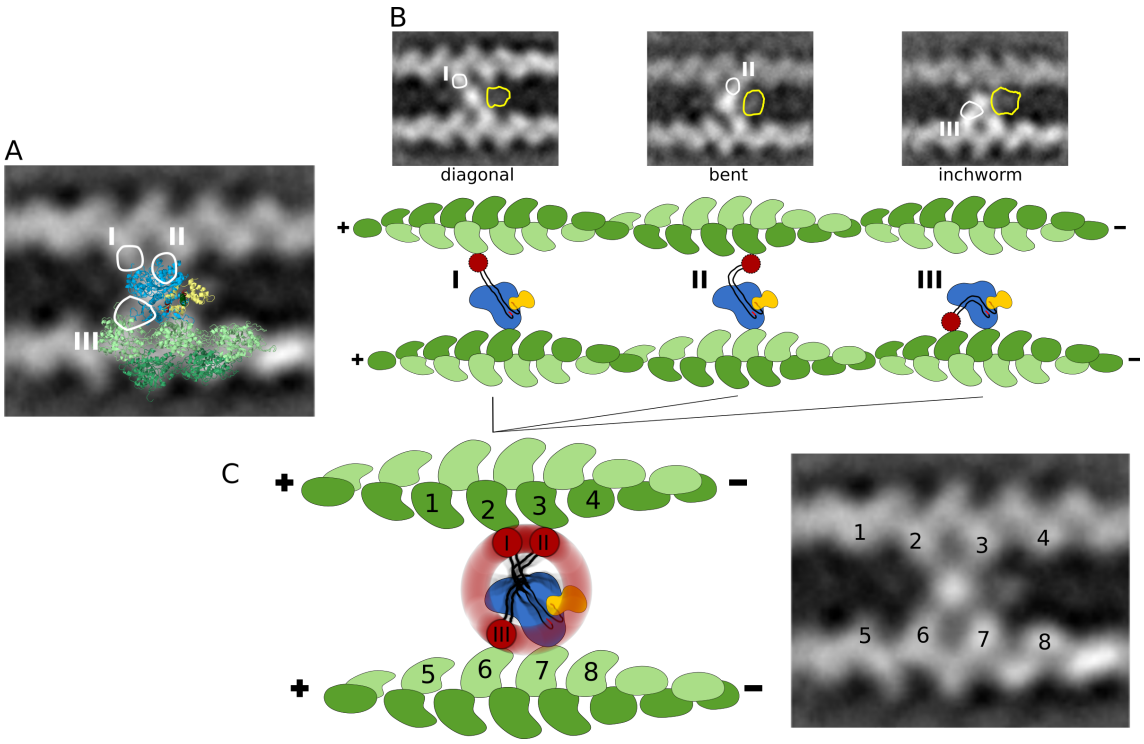


Figure 6.31: Summary of SPA on myosin IX in the presence of actin and absence of calcium. **A** Superposition of EM polarity aligned EM average, crystal structure and outlines from the variance analysis, indicating the three additional binding sites I-III. **B** Classification into the three conformations with additional outlines (yellow) of the assumed position of calmodulin. The cartoon below illustrates the realization of these conformations. **C** A superposition of conformation I,II and III explains the presence of a probable search radius of the unique insert, resulting in only three possible binding positions on actin: position 2 for conformation I, 3 for II and 6 for III. Position 7 is reserved for the primarily actin binding site of myosin IX.

In order to explain the three discovered conformations of how myosin IX crosslinked actin in the absence of calcium (section 6.15), the optimized crystal structure (Figure 6.18) was converted into a cartoon as illustrated in Figure 6.30 **A**. This cartoon does not contain the approx. 207 aa from the special insert, as it could not be resolved directly, neither by crystallography, nor by microscopy. On the other hand, as it was discovered that the second binding site of myosin IX was located within this insert (see section 6.1), it made sense to assume that, due to its length and flexibility, this part of myosin IX was able to reach out and search for actin monomers close by to attach. In **B** a small number of possible arrangements of the flexible loop are illustrated. In combination with the previous results revealed in this studies, this seemingly undefined search-behavior could be characterized. Figure 6.31 **A** shows a superposition of 1) the global polarity aligned EM average, 2) the optimized crystal structure of myosin IE including the modeled calmodulin and 3) the outlines of the variance image indicating the second actin binding sites as well as the position of the calmodulin. Splitting this superposition into the three discovered conformations (**B**) it was possible to indirectly characterize the structural uniqueness of myosin IX and in particular of the special insert in the catalytic domain by exploring EM intensities belonging to the second binding site. Furthermore, the yellow outlines, which were separately calculated for all three conformations, indicated similar positions of the calmodulin. The cartoons below summarize and explain the behavior of the special insert. Its flexibility seemed to be more restricted than shown in Figure 6.30 **B**, resulting in the three identified conformations: diagonal (I), bent (II) and inchworm (III). If these three conformations were superimposed (**C**), there seemed to be a distinct search radius of the insert for possible actin binding events. The existence of such a search radius (indicated by the blurred ring) explained why only the three identified conformations were found and no second binding to other actin monomers (such as number 1, 4, 5 or 8 in **C**) did occur. The cartoon of the actin filaments comprised both information, the polarity and phase. While the polarity is indicated by asymmetric actin monomers, the phase can be identified by looking at the two (pseudo)-helices of actin, colored in dark and lightgreen. The position of each helix was the same for both filaments, equivalent to an alignment in phase.

Based on these findings, the following model illustrates how a monomeric myosin is able to self-organize actin networks (Figure 6.32):
(1) A free myosin IXa MD monomer in solution eventually binds to an actin filament with its conserved primarily actin binding site. Once bound to actin, the flexible insert in loop2 with its second binding is able to freely move around in search of other actin monomers. (2) In the presence of a second actin filament in the right polarity (parallel and in phase to the first one), myosin IX is able to form a crosslink between two fil-

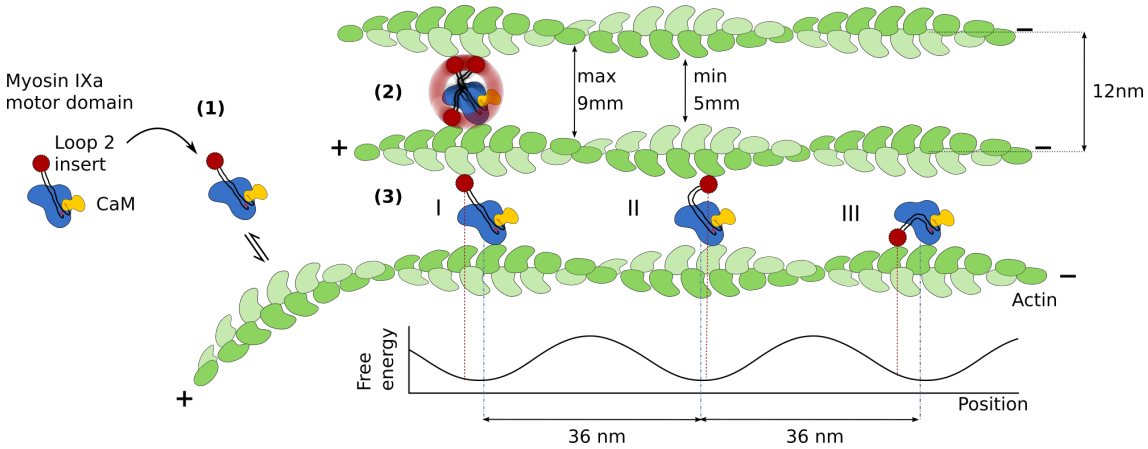


Figure 6.32: Suggested model of how a monomeric myosin self-organizes actin networks: **(1)** A free myosin IXa MD monomer binds actin and, if present, establishes a crosslink with another actin filament **(2)**, aligning them parallel and in phase. From here, other myosin molecules are recruited to extend the crosslink. **(3)** Myosin can adopt, apart from the two crosslinking conformations I and II, a third conformation that does not actually connect two actin filaments, but enables the motor to move processively and seek for its preferred binding site. The position of the crosslinks is determined by minimizing the free energy.

aments, acting as a seed for further myosin molecules to bind. (3) In this state, other myosin molecules can be recruited to join and extend this crosslinking process, which becomes the more stable, the more myosin molecules are involved. The position of new crosslinks is determined by the spatial arrangement of F-actin, in particular by its periodicity of 36 nm and spacing of 12 nm. This creates a free energy landscape which is minimized in the indicated positions. Myosin IXa can adopt three distinct conformations within these crosslinks, dependent on the sterical restrictions of myosin and actin, which together create the energy landscape shown in the cartoon. As the statistics showed (table 6.4), the diagonal conformation (I) is the most preferred one, followed by the bent conformation (II). With increasing stability and size of the crosslinks, the inchworm conformation (III) - which in fact does not crosslink two filaments - becomes also possible. This also explains, why no single myosin molecules bound to single actin filaments were observed (see section 6.2.1.1), as the presence of a second actin filament is required for myosin to adopt a stable conformation.

On the cellular level, the initiation of highly coordinated actin-myosin-networks introduces orientated tracks at specific sites in the cell. For one thing, keeping in mind the previously mentioned RhoGAP domain located in the tail of myosin IX, the induced networks could focus RhoGAP activity at definite locations in the cell. For another, these lattices can be used by a number of other myosin classes as orientated tracks and also introduce force-sensing mechanisms into the cytoskeleton, regulated by myosin. Such mechanisms are involved in cell polarization and collective cell migration.

6.3.2 Myosin IX - actin - crosslinks in the absence and presence of calcium

It was shown that the motor domain of myosin IXa monomers was able to form crosslinks with actin filaments [110] via a second actin-binding site. The analysis using the electron microscopy revealed striking information about the polarity of the actin filaments and the shape of the crosslinking myosin. As presented in Figure 6.33 and the two graphs in Figure 6.34 and 6.35, without calcium, myosin IX monomers crosslinked actin filaments almost exclusively **parallel with respect to the polarity and phase of the filaments**. The measured **periodicity of the crosslinks of 36 nm** matched precisely the helical pseudo repeat of F-actin. **The myosin itself was present in three distinct conformations**: diagonal (I), bent (II) and inchworm (III), where only the diagonal and bent conformation contributed directly to the formation of crosslinks via the binding of the two actin-binding sites of myosin IX to different actin filaments. The inchworm shape, bridging between two adjacent actin monomers of the same filament, was the less abundant. While the myosin in this state does not crosslink actin filaments, its role could be more concentrated on finding the right binding position. Although it could not be shown yet that myosin IX was able to actually move processively along single actin filaments, it has been reported for other myosin IX isoforms [102–107]. On the other hand, the slow ATPase activity of myosin IX and thus low speed of processive movement suggest processivity, which however can be challenging to detect, as the movement might lie below the resolution limit of the measuring-device, such as optical trap[8].
In this sense, myosin IX could be able to move in an inchworm-type processive movement along actin in order to seek out the preferred binding position to establish a crosslink with another actin filament. Given the sterical hindrance of long actin filaments, a slow movement would be preferred, which correlates with the reported [110] relatively slow ATPase of myosin IX (compared with other myosin and kinesin classes (7)). Furthermore, the at the beginning of this chapter mentioned controversies about the directionality of myosin IX (plus-end vs. minus-end directed) could be explained by the hypothesis that this myosin class is in fact able to move in both directions on actin filaments. This bi-directionality of myosin IX might increase the efficiency for searching its preferred binding position. Together with the result that only approx. 25% were found in this inchworm conformation and the remaining majority took part in the actual crosslinking of actin filaments (see the graph in Figure 6.35), the role of myosin IX however is more structural in the sense that it is more involved in the organization of the actin cytoskeleton, rather than in the transport of cargo. Moreover, the experiments presented here

[8]The optical trap is a widely used technique for the analysis of processive movement of myosin

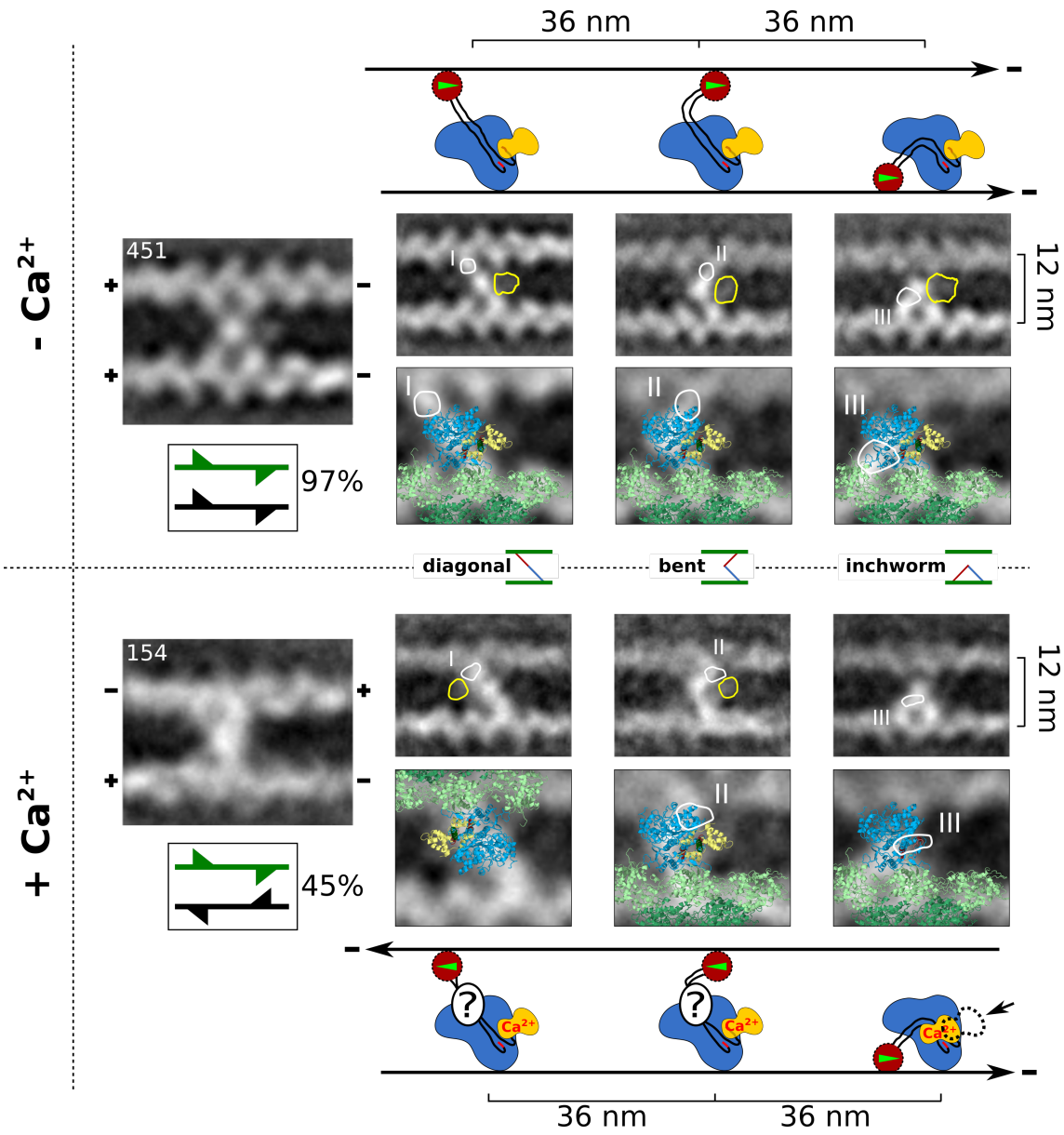


Figure 6.33: Comparison of myosin IX - actin - crosslinks in the absence and presence of calcium. The general measurements of the crosslinks, a distance of 36 nm between the myosin molecules and an interfilament distance of approx. 12 nm was not influenced by calcium. In the absence of calcium, the majority of crosslinks were comprised of parallel actin filaments and in phase. At high calcium, the relation changed to antiparallel and out of phase. Apart from overall similar conformations, the calcium-calmodulin in the inchworm condition was presumably hidden.

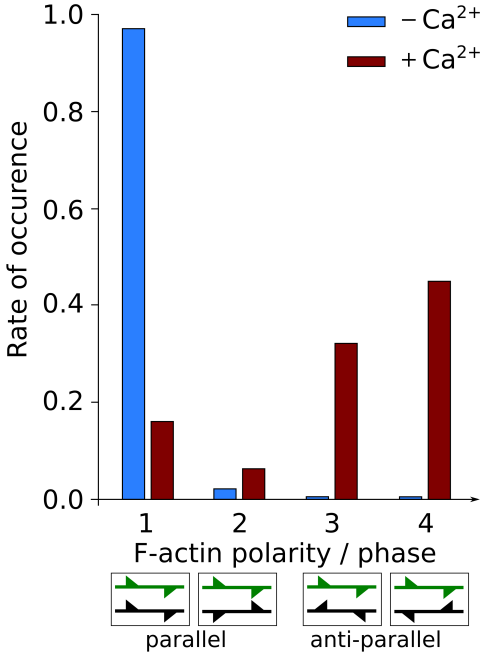


Figure 6.34: Statistical distribution of calcium dependent F-actin polarity and phase

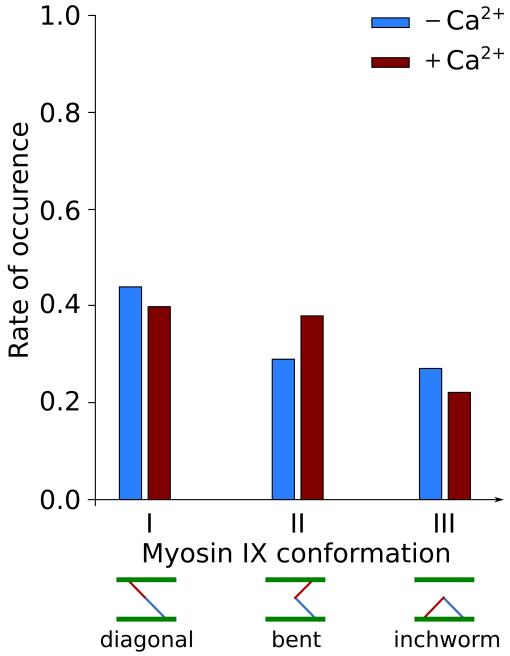


Figure 6.35: Statistical distribution of calcium dependent myosin IX-conformations

were carried out in the absence of any nucleotide, which shows that this searching was not necessary for the formation of actin-bundles in general.
As reported, the myosin IXa motor domain was able to bind calmodulin at the N-terminal part of its loop2, where the approx. 207 aa long unique insert is located. **This calmodulin could be identified in the electron microscopic data**. The interpretation of the EM data by comparing it with a crystal structure model of myosin IE bound to actin (including a calmodulin modeled onto the N-terminal of loop2 of myosin IE) revealed the possible structural location of this calmodulin. In Figure 6.33 the yellow outlines indicate the corresponding intensities, being located on the right hand side of the central myosin mass. The revealed characteristics about the polarity of actin, the structure of myosin and the position of the CaM are illustrated in the cartoons (black arrow: F-actin, blue: myosin IXa catalytic domain, yellow: calmodulin, black lines and red circle: special insert in loop2).

In a complete reversal, the polarity of the crosslinked actin filaments in the presence of calcium (pCa 4) was anti-parallel and out of phase (approx. 45%), although the result was not as significant as without calcium. The lower part of Figure 6.33 depicts the behavior of myosin IXa under the influence of calcium. Although the polarity changed, **the same three conformations of myosin were found and in similar ratios**. Here again, the diagonal conformation was dominating, followed by the diago-

nal. The inchworm conformation was only present in 22% of the data. In contrast to the non-calcium condition, the EM intensity presumably representing the bound calmodulin was only present in the diagonal and bent conformation. The crystal structure in the diagonal conformation was turned by 180° in order to cover the mass situated on the other side. In the inchworm conformation, no such mass could be identified. However, given the strong binding of CaM in the absence of calcium (seen in the EM data in all conformations), the calmodulin should be present as well in the plus-calcium-data. However, calcium-induced structural changes of calmodulin have been reported (compare the results on myosin VI, chapter 5) and might have occurred in this case as well. These changes could have shifted the position of the calmodulin in a way that it lined up with the catalytic domain. In one line with the rest of the myosin molecule, it would not be seen in the TEM images. The cartoons illustrate this behavior (see dotted line in the $+Ca^{2+}$, inchworm cartoon image in Figure 6.33). The illustrated Ca^{2+}-CaM molecule in the inchworm conformation only give a rough idea of why CaM is not seen in the EM. As no more information about the actual position of the CaM could be revealed, it was left out of the crystal structures model.

On the other hand, the observed intensity in the diagonal and bent conformation had comparable locations (yellow outlines) to the one without calcium. The fact that the majority of crosslinked actin filaments were anti-parallel in the presence of calcium, supports the hypothesis that calcium-induced structural changes of the calmodulin bound to loop2 take place in these conformations too. The changes might have longer range effects on the structure of the remaining insert as well (the question marks in the $+Ca^{2+}$, diagonal and bent cartoon images in Figure 6.33 indicate this), causing the second actin binding site to be altered in terms of polarity. This effect is indicated by the green arrows within the actin binding part of the loop2-inserts (red circles). The direction of these arrows reflects the polar binding to actin (pointing towards the minus end). Without calcium all green arrows point towards the same direction (right) as the minus end of the lower actin filament, resulting in the upper filament to be aligned parallel. In the presence of calcium however, the binding of the insert to actin has changed, as indicated by the arrows now pointing towards the left, causing the second actin filament to be crosslinked with opposite polarity. This however was only the case for the diagonal and bent conformation. In the inchworm conformation, the binding to actin was unchanged as only one filament is involved in this case. Interestingly it was this conformation with seemingly unchanged actin binding behavior, that showed an altered CaM position. Nevertheless, structural changes of CaM under calcium-influence probably occurred in the other two conformations as well but were not resolved in these averages. The differing position of calmodulin in all three conformations in general im-

plies a structural dependency on the actin monomers that myosin IX binds to and hence complex underlying processes.

While without calcium all myosin IX molecules were exposed to the same conditions and could crosslink actin filaments in a parallel way, the addition of calcium provoked a competition between myosin IX molecules having an apo- or Ca^{2+}-calmodulin attached (assuming that not 100% of calmodulin had calcium bound). Whenever a small number of apo-myosin IX molecules was able to establish crosslinks between a parallel actin filaments, it was likely to grow by recruiting more apo-myosin IX molecules. Motors with their calmodulin in a Ca^{2+}-bound state might bind to the crosslinked actin filaments in the inchworm conformation and d out, but wont be able to bind to the same (parallel) actin filament. Only if another actin filament is within reach in the "correct" polarity (anti-parallel), the myosin molecule is able to crosslink the two filaments. The binding of more Ca^{2+}-myosin IX motors to these filaments would form more crosslinks, extending the length of this actin-myosin-construct further. This suggests that the formation of crosslinks itself was not influenced by myosin molecules in different calcium-states, but rather the ratio of parallel to anti-parallel crosslinks. The resulting mixed populations of parallel and anti-parallel actin crosslinks might be the reason for the more distributed polarities observed for the sample of myosin IX in the presence of calcium (see red graph in Figure 6.34).
Although first intriguing structural insights could be revealed of myosin IX crosslinking actin filaments using negative stain EM in combination with other biochemical methods, many questions are still unanswered, for instance whether mixed polarity populations within big bundles build from several actin filaments occur or if some form of regulation keeps the polarity of all filaments in one bundles the same.

6.3.3 Why is the structure of myosin IX just a blob?

As it became obvious in the previous sections, the crosslinking myosin motor domain itself had a rather blob-like structure, although structural properties of the catalytic domain could be resolved by negative stain EM. The elaborated analysis on myosin VI (chapter 5) for example demonstrated, that this form of microscopy in combination with SPA was sufficient to achieve resolutions of under 20 Å and thus could resolve the cleft between the upper and lower 50 kDa domain, the converter region, the SH3 domain and in general the asymmetric shape of the motor domain. However, even though the overall resolution of the myosin IX motor domain presented here was lower (with actin being the best resolved parts, approx. 26 Å), this could only explain smaller structures to be unresolved, such as the SH3 domain. The overall asymmetric shape of the molecule on

the other hand should be identifiable in the EM data.

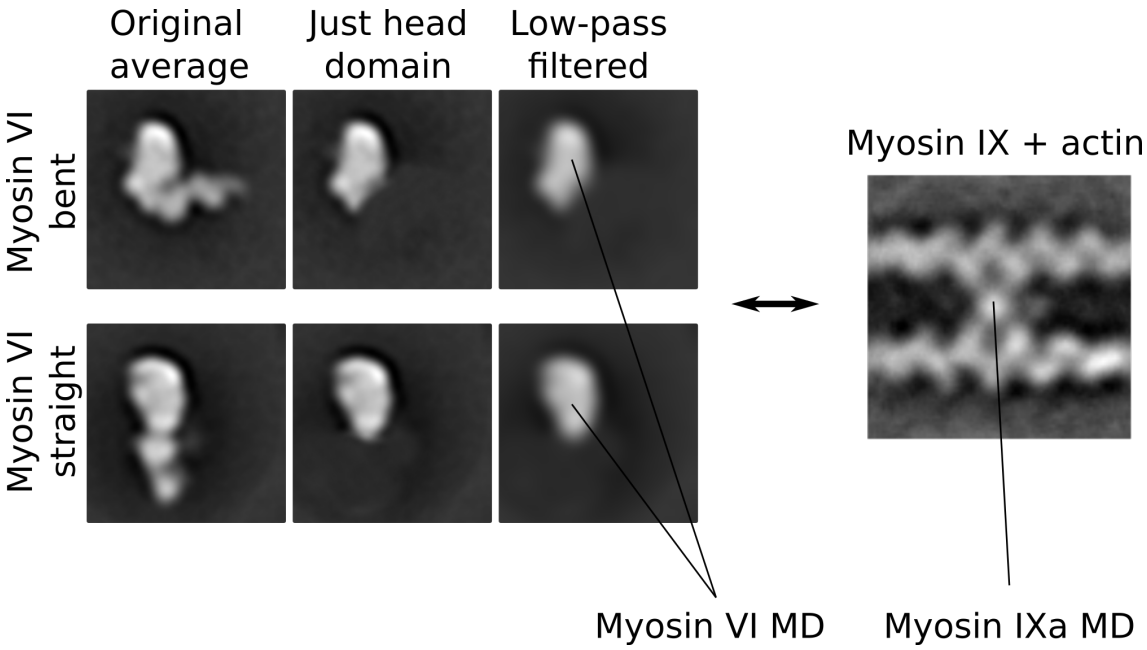


Figure 6.36: Comparison of 18 Å resolution EM averages of the myosin VI head domain (low-pass filtered to simulate the myosin IX resolution of approx. 30-40 Å) with the myosin IX head domain. While myosin VI indicates that asymmetric features should be still visible even at low resolutions, the EM average of myosin IX crosslinks however displays this domain as a single blob. Scale: images are 30 nm wide.

As a comparison, Figure 6.36 shows the catalytic domain of the bent and straight conformation of myosin VI (Figure 5.5 in chapter 5) after the application of a low-pass filter to simulate lower resolution. Many characteristic features became washed out, but the asymmetry shape could be still resolved. In contrast, for myosin IX, the catalytic domain was represented mainly by a blob, which also was significantly smaller than the head of myosin VI. However, it is known that the structure of the catalytic domain is highly conserved in all myosin classes and therefore should show similar features in both EM averages.

One way to explain these observations is to consider these structures not in 2D but in three dimensions. While single molecules such as myosin or single constructs like actin filaments are strongly absorbed to the EM grid surface and with respect to their size lie relatively flat on the grid, the situation for bigger complexes is different. Complexes such as myosin - actin - crosslinks might withstand the forces during staining and preserve a larger three-dimensional structure. This has to be taken into account when 2D projections of such structures produced by the TEM are interpreted.

One possible 3D arrangement of the crosslinks formed by actin and myosin IX is illustrated as an overview in Figure 6.37. Actin filaments are colored in green, the myosin IX motor domain in blue and the uniquely extended loop2 (due to the special insert) in red. The four actin monomers that take part in the crosslinks are also color-coded:

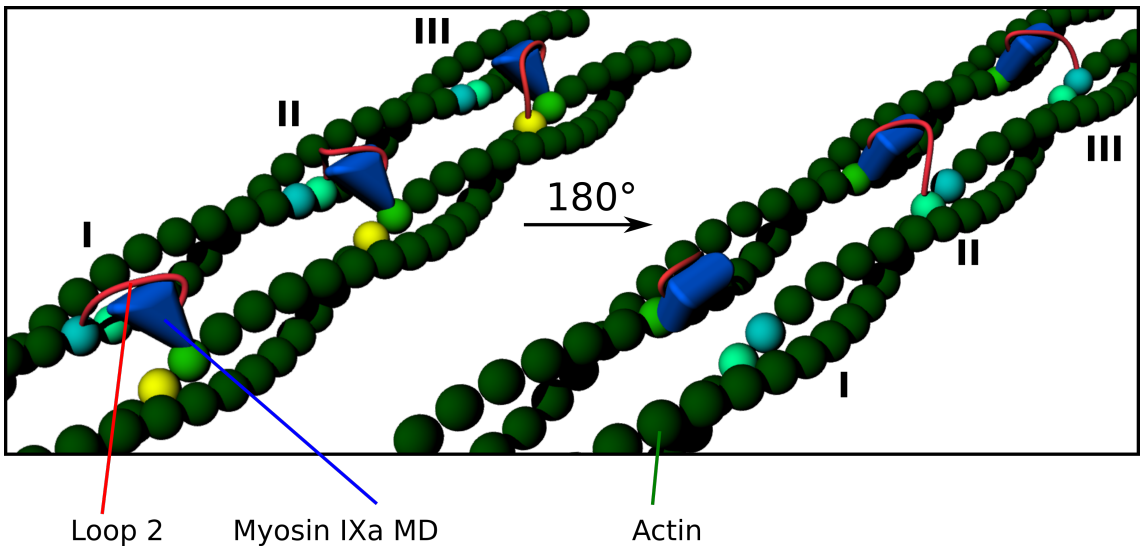


Figure 6.37: Possible three dimensional arrangement of how myosin IX molecules crosslink actin filaments. Shown are the three conformations in two views, related by a 180° rotation. Color code: myosin IX motor domain in blue, loop2 with insert in red, actin in green, actin monomer of primarily actin binding site in light green, actin monomers for conformation I,II and III: light blue, turquoise and yellow respectively.

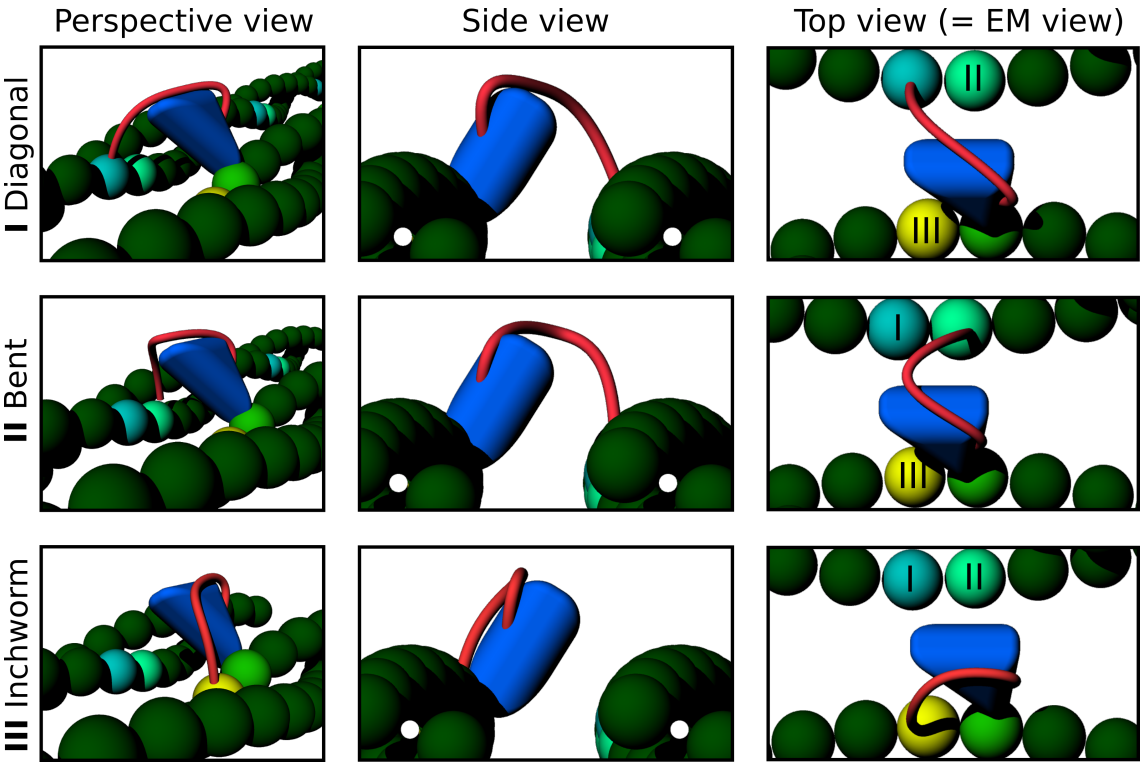


Figure 6.38: Possible simplified three dimensional arrangement of myosin crosslink in a perspective, side and top view. Compared to if myosin would lie flat on the surface, the 2D projection view in the TEM (equals top view) of the orientation suggested here would make the head domain (central mass in EM averages) appear rather structureless, i.e. blob-like.

primarily actin binding site: light green, second actin binding sites on adjacent filament (for diagonal (**I**) and bent (**II**) conformations): light blue and turquoise and second actin binding site on same filament (for inchworm (**III**) conformation): yellow. In Figure 6.38, the three identified conformations, diagonal, bent and inchworm are shown in different views (perspective, side and top). These cartoons suggest an orientation of the crosslinking myosin that was rotated by approx. 45° upwards, leading to an projection view in the TEM that is usually not seen if myosins are without any restriction. A single myosin bound to one actin filaments would most likely also not adopt such an orientations as the staining and drying process would force the myosin to be adsorbed to the surface. In the case of myosin IX however, the linking of the loop2 to another actin monomer might stabilize the construct in a way that the protruding orientation could be preserved.
This has several effects on the imaging with the TEM: On the one hand, the 2D projection view of the motor domain that is rotated upwards results in the superposition of many structural features. In the final EM images these features would therefore be hidden. Furthermore, in this orientation the catalytic domain appears more symmetric than if it were to lie flat on the surface, resulting in the blob-like shape in the electron micrographs. And finally, the space that is created between the filaments and under the molecule would be filled by stain solution. This would drastically decrease the contrast in this area and consequently alter the size and shape of the observed intensity representing the molecule. However, the structure of myosin IX was not completely without distinct features. The loop2 was, due to its relatively long size, able to crosslink to other actin monomers, giving rise to the additional intensities observed at the three second actin binding positions (I, II and III).

6.4 Methods

This section will describe how myosin IXa MD (In the following called myosin IX) and F-actin samples were prepared for negative staining, followed by the staining process itself. The focus will lie on the TEM image acquisition of myosin IX molecules bound to actin filaments and the subsequent single particle image processing methods used to analyze the digital data.

6.4.1 Preparation and negative staining of myosin IX bound to F-actin

The electron microscopy allows for resolving structural information in the sub-nanometer range. This high resolution is a result of three factors: the resolving power of the electron microscope itself, the following image processing of big data sets and the degree of purity of the sample. While the maximal resolution of the EM is usually fixed, the two latter parameters are highly influenced by the experimentalist. In order to successfully analyze negative stain images of single myosin IX molecules bound actin filaments, a few requirements have to be fulfilled: On the electron micrographs the actin filaments need to be straight for the length of at least one repeat (72 nm) on which a myosin molecule is bound. While one straight actin repeat might be sufficient, longer straight segments are preferred for the sake of efficiency of the image processing. This can be achieved on one hand by gentle handling of the sample (i.e. careful pipetting during the sample preparation and careful application onto the EM grid during staining[9]) and by choosing the right concentrations on the other hand. The ideal concentrations were crucial and had to be determined by try and error, as the expressed and purified myosin IX protein was not stable for a longer period of time (approx. 1-2 days) and each purification process usually resulted in slightly different final concentrations of the protein. An ideal concentration of the actin - myosin IX - sample provides images with clearly identifiable actin filaments (no actin-aggregates) and single myosin molecules bound to F-actin. The myosin concentration should be as low as possible to ensure that only **single** molecules are bound to actin and no myosin-aggregates can form and at the same time as high as possible to increase the efficiency of the particle picking. Furthermore, the successful formation of actin - myosin - crosslinks depends strongly on the concentration-ratio between F-actin and myosin. Relative high concentrations of myosin were needed to form crosslinks in the first place. Choosing these concentrations too high however led to the formation of aggregates, making the image processing challenging. And finally,

[9] One way to decrease the forces on the sample during pipetting is by cutting the first bit of the pipette tip in order to increase its radius.

the staining method itself had a strong influence on the quality of the images. The negative staining protocol needed to be adjusted to the sample of interest and optimized by try and error.

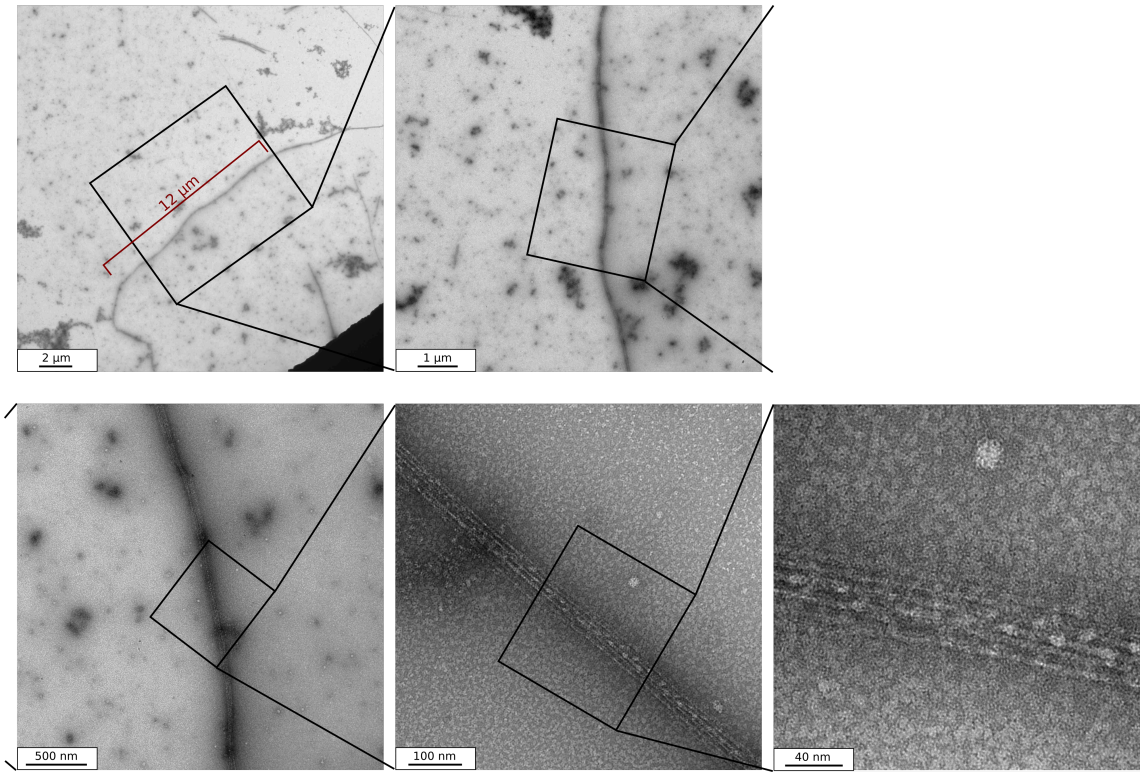


Figure 6.39: Magnification-series of one actin-myosin crosslink EM micrographs shows a crosslink that was straight for a length of almost 12 µm

Figure 6.39 shows a magnification-series of one actin-myosin crosslink that was relatively straight over the length of 12 µm. While the length and thickness of this filament seemed to be ideal for particle picking, its quality however was not sufficient. The micrograph with highest magnification showed a very high myosin background and therefore poor quality of the actual crosslinks. Furthermore, too many actin filaments were involved in this crosslink which increases the possibility of filaments lying on top of each other. The molecular biological methods for protein expression and purification of myosin IX and cosedimentation with F-actin were done according to published protocols [110]. As mentioned above, final preparations of myosin IX were not stable for a long time and freezing them would also result in loss of functionality, in contrast to e.g. myosin VI, which could be stored at -80 °C without being damaged (see chapter chap:vi, section 5.5.1). The sample had to be freshly prepared for each experiment and could be stored only at 4 °C for a maximum time of 3 days[10]. Moreover, prepared samples of

[10]The preparation of highly purified myosin IX and cosedimentation with F-actin was subject of the

myosin IX together with actin were even more unstable and started to form aggregates rapidly. These aggregates could be removed by high speed centrifugation. The supernatant was then almost free from aggregates for some time. However, 3h were enough for myosin to form aggregates again, which meant that staining had to be done shortly after centrifugation.

Buffer	mM
Mops (pH 7.2)	20
NaCl	100
EGTA	0.2
DTT	1

Table 6.9: Myosin IX / Actin buffer for negative stain EM.

In its optimized conditions, nucleotide-free myosin IX molecules were diluted together with F-actin to final concentrations of 1 μM and 0.5 μM respectively, using the buffer listed in Table 6.9. In order to prevent aggregation, the mixed myosin IX samples had to be stained within 30 min after preparation. Sticking to this time-rule, the sample was applied to a carbon-coated copper grid[11] that had been hydrophilized immediately prior to staining. Hydrophilization was done by glow-discharging the grids in a plasma cleaner for 60 seconds at 24W using argon gas[12]. Several negative staining methods have been tested and for the study of myosin IX bound to F-actin, the optimal staining method was the following:
The EM grid was picked up (coated side facing up) and fixed using a reverse tweezer. A 5 μl drop sample was applied directly to the coated side of the grid and allowed to rest for 2 min. Immediately after blotting the sample using a filter paper, a first drop of 5 μl freshly prepared 2% Uranylformat (see section 2.4) was applied to the same side of the grid and blotted away instantly. Subsequently, a second drop of 5 μl UF was applied and after a staining time of 10 seconds any excess solution was being blotted away using a filter paper. Finally, the stained grid had to completely dry on air for approx. 30 min to prevent any liquid being transferred to the high vacuum in the TEM.

research of my colleges.

[11] ECF400-Cu-50: TEM Grids, Carbon Film coated, approx. 5-6 nm (standard thickness), 400 Mesh, Cu, 50 pieces; Science Services

[12] Chair for Experimental Physics, Prof. Joachim Rädler, Physics Department, Ludwig-Maximilian-Universität Munich

6.4.2 SPA workflow for EM images of myosin IX bound F-actin

This section will describe the single particle analysis workflow to analyze EM images of F-actin crosslinked by single myosin IX molecules. The first step was to create a stack of images with sufficient resolution in order to obtain the first global average of the myosin - actin - crosslinks, which then could be analyzed further, resulting in first structural information about the linking myosin. Following this, the polarity and phase of the crosslinked actin filaments was examined and by taking these results into account the structure of the crosslinking myosin molecules was recalculated and studied in more detail. Eventually, interpretation of the EM data was performed by crystal structure projection matching with various crystal structures.
Particle picking was carried out using the particle picking protocol ”e2boxer” implemented in EMAN2 [45] and the following SPA (alignment and classification and statistics) was done completely with my personally developed and customized scripts based on and using the SPIRE / SPIDER software suite [24, 44, 60]. The set of scripts needed for this workflow are shown in the supplementary (chapter IV, figures 6.1-6.4). The following description will refer to the names of the corresponding scripts (extension ”.spm”) shown in these figures.

6.4.2.1 Acquisition of micrographs and particle picking

The EM grids prepared as described in section 6.4.1 containing the actin - myosin - sample were recorded using a Philips CM 100 TEM[13] operating at 100 kV and micrographs were manually imaged using a CCD camera at a magnification of 42k, which corresponded to a resolution of 0.332886 nm/px in the case of the microscope in use. In total, 146 micrographs of the best grid (with respect to concentration, stain and overall quality of the grid) were acquired and used for further processing. The finding of proper areas to record an image had to be done carefully as the final results will naturally depend heavily on the quality of the chosen micrographs.

As one can see in an example micrograph in Figure 6.40, the number of crosslinked actin filaments can vary considerably (more example micrographs showing overviews, single crosslinks, double crosslinks or crosslinks with more then two actin filaments are given in the supplementary material, section 6.5, figures 6.63-6.73). Actin bundles with fewer actin filaments involved should be preferred, such as the one marked in green. Furthermore, in order to perform proper image processing of negatively stained micrographs, all acquired images should provide similar depths of stain. While this depth

[13] Hendrick Dietz, Laboratory for Biomolecular Nanotechnology, Technische Universität Munich

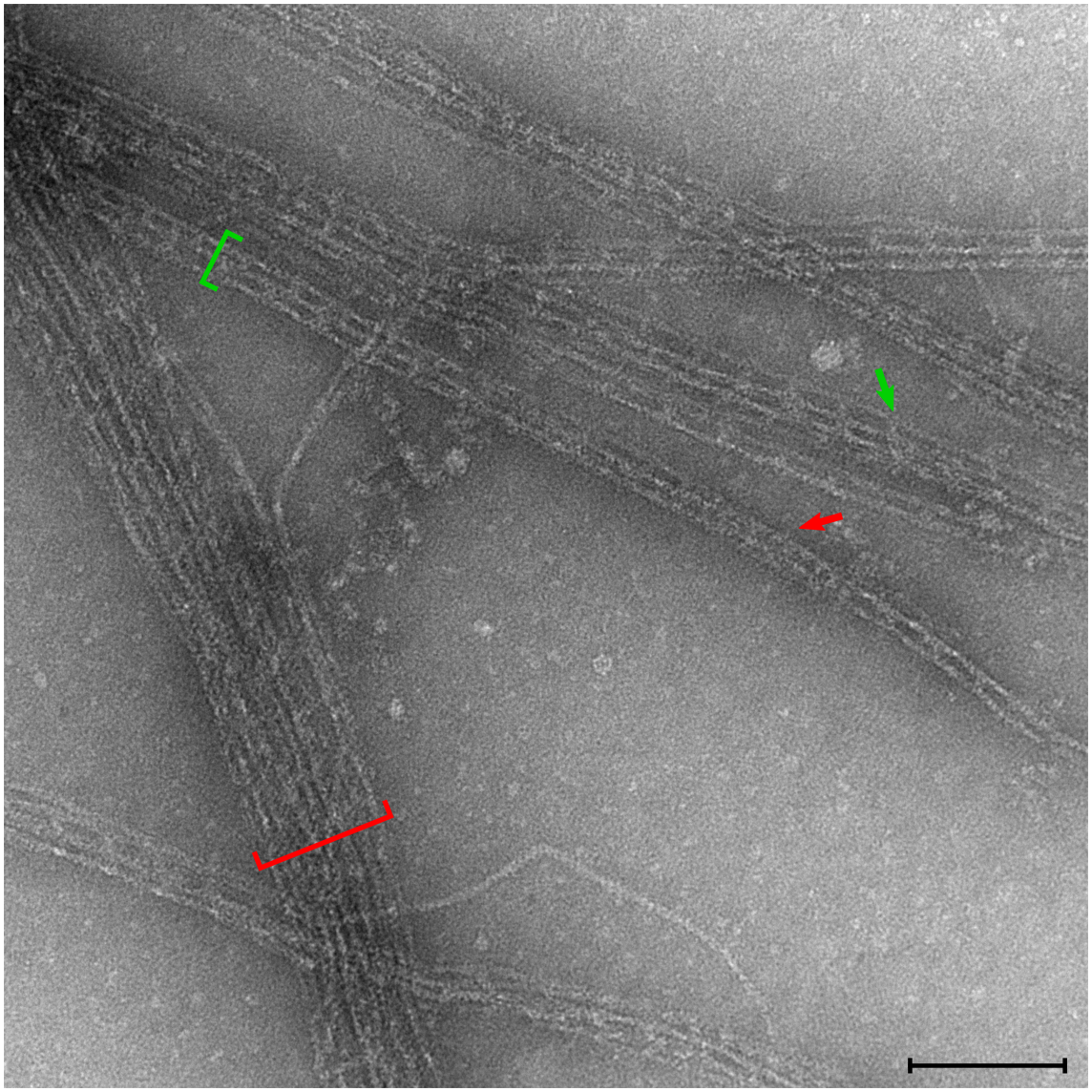


Figure 6.40: Raw micrograph of myosin IX - actin - crosslinks. The number of actin filaments within one bundle varied a lot. Bundles with less filaments involved were preferred (marked in green) and bundles with a large number of filaments should be excluded from the analysis (red) as actin filaments might lie on top of each other. Scale bar: 100 nm

could be directly observed in the case of single myosin molecules as a result of their small size (seen by darker / brighter areas or lake-like patterns, see section 2.4), bigger complexes such as the actin-crosslinks studied here might stick out of the stain, even though the stain depth seemed to be ideal (see crosslink marked in red). Therefore, one has to make sure to record only images with a "single layer" of crosslinks, i.e. all actin filaments within such crosslinks are bound to the grid surface and no crosslink events are in the direction perpendicular to the grid-plane, as it was the case in the section marked in red. And finally, as the crosslinking myosin molecules will be analyzed, they should be clearly visible without any background objects such as unbound myosin or other undefined objects. Examples of suitable and unusable myosin crosslinks are indicated by the green and red arrows, respectively.
Using EMAN2 combined with the SPIDER script *Particles/particlepicking_eman_spider.spm*, images of single myosin molecules crosslinking actin filaments were picked in a completely manual manner by storing the coordinates of the crosslinking object. In contrast to other data sets such as single objects, for the detection of usable crosslinking myosin no (half-)automated method could be applied, as there were too many parameters that had to be taken into account: stain quality, number of actin filaments involved, number of background objects, straightness of the actin filaments and quality of the crosslinking myosin molecules. In order to analyze the distance between the crosslinks, the picked image size was chosen to be 300x300px (100x100 nm), which ensured the presence of three consecutive crosslinks within one image and furthermore three actin-pseudo-repeats. However, for further processing (alignment, classification, polarity determination, conformation examination and crystal structure matching), each image was windowed again to a size of 90x90px, which included only one pseudo repeat and one crosslinking myosin. The main purpose for this was an increase of accuracy of the actin-alignment when the image covered smaller sequences. Small deviations from an ideal straightness of the actin filament could be compensated by this procedure.

Figure 6.41 illustrates the impact of this effect. **A** shows the global average of the 300x300px picked images. Obviously, this image does not contain any structural information as the images were not aligned at this point. For the alignment, the data set was windowed to smaller images of 90x90px in size (**B**) prior to the alignment, resulting in an average with great details especially of both actin-filaments (**C**). If the same alignment-parameters are applied to the original, 300x300px images, one can clearly see the drastic decrease of resolution with increasing distance from the center (**D**), indicating the above mentioned deviations from absolute straightness. In contrast, if the unaligned images (**A**) were directly aligned without any additional windowing step in between, the actin filaments were much better resolved towards the edges of the images as well as the outer

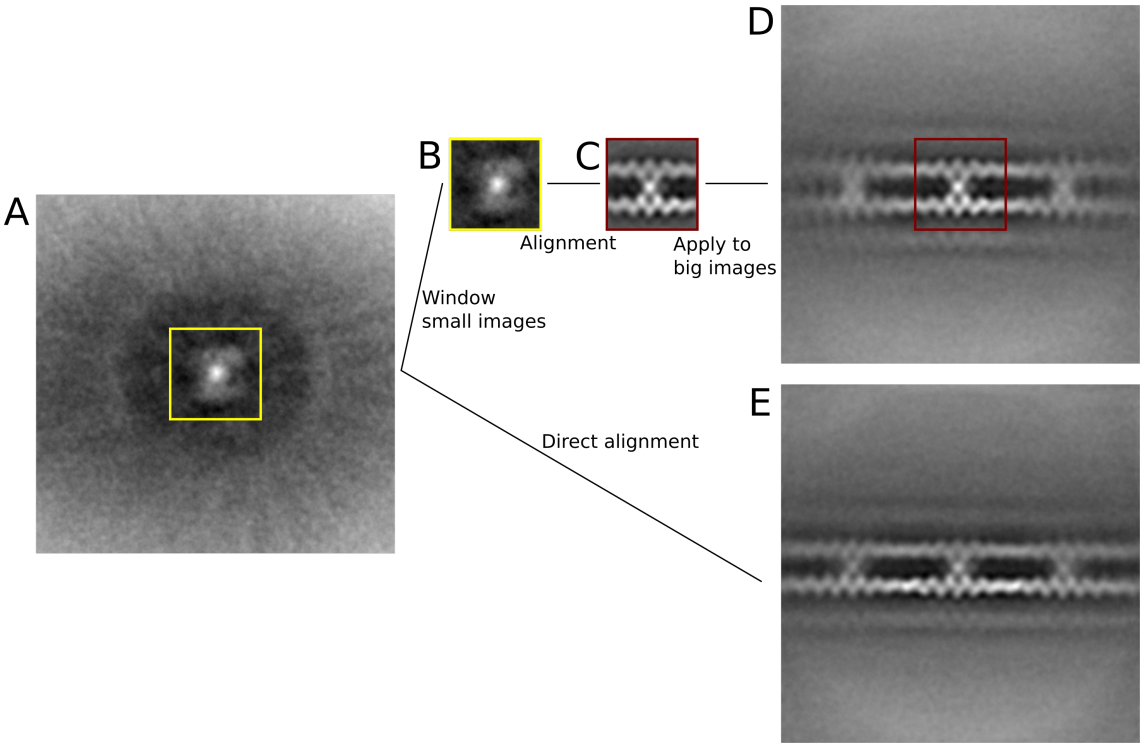


Figure 6.41: Difference between alignment of small and large images. **A** Unaligned data set. **B,C** Alignment of small, windowed images result in good resolution averages. **D** Application of alignment parameter on initial big images showed a decrease of resolution towards the edges, revealing that the actin filaments were not perfectly straight. **E** Direct alignment of big images showed better resolution towards the edges, however on the cost of overall resolution (note the blurred upper filament).

two crosslinking myosin molecules (**E**). The overall resolution however was lower than in case **D**. In particular, note the blurred appearance of the upper filament and also the crosslinking myosin.

6.4.2.2 General shape of actin - myosin - crosslinks

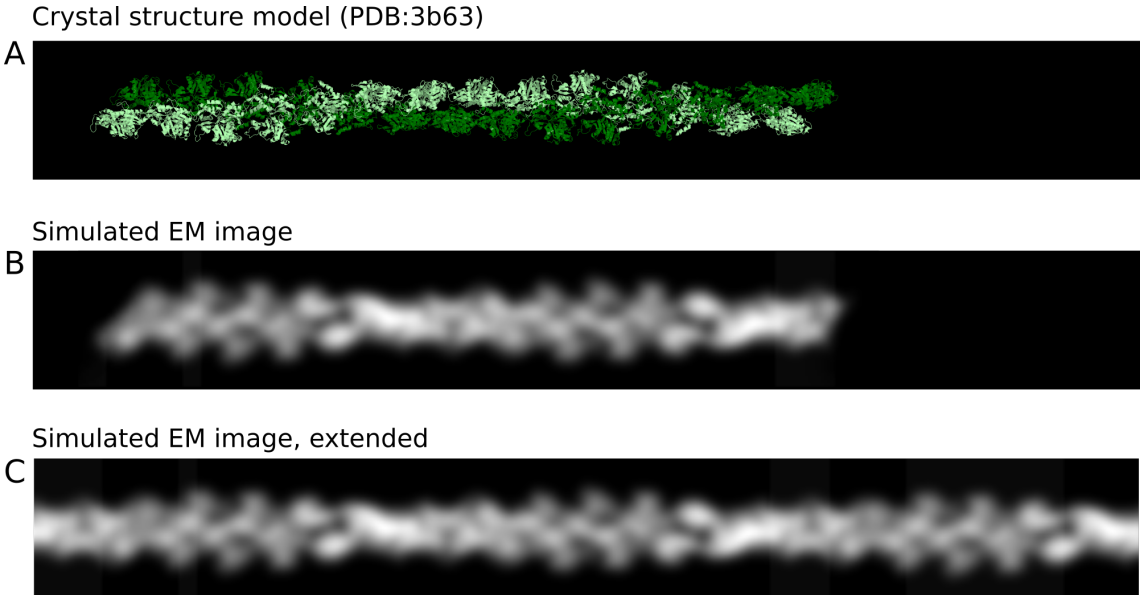


Figure 6.42: Crystal structure of filamentous actin comprising one actin repeat (PDB: 3b63 [138]) in ribbon representation (**A**), 2D low pass filtered (**B**) and extended to match the covered length in the EM averages (**C**).

In order to analyze the distance between the crosslinking myosin IX molecules, a first global average was needed that contained more than one crosslinking molecule. As mentioned above, images of the size of 300x300px were appropriate for this purpose as they contain three successive crosslinks. Images of this size however were challenging to align without any reference, which is why in a first step the data set was reference-based aligned using SPIDER. The reference (shown in Figure 6.42 **C**) was a 2D projection of a single actin filament crystal structure model (PDB: 3B63 [138], long axis of filament parallel to image plane), which had been scale-adjusted and low-pass filtered to simulate negative stain EM images (**A** and **B**). As the actin-segment in **A** and **B** represented only one entire actin-repeat, the image was extended by standard image processing (GIMP) to match the size of the filament in the EM data (**C**). The advantage of using a crystal structure model is that all main characteristics of F-actin, i.e. polarity and phase, were represented. This will become even more important in the later analysis of the polarity. Furthermore, by choosing different radii for the low-pass filter, the strength of these features could be controlled.

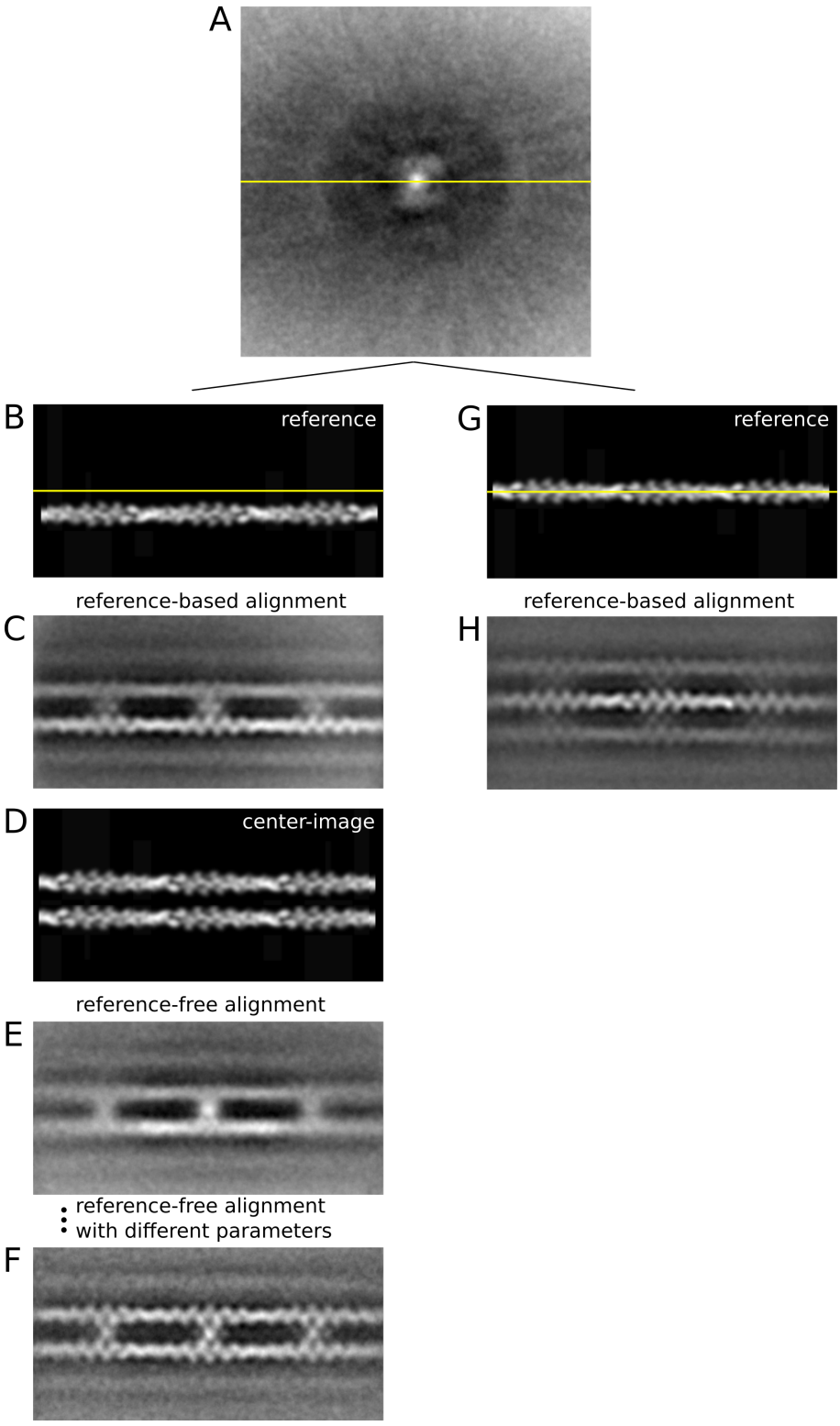


Figure 6.43: Creation of a first crosslink average. **A** Unaligned data set. **B,C** Reference based alignment with the simulated actin filament slightly offset from the center. This keeps the myosin centered during the alignment. **D,E** Reference-free alignment of aligned data set with a simulated crosslink reference image to keep the alignment centered. **F** Various more rounds of reference-free alignment with different parameters (x-, y-shift restrictions) led to the first global average. **G,H** Alignment with a centered actin filament as reference results in aligned images with actin in the center and myosin offset to both sides.

Using the reference image generated in such a fashion, the data set was reference-based aligned (script: *Alignment/align_apshc.spm*) as shown in Figure 6.43 (all images apart from **A** have been cropped for the purpose of illustration). Since at this point nothing was known about neither the distance between the crosslinked actin filaments, nor the polarity or phase of them, only one single actin filament could be used as a reference. Furthermore, the reference actin filament was chosen to be slightly offset from the middle line (which is shown in yellow) to keep the crosslinking myosin in the center of the images (**B**). As a result, this alignment led to an average that showed one of the filaments in great detail, the other however appeared blurred mainly due to different shifts in the x-direction (**C**). If the reference filament was chosen to be centered itself (**G**), the aligned image comprised one centered actin filament with good resolution and two adjacent filaments on each side poorly resolved. This was an effect of the alignment of all crosslink images to one polar actin filament lying in the middle and applying no angular restriction. In favor of the later analysis of the crosslinking myosin, the image processing was continued using the "myosin-centered images" (**C**). In a second step, the data set was reference-free aligned (script: *Alignment/align_apsr.spm*) using a simulated EM image of two parallel actin filaments (called here "center-image") to keep the alignment centered (see section 3.1.2 for more information about "references" in reference-free alignment). This *center-image* (**D**) was created by duplicating and shifting the actin filament. As this image was no real reference, the polarity and phase could be arbitrary. The resulting images (**E**) appeared very blurred in all features (actin and myosin) and therefore underwent various more iterations of reference-free alignment with different parameters (mainly the maximum of allowed x- and y-shifts) in order to improve the resolution. For all alignment steps a mask was used in order to exclude background noise. The shape of the mask as well as the alignment parameters, in particular the maximum of allowed shifts, were crucial for the successful alignment. The effect of these parameters is discussed in section 6.3. Using the final class average (**F**), the distance between the actin-filaments as well as the distance between the crosslinking myosin IX monomers could be measured.

Following the creation of the global average, the crosslinking myosins were analyzed. For this, all aligned images were windowed to a size of 90x90px and once more reference-free aligned as mentioned before. As the cross-like shape of the linking myosin suggested a binding to one or more of the four actin monomers, the new data set was classified with respect to these binding positions using a custom mask as shown in Figure 6.44 **A**, where the four positions were denoted with the numbers 1-4. The classification was performed using the Hierarchical Ascendant Classification (HAC) method and classes were displayed in the form of a binary tree (see 3.2.4) using the scripts *Classifi-*

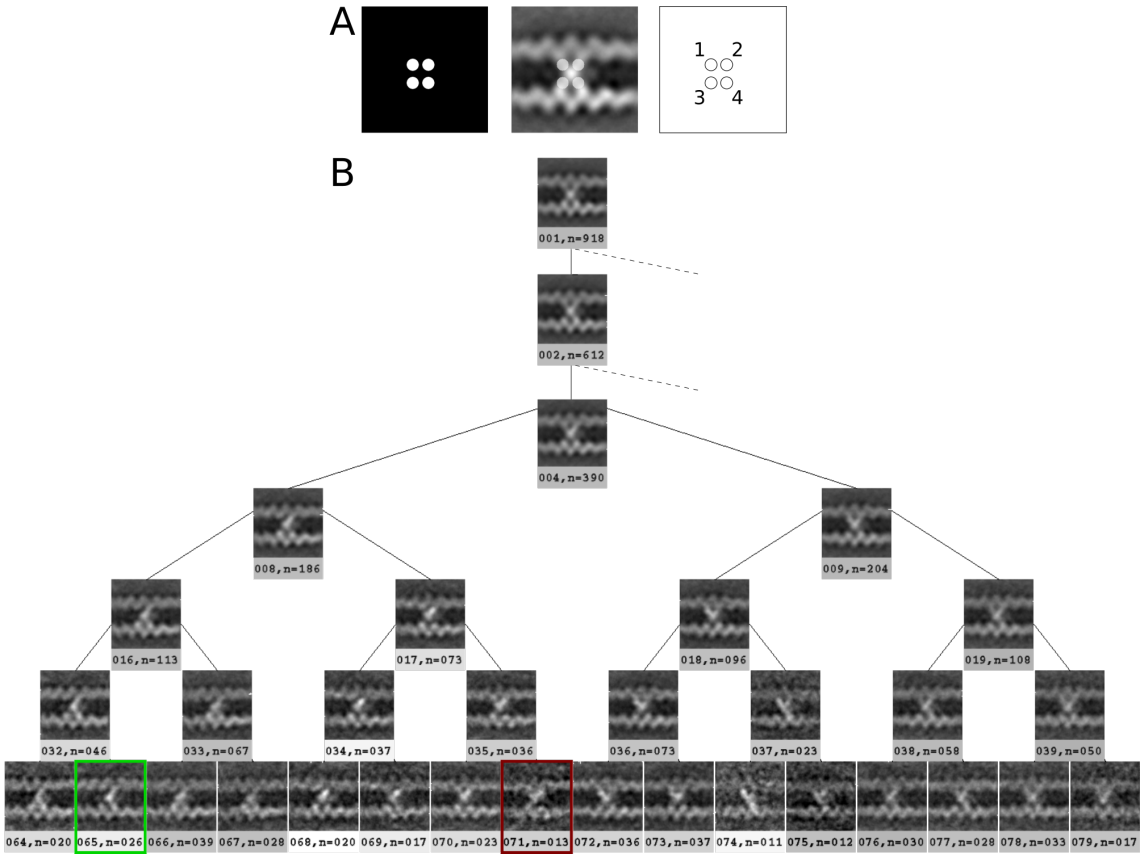


Figure 6.44: **A** Classification of an aligned and windowed data set with respect to binding to one of the actin monomers 1-4. **B** Tree-representation of classification in order to identify classes with different binding patterns (e.g. class marked in green). Some classes could not be assigned to any pattern, such as the class marked in red. Scale: Images are 30 nm wide.

cation/cas_apsr.spm,Tree.spm. Figure 6.44 **B** shows an example of one of the branches. Note that even at the highest level (lowest row in the figure), where the class averages contain only 10-30 images (here only 16 of 64 classes are shown), the structural heterogeneity was still very high. This required a careful and in particular manual inspection of all class averages. While some showed clear conformations (such as class 65, marked in green: "C" shape, myosin binding to position 2 and 4), the resolution of other class averages was not sufficient enough to identify any binding behavior and were therefore excluded from further processing (such as class 71, marked in red, binding behavior not discernible, poor actin structure). The remaining classes were combined and sorted according to three discovered binding positions of myosin: the diagonal crosslink-, bent crosslink- and inchworm-conformation, each present in different, mirror-related appearances.

6.4.2.3 Polarity and phase of actin-crosslinks

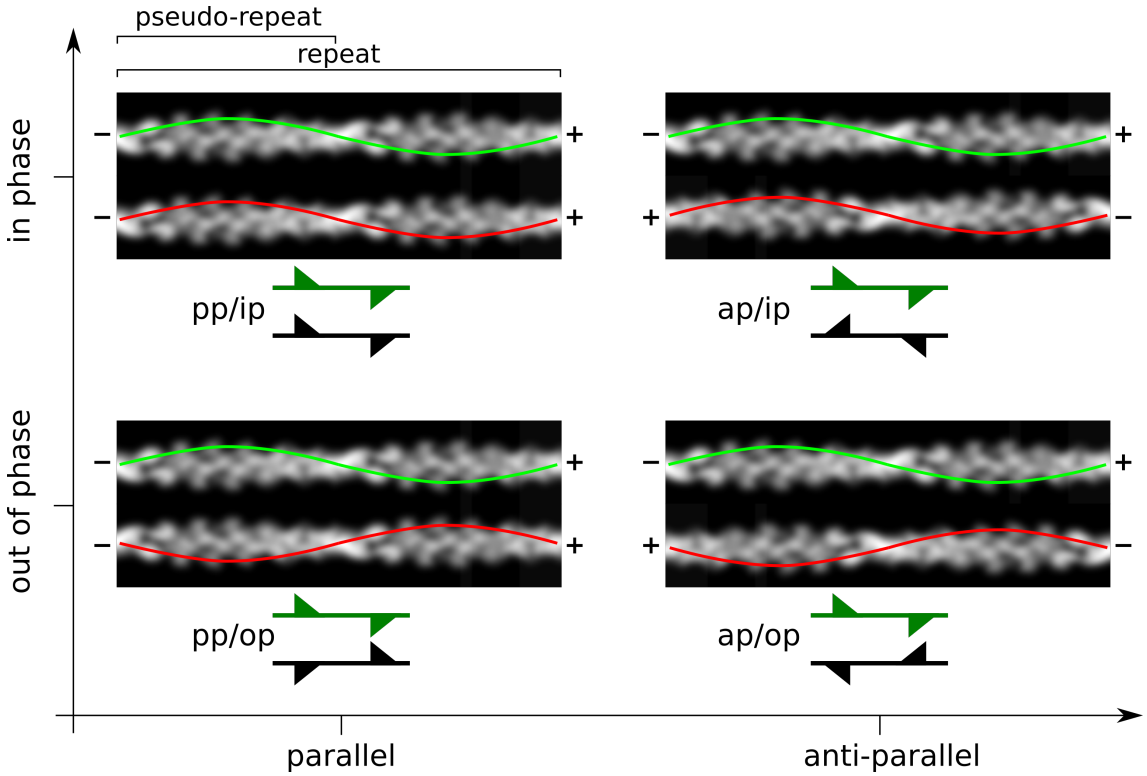


Figure 6.45: Possible arrangements of two actin filaments within one crosslink with respect to their polarity (+,-) and phase (red and green curves) illustrated using simulated crosslinks.

The minimal form of an actin - myosin - crosslink are two single actin filaments linked by myosin. As described in section 1.2, F-actin is a polar filament, it has a defined plus and minus end. If two filaments are orientated next to each other, they can

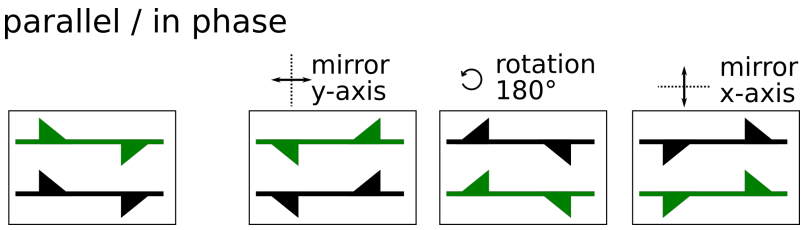


Figure 6.46: Each polarity / phase conformation can be present in four orientations as illustrated on the example of a parallel / in phase crosslink.

be arranged in different ways, as illustrated in Figure 6.45: They can have either the same polarities, termed *parallel* (p), or opposite *anti-parallel* polarities (ap). Furthermore, they can be arranged with arbitrary offsets. As one actin-repeat includes two pseudo-repeats (and one crossover), these offsets can result in the two filaments being *in phase* (ip), e.g. no offset, or *out of phase* (op), which equals to an offset of the length of one pseudo-repeat. Together with the restriction that all crossovers are aligned (as seen in the average **F** in Figure 6.43), this gives four main permutations: pp/ip, pp/op, ap/ip, ap/op. Moreover, since the actin filaments themselves were absorbed to the EM grid in arbitrary orientations, the picked images of crosslinks were randomly orientated as well. Therefore, even if all images were horizontally aligned (with respect to the actin filaments), each of the four permutations could exist in four different orientations itself, as illustrated in Figure 6.46 on the example of the *parallel / in phase* conformation. In total this makes 16 different variations of crosslink images that were analyzed with the following SPA methods.

The actin-images shown in the figure here (6.45) are low-pass filtered projections of an actin crystal structure (PDB: 3B63 [138]). While the polarity and phase of such simulated crosslinks could be identified directly, the raw negative stain images were much more challenging to analyze, as shown in Figure 6.47. The resolution and SNR made digital image processing unavoidable. Furthermore, in order to analyze the actin - myosin - crosslinks correctly, the data set was optimized beforehand by discarding images that did not have sufficient resolution (see previous section). The chosen 90x90px px EM images of actin - myosin - crosslinks (containing one pseudo-repeat and one crosslinking myosin) were then compared with the simulated negative stain EM images of F-actin shown in the previous figures. As it will become clear in the following part, it was important to explore the properties of each actin filament independently. Apart from this the binding-behavior of the crosslinking myosin could be used to bias the polarity analysis.

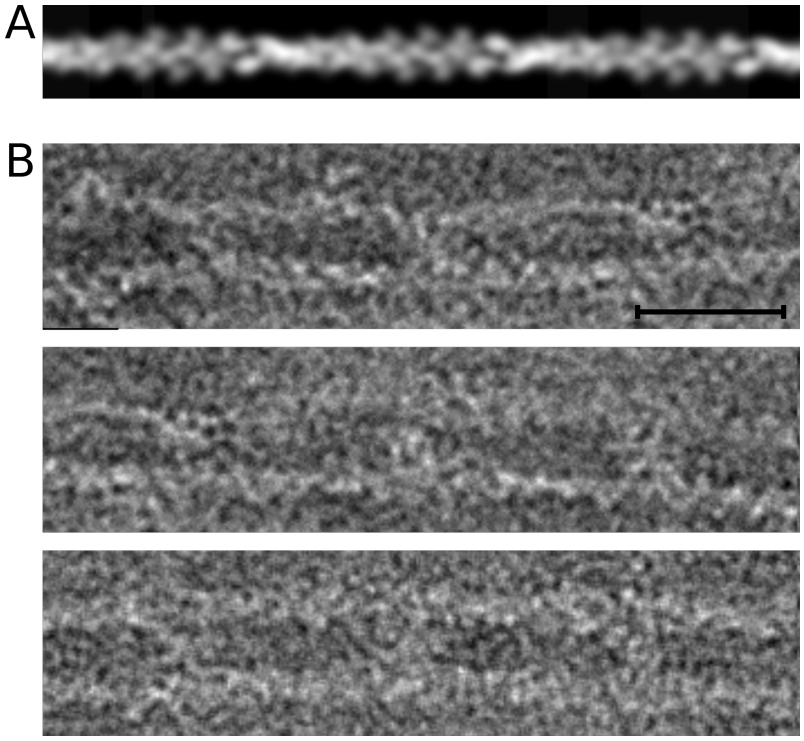


Figure 6.47: Simulated crosslinks (**A**) and real crosslink images from raw micrographs (**B**). Scale bar: 20 nm

6.4.2.4 Workflow for determination of polarity and phase of actin - myosin IX - crosslinks

Prealignment of the data set

As a first step to bias the polarity / phase assignment, the chosen data set of prealigned actin - myosin - crosslinks was classified with respect to the myosin binding behavior as described in section 6.4.2.2. The found classes were then adjusted by mirroring at the x- and / or y-axis in a way that the direction of myosin binding to actin was in agreement within all classes. Figure 6.48 shows the global class average (**A**), the classification mask (**B**), the six found shapes (**C**), and their three mirror related orientations (**D**). The number on the top left indicates the number of images contained in the average, the bottom left number indicates the class and the labeling 1A-6B refers to the notation of the following Figure 6.50.To choose classes to be used for the further steps, the binding orientation of myosin was set arbitrary to emerge diagonal from bottom left to top right, as indicated in the cartoon. All classes that show this feature are marked with a black frame.

The application of various mirrorings in order to ”align” the myosin binding was legit as explained in the illustration in Figure 6.49. As it is known, myosin binds to actin in a highly conservative and polar way. Therefore, if in the EM average a polar structure

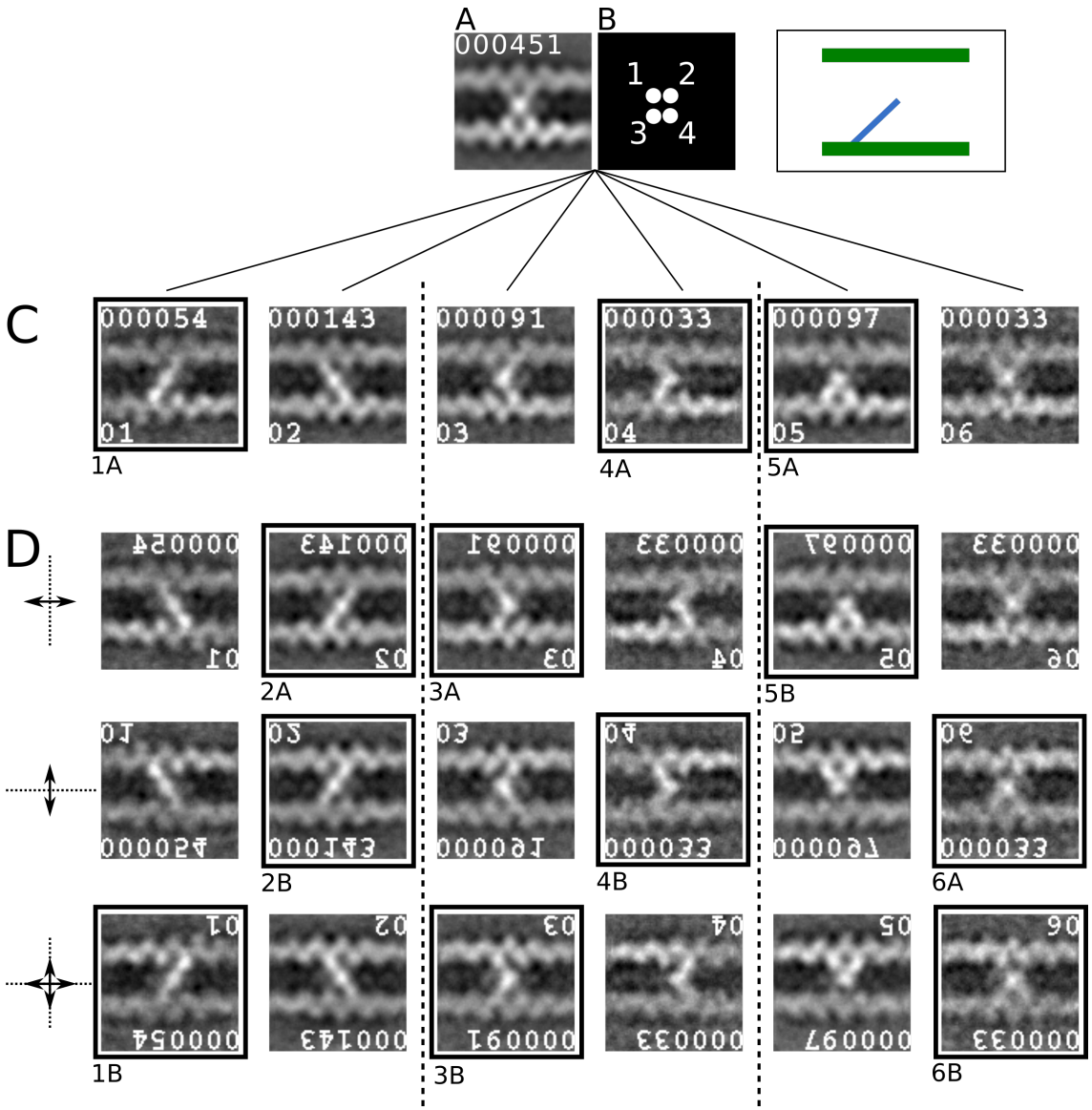


Figure 6.48: Creation of prealigned crosslink data set step 1: Aligned, "good" data set (only images with sufficient resolution, 451 from 981) (**A**), classification mask with indicated binding sites 1-4 (**B**) and the six found shapes **C**. **D** Completing the possible orientations by applying rotations and mirroring with the aim to combine the images to a "prealigned" data set with all myosin bound to binding site 2. The upper left number denotes the number of images contained in the averages and the labeling 1A-6B refers to the notation of the following Figure 6.50.

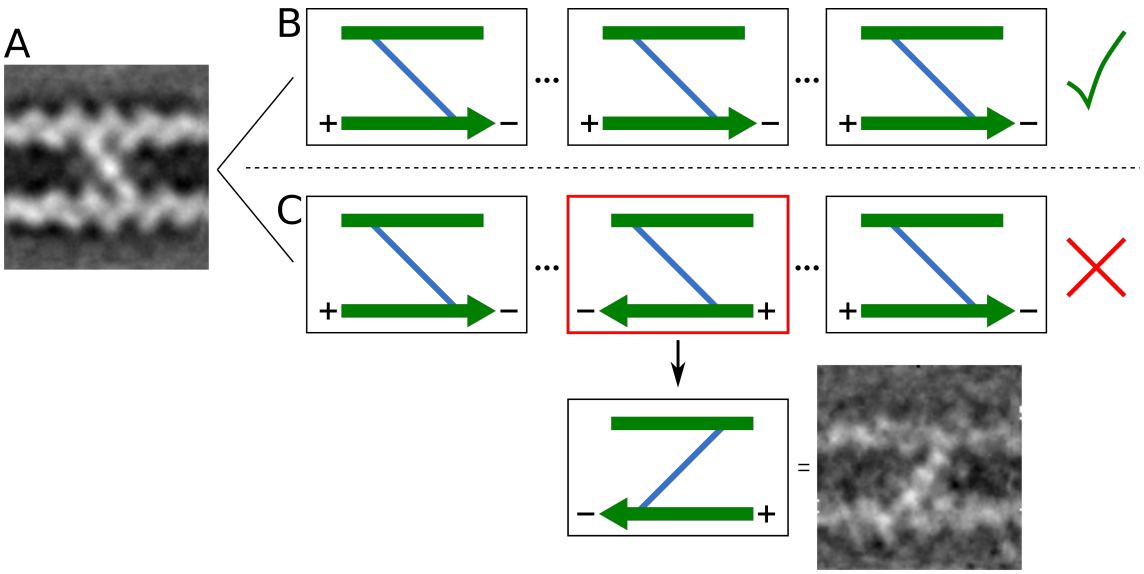


Figure 6.49: Why aligning data set with respect to the myosin was legit: **A** Class average of one of the myosin conformations (diagonal from bottom right to top left). **B,C** Cartoons to illustrate images contained in this class (actin in green, myosin in blue). Case **B**: all actin filaments have the same polarity. Case **C**: If the actin in one of the images would have the opposite polarity, the myosin (due to its polar binding to actin) would follow this change and look distinctively different from the class average. This image within the class would have been assigned to another class.

of myosin could be determined, the underlying actin filaments will have a consistent polarity in all of the contained raw images. In the example EM average **A**, the binding of myosin could be clearly identified as being diagonal from bottom right to top left. The cartoons on the right illustrate examples from EM images contained in the average (green: actin, blue: myosin). As indicated, the polarity of actin was in fact not known at this point. However, for the purpose of this illustration only, the relative polarities between all images was important. In case **B**, the actin filaments in all images point towards the same direction. In case **C** on the other hand, some of the images comprise the same myosin structure but a different actin polarity. This however could not be possible due to the myosin binding behavior and its structural arrangement would have to change as indicated. In this case the EM image looks distinctively different from the other images, which means that it would not be classified into the class shown in **A** in the first place.

In the next step, summarized in Figure 6.50, all 12 class averages (from 6 different classes, see Figure 6.48) were processed as follows: **A** Since two class-pairs each represent one of the three conformations (class01+class02: diagonal, class03+class04: bent, class05+class06: inchworm), these pairs were first combined separately (script: *Classification/combineclasses_mr.spm*). For four classes representing one conformation, four different combinations were possible, as only combinations of two class averages from different classes were constructive (**B**). Following this, all classes were combined by averag-

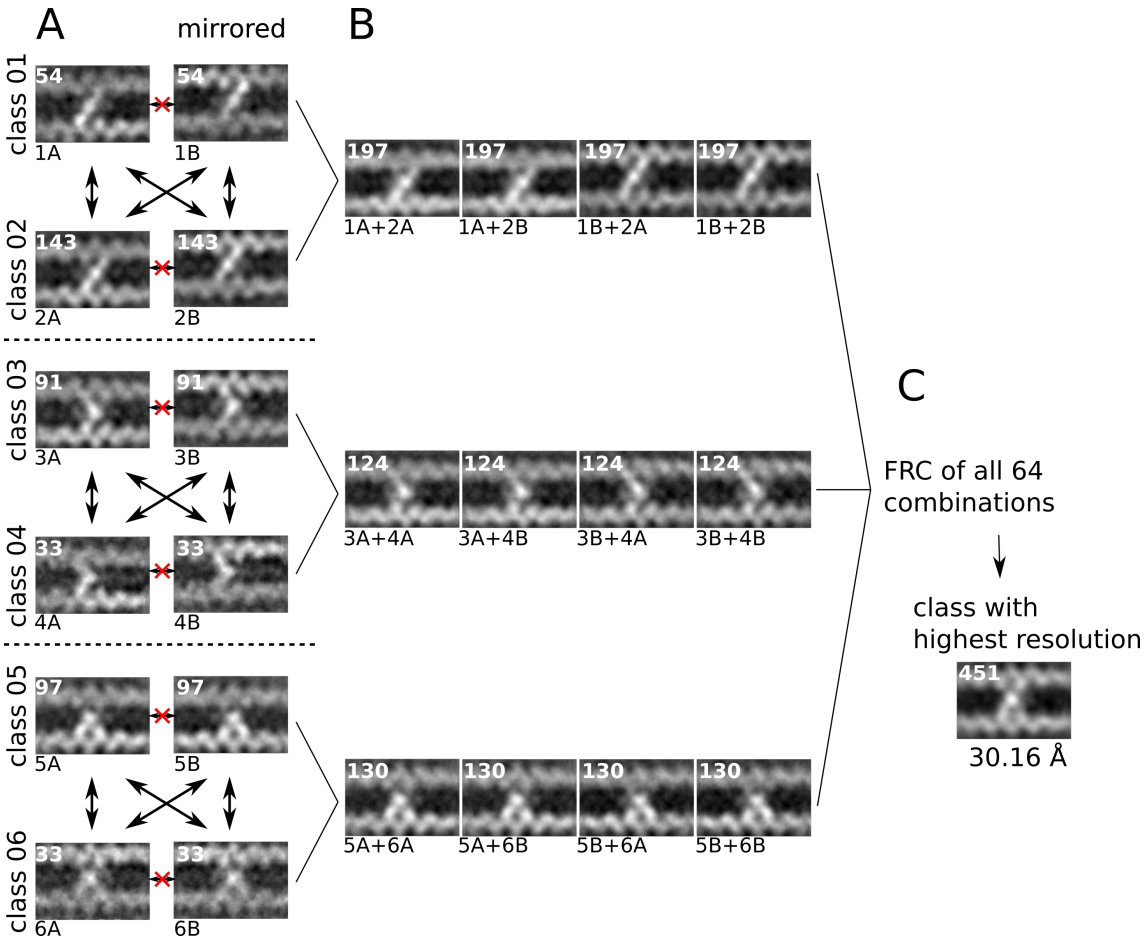


Figure 6.50: Creation of prealigned crosslink data set step 1: **A** Two classes each comprised a similar crosslink structure (class01 and class02, class03 and class04, class05 and class06). Furthermore, each class has a mirrored version that looks similar (”normal” and ”mirrored” images are labeled with 1A-6A and 1B-6B respectively). **B** Adding up two similar classes (e.g. class01 and class02) could be done in the four combinations marked by black double-arrows (the crossed out arrows indicate combinations that were not done, as they would add the same images) **C** The combination of one of the images from each set resulted in 64 classes, where the combination with the best resolution is shown (resolution 30.16 Å.

ing one of each of the three classes, making a total number of $N = 4^3 = 64$ combinations. In order to quantify the quality of these final classes, the resolution was determined by measuring the FRC (cutoff value: 0.5, see section 3.3) and the highest ranking class average was used for the next steps, in this case the combination 1A+2B+3A+4B+5A+6A with a resolution of 30.16 Å (**C**).
The purpose of these first two steps was to ”prealign” the data set by mirroring on the x- and / or y-axis in order to have one consistent binding position of the myosin IX molecule within all images. The actual analysis of the polarity and phase assignment of the crosslink-data set will be described in the following using an artificially created data set:

Polarity and phase assignment, generic data set

The workflow for the polarity and phase assignment will be illustrated on the example of artificially created images that simulate myosin - actin - crosslinks as shown in Figure 6.51. To establish such a generic data set, two simulated negative stain EM images of F-actin (low-pass filtered projections of the F-actin crystal structure PDB: 3b63, see above) were arranged parallel to each other and a Gaussian blob was placed in the middle to imitate actin - myosin - crosslinks (**A**). By mirroring one or both of the filaments at the x- and / or y-axis, all 16 different crosslink-conformations could be reproduced. Furthermore, all images were overlaid with random noise to more realistically simulate the unique properties of negative stain micrographs (i.e. SNR). Also two different versions of ”bad” images were created by making one or both of the filaments featureless[14] (**B**). These 20 artificial crosslink images were then combined in random numbers to create a stack of 100 images[15] (**C**). As the molecules in the real experiment are naturally lying in arbitrary rotational orientations on the grid and the manual particle picking usually contains small errors (with respect to the accuracy of picking the center of the crosslink), the images of the generic data set were also rotated and shifted randomly to create the final generic data set of actin - myosin - crosslinks (**D**). This created a set of images that could be used to demonstrate the polarity analysis. The usage of such generic data has two advantages. First, the workflow will be easier to follow as the features of interest (polarity and phase) can be seen in the raw images already. And second, more importantly, since the properties of each image in the input data set is known, the workflow can be checked for errors by comparing the results.

[14]Featureless actin filaments were created by averaging over four different filaments with different polarities and applying another low-pass filter.

[15]The number of 100 images was chosen so that later results in percentage can be directly related to the number of images.

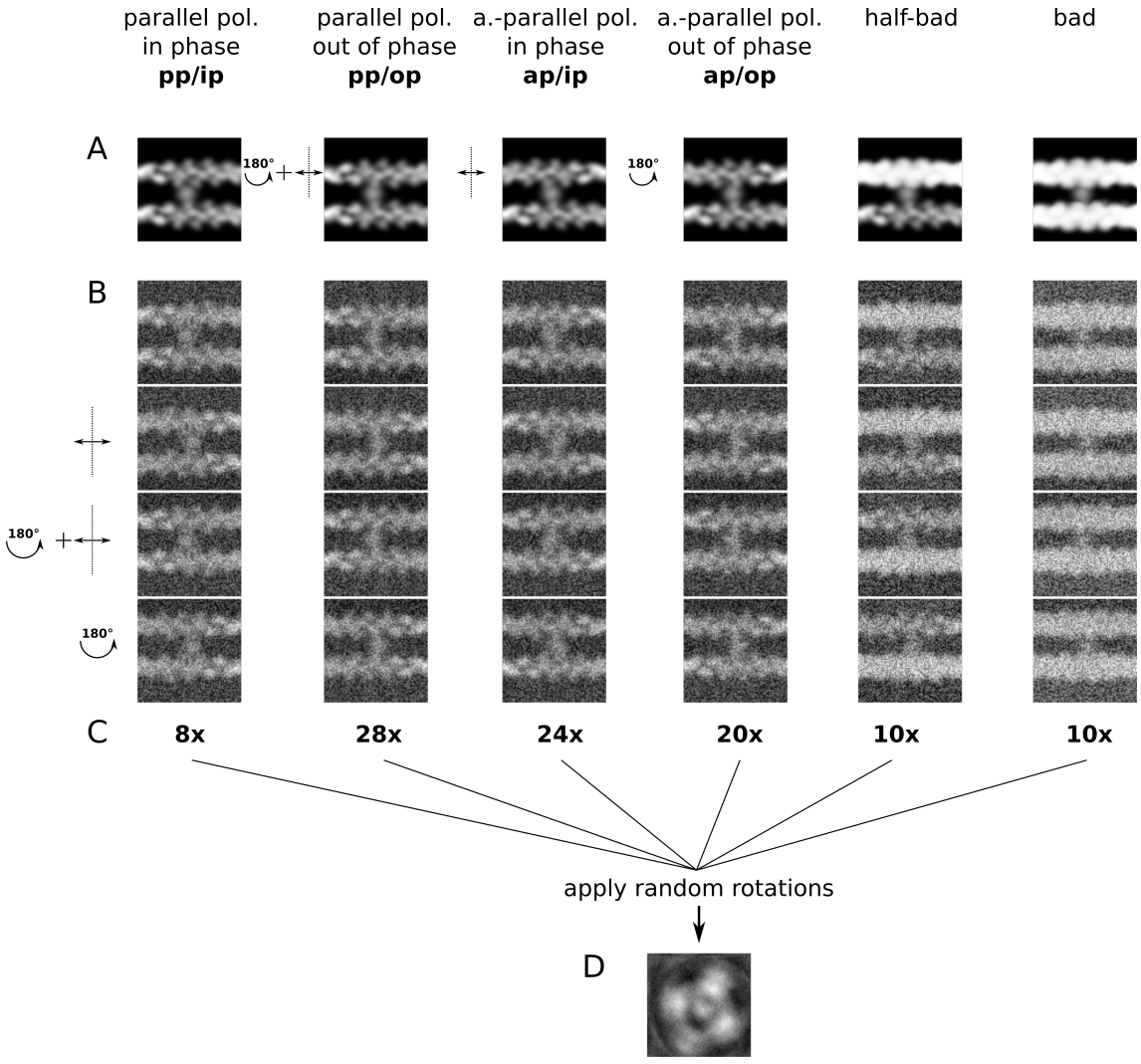


Figure 6.51: Polarity analysis on generic data set, step 1: Artificial data set to verify and illustrate the polarity assignment method. **A** The generic data set included all four polarity / phase combinations, one half bad and one bad class. These images, each present in four orientations, were overlaid with noise and combined in arbitrary numbers and with the application of different rotations, resulting in a stack with 100 images.

As described before, the data set has to be aligned so that the crosslinks can be analyzed properly. This first reference free alignment was applied to the generic data set and the result is shown in 6.52 **A**. As expected, the general shape of the crosslinks can be seen (two filaments and an intensity in the middle), the polarity and phase however were hidden due to averaging. In order to examine these properties, filaments within one crosslink had to be analyzed separately. The basic idea behind this analysis was to align the real images of actin filaments with the simulated images and to investigate the rotations and mirrorings that had to be applied in order to achieve the best fit. While the distance between the actin filaments could be measured from the global average and would make the alignment with completely simulated myosin - actin - crosslink - images possible, the relative shift between the filaments was unknown. Therefore, a separate analysis of each filament within one crosslink was crucial. One way to do so was by applying a mask that covered only one filament and then aligning the masked image reference-based with the simulated actin image of a certain polarity. However, due to the underlying algorithms of the reference-based alignment method of SPIDER (Operation: AP SHC), it was necessary to first shift the images so that the filament of interest was centered. The distance from the center was measured and applied by using the script *Alignment/crosslink_shift.spm*. This script will create two image sets with either the lower or upper filament in the center (Figure 6.52 **B**). For the following alignment a mask covering only the centered filament and also excluding the crosslinking myosin was applied. Each of the two masked data sets was then two times reference-based aligned (script: *Alignment/align_apshc.spm*) with the simulated actin filament in two different polarities as references (**C**). Arbitrary rotations as well as mirroring on the y-axis were allowed during the alignment. Underneath the two references in **C**, the corresponding aligned averages of the (generic) data set are shown. The two graphs below show the angles that were necessary to align the images to the reference (**D**) and whenever mirroring on a vertical axis was needed. The combination of both information indicated the polarity and phase of each single filament. If e.g. the necessary rotation was in the range of 180° and no mirroring on the y-axis had to be applied, both, the polarity and phase would be different from the reference. If only mirroring was necessary, the polarity would be different, the phase however would be the same. If both, rotation and mirroring had to be applied, the polarity would be the same again but with a different phase (always compared to the reference). Note that a rotation of approx. 180° was equivalent to mirroring at the x- and y-axis and a 180° rotation in combination with a mirroring at the y-axis was equal to a mirroring at the x-axis.
With the description above it would be sufficient to use only one single reference to

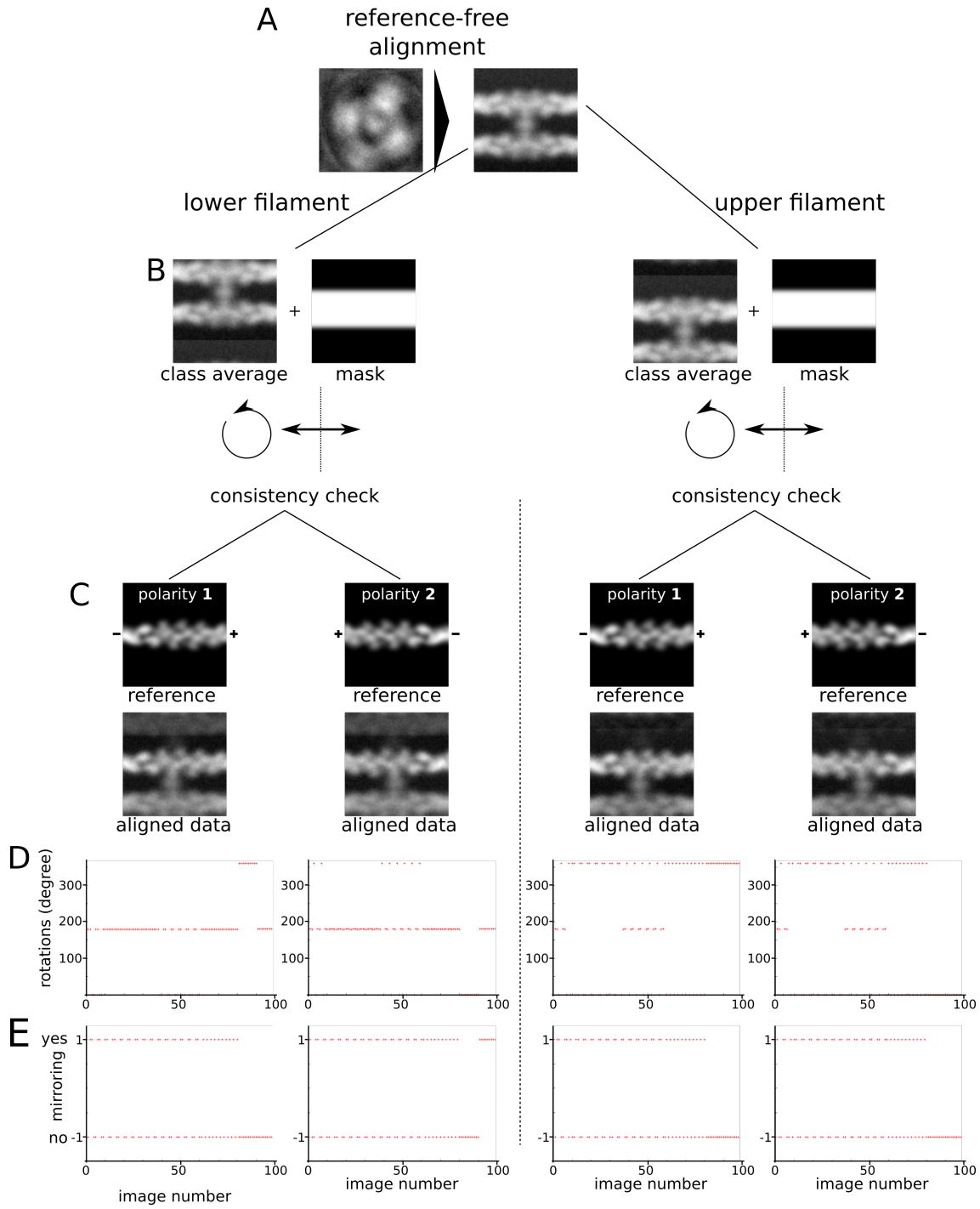


Figure 6.52: Polarity analysis on generic data set, step 2: consistency check. **A,B** Separate analysis of lower and upper filaments in the crosslink images. **C** Each image was aligned with a simulated actin filament in two polarities as a reference to examine the quality of this filament in terms of consistency. **D,E** Optimal rotation and mirror parameter applied during the alignments.

assign the polarity and phase to each image. However, the implementation of the alignment with two references was important in order to check each image for consistency. This means that an image was considered to be consistent, if the alignment with two references of different polarity resulted in the corresponding alignment parameter. If the alignment to the reference with *polarity 1* required an approx. 180° rotation, the alignment with *polarity 2* needed to be in the range of 0 / 360° to consider the filament as being *consistent*[16]. Therefore, the next step was to compare the alignments with the two references for each filament (lower and upper) separately.

Figure 6.53 illustrates the subsequent steps: Shown in **A** is again the initial global average and in **B** the reference images for the alignment. The graphs in **C** combine the alignment parameter for each of the four alignments and assign values (without units) to all images for each possibility as listed in Table 6.10:

y-value (without unit)	Alignment parameter
0	rotation in the range of 0 / 360°; no mirroring
1	rotation in the range of 180°; no mirroring
2	rotation in the range of 0 / 360°; mirroring at y-axis
3	rotation in the range of 180°; mirroring at y-axis

Table 6.10: Values for graphs in Figure 6.53 **C**

Each image was assigned with two values and the combination of these two values led to the next graphs (**D**) that show the percentage of images with the corresponding polarity and phase relations (see also Table 6.11). As described above, images that led to the same polarity although aligned with references of different polarity were considered to be inconsistent and not used for further processing in order to not falsify the results.

	lower filament	**upper filament**
different polarity different phase	90%	80%
different polarity same phase	0%	0%
same polarity same phase	10%	20%
different polarity (different + same) phase	90%	80%

Table 6.11: Values for graphs in Figure 6.53 **D**

At this point we can check if the results were in agreement with the expectations. As already shown in Figure 6.51, the generic data set contained 20 bad images, 10 of

[16]For better understanding the mirror-parameter was left out in this example

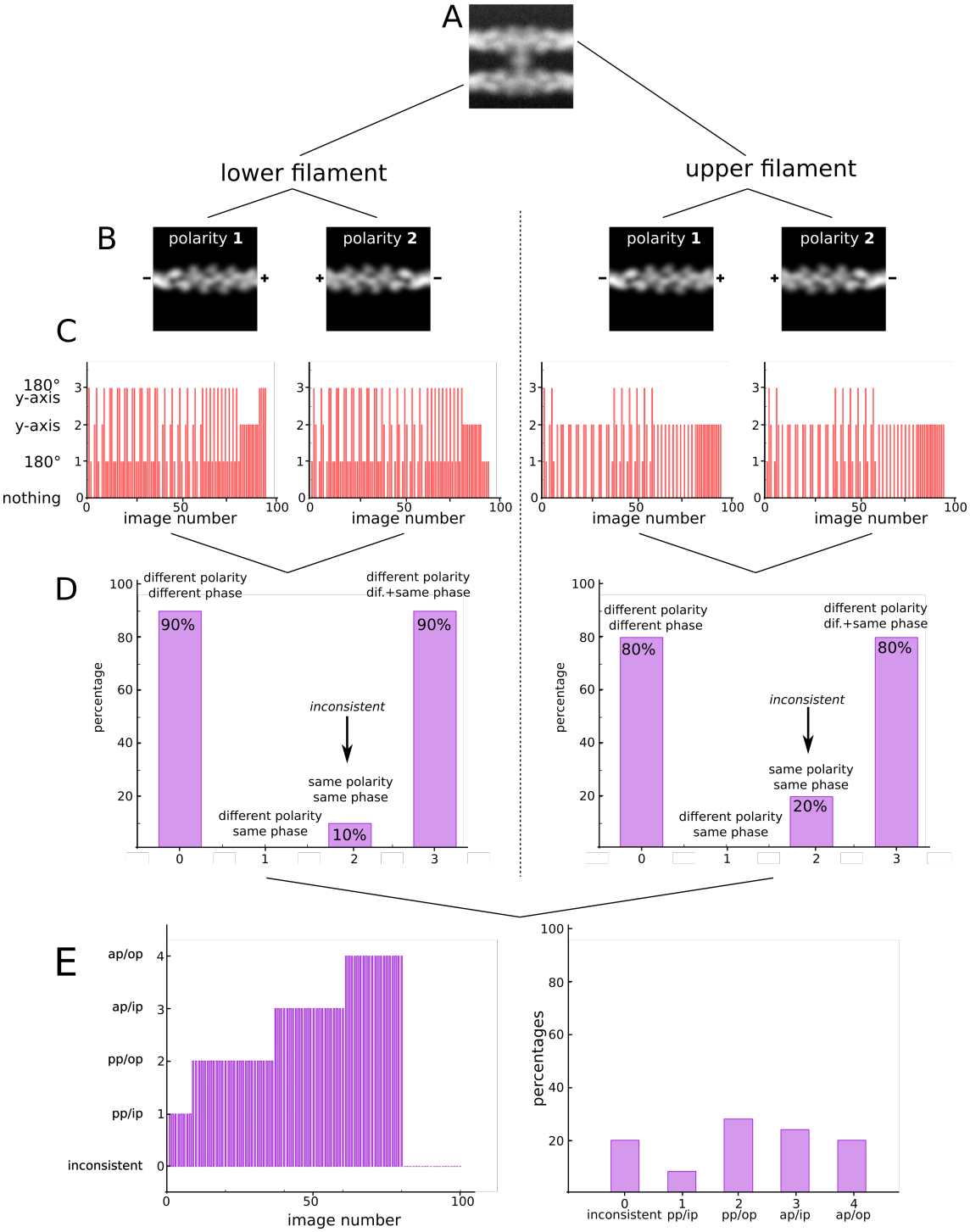


Figure 6.53: Polarity analysis on generic data set, step 3: **C** Combined and simplified alignment parameter from the alignment of the lower and upper filaments to the reference images **B** with polarity 1 and 2. The values indicate for each images whether no change, an approx. 180° rotation, mirroring at the y-axis or rotation+mirroring was necessary. **D** Comparison of alignments with polarity 1 and 2 of lower and upper filaments. If the alignment parameter reflected the two polarities of the reference images (different polarity), the filament was considered to be consistent. Filaments that were aligned in the same orientation to both references were denoted as inconsistent. **E** Comparison of lower and upper filaments. Only if both filaments within one crosslink image were consistent, the whole EM image was considered to be consistent.

which had both filaments in low resolution and 10 only the upper filament. The graphs in Figure 6.53 **D** reproduced this by assigning 10% of the images from the lower filament and 20% of the upper filaments as inconsistent (same polarity, same phase, indicated by the arrow).
Once the assignments for each of the filaments were identified, the final step was to combine and compare the alignment parameters (shown in the four graphs in **C**) of both, the lower and upper filaments. Here, another quality check was implemented which allowed only those crosslink images to be used for the final polarity / phase assignment, where both filaments of one crosslink image were consistent. The comparison of the parameters led to the final assignment of the crosslinks, which is shown in the graphs in **E**. If no rotation or mirroring was necessary to fit both filaments to one reference, the crosslinks were "parallel and in phase". If one of the filaments had to be mirrored at the y-axis, the crosslinks were termed "antiparallel and in phase". If a rotation in the range of 180° of one filament was necessary, the crosslink image was assigned "parallel and out of phase". And finally, necessary mirroring and rotation resulted in "antiparallel and out of phase" crosslinks. Obviously, there were more combination that had to be considered. Mirroring of the lower filament on the y-axis for example led only to antiparallel polarity, if the upper filament was not mirrored as well. In case that both filaments were mirrored at the y-axis, the crosslink image was again considered to be parallel. All possible combinations with the corresponding assignment are listed in Table 6.12 and Table 6.13 gives the resulting image numbers / percentages for the polarity / phase relations of the data set.

		upper filament			
		0	1	2	3
lower filament	0 (nothing)	pp/ip	pp/op	ap/ip	ap/op
	1 (180° rotation)	pp/op	pp/ip	ap/op	ap/ip
	2 (mirror y-axis)	ap/ip	ap/op	pp/ip	pp/op
	3 (rotation+mirroring)	ap/op	ap/ip	pp/op	pp/ip

Table 6.12: Relation between polarity / phase and applied rotations / mirroring during the alignment of real EM images to a simulated actin reference image.

The final results are depicted in Figure 6.54, showing the distribution of all five possibilities. As expected, the percentages were in agreement with the original data set (20 bad images, 8 images pp/ip, 28 images pp/op, 24 images ap/ip and 20 images ap/op). The corresponding class averages (**A**) however did not show these features as for each variation there were still four orientations possible (see Figure 6.51) over which was averaged in the classes shown here.

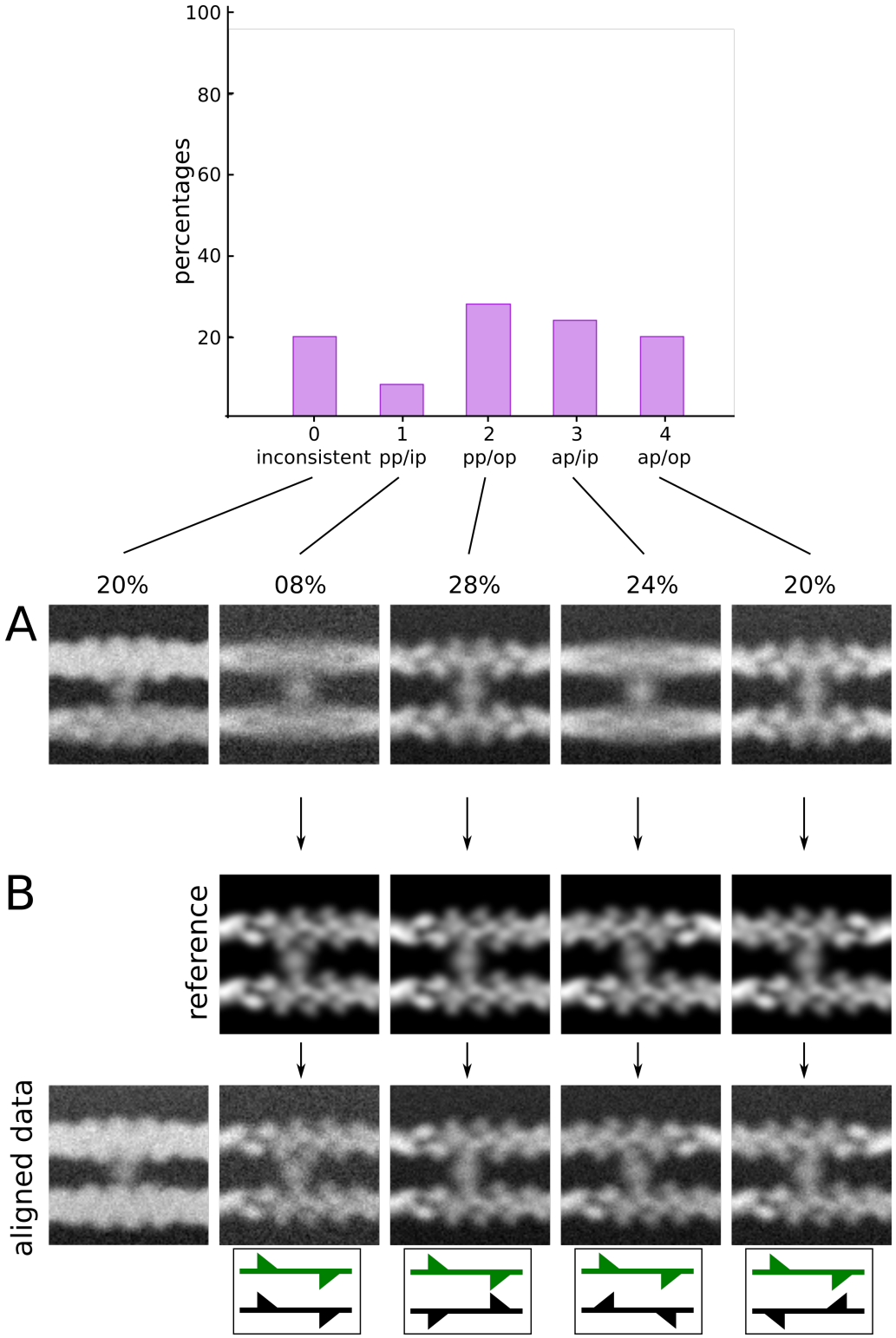


Figure 6.54: Polarity analysis on generic data set, step 4: Polarity and phase assignment. **A** Corresponding class averages for each case. The resulting percentages were in agreement with the original generic data set. The averages still appeared blurred. But alignment with the corresponding reference image reproduced the classes of the initial data set.

		all images (100 img)	only consistent images (80 img)
inconsistent	20 img	20%	
pp/ip	08 img	08%	10%
pp/op	28 img	28%	35%
ap/ip	24 img	24%	30%
ap/op	20 img	20%	25%

Table 6.13: Values for graphs in Figure 6.53 **E**

Recalculation of crosslinks, taking polarity and phase into account:

Using the information about the polarity and phase of each crosslink image, the corresponding images from the initial data set were aligned with their assigned reference as demonstrated in Figure 6.54 **B**[17].
Switching back to the real EM data (Figure 6.48 **A**), all images - being now "polarity-aligned" to their corresponding crosslink-reference - were once more classified with respect to the myosin-binding position using the same mask (Figure 6.48 **B**). This classification led then to the final class averages showing the three distinct myosin conformations.

6.4.2.5 Interpretation of the actin - myosin IX - crosslinks

Identification of the actin-myosin binding by calculating the variance

In order to identify the EM intensities and to probe which parts represented actin filaments and which belonged to the crosslinking myosin, the variances (see section 3.1.2) between a simulated crosslink image (two actin filaments at the correct distance, polarity, phase and *without* any simulated myosin in the middle) and the real EM averages were calculated. Figure 6.55 illustrates the workflow of this process. **A** and **B** show one more time the global average of the prealigned data set[18], which has been aligned with the simulated crosslink image mentioned before (without myosin), the mask that was used for the classification of the aligned data and the data set classified into the three found conformations. These images were then used to calculate the variance with the simulated image shown in **C**, which was carried out by applying the operation *AS R* in SPIDER. The variance between the global average and the model showed a pattern that identified all intensities arising from the crosslinking myosin. While the variance between the global average and the simulated actin filaments identified several differ-

[17]The alignment of inconsistent images could be done only reference free as no polarity and / or phase had been assigned

[18]The prealignment refers to the creation of an average with one common myosin binding conformation position at the lower filament

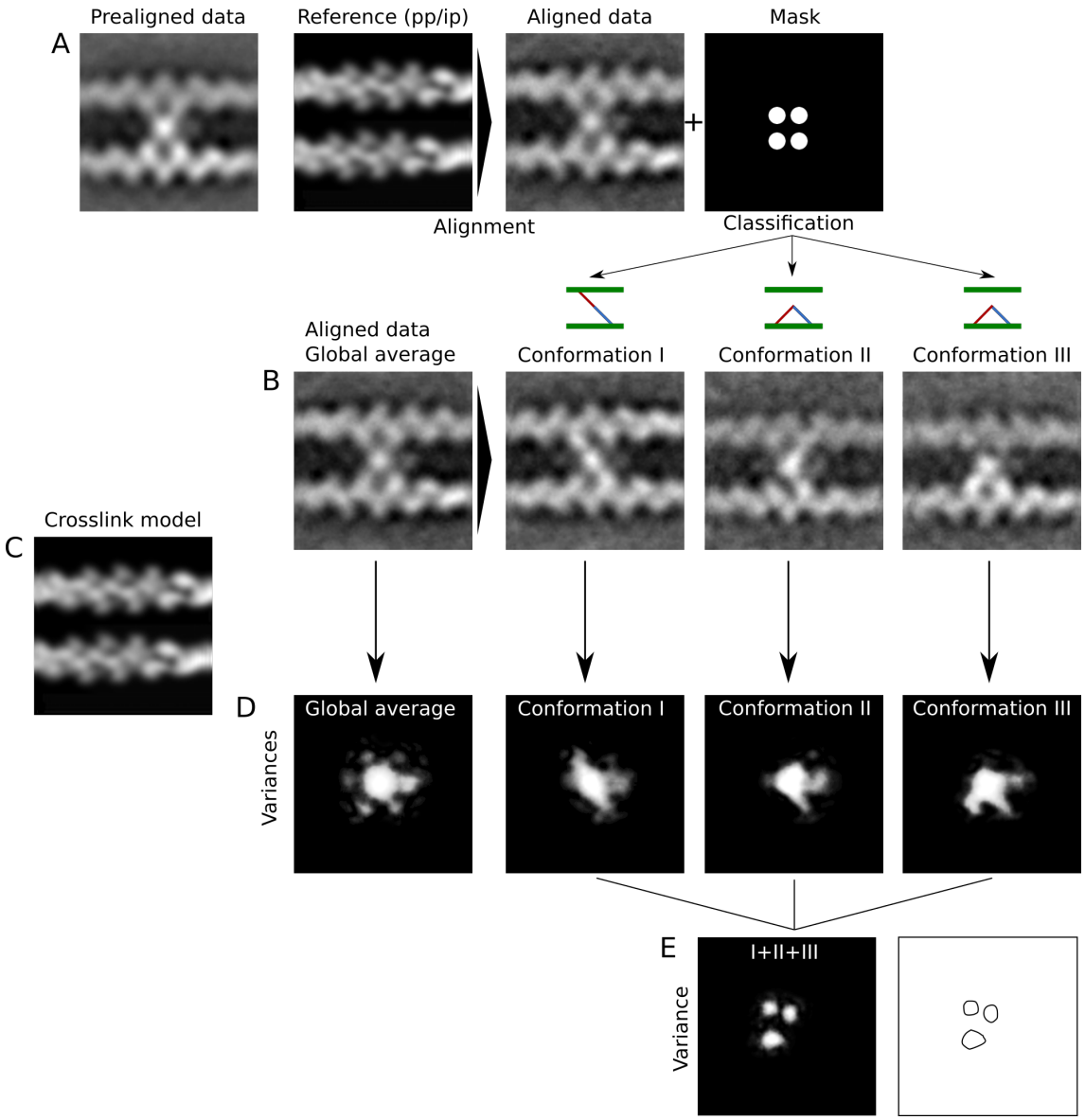


Figure 6.55: Variance analysis for identification of the actin binding position: **A** Prealigned data set was reference based with a generic crosslink and classified with respect to the four actin binding positions. **B** Aligned global average and three classes showing conformation I, II and III. Variance calculation of each of these images (**B**) with the generic crosslink (**C**) identified which parts belong to actin, which to myosin (**D**). Variance calculation of the variance images of the three conformation (**E**) identified the positions of the varying binding site. These variance image was converted into outlines for the purpose of superposition.

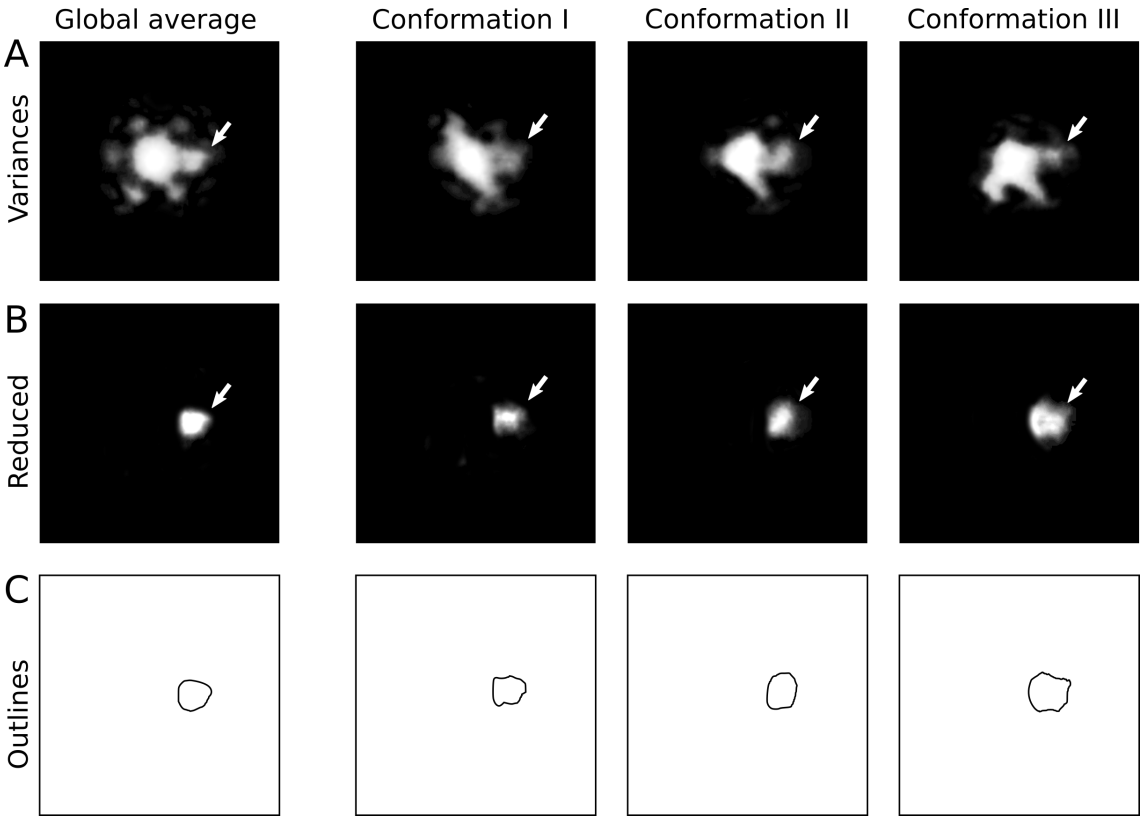


Figure 6.56: Reduction of variance images (**A**) to the consistent blob on the right hand side of the central intensity (see arrows) (**B**) and conversion into outlines (**C**). These outlines in combination with crystal structures were used to identify parts of the original EM averages.

ent spots, the variances between each of the three conformations with the simulated image were more specific (**D**). Another interimage variance calculation between these three variances (I, II and III) specified further the differences between the three conformations (**E**). The intensity distribution of this variance image was then converted into simple outlines for the purpose of superpositioning them onto the EM average. Apart from concentrating on the differences between the three conformations, there were as well similarities except the standard actin binding site. As indicated by the arrows in Figure 6.56 **A**, the intensity blob on the right side was consistent in all conformations. For later identification of this mass the variance images have been reduced to this area (**B**) and converted into outlines (**C**).

Figure 6.57 shows the same variance-analysis for the data set of myosin IX - crosslinks in the presence of calcium. The variance of the global average (**A**) was classified into the three variances for each conformation, which were then compared with each other by calculating the variance once more (**B**). In analogy to Figure 6.56, the intensity on one of the sides of the central spot was cropped and contrast improved. This additional mass was only visible in conformation I and II (**C**).

Interpretation of the crosslinks using crystal structure projection matching

In order to interpret the EM averages of actin filaments crosslinked by myosin IX, the images were compared with a crystal structure of an actin-myosin complex. As there were no structures of myosin IX available, neither on its own, nor as a complex with actin, a structure of another myosin had to be used instead. A seemingly ideal available crystal structure was the actin-myosin IE complex, which comprised five actin monomers, tropomyosin, and three myosin IE heavy chains, each bound to a different actin monomer (PDB: 4A7F [139]), as shown in Figure 6.58 **A**. In contrast to other structures, such as myosin V (PDB: 1OE9 [140], **B**), the myosin IE structure here included only the catalytic domain without any lever arm. Furthermore, it did not contain the N-terminal SH3 domain. These were optimal conditions, as the myosin IX construct used for the project also did not include the N-terminal SH3 domain. However, the myosin IX comprised a 207 amino acid long and calmodulin binding special insert in its loop2, which was not covered by the myosin IE structure. Also, the actin filaments in the crosslinks were linked by only one myosin molecule (at least for the distance of one pseudo repeat). Therefore the model had to be adjusted in order to use it for the analysis. From the three myosin IE molecules, two were deleted and also the tropomyosin. This generated three models, each having the myosin IE bound to a different position (**C**, *a1-a3*). Furthermore, in order to extend the length of the actin filament of the model (as

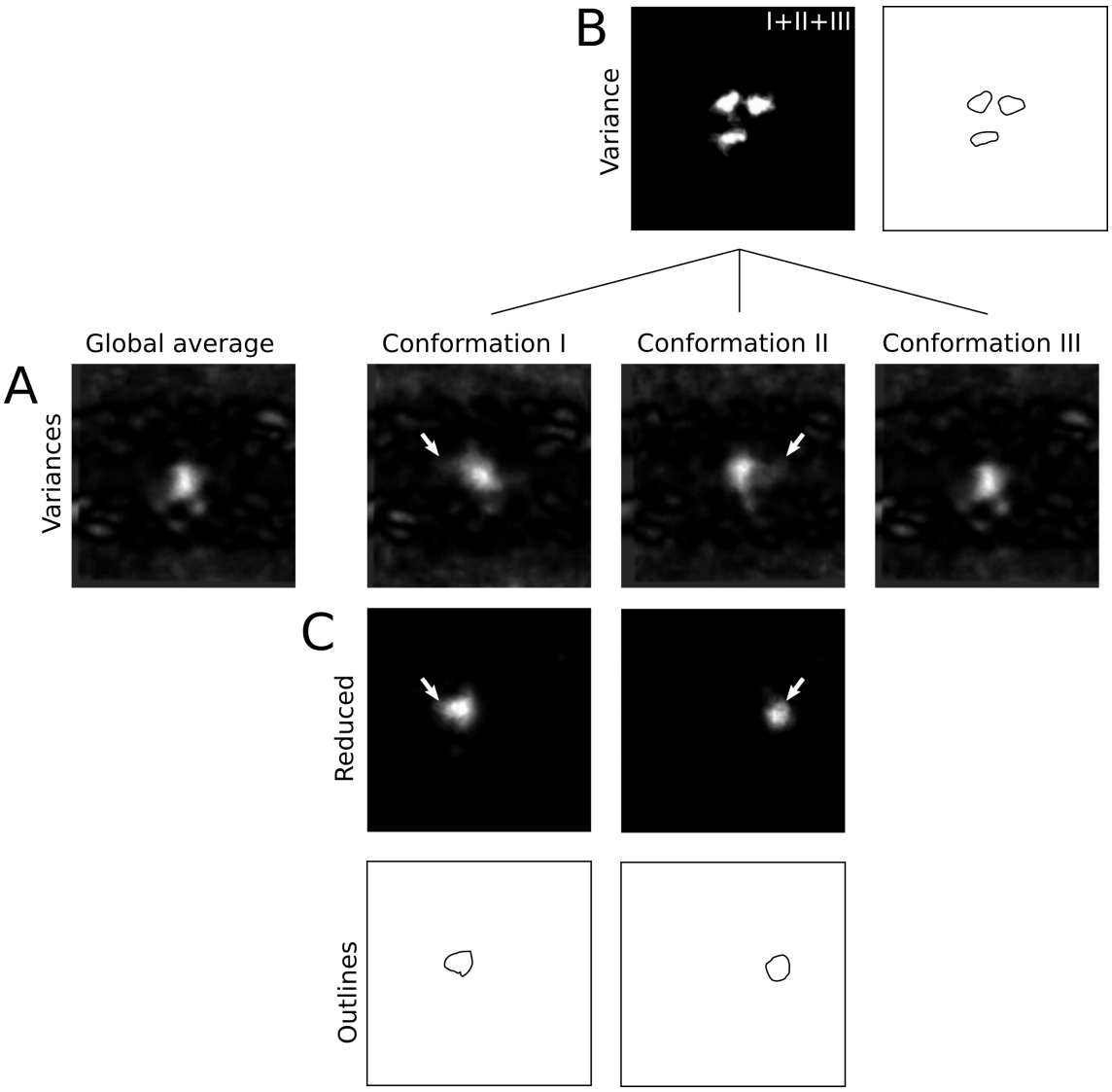


Figure 6.57: Variance analysis for identification of the actin binding positions in the presence of calcium in analogy to Figure 6.55: **A** Variance calculation of the global average and each of the three conformations. each of these images with the generic crosslink identified which parts belong to actin, which to myosin. **B** Variance calculation of the variance images of the three conformations. This variance image was converted into outlines for the purpose of superposition. **C** Reduction of variance images to the blob on the left (conformation I) and right side (conformation II) and conversion into outlines.

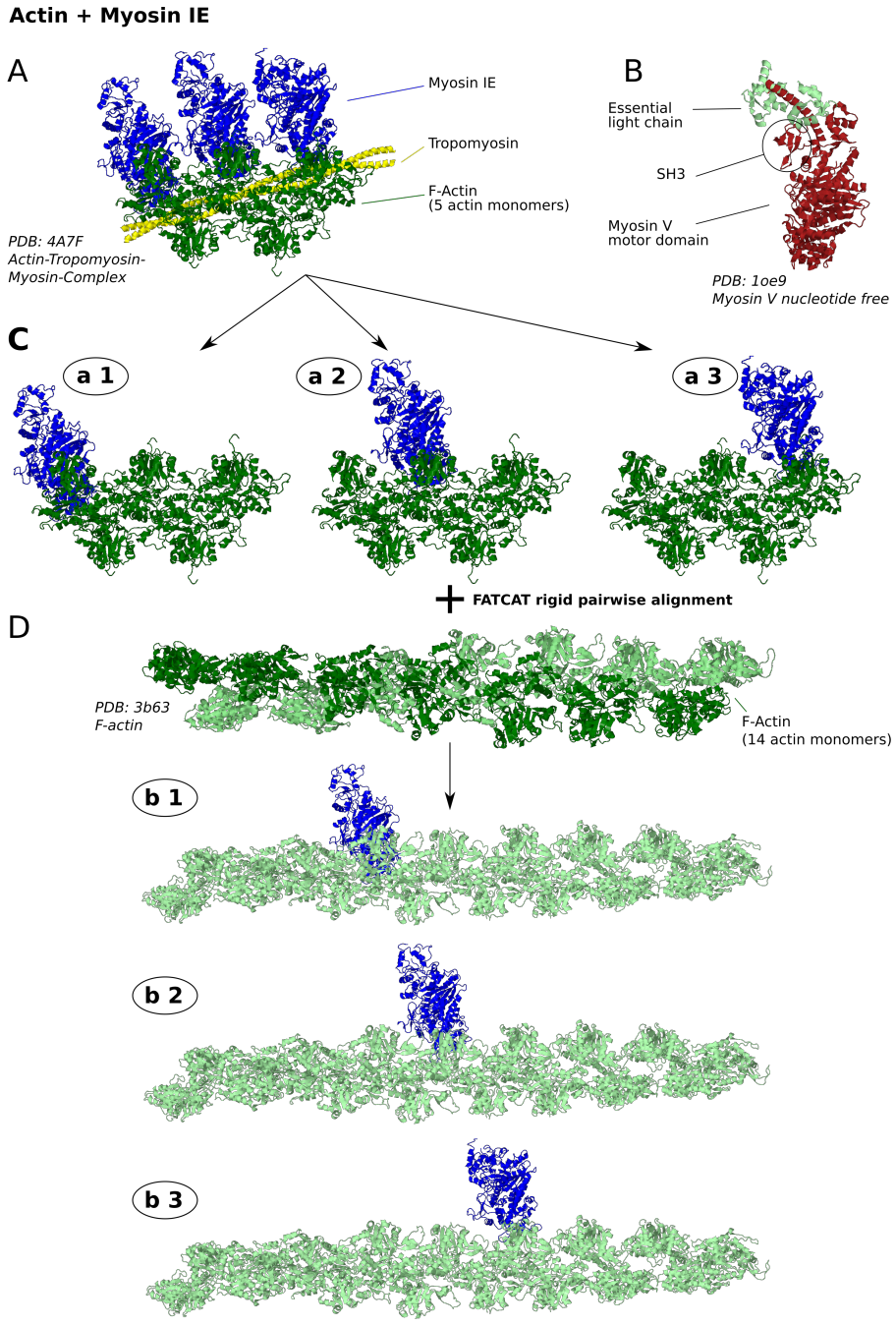


Figure 6.58: Suitable crystal structure to analyze myosin IX - actin - crosslinks. **A** Complete structure of myosin IE bound to actin (PDB: 4A7F [139]) with five actin monomers, tropomyosin and three myosin molecules. Other available crystal structures would be e.g. myosin V (**B**). This structure however included the lever arm and the SH3 domain, which was not contained in the myosin IX construct used here.**C** Deletion of tropomyosin and two myosin molecules each resulted in three structures a1-a3. **D** FATCAT pairwise alignment with the 14 monomer long F-actin structure (PDB: 3b63 [138]) created another set of models (b1-b3) suitable for the crystal structure projection matching of crosslinks.

more than 5 actin monomers are seen in the filaments in the EM data), another crystal structure of F-actin comprising 14 actin monomers (PDB: 3B63 [138]) was aligned with each of the three modified actin-myosin IE structures (**D**, *b1 b3*). This alignment was done using the FATCAT rigid pairwise alignment [58]. The corresponding P-values for the alignments were 2.75E-07 in each case. In total, this created six different models of the acto - myosin IE - complex.

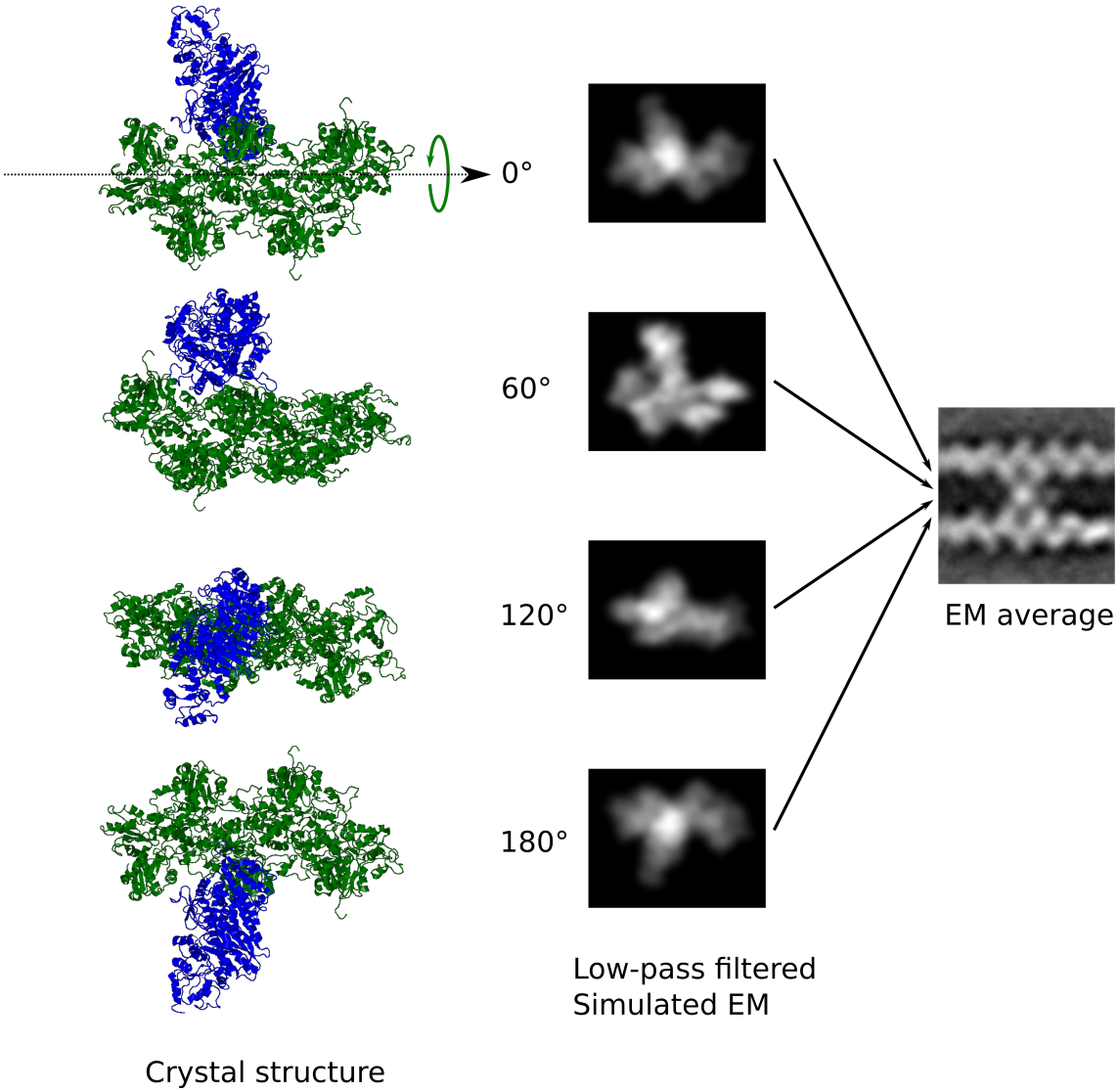


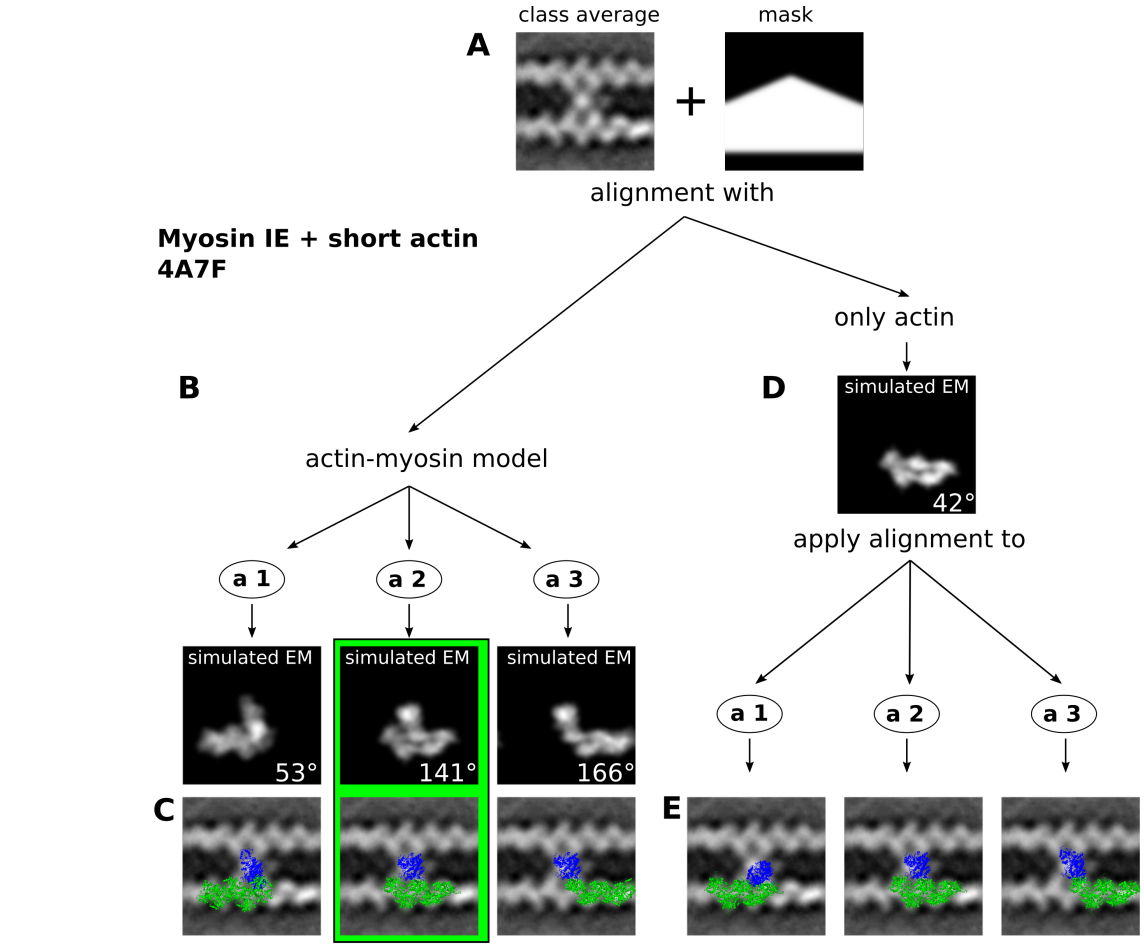
Figure 6.59: Low-pass filtered 2D projections by rotating the models around the long axis of actin in 1° steps. More projections were not needed as the filaments could be assumed to lie flat on the EM surface.

These models were then used to analyze the EM averages by crystal structure projection matching based on the description in section 4.2 and using the scripts *PDB Fitting/pdbtospi.spm, findorientation.spm, writejmolmacros.spm*: A set of simulated EM images was calculated for each model containing 360 low pass-filtered 2D projections, each

related by 1° rotations around the F-actin axis (see Figure 6.59). Since it could be assumed that the crosslinked actin filaments were lying flat on the EM grid surface, it was not necessary to calculate the complete set of 799 projections, in contrast to the myosin VI project (see chapter 5, section 5.5.2.3).

To identify the best matching of the six different models, all 360 projections of each model were aligned with the EM average as illustrated in the two figures 6.60 and 6.61 for both sets of crystal structures, *a1-3* (myosin IE with short actin filament) and *b1-3* (myosin IE with long actin filament) respectively[19]: Using a mask that covered only the lower filament and the crosslinking myosin molecule (**A**), the EM average was aligned directly with the modeled structures (**B**). Shown are the structures in the orientation that yielded the highest cross correlation coefficient during the alignment. The crystal structure was then in its original resolution overlaid onto the real EM average (**C**). Apart from the alignment with the whole modeled structure, the EM image was also aligned with only the actin filament (hidden myosin molecules), in order to focus the alignment completely on the orientation of the actin filament. Once the optimized orientation was found (**D**), the myosin molecules were put back onto the actin model and the resulting structure overlaid onto the EM average (**E**). For the calculation of the cross correlation coefficients one can make use of various masks with the purpose of focusing the similarity of specific parts of the crystal structure. The masks used here are shown in **F**. The first mask was equivalent to the mask used for the alignment. Mask two and three were similar but cut out the edges of the actin filaments as at these positions the resolution was significantly lower than for the rest of the image. Mask four cuts out all image-parts except the crosslinking myosin and therefore focuses the ccc-calculation on myosin alone. In both figures included are the final ccc-values for all modeled structures / masks. For each mask, the underlined numbers indicate the maximum ccc within each case (**B** and **D**), while the bold numbers indicate the maximum value for both cases together. Further, the outlined values highlight the maximum ccc when taking both sets (figures 6.60 and 6.61) into account.

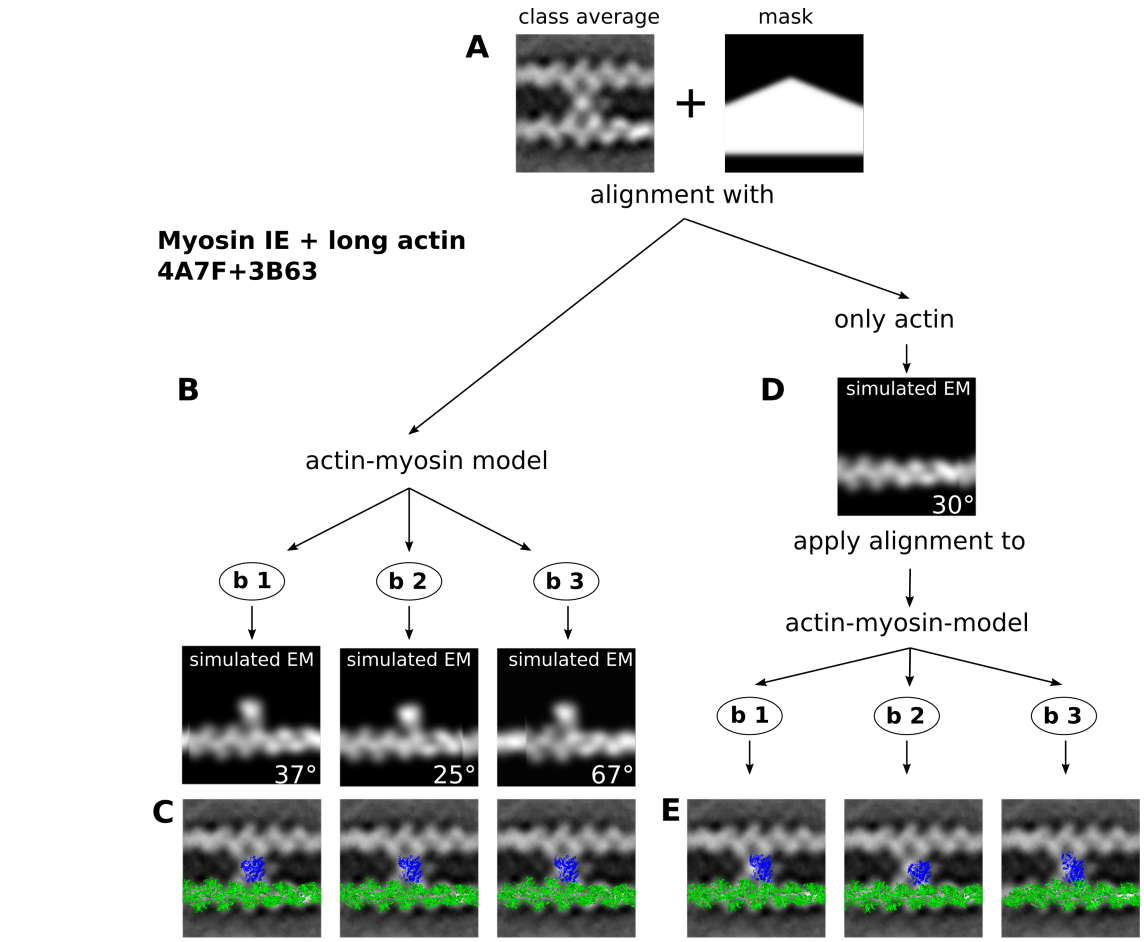
[19] The workflow will be described only once as it was analogous for both sets of images



cross correlation coefficients

F						
masks for ccc-calculation	mask 1	0.51560	0.55862	**0.59142**	0.23964	0.44693	0.27830
	mask 2	0.33881	**0.63021**	0.56861	0.24252	0.38335	0.29334
	mask 3	0.42688	**0.70298**	0.61860	0.26797	0.42510	0.32236
	mask 4	0.06754	**0.59582**	0.53011	0.09945	0.29944	0.20608

Figure 6.60: Crystal structure projection matching of models a1-a3 to the masked (**A**) crosslink EM average was performed in two ways: **B,C** Alignment and superposition of the entire structure or **D,E** alignment of only the actin and later recreation / superposition of the entire model. **F** Corresponding cross correlation coefficients using different masks for the calculation. Underlined values: maximum values within each alignment approach (entire model or just actin), bold values: maxima taking both alignment methods together, framed values: taking values from next figure into account.



cross correlation coefficients

F

masks for ccc-calculation

0.82235	0.80480	**0.82263**	0.54114	0.77921	0.47515
0.85059	0.83189	0.85047	0.54136	0.78073	0.48667
0.86882	0.85319	**0.87806**	0.53203	0.80262	0.50174
0.58167	0.56856	**0.58711**	-0.14175	0.53826	0.33736

Figure 6.61: Crystal structure projection matching of models b1-b3 to the masked (**A**) crosslink EM average in two ways: **B,C** Alignment and superposition of the entire structure or **D,E** alignment of only the actin and later recreation / superposition of the entire model. **F** Corresponding cross correlation coefficients using different masks for the calculation. Underlined values: maximum values within each alignment method (entire model or just actin), bold values: maxima taking both alignment methods together, framed values: taking values from previous figure into account.

This revealed, that for the case when only actin was used for the alignment, the model with myosin IE sitting in the middle yielded the best fit. On the other hand, the alignment with the whole structure directly resulted in even higher ccc-values in general. When the two sets were compared, for masks 1-3 the method of using the longer actin filaments worked the best, resulting in the highest ccc-values. However, as this analysis was supposed to be focused on the structure of the myosin, the most reasonable mask was number 4. In this case, the modeled structure with one myosin IE molecule sitting in the middle of 5 actin monomers resulted in the best fit (as highlighted by the green outline in Figure 6.60 **F**). This crystal structure in its optimized orientation was then used for the further analysis.

Modeling the calmodulin onto the N terminus of loop2

The myosin IX MD includes a special insert in loop2 that could be seen in the EM average as an intensity next to the mass representing the catalytic domain of myosin IX. The crystal structure of myosin IE used to interpret the EM data however did not comprise such an insert. Therefore, the best fitting model (see previous section) was modified in the following way (illustrated in Figure 6.62): The N- and C-terminus of loop2 of Myosin IE could be localized at aa 541-546 and aa 563-577 respectively [139]. As the structure of the loop itself (marked in red in figure 6.62, see inset for more details) was undefined, the preceding and following helices were used for the modeling. Furthermore it was shown that the motor domain of myosin IX was able to bind calmodulin at the N-terminus of the loop2 insert (IXa-MD-CBS) [106] [110]. The spacial arrangement however of the loop2 with its insert, and the calmodulin were unknown and therefore a structure of calmodulin (using the second calmodulin of the myosin VI structure (yellow) including the corresponding alpha-helix of the lever arm (green, PDB: 3GN4 [81]) was modeled onto the loop2 N-terminus by aligning the preceding helix of myosin IE with the alpha-helix of the calmodulin-structure. This was done using Swiss-PDBViewer [57]. In **B** shown is the final structure in two views, related by a 90° rotation.

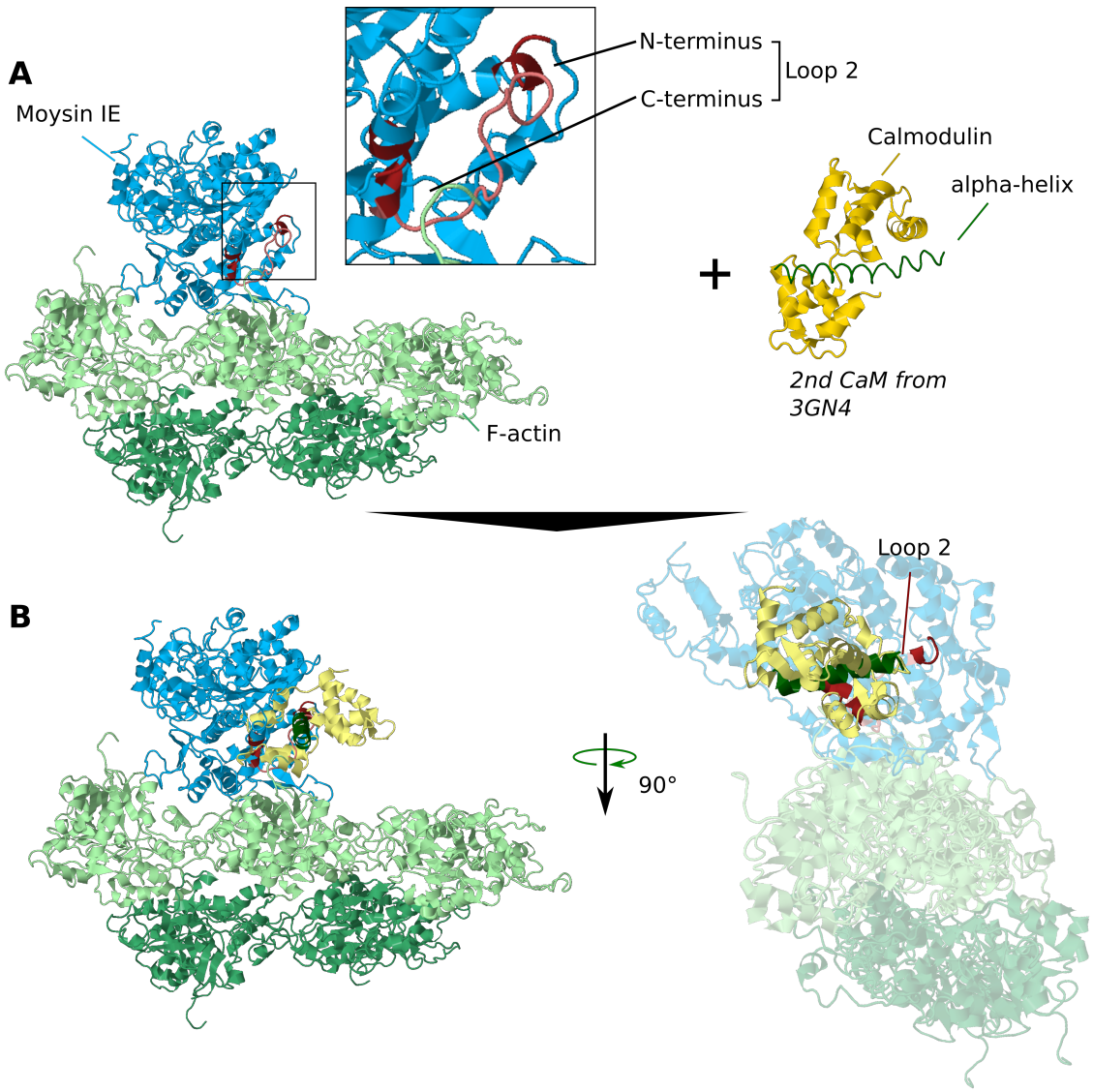


Figure 6.62: Modeling calmodulin onto loop2 of myosin IE: **A** Myosin IE - actin - model with colored N- and C-terminus of its loop2 and a model of the second calmodulin from PDB: 3GN4 [81]. **B** Combined model with added calmodulin in a way that its α-helix of the target peptide (green) continued the N-terminus of loop2.

6.5 Supplementary material

The following section contains the supplementary material that has been referred to in this chapter. Representative negative stain electron micrographs show the general network of (calcium free) myosin IX - actin - crosslinks, as well as single bundles with two, three or more actin filaments involved:

1) Overview

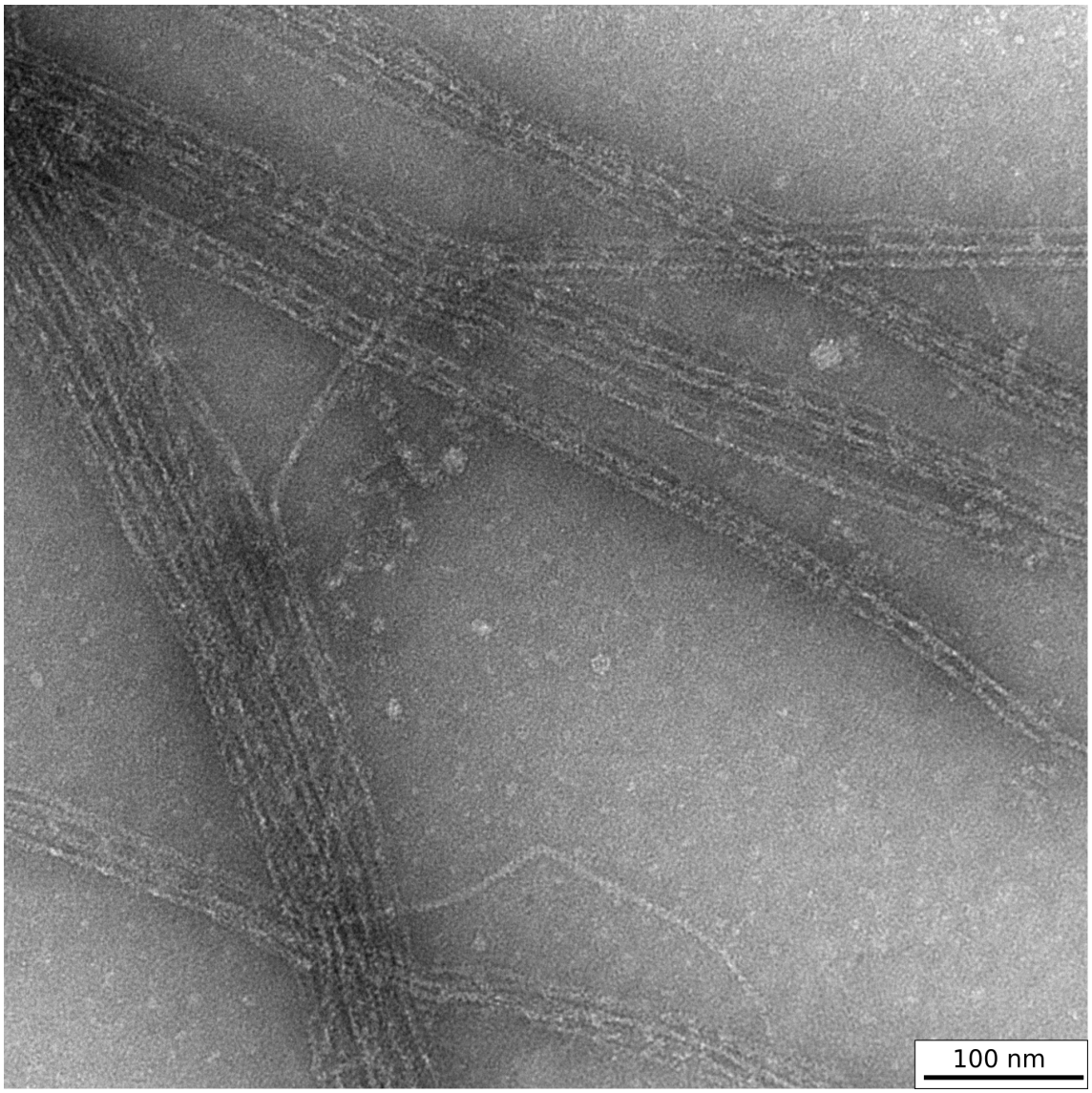


Figure 6.63: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: This overview image shows relatively straight bundles of a varying number of actin filaments crosslinked by myosin IX monomers . Scale bar: 100 nm

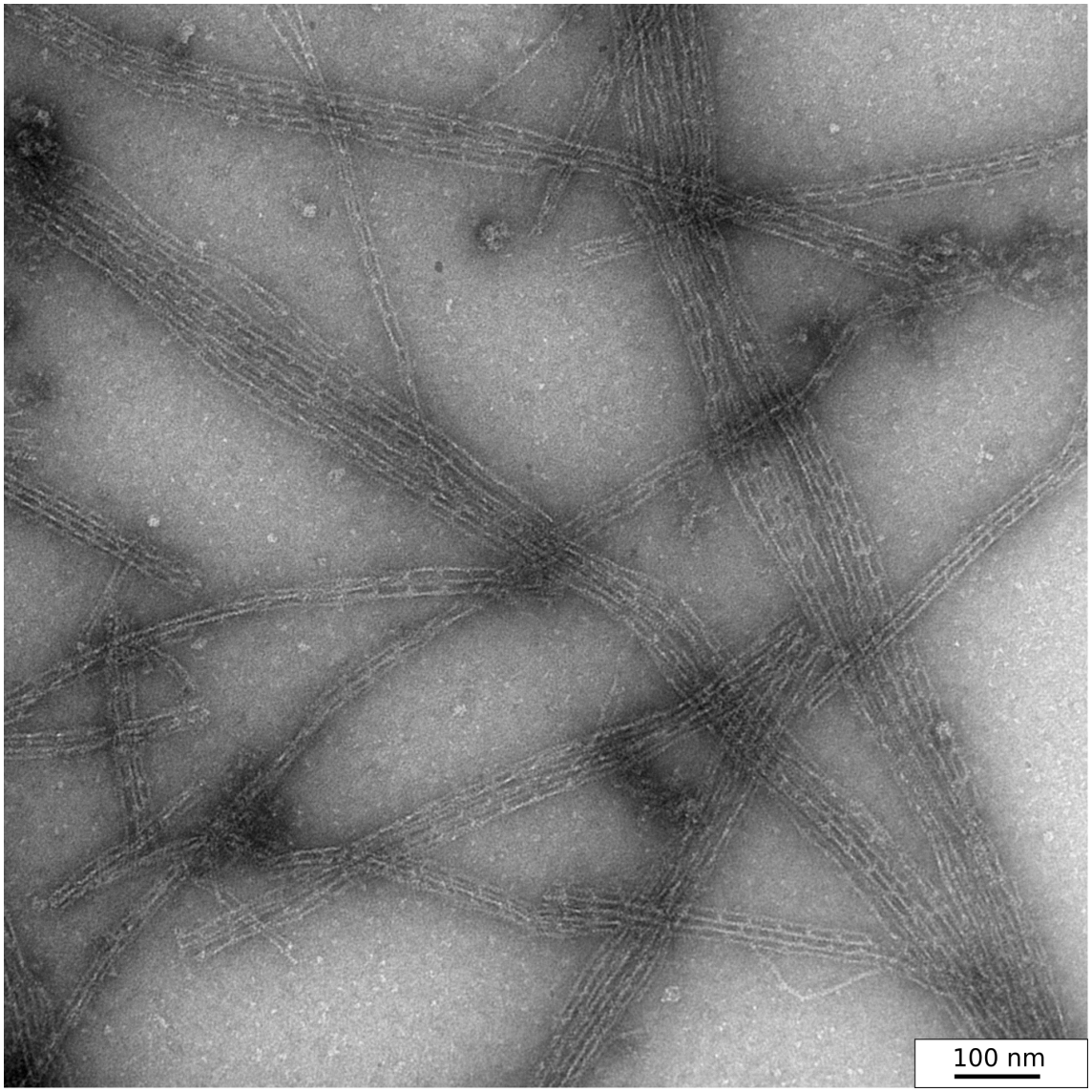


Figure 6.64: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: This overview image shows relatively straight bundles of a varying number of actin filaments crosslinked by myosin IX monomers . Scale bar: 100 nm

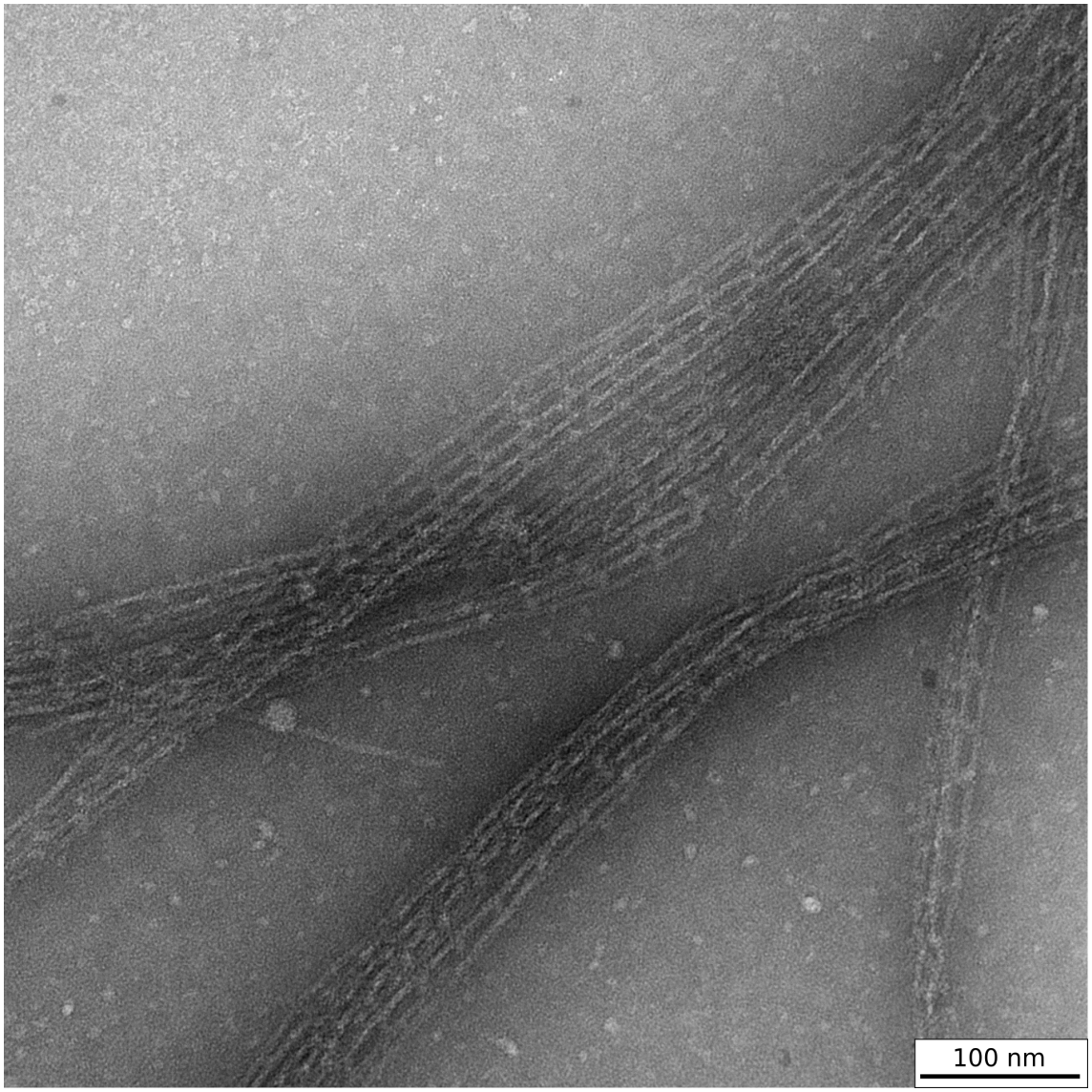


Figure 6.65: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: This overview image shows relatively straight bundles of a varying number of actin filaments crosslinked by myosin IX monomers . Scale bar: 100 nm

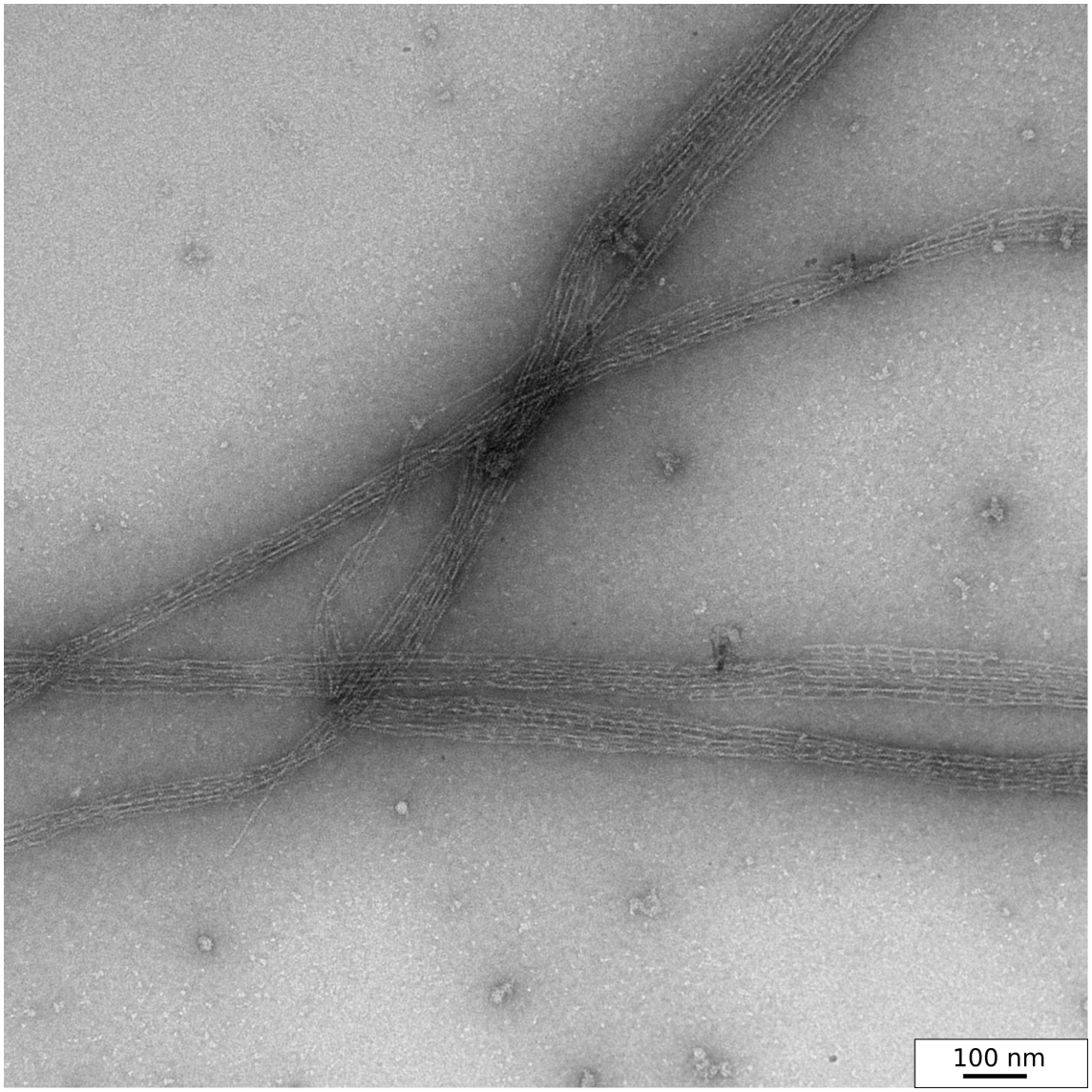


Figure 6.66: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: This overview image shows relatively straight bundles of a varying number of actin filaments crosslinked by myosin IX monomers . Scale bar: 100 nm

2) Crosslinks with two actin filaments

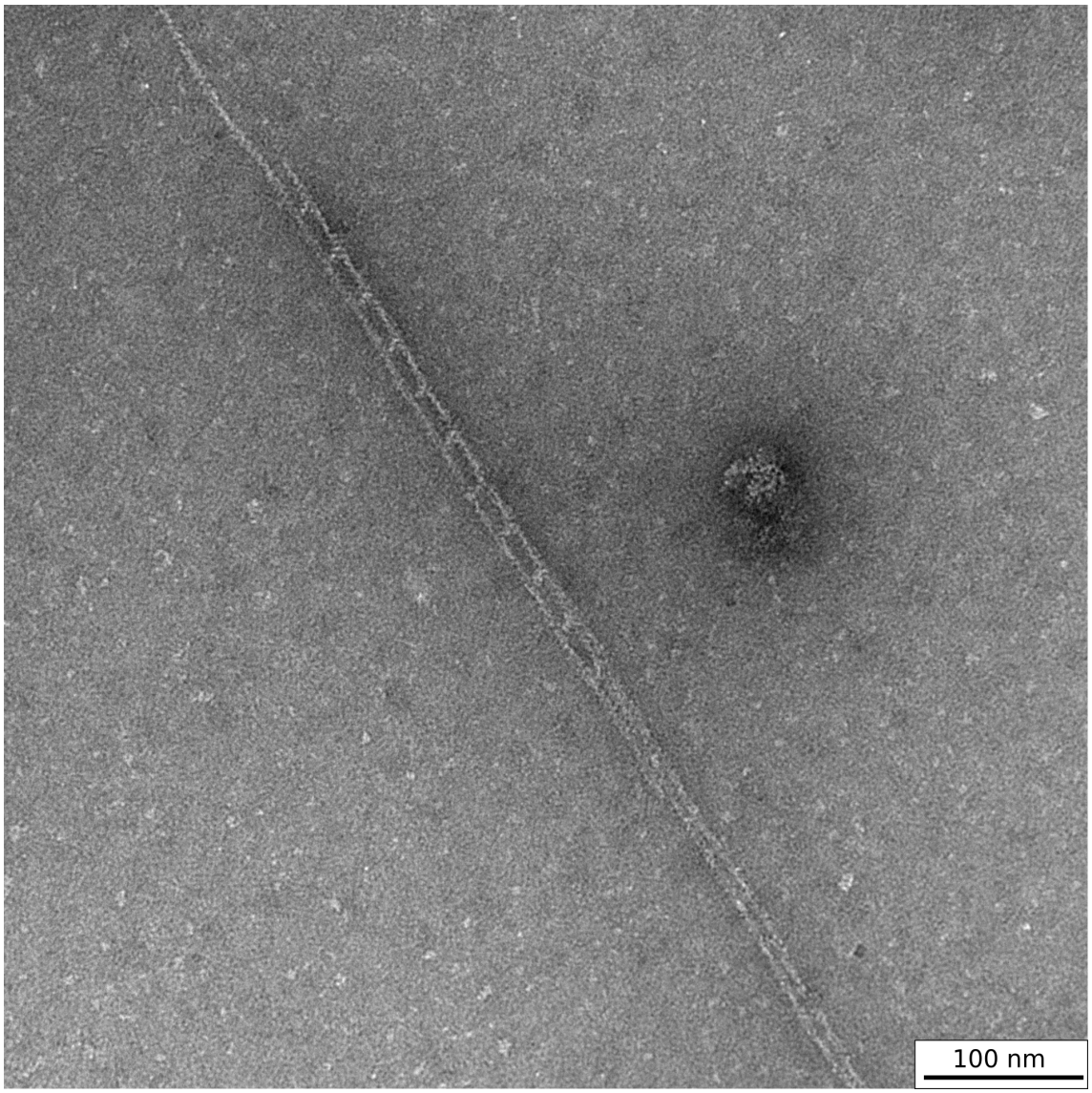


Figure 6.67: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: Two actin filaments crosslinked by myosin IX monomers in a regular pattern. Scale bar: 100 nm

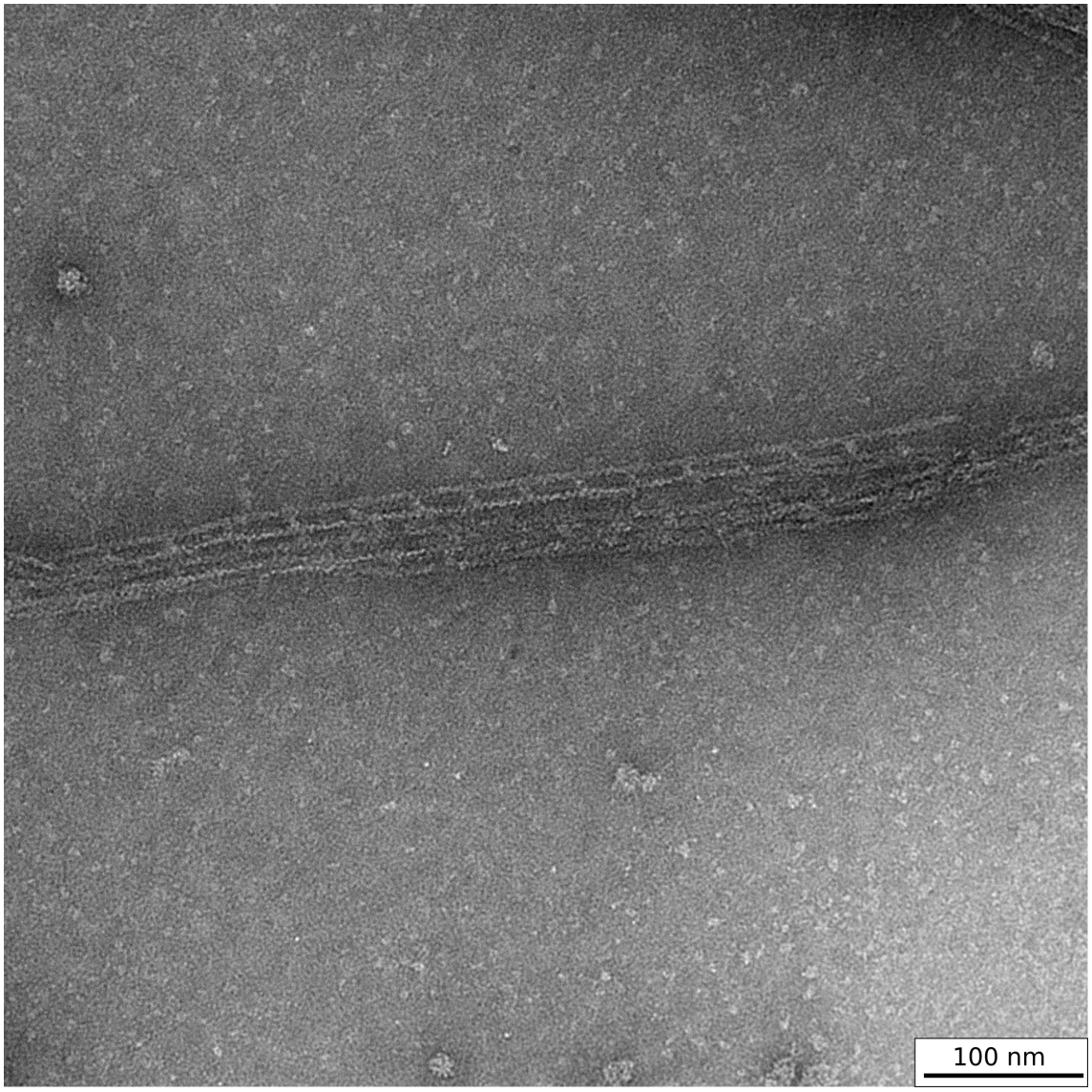


Figure 6.68: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: Two actin filaments crosslinked by myosin IX monomers in a regular pattern. Scale bar: 100 nm

3) Crosslinks with three actin filaments

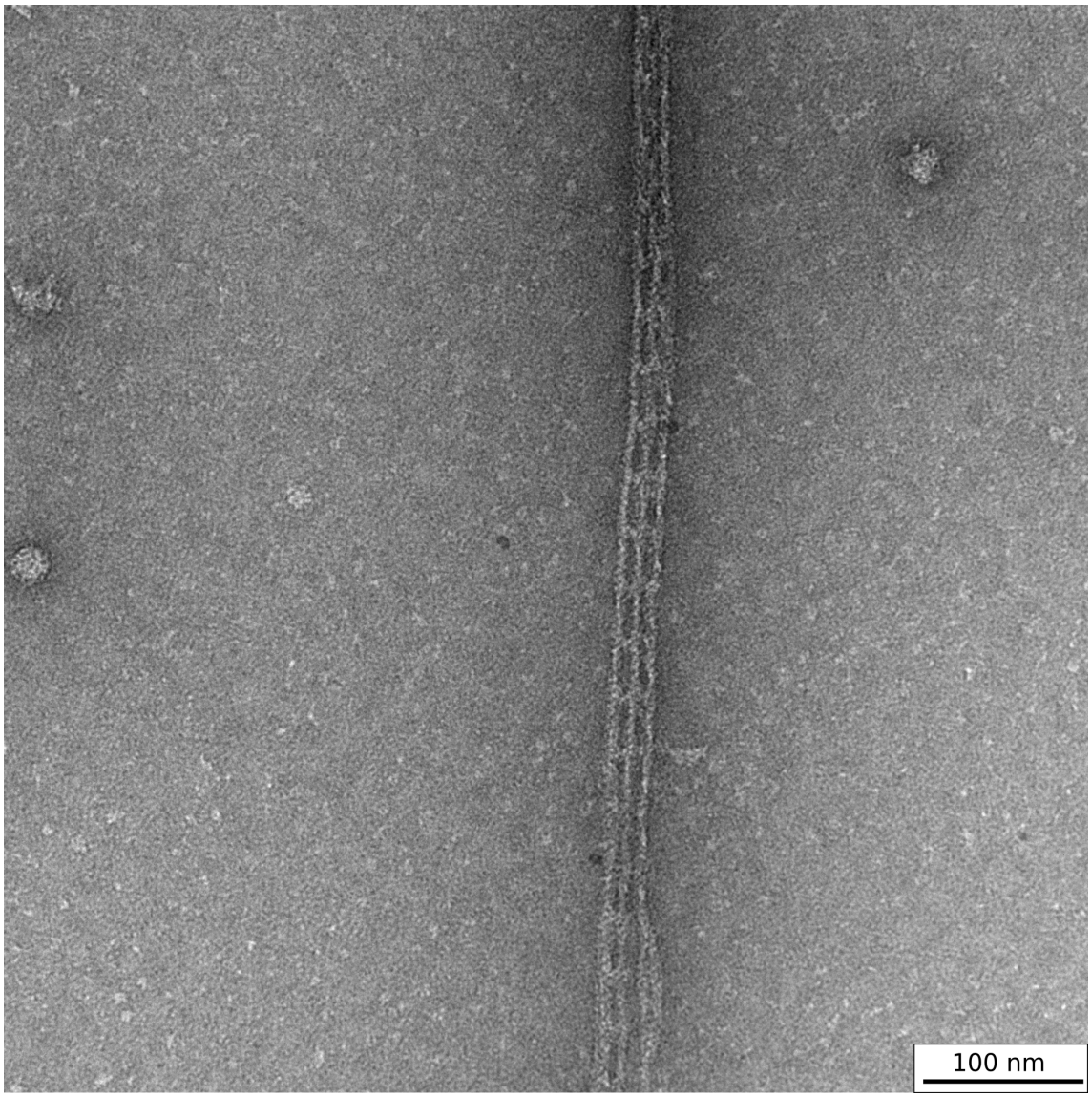


Figure 6.69: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: Three actin filaments crosslinked by myosin IX monomers in a regular pattern. Scale bar: 100 nm

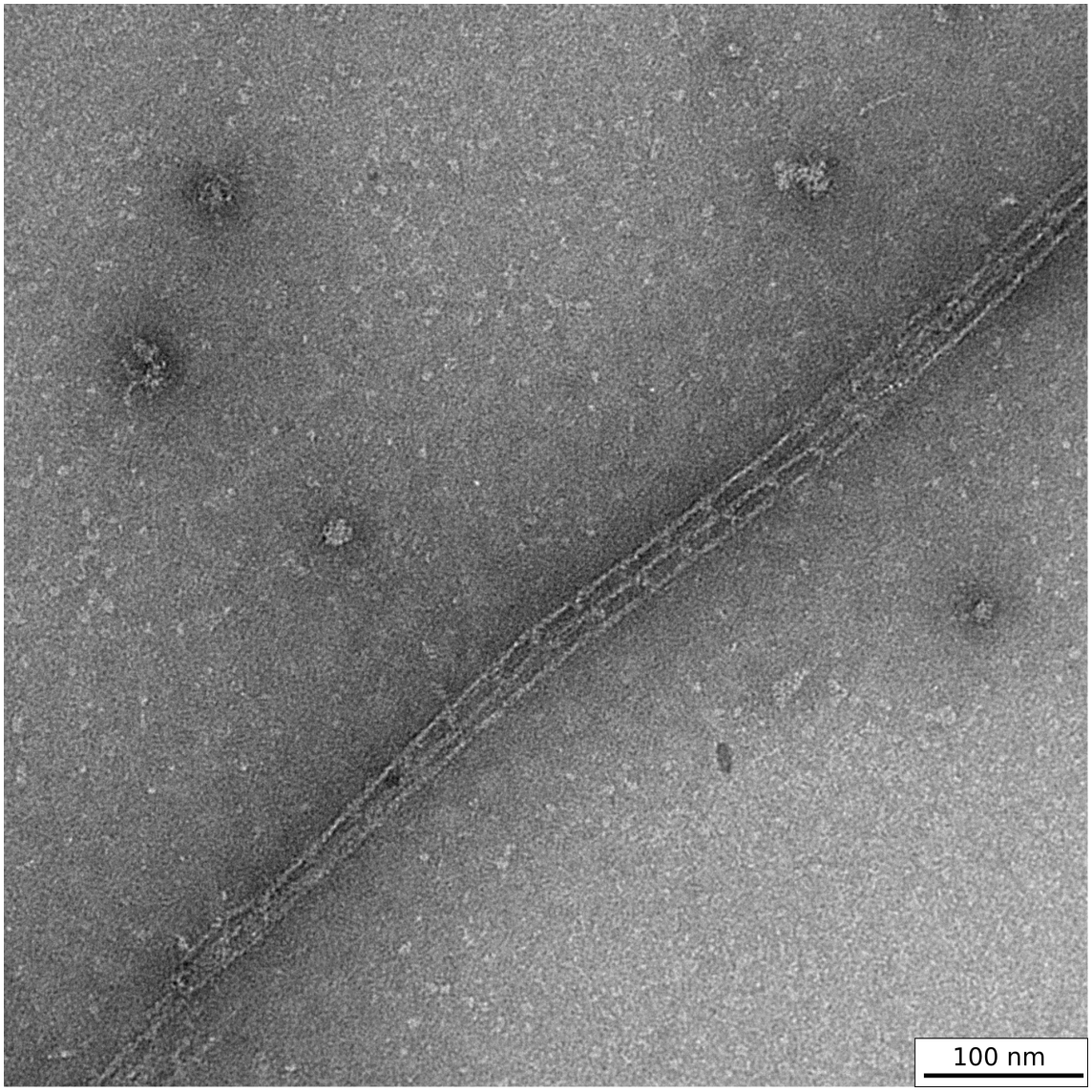


Figure 6.70: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: Three actin filaments crosslinked by myosin IX monomers in a regular pattern. Scale bar: 100 nm

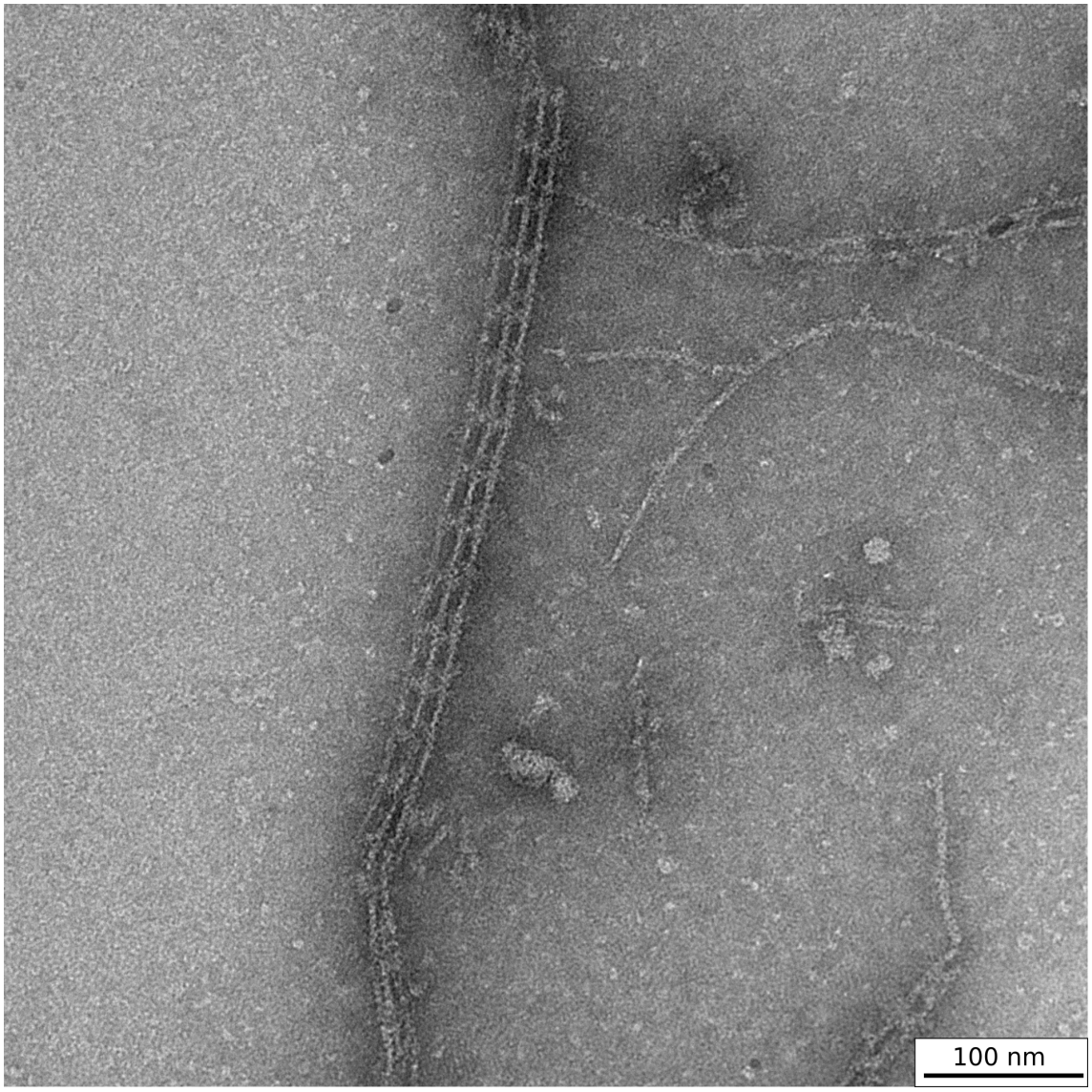


Figure 6.71: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: Three actin filaments crosslinked by myosin IX monomers in a regular pattern. Scale bar: 100 nm

4) Crosslinks with more then three actin filaments

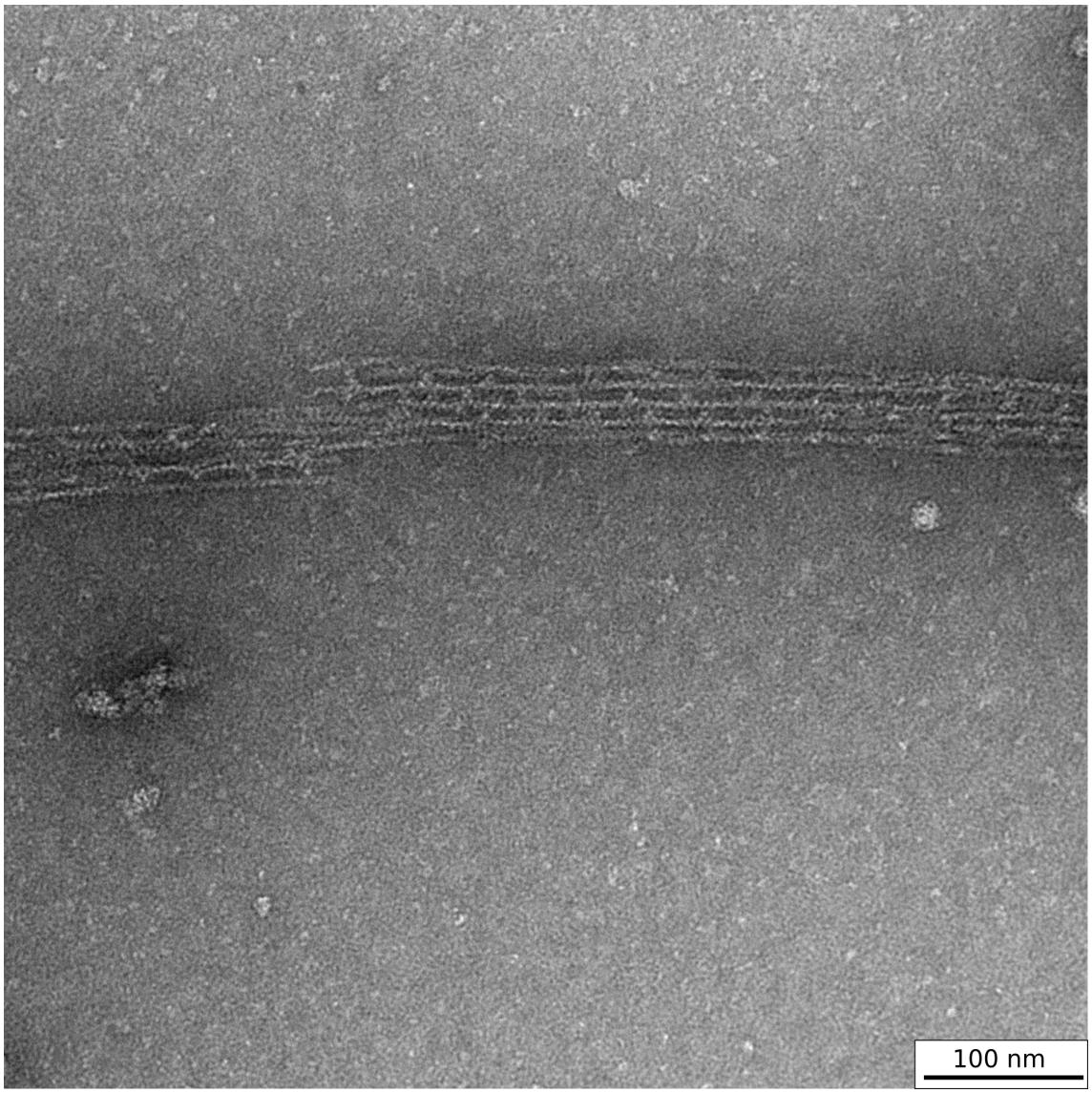


Figure 6.72: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: More then three actin filaments crosslinked by myosin IX monomers in a regular pattern. Scale bar: 100 nm

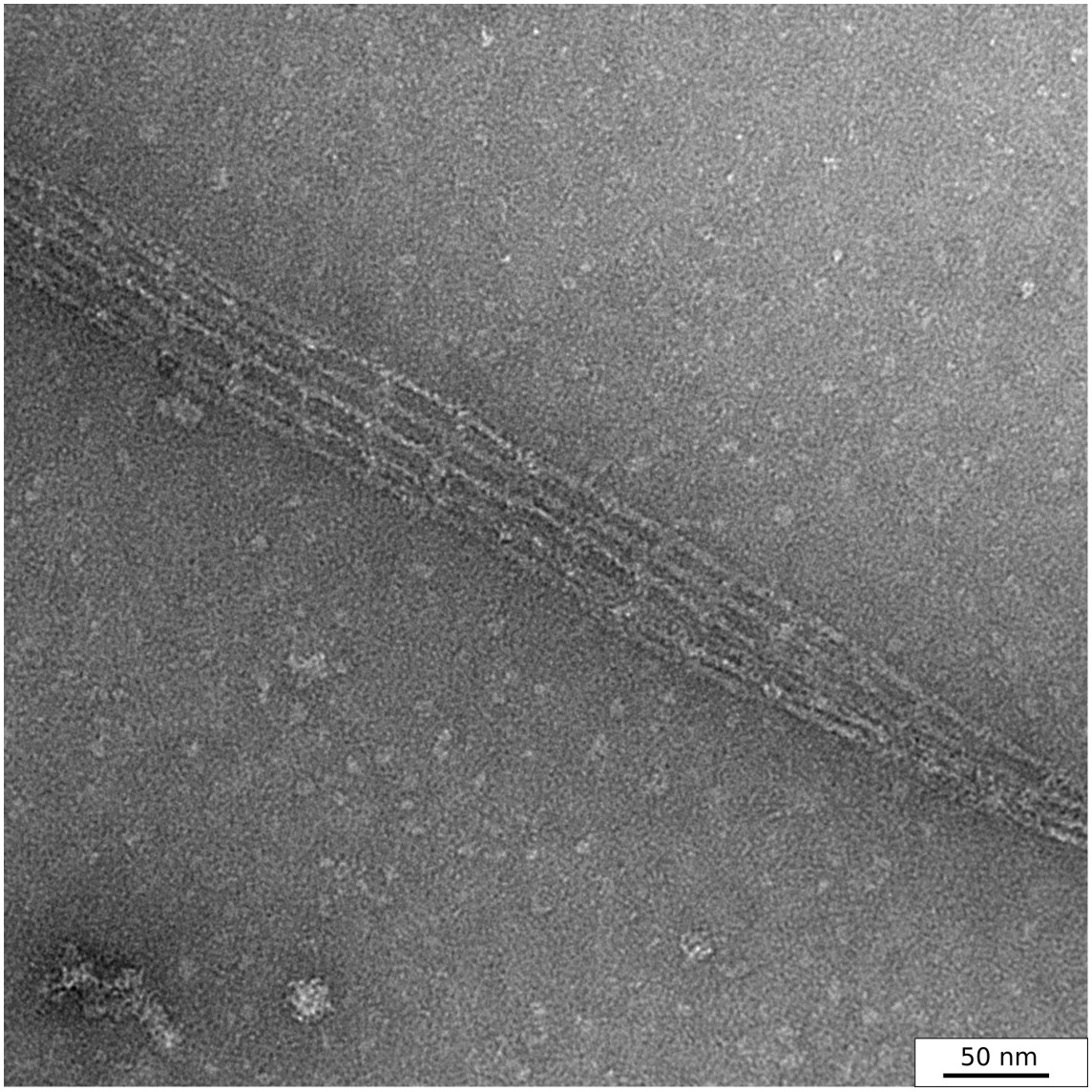


Figure 6.73: Raw micrograph of (calcium free) myosin IX - actin - crosslinks: More then three actin filaments crosslinked by myosin IX monomers in a regular pattern. Scale bar: 100 nm

Part III

Conclusion and Outlook

General conclusion and Outlook

The central aim of this work was to develop methods for the single particle analysis of negatively stained electron micrographs in order to investigate the structural properties of various myosins and their associated binding partners. These protocols were constructed on the framework of the terminal based image processing software SPIDER, which provides a large number of various alignment, classification and miscellaneous algorithms for this purpose. The successful application of such workflows on negative stain EM data sets could resolve the two myosin classes VI and IX up to 18 Å resolution, demonstrating the power of single particle analysis and the importance of careful evaluation of the results.
The studies presented here revealed for the first time striking new insights into the structure of single parts of myosin (second calmodulin bound to the lever arm of myosin VI and regulated by calcium) as well as structures of bigger complexes such as actin-myosin networks established by the myosin IX motor domain, promoting further studies and laying the foundation for higher resolution approaches including cryo-EM.

Apart from the elaborated applications, the workflows were written in a way that they can be comfortably implemented into the recently developed graphical user interface of SPIDER (SPIRE) and therefore easily transferred to similar data sets. Furthermore, keeping in mind that in its essence the single particle analysis is just image processing, many of the methods can be applied to other fields of research. The diverse approaches in our group for example include the recent findings of lipid-bilayer ring-like pattern formation induced by myosin VI and imaged with Stochastic Optical Reconstruction Microscopy (STORM). Although having a high resolution itself, a different kind of information can be obtained by applying single particle analysis on such data sets.

Many questions in the intriguing field of molecular motors are still unanswered. One focus of future experiments will be the investigation of the studied myosin classes in different conditions: How is the two-step regulation mechanism of myosin VI structurally influenced by different binding partners? Can we resolve the flexible tail of myosin

VI? What is the structure of myosin IX in the absence of actin and what are its properties in the presence of nucleotide? Is the suggested spatial arrangement of the calmodulin bound to the lever arm of myosin IX resolvable in three dimensional EM? What is the underlying structure of actin crosslinks formed by other myosin, such as myosin XXI dimers? And what are many other structural characteristics of unresolved myosin classes?

To answer these questions, the application of different and novel approaches and techniques, based on the ones developed in this work, will be inevitable. Eventually, three dimensional reconstructions of full length myosin in a near native state using cryo-EM could reveal new insights. The problem in this context is however the small size of myosin VI, which is hence challenging to image with conventional cryo-EM. The low SNR makes it almost impossible to localize individual objects, necessary for the first and critical step in SPA: the particle picking.

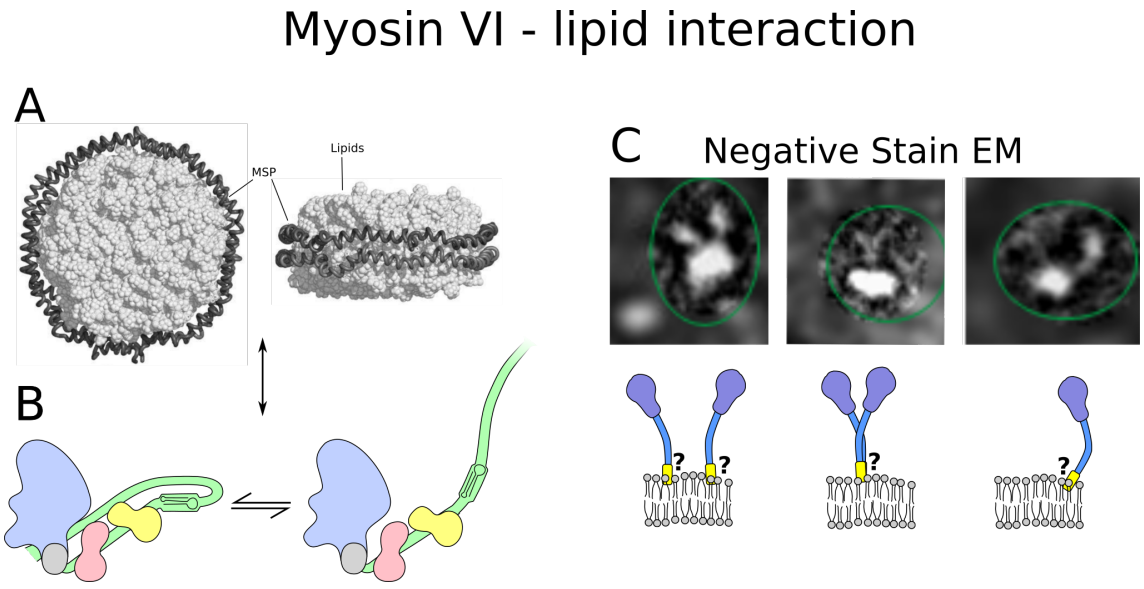


Figure 6.74: Model for myosin lipid interaction. Myosin was discovered to bind to lipid membrane without linker proteins. Using nanodiscs (**A**, artificially created phospholipid bilayers of controllable size), the formation of myosin-disc-complexes could help to identify single myosin molecules in cryo-EM, which alone are usually too small to be seen. **C** First preliminary results using negative stain EM to verify the successful complex-formation. The images shown here are from raw micrographs a possible arrangement is shown below. How the myosin in detail interacts with lipids bilayers is unknown.

To overcome this, one could let myosin VI bind to one of its binding partners. These are among others actin or via binding partners different cargo. However, our group showed that without the need of any linking molecules, myosin VI alone binds stable to lipid membrane. Using this as a background, one could make use of so called nanodiscs, artificially created phospholipid bilayers of controllable size (usually 5-15 nm) (see Figure 6.74 **A**). In the hope of the formation of individual nanodisc-myosin units (cartoon of myosin in **B**), the image processing of cryo-EM data would be guided by the detectable

discs. In order to verify the binding of myosin to the discs in the first place, negative stain EM seems to be an ideal choice. Preliminary results of this strategy (in collaboration with my colleges) are shown in **C**, where possible complexes could be observed. Note that these images are raw micrographs, as the probability of binding events was too low to apply single particle analysis. The cartoon underneath illustrates a possible arrangement. As mentioned, the yield of created complexes was low and further experiments will have the aim to improve that as a crucial requirement for the imaging in cryo-EM. Once this approach will be optimized, it might act as a model system for other small objects to be imaged in 3D by cryo-EM.

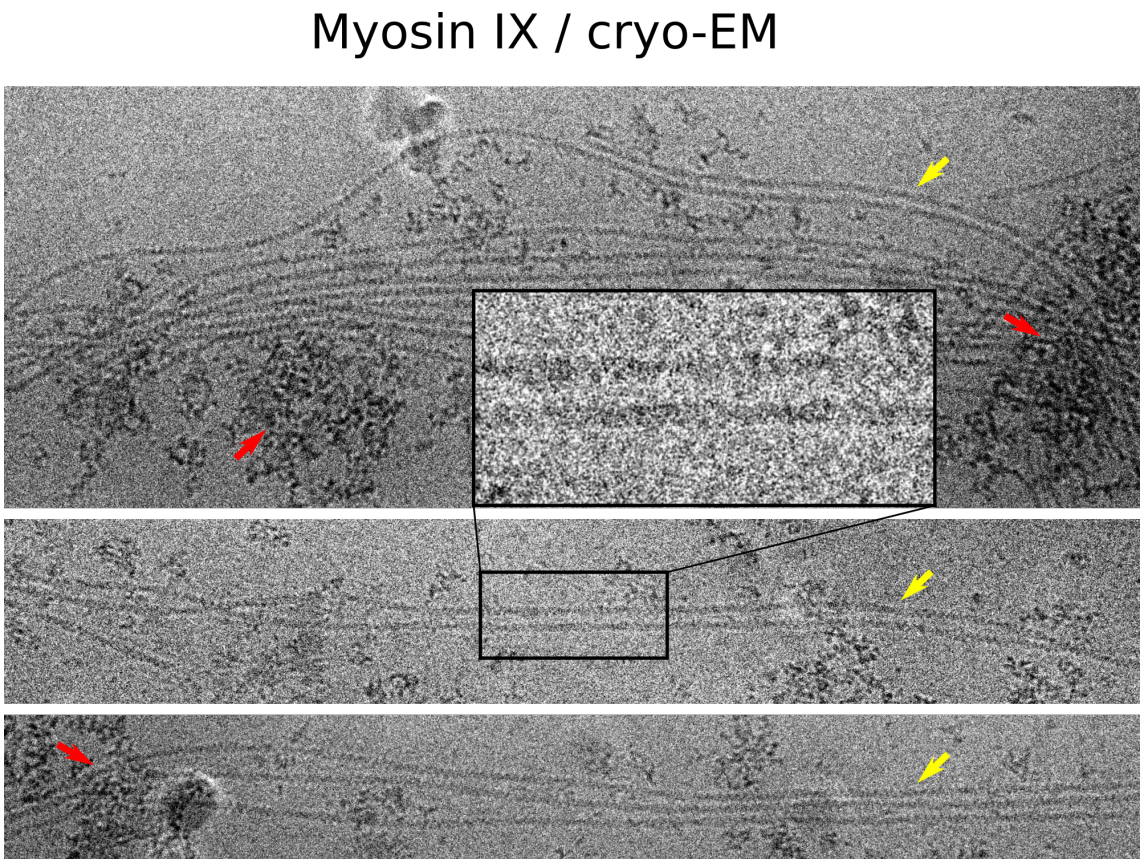


Figure 6.75: First cryo-EM images of myosin IX crosslinking actin filaments. While the actin can be seen and its arrangement suggests the presence of crosslinks, no myosin molecule could be observed.

As mentioned, using actin as a binding partner of myosin might allow for the imaging of myosin - actin - complexes in cryo-EM. Based on the results of myosin IX forming highly ordered and cooperative crosslinks with actin filaments, first images were taken with a cryo-EM (FEI Tecnai)[20], as shown in Figure 6.75. In these raw micrographs, the actin filaments (yellow arrows) are clearly visible with a surprising good SNR. Neverthe-

[20]Hendrick Dietz, Laboratory for Biomolecular Nanotechnology, Technische Universität Munich

less, while the high degree of parallelism of some actin filaments indicates the presence of crosslinks[21], no crosslinking myosin molecule could be identified in these images. The red arrows indicate myosin aggregates and confirm the general presence of myosin IX in the sample. Here, image processing will need to prove its power once more in future experiments to hopefully resolve a higher resolution of the crosslinking myosin IX. Another, likely possibility is that crosslinks were in fact established initially, but due to some effects during the freezing process they dissolved, leaving the bare actin filaments in the crosslinking pattern behind.

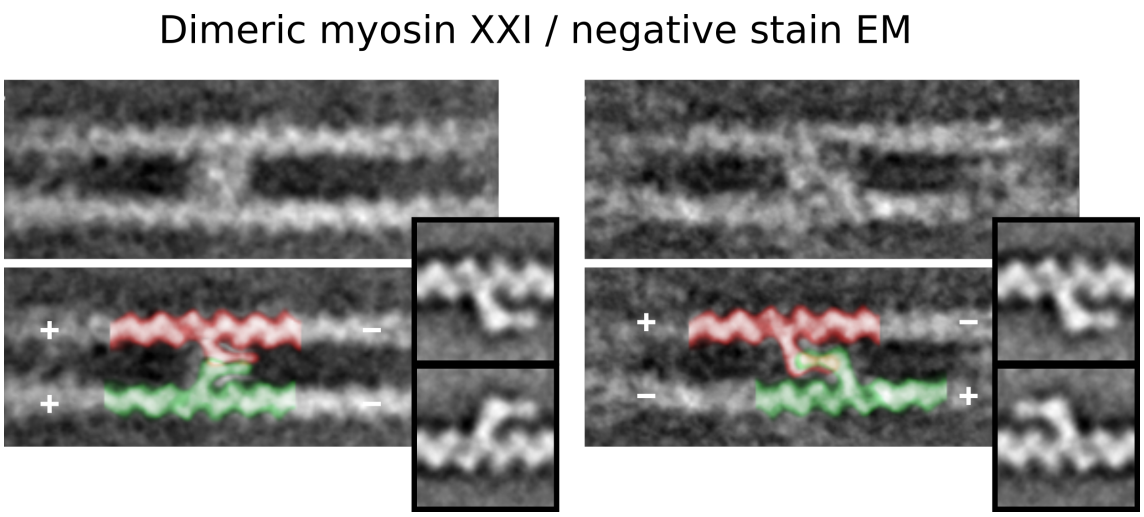


Figure 6.76: Myosin XXI can crosslink actin filaments in the form of a dimer. Imaged by negative stain EM.

It was mentioned that myosin XXI was also discovered to have crosslinking functions. Myosin XXI is the only myosin found in the Leishmania parasite and therefore required to fulfill the broad range of tasks in the cell that are usually distributed to a large number of different myosin classes. First structural experiments with this unique myosin were done by negative stain as shown in Figure 6.76. The small achieved data set did not have enough resolution to properly describe the class averages. However, the structure of the linking myosin XXI dimer (a XXI monomer would not be able crosslink filaments as it has, like most of myosin, only one actin binding site) suggests that there might be mixed populations of parallel and anti-parallel actin crosslinks. These preliminary results are promising and need further studies to investigate the organization, polarity, structure of the myosin and regulation in more detail.

Coming back to myosin IX but extending the point of view, it would be interesting to investigate myosin IX in the absence of actin in order to resolve structural information of this myosin class itself, in particular of the special insert in loop 2 (see Figure 6.77). The

[21] Actin filaments are highly unlikely to arrange in such way without crosslinkers.

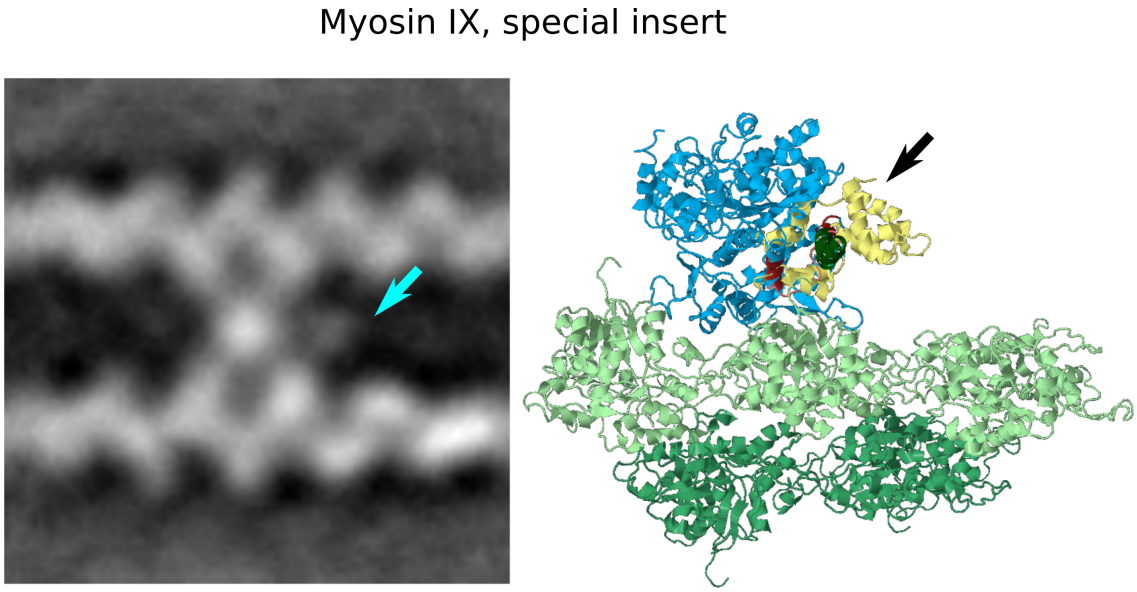


Figure 6.77: Special insert in loop 2 of myosin IXa remains to be resolved in high resolution.

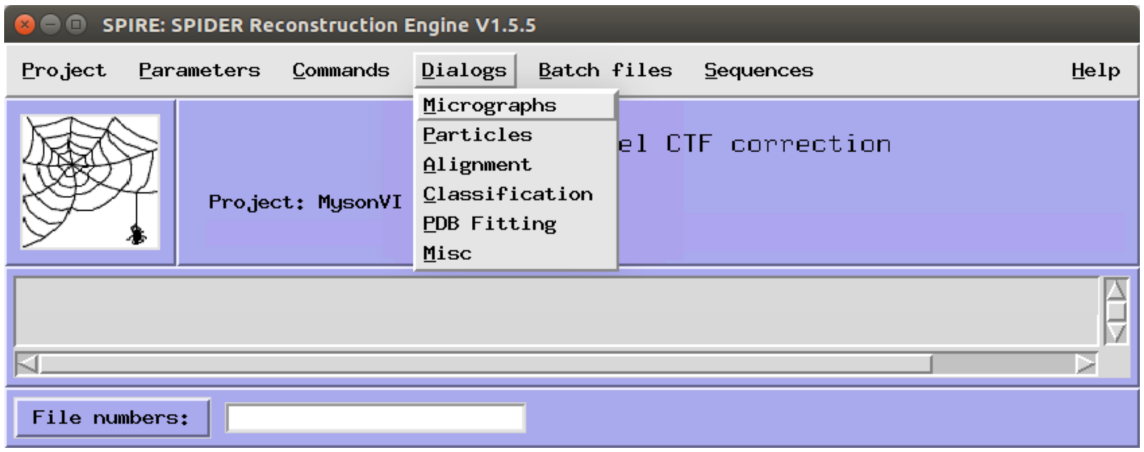
initial, promising approach will be negative stain to reveal first characteristics. Eventually, based on these results with either further technological advances (allowing to study smallest molecules in cryo-EM) or with the help of the mentioned nanodisc-complex, the investigation of single myosin IX molecules in cryo-EM might reveal intriguing insights. And finally, my colleges are working intensively on expressing myosin IX in its full length (until now only the minimal motor domain and a S1 construct were subject of our research) to study the molecule as a whole. The structural analysis using EM and SPA on FL myosin IX will of high interest. Apart from that, the insert expressed and purified on its own would be highly interesting to structurally investigate, for example by crystallography.

Part IV

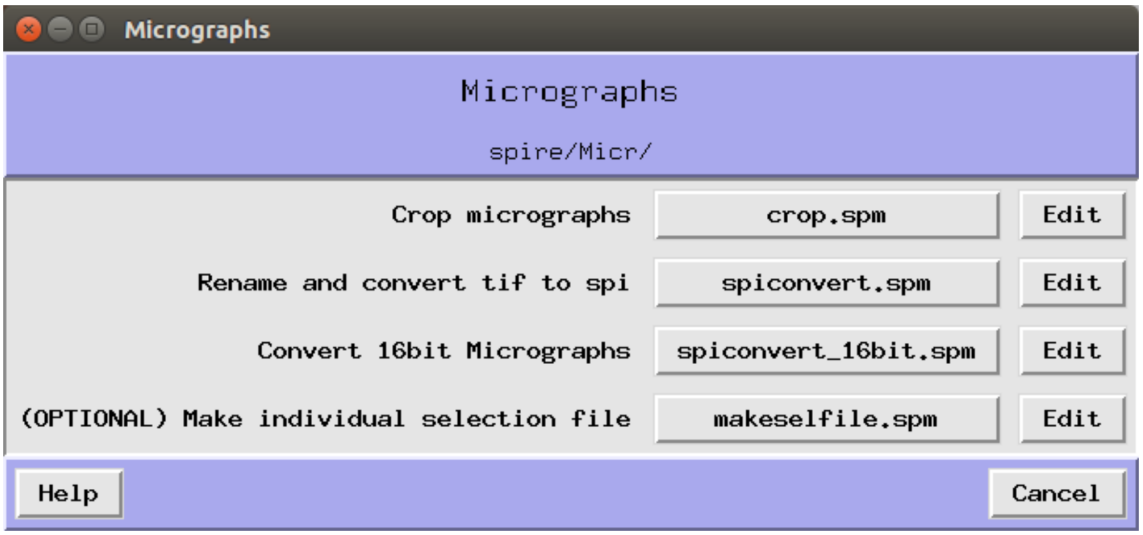
Appendix

Developed SPIRE workflow for the single particle analysis of myosin VI and myosin XI

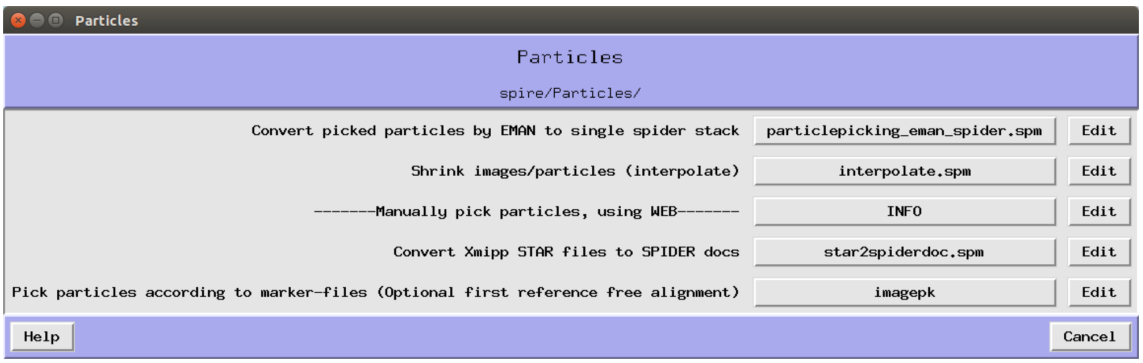
The following screen captures show the SPIRE GUI that has been used for the analysis of negative stain electron micrographs of myosin VI and myosin XI. In the main window, the workflow for single particle analysis is presented in the form of various dialogs, each of which contains the customized scripts:



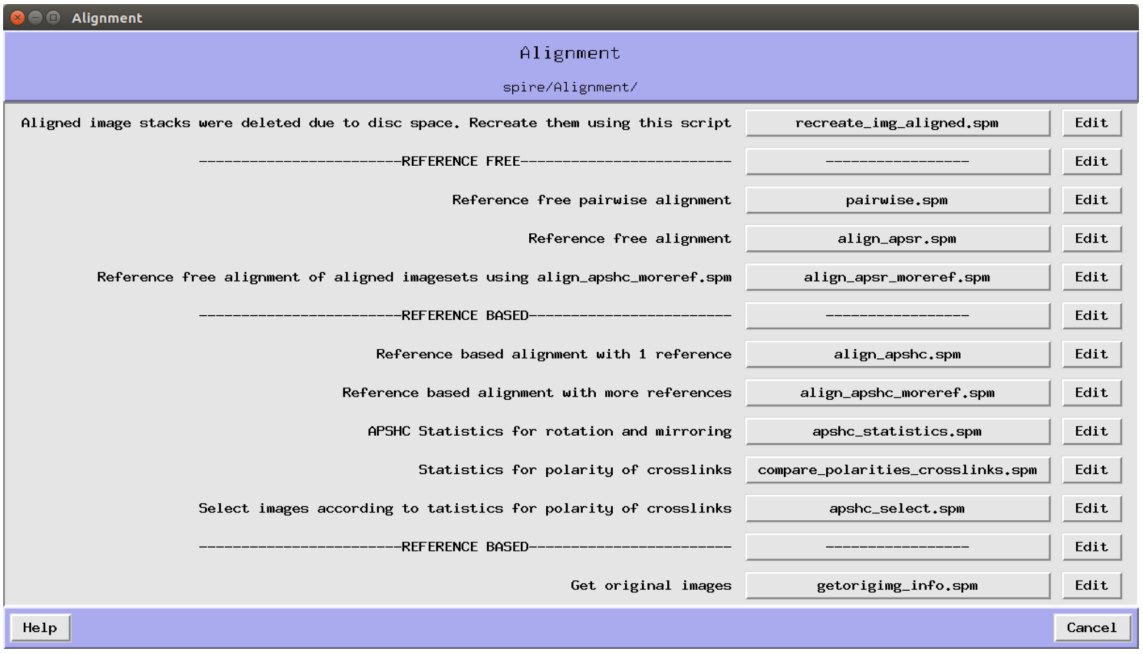
Supplementary Figure 6.1: SPIRE GUI: developed workflow for single particle analysis: Main window



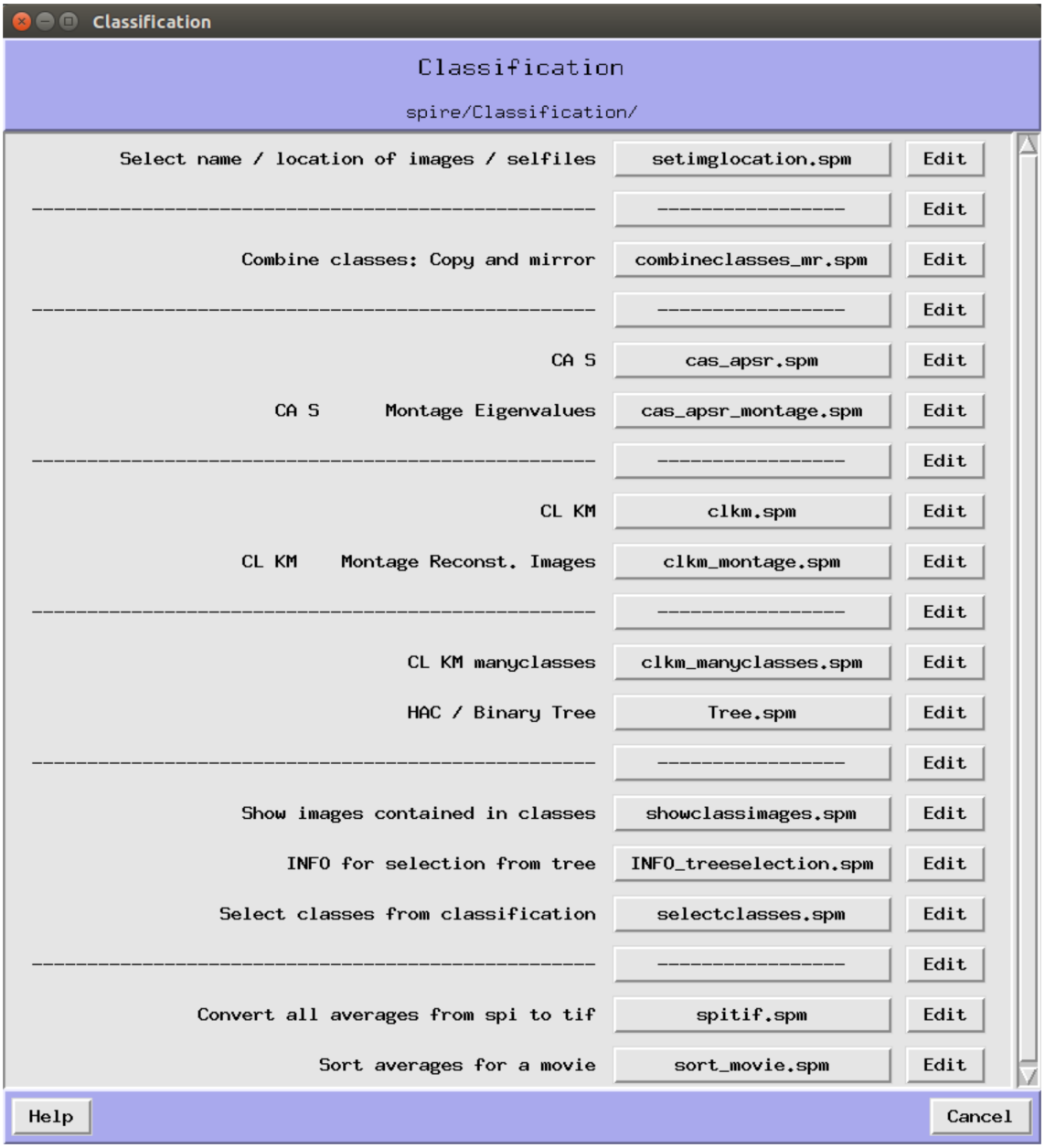
Supplementary Figure 6.2: SPIRE GUI: developed workflow for single particle analysis: ”Micrographs” dialog with scripts to preprocess the micrographs.



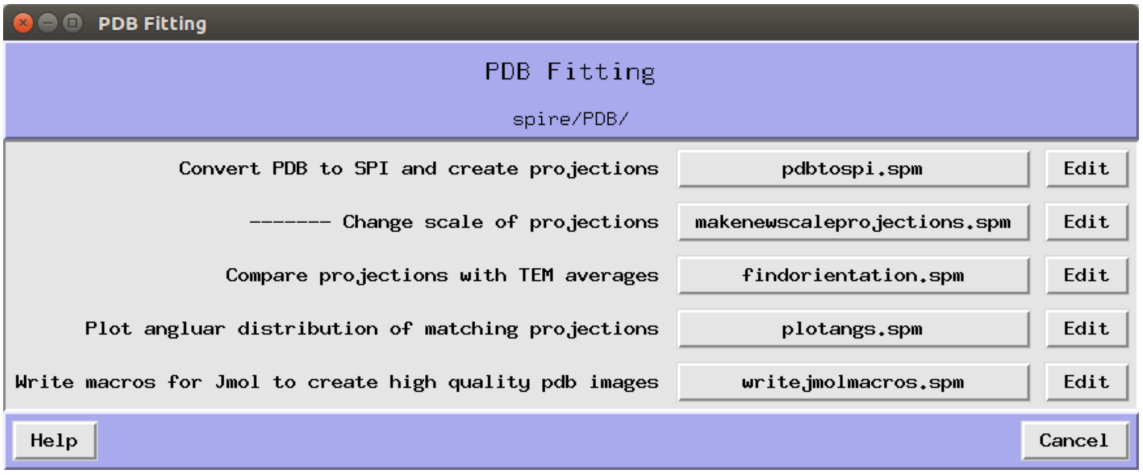
Supplementary Figure 6.3: SPIRE GUI: developed workflow for single particle analysis: ”Particles” dialog with scripts to pick and process particles.



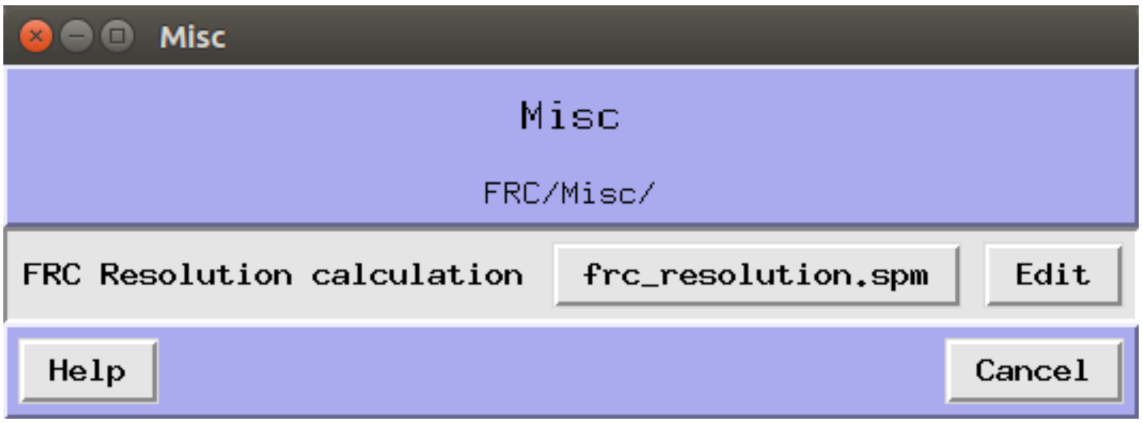
Supplementary Figure 6.4: SPIRE GUI: developed workflow for single particle analysis: ”Alignment” dialog with scripts to apply various alignment methods.



Supplementary Figure 6.5: SPIRE GUI: developed workflow for single particle analysis: "Classification" dialog with scripts to classify the (aligned) EM data.



Supplementary Figure 6.6: SPIRE GUI: developed workflow for single particle analysis: "PDB Fitting" dialog with scripts to process crystal structure models and apply crystal structure projection matching.



Supplementary Figure 6.7: SPIRE GUI: developed workflow for single particle analysis: "Miscellaneous" dialog with the script to determine the resolution of EM averages.

Bibliography

[1] G. Karp, *Molekulare Zellbiologie*. Springer, 2005.

[2] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. E. Darnell, *Molekulare Zellbiologie*. Heidelberg: Spektrum Akademischer Verlag, 4 ed., Sept. 2001.

[3] E. D. Korn, M. F. Carlier, and D. Pantaloni, "Actin polymerization and ATP hydrolysis," *Science (New York, N.Y.)*, vol. 238, pp. 638–644, Oct. 1987.

[4] K. C. Holmes, D. Popp, W. Gebhard, and W. Kabsch, "Atomic model of the actin filament," *Nature*, vol. 347, pp. 44–49, Sept. 1990.

[5] W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai, and K. C. Holmes, "Atomic structure of the actin: DNase I complex," *Nature*, vol. 347, pp. 37–44, Sept. 1990.

[6] P. B. Moore, H. E. Huxley, and D. J. DeRosier, "Three-dimensional reconstruction of F-actin, thin filaments and decorated thin filaments," *Journal of Molecular Biology*, vol. 50, pp. 279–295, June 1970.

[7] D. Woodrum, S. Rich, and T. Pollard, "Evidence for biased bidirectional polymerization of actin filaments using heavy meromyosin prepared by an improved method," *The Journal of Cell Biology*, vol. 67, pp. 231–237, Oct. 1975.

[8] J. A. Spudich and S. Watt, "The regulation of rabbit skeletal muscle contraction. I. Biochemical studies of the interaction of the tropomyosin-troponin complex with actin and the proteolytic fragments of myosin," *The Journal of Biological Chemistry*, vol. 246, pp. 4866–4871, Aug. 1971.

[9] J. D. Pardee and J. A. Spudich, "Purification of muscle actin," *Methods in Cell Biology*, vol. 24, pp. 271–289, 1982.

[10] P. Dancker, I. Löw, W. Hasselbach, and T. Wieland, "Interaction of actin with phalloidin: polymerization and stabilization of F-actin," *Biochimica Et Biophysica Acta*, vol. 400, pp. 407–414, Aug. 1975.

[11] O. Warburg and W. Christian *Biochem.*, vol. Z310, pp. 384–421, 1942.

[12] "Panreac AppliChem." https://www.applichem.com/start.html, Sept. 2017.

[13] J. A. Cooper, "Effects of cytochalasin and phalloidin on actin," *The Journal of Cell Biology*, vol. 105, pp. 1473–1478, Oct. 1987.

[14] F. Odronitz and M. Kollmar, "Drawing the tree of eukaryotic life based on the analysis of 2,269 manually annotated myosins from 328 species," *Genome Biology*, vol. 8, no. 9, p. R196, 2007.

[15] W. M. Bement, T. Hasson, J. A. Wirth, R. E. Cheney, and M. S. Mooseker, "Identification and overlapping expression of multiple unconventional myosin genes in vertebrate cell types," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, pp. 6549–6553, July 1994.

[16] H. Kuboniwa, N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax, "Solution structure of calcium-free calmodulin," *Nature Structural Biology*, vol. 2, pp. 768–776, Sept. 1995.

[17] H. Y. Park, S. A. Kim, J. Korlach, E. Rhoades, L. W. Kwok, W. R. Zipfel, M. N. Waxham, W. W. Webb, and L. Pollack, "Conformational changes of calmodulin upon Ca2+ binding studied with a microfluidic mixer," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 542–547, Jan. 2008.

[18] C. Veigel, F. Wang, M. L. Bartoo, J. R. Sellers, and J. E. Molloy, "The gated gait of the processive molecular motor, myosin V," *Nature Cell Biology*, vol. 4, pp. 59–65, Jan. 2002.

[19] P.-D. Coureux, H. L. Sweeney, and A. Houdusse, "Three myosin V structures delineate essential features of chemo-mechanical transduction," *The EMBO journal*, vol. 23, pp. 4527–4537, Nov. 2004.

[20] D. N. Blauch, "Spectrophotometry: Basic Principles." http://www.chm.davidson.edu/vce/spectrophotometry/Spectrophotometry.html, June 2017.

[21] D. R. Caprette, "Introduction to SDS-PAGE," Sept. 2017.

[22] F. Tore and E. Gösta, "Nobel Lectures, Physics 1981-1990," 1993.

[23] E. Ruska, *The Early Development of Electron Lenses and Electron Microscopy*. Hirzel, 1980.

[24] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state.* Oxford: Oxford University Press, Mar. 2006.

[25] Authoradmin, "Ernst Ruska and Max Knoll: the origin of electron microscopy," May 2015.

[26] D. T. Caceci, "Appendix: Electron Microscopes." http://www.doctorc.net/Labs/Lab2/Appendix/ems.htm, Sept. 2017.

[27] P. A. Penczek, "Resolution measures in molecular electron microscopy," *Methods in Enzymology*, vol. 482, pp. 73–100, 2010.

[28] D. B. Williams, *Transmission Electron Microscopy - A Textbook for Materials Science.* Springer, 2004.

[29] J. Rodenburg, "Learn to use TEM." http://www.rodenburg.org/guide/t300.html, June 2017.

[30] E. V. Orlova and H. R. Saibil, "Structural Analysis of Macromolecular Assemblies by Electron Microscopy," *Chemical Reviews*, vol. 111, pp. 7710–7748, Dec. 2011.

[31] J. Gu and P. E. Bourne, *Structural Bioinformatics*, vol. 143:170. Hoboken, NJ: John Wiley & Sons, 2 ed., Apr. 2009.

[32] P. R. Hazelton and H. R. Gelderblom, "Electron Microscopy for Rapid Diagnosis of Emerging Infectious Agents," *Emerging Infectious Diseases*, vol. 9, pp. 294–303, Mar. 2003.

[33] E. Kellenberger and J. Kistler, "The Physics of Specimen Preparation," in *Unconventional Electron Microscopy for Molecular Structure Determination* (W. Hoppe and R. Mason, eds.), Advances in Structure Research by Diffraction Methods / Fortschritte der Strukturforschung mit Beugungsmethoden, pp. 49–79, Vieweg+Teubner Verlag, 1979.

[34] S. A. Burgess, M. L. Walker, K. Thirumurugan, J. Trinick, and P. J. Knight, "Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules," *Journal of Structural Biology*, vol. 147, pp. 247–258, Sept. 2004.

[35] P. N. Unwin, "Electron microscopy of the stacked disk aggregate of tobacco mosaic virus protein. II. The influence of electron irradiation of the stain distribution," *Journal of Molecular Biology*, vol. 87, pp. 657–670, Aug. 1974.

[36] M. Ohi, Y. Li, Y. Cheng, and T. Walz, "Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy," *Biological Procedures Online*, vol. 6, pp. 23–34, Mar. 2004.

[37] M. Adrian, J. Dubochet, J. Lepault, and A. W. McDowall, "Cryo-electron microscopy of viruses," *Nature*, vol. 308, pp. 32–36, Mar. 1984.

[38] J. L. S. Milne, M. J. Borgnia, A. Bartesaghi, E. E. H. Tran, L. A. Earl, D. M. Schauder, J. Lengyel, J. Pierson, A. Patwardhan, and S. Subramaniam, "Cryo-electron microscopy–a primer for the non-microscopist," *The FEBS journal*, vol. 280, pp. 28–45, Jan. 2013.

[39] S. Wu, A. Avila-Sakar, J. Kim, D. S. Booth, C. H. Greenberg, A. Rossi, M. Liao, X. Li, A. Alian, S. L. Griner, N. Juge, Y. Yu, C. M. Mergel, J. Chaparro-Riggers, P. Strop, R. Tampé, R. H. Edwards, R. M. Stroud, C. S. Craik, and Y. Cheng, "Fabs Enable Single Particle cryoEM Studies of Small Proteins," *Structure*, vol. 20, pp. 582–592, Apr. 2012.

[40] O. L. Krivanek and P. E. Mooney, "Applications of slow-scan CCD cameras in transmission electron microscopy," *Ultramicroscopy*, vol. 49, pp. 95–108, Feb. 1993.

[41] J. Frank, A. Verschoor, and M. Boublik, "Computer averaging of electron micrographs of 40s ribosomal subunits," *Science (New York, N.Y.)*, vol. 214, pp. 1353–1355, Dec. 1981.

[42] W. O. Saxton and J. Frank, "Motif detection in quantum noise-limited electron micrographs by cross-correlation," *Ultramicroscopy*, vol. 2, pp. 219–227, Jan. 1976.

[43] M. van Heel and J. Frank, "Use of multivariate statistics in analysing the images of biological macromolecules," *Ultramicroscopy*, vol. 6, no. 2, pp. 187–194, 1981.

[44] J. Frank, B. Shimkin, and H. Dowse, "Spider—A modular software system for electron image processing," *Ultramicroscopy*, vol. 6, pp. 343–357, Jan. 1981.

[45] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "EMAN2: an extensible image processing suite for electron microscopy," *Journal of Structural Biology*, vol. 157, pp. 38–46, Jan. 2007.

[46] P. Penczek, M. Radermacher, and J. Frank, "Three-dimensional reconstruction of single particles embedded in ice," *Ultramicroscopy*, vol. 40, pp. 33–53, Jan. 1992.

[47] L. Lebart and etc, *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. New York: John Wiley & Sons Inc, May 1984.

[48] J. H. J. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol. 58, pp. 236–244, Mar. 1963.

[49] W. O. Saxton and W. Baumeister, "The correlation averaging of a regularly arranged bacterial cell envelope protein," *Journal of Microscopy*, vol. 127, pp. 127–138, Aug. 1982.

[50] S. J. Ludtke, P. R. Baldwin, and W. Chiu, "EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions," *Journal of Structural Biology*, vol. 128, pp. 82–97, Dec. 1999.

[51] S. H. W. Scheres, "A Bayesian View on Cryo-EM Structure Determination," *Journal of Molecular Biology*, vol. 415, pp. 406–418, Jan. 2012.

[52] J. M. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. Vargas, J. M. Carazo, and C. O. S. Sorzano, "Xmipp 3.0: An improved software suite for image processing in electron microscopy," *Journal of Structural Biology*, vol. 184, pp. 321–328, Nov. 2013.

[53] J. M. de la Rosa-Trevín, A. Quintana, L. Del Cano, A. Zaldívar, I. Foche, J. Gutiérrez, J. Gómez-Blanco, J. Burguet-Castell, J. Cuenca-Alba, V. Abrishami, J. Vargas, J. Otón, G. Sharov, J. L. Vilas, J. Navas, P. Conesa, M. Kazemi, R. Marabini, C. O. S. Sorzano, and J. M. Carazo, "Scipion: A software framework toward integration, reproducibility and validation in 3d electron microscopy," *Journal of Structural Biology*, vol. 195, pp. 93–99, July 2016.

[54] W. T. Baxter, A. Leith, and J. Frank, "SPIRE: The SPIDER Reconstruction Engine," *Journal of Structural Biology*, vol. 157, pp. 56–63, Jan. 2007.

[55] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, Jan. 2000.

[56] Jmol, "Jmol: an open-source Java viewer for chemical structures in 3d." http://www.jmol.org/, June 2017.

[57] N. Guex and M. C. Peitsch, "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling," *Electrophoresis*, vol. 18, pp. 2714–2723, Dec. 1997.

[58] Y. Ye and A. Godzik, "Flexible structure alignment by chaining aligned fragment pairs allowing twists," *Bioinformatics (Oxford, England)*, vol. 19 Suppl 2, pp. ii246–255, Oct. 2003.

[59] P. A. Penczek, R. A. Grassucci, and J. Frank, "The ribosome at improved resolution: new techniques for merging and orientation refinement in 3d cryo-electron microscopy of biological particles," *Ultramicroscopy*, vol. 53, pp. 251–270, Mar. 1994.

[60] A. Leith, W. Baxter, and J. Frank, "Use of SPIDER and SPIRE in image reconstruction," *International Tables for Crystallography*, vol. F, pp. 620–623, Apr. 2012.

[61] A. L. Wells, A. W. Lin, L.-Q. Chen, D. Safer, S. M. Cain, T. Hasson, B. O. Carragher, R. A. Milligan, and H. L. Sweeney, "Myosin VI is an actin-based motor that moves backwards," *Nature*, vol. 401, pp. 505–508, Sept. 1999.

[62] K. B. Avraham, T. Hasson, K. P. Steel, D. M. Kingsley, L. B. Russell, M. S. Mooseker, N. G. Copeland, and N. A. Jenkins, "The mouse Snell's waltzer deafness gene encodes an unconventional myosin required for structural integrity of inner ear hair cells," *Nature Genetics*, vol. 11, pp. 369–375, Dec. 1995.

[63] O. Pylypenko, L. Song, A. Shima, Z. Yang, A. M. Houdusse, and H. L. Sweeney, "Myosin VI deafness mutation prevents the initiation of processive runs on actin," *Proceedings of the National Academy of Sciences*, vol. 112, pp. E1201–E1209, Mar. 2015.

[64] C. Yu, J. Lou, J. Wu, L. Pan, W. Feng, and M. Zhang, "Membrane-induced Lever Arm Expansion Allows Myosin VI to Walk with Large and Variable Step Sizes," *Journal of Biological Chemistry*, vol. 287, pp. 35021–35035, Oct. 2012.

[65] F. Buss and J. Kendrick-Jones, "Multifunctional myosin VI has a multitude of cargoes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 5927–5928, Apr. 2011.

[66] A. A. Olama, Al, "A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer," *Nature Genetics*, vol. 46, pp. 1103–1109, Oct. 2014.

[67] T. A. Dunn, S. Chen, D. A. Faith, J. L. Hicks, E. A. Platz, Y. Chen, C. M. Ewing, J. Sauvageot, W. B. Isaacs, A. M. De Marzo, and J. Luo, "A novel role of myosin VI in human prostate cancer," *The American Journal of Pathology*, vol. 169, pp. 1843–1854, Nov. 2006.

[68] H. Naora and D. J. Montell, "Ovarian cancer metastasis: integrating insights from disparate model organisms," *Nature Reviews. Cancer*, vol. 5, pp. 355–366, May 2005.

[69] H. Yoshida, W. Cheng, J. Hung, D. Montell, E. Geisbrecht, D. Rosen, J. Liu, and H. Naora, "Lessons from border cell migration in the Drosophila ovary: A role for myosin VI in dissemination of human ovarian cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 8144–8149, May 2004.

[70] C. Wei, X. Wang, M. Chen, K. Ouyang, L.-S. Song, and H. Cheng, "Calcium flickers steer cell migration," *Nature*, vol. 457, pp. 901–905, Feb. 2009.

[71] A. J. Ridley, M. A. Schwartz, K. Burridge, R. A. Firtel, M. H. Ginsberg, G. Borisy, J. T. Parsons, and A. R. Horwitz, "Cell migration: integrating signals from front to back," *Science (New York, N.Y.)*, vol. 302, pp. 1704–1709, Dec. 2003.

[72] A. Malgaroli, D. Milani, J. Meldolesi, and T. Pozzan, "Fura-2 measurement of cytosolic free Ca2+ in monolayers and suspensions of various types of animal cells," *The Journal of Cell Biology*, vol. 105, pp. 2145–2155, Nov. 1987.

[73] A. Miyawaki, O. Griesbeck, R. Heim, and R. Y. Tsien, "Dynamic and quantitative Ca2+ measurements using improved cameleons," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, pp. 2135–2140, Mar. 1999.

[74] D. A. Tumbarello, J. Kendrick-Jones, and F. Buss, "Myosin VI and its cargo adaptors - linking endocytosis and autophagy," *Journal of Cell Science*, vol. 126, pp. 2561–2570, June 2013.

[75] M. A. Geeves, M. R. Webb, C. F. Midelfort, and D. R. Trentham, "Mechanism of adenosine 5'-triphosphate cleavage by myosin: studies with oxygen-18-labeled adenosine 5'-triphosphate," *Biochemistry*, vol. 19, pp. 4748–4754, Oct. 1980.

[76] D. A. Sahlender, R. C. Roberts, S. D. Arden, G. Spudich, M. J. Taylor, J. P. Luzio, J. Kendrick-Jones, and F. Buss, "Optineurin links myosin VI to the Golgi complex and is involved in Golgi organization and exocytosis," *The Journal of Cell Biology*, vol. 169, pp. 285–295, Apr. 2005.

[77] L. M. Bond, A. A. Peden, J. Kendrick-Jones, J. R. Sellers, and F. Buss, "Myosin VI and its binding partner optineurin are involved in secretory vesicle fusion at the plasma membrane," *Molecular Biology of the Cell*, vol. 22, pp. 54–65, Jan. 2011.

[78] J. Ménétrey, A. Bahloul, A. Wells, C. Yengo, C. Morris, H. L. Sweeney, and A. Houdusse, "The structure of the myosin VI motor reveals the mechanism of directionality reversal," *Nature*, vol. 435, pp. 779–785, June 2005.

[79] T. Hasson and M. S. Mooseker, "Porcine myosin-VI: characterization of a new mammalian unconventional myosin.," *The Journal of Cell Biology*, vol. 127, pp. 425–440, Oct. 1994.

[80] A. Bahloul, G. Chevreux, A. L. Wells, D. Martin, J. Nolt, Z. Yang, L.-Q. Chen, N. Potier, A. Van Dorsselaer, S. Rosenfeld, A. Houdusse, and H. L. Sweeney, "The unique insert in myosin VI is a structural calcium-calmodulin binding site," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 4787–4792, Apr. 2004.

[81] M. Mukherjea, P. Llinas, H. Kim, M. Travaglia, D. Safer, J. Ménétrey, C. Franzini-Armstrong, P. R. Selvin, A. Houdusse, and H. L. Sweeney, "Myosin VI dimerization triggers an unfolding of a three-helix bundle in order to extend its reach," *Molecular Cell*, vol. 35, pp. 305–315, Aug. 2009.

[82] M. Peckham, "Coiled coils and SAH domains in cytoskeletal molecular motors," *Biochemical Society Transactions*, vol. 39, pp. 1142–1148, Oct. 2011.

[83] C. Batters, D. Brack, H. Ellrich, B. Averbeck, and C. Veigel, "Calcium can mobilize and activate myosin-VI," *Proceedings of the National Academy of Sciences*, vol. 113, pp. E1162–E1169, Jan. 2016.

[84] R. S. Rock, S. E. Rice, A. L. Wells, T. J. Purcell, J. A. Spudich, and H. L. Sweeney, "Myosin VI is a processive motor with a large step size," *Proceedings of the National Academy of Sciences*, vol. 98, pp. 13655–13659, Nov. 2001.

[85] H. L. Sweeney and A. Houdusse, "What can myosin VI do in cells?," *Current Opinion in Cell Biology*, vol. 19, pp. 57–66, Feb. 2007.

[86] M. Y. Ali, G. G. Kennedy, D. Safer, K. M. Trybus, H. L. Sweeney, and D. M. Warshaw, "Myosin Va and myosin VI coordinate their steps while engaged in an in vitro tug of war during cargo transport," *Proceedings of the National Academy of Sciences*, vol. 108, pp. E535–E541, Aug. 2011.

[87] L. M. Bond, S. D. Arden, J. Kendrick-Jones, F. Buss, and J. R. Sellers, "Dynamic exchange of myosin VI on endocytic structures," *The Journal of Biological Chemistry*, vol. 287, pp. 38637–38646, Nov. 2012.

[88] D. Phichith, M. Travaglia, Z. Yang, X. Liu, A. B. Zong, D. Safer, and H. L. Sweeney, "Cargo binding induces dimerization of myosin VI," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 17320–17324, Oct. 2009.

[89] I. Lister, S. Schmitz, M. Walker, J. Trinick, F. Buss, C. Veigel, and J. Kendrick-Jones, "A monomeric myosin VI with a large working stroke," *The EMBO journal*, vol. 23, pp. 1729–1738, Apr. 2004.

[90] B. J. Spink, S. Sivaramakrishnan, J. Lipfert, S. Doniach, and J. A. Spudich, "Long single alpha-helical tail domains bridge the gap between structure and function of myosin VI," *Nat Struct Mol Biol*, vol. 15, pp. 591–597, May 2008.

[91] C. A. Morris, A. L. Wells, Z. Yang, L.-Q. Chen, C. V. Baldacchino, and H. L. Sweeney, "Calcium functionally uncouples the heads of myosin VI," *The Journal of Biological Chemistry*, vol. 278, pp. 23324–23330, June 2003.

[92] M. Yoshimura, K. Homma, J. Saito, A. Inoue, R. Ikebe, and M. Ikebe, "Dual Regulation of Mammalian Myosin VI Motor Function," *Journal of Biological Chemistry*, vol. 276, pp. 39600–39607, Oct. 2001.

[93] C. F. Song, K. Sader, H. White, J. Kendrick-Jones, and J. Trinick, "Nucleotide-Dependent Shape Changes in the Reverse Direction Motor, Myosin VI," *Biophysical Journal*, vol. 99, pp. 3336–3344, Nov. 2010.

[94] M. Terrak, G. Rebowski, R. C. Lu, Z. Grabarek, and R. Dominguez, "Structure of the light chain-binding domain of myosin V," *Proceedings of the National Academy of Sciences*, vol. 102, pp. 12718–12723, Sept. 2005.

[95] I. Rayment, W. R. Rypniewski, K. Schmidt-Bäse, R. Smith, D. R. Tomchick, M. M. Benning, D. A. Winkelmann, G. Wesenberg, and H. M. Holden, "Three-dimensional structure of myosin subfragment-1: a molecular motor," *Science (New York, N.Y.)*, vol. 261, pp. 50–58, July 1993.

[96] M. L. Walker, S. A. Burgess, J. R. Sellers, F. Wang, J. A. Hammer, J. Trinick, and P. J. Knight, "Two-headed binding of a processive myosin to F-actin," *Nature*, vol. 405, pp. 804–807, June 2000.

[97] S. A. Burgess, M. L. Walker, H. Sakakibara, P. J. Knight, and K. Oiwa, "Dynein structure and power stroke," *Nature*, vol. 421, pp. 715–718, Feb. 2003.

[98] M. Khoshouei, M. Radjainia, W. Baumeister, and R. Danev, "Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate," *Nature Communications*, vol. 8, June 2017.

[99] M. Krendel and M. S. Mooseker, "Myosins: tails (and heads) of functional diversity," *Physiology (Bethesda, Md.)*, vol. 20, pp. 239–251, Aug. 2005.

[100] J. Ménétrey, T. Isabet, V. Ropars, M. Mukherjea, O. Pylypenko, X. Liu, J. Perez, P. Vachette, H. L. Sweeney, and A. M. Houdusse, "Processive steps in the reverse direction require uncoupling of the lead head lever arm of myosin VI," *Molecular Cell*, vol. 48, pp. 75–86, Oct. 2012.

[101] J. Schindelin, C. T. Rueden, M. C. Hiner, and K. W. Eliceiri, "The ImageJ ecosystem: An open platform for biomedical image analysis," *Molecular Reproduction and Development*, vol. 82, pp. 518–529, July 2015.

[102] K. Elfrink, W. Liao, U. Pieper, S. J. Oeding, and M. Bähler, "The Loop2 Insertion of Type IX Myosin Acts as an Electrostatic Actin Tether that Permits Processive Movement," *PLOS ONE*, vol. 9, p. e84874, Jan. 2014.

[103] S. Struchholz, K. Elfrink, U. Pieper, G. Kalhammer, U. Honnert, A. Grützner, W. A. Linke, W. Liao, and M. Bähler, "Functional Role of the Extended Loop 2 in the Myosin 9b Head for Binding F-actin," *Journal of Biological Chemistry*, vol. 284, pp. 3663–3671, Feb. 2009.

[104] A. Inoue, J. Saito, R. Ikebe, and M. Ikebe, "Myosin IXb is a single-headed minus-end-directed processive motor," *Nature Cell Biology*, vol. 4, pp. 302–306, Apr. 2002.

[105] P. L. Post, M. J. Tyska, C. B. O'Connell, K. Johung, A. Hayward, and M. S. Mooseker, "Myosin-IXb is a single-headed and processive motor," *The Journal of Biological Chemistry*, vol. 277, pp. 11679–11683, Apr. 2002.

[106] W. Liao, K. Elfrink, and M. Bähler, "Head of Myosin IX Binds Calmodulin and Moves Processively toward the Plus-end of Actin Filaments," *The Journal of Biological Chemistry*, vol. 285, pp. 24933–24942, Aug. 2010.

[107] T. Kambara and M. Ikebe, "A unique ATP hydrolysis mechanism of single-headed processive myosin, myosin IX," *The Journal of Biological Chemistry*, vol. 281, pp. 4949–4957, Feb. 2006.

[108] T. Omelchenko and A. Hall, "Myosin-IXA regulates collective epithelial cell migration by targeting RhoGAP activity to cell-cell junctions," *Current biology: CB*, vol. 22, pp. 278–288, Feb. 2012.

[109] J. M. Wood and M. F. Olson, "Collective migration: spatial tension relief," *Current biology: CB*, vol. 22, pp. R125–127, Feb. 2012.

[110] D. Saczko-Brack, E. Warchol, B. Rogez, M. Kröss, S. M. Heissler, J. R. Sellers, C. Batters, and C. Veigel, "Self-organization of actin networks by a monomeric myosin,"

Proceedings of the National Academy of Sciences, vol. 113, pp. E8387–E8395, Dec. 2016.

[111] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. Garland Science, 4th ed., 2002.

[112] "Actin filaments support a variety of structures in a cell. | Learn Science at Scitable." https://www.nature.com/scitable/content/actin-filaments-support-a-variety-of-structures-14706332, Sept. 2017.

[113] R. Nambiar, R. E. McConnell, and M. J. Tyska, "Myosin motor function: the ins and outs of actin-based membrane protrusions," *Cellular and molecular life sciences: CMLS*, vol. 67, pp. 1239–1254, Apr. 2010.

[114] F. Buss and J. Kendrick-Jones, "How are the cellular functions of myosin VI regulated within the cell?," *Biochemical and Biophysical Research Communications*, vol. 369, pp. 165–175, Apr. 2008.

[115] C. Le Clainche and M.-F. Carlier, "Regulation of actin assembly associated with protrusion and adhesion in cell migration," *Physiological Reviews*, vol. 88, pp. 489–513, Apr. 2008.

[116] L. Blanchoin, R. Boujemaa-Paterski, C. Sykes, and J. Plastino, "Actin dynamics, architecture, and mechanics in cell motility," *Physiological Reviews*, vol. 94, pp. 235–263, Jan. 2014.

[117] "Microtubules, Filaments | Learn Science at Scitable." https://www.nature.com/scitable/topicpage/microtubules-and-filaments-14052932, Sept. 2017.

[118] M. Melani, K. J. Simpson, J. S. Brugge, and D. Montell, "Regulation of cell adhesion and collective cell migration by hindsight and its human homolog RREB1," *Current biology: CB*, vol. 18, pp. 532–537, Apr. 2008.

[119] Y. Arboleda-Estudillo, M. Krieg, J. Stühmer, N. A. Licata, D. J. Muller, and C.-P. Heisenberg, "Movement directionality in collective migration of germ layer progenitors," *Current biology: CB*, vol. 20, pp. 161–169, Jan. 2010.

[120] M. Abouhamed, K. Grobe, I. V. L. C. San, S. Thelen, U. Honnert, M. S. Balda, K. Matter, and M. Bähler, "Myosin IXa regulates epithelial differentiation and its deficiency results in hydrocephalus," *Molecular Biology of the Cell*, vol. 20, pp. 5074–5085, Dec. 2009.

[121] E. Chieregatti, A. Gärtner, H. E. Stöffler, and M. Bähler, "Myr 7 is a novel myosin IX-RhoGAP expressed in rat brain," *Journal of Cell Science*, vol. 111 (Pt 24), pp. 3597–3608, Dec. 1998.

[122] M. Lorenz and K. C. Holmes, "The actin-myosin interface," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 12529–12534, July 2010.

[123] M. A. Geeves and K. C. Holmes, "The molecular mechanism of muscle contraction," *Advances in Protein Chemistry*, vol. 71, pp. 161–193, 2005.

[124] M. Bähler, "Are class III and class IX myosins motorized signalling molecules?," *Biochimica Et Biophysica Acta*, vol. 1496, pp. 52–59, Mar. 2000.

[125] A. D. Mehta, R. S. Rock, M. Rief, J. A. Spudich, M. S. Mooseker, and R. E. Cheney, "Myosin-V is a processive actin-based motor," *Nature*, vol. 400, pp. 590–593, Aug. 1999.

[126] M. Nishikawa, S. Nishikawa, A. Inoue, A. H. Iwane, T. Yanagida, and M. Ikebe, "A unique mechanism for the processive movement of single-headed myosin-IX," *Biochemical and Biophysical Research Communications*, vol. 343, pp. 1159–1164, May 2006.

[127] V. Nalavadi, M. Nyitrai, C. Bertolini, N. Adamek, M. A. Geeves, and M. Bähler, "Kinetic Mechanism of Myosin IXB and the Contributions of Two Class IX-specific Regions," *Journal of Biological Chemistry*, vol. 280, pp. 38957–38968, Nov. 2005.

[128] A. Hall, "Rho GTPases and the actin cytoskeleton," *Science (New York, N.Y.)*, vol. 279, pp. 509–514, Jan. 1998.

[129] F. van den Boom, H. Düssmann, K. Uhlenbrock, M. Abouhamed, and M. Bähler, "The Myosin IXb motor activity targets the myosin IXb RhoGAP domain as cargo to sites of actin polymerization," *Molecular Biology of the Cell*, vol. 18, pp. 1507–1518, Apr. 2007.

[130] G. Kalhammer, M. Bähler, F. Schmitz, J. Jöckel, and C. Block, "Ras-binding domains: predicting function versus folding," *FEBS letters*, vol. 414, pp. 599–602, Sept. 1997.

[131] V. Schaller, C. Weber, C. Semmrich, E. Frey, and A. R. Bausch, "Polar patterns of driven filaments," *Nature*, vol. 467, pp. 73–77, Sept. 2010.

[132] C. A. Weber, R. Suzuki, V. Schaller, I. S. Aranson, A. R. Bausch, and E. Frey, "Random bursts determine dynamics of active filaments," *Proceedings of the National Academy of Sciences*, vol. 112, pp. 10703–10707, Aug. 2015.

[133] "Types of actin filament crosslinking proteins." https://www.mechanobio.info/figure/1384245010175/, Sept. 2017.

[134] S. Jansen, A. Collins, C. Yang, G. Rebowski, T. Svitkina, and R. Dominguez, "Mechanism of actin filament bundling by fascin," *The Journal of Biological Chemistry*, vol. 286, pp. 30087–30096, Aug. 2011.

[135] N. Volkmann, D. DeRosier, P. Matsudaira, and D. Hanein, "An Atomic Model of Actin Filaments Cross-Linked by Fimbrin and Its Implications for Bundle Assembly and Function," *The Journal of Cell Biology*, vol. 153, pp. 947–956, May 2001.

[136] C. M. Hampton, D. W. Taylor, and K. A. Taylor, "Novel Structures for alpha-Actinin:F-Actin Interactions and their Implications for Actin–Membrane Attachment and Tension Sensing in the Cytoskeleton," *Journal of Molecular Biology*, vol. 368, pp. 92–104, Apr. 2007.

[137] C. Batters, K. A. Woodall, C. P. Toseland, C. Hundschell, and C. Veigel, "Cloning, Expression, and Characterization of a Novel Molecular Motor, Leishmania Myosin-XXI," *Journal of Biological Chemistry*, vol. 287, pp. 27556–27566, June 2012.

[138] Y. Cong, M. Topf, A. Sali, P. Matsudaira, M. Dougherty, W. Chiu, and M. F. Schmid, "Crystallographic Conformers of Actin in a Biologically Active Bundle of Filaments," *Journal of Molecular Biology*, vol. 375, pp. 331–336, Jan. 2008.

[139] E. Behrmann, M. Müller, P. A. Penczek, H. G. Mannherz, D. J. Manstein, and S. Raunser, "Structure of the Rigor Actin-Tropomyosin-Myosin Complex," *Cell*, vol. 150, pp. 327–338, July 2012.

[140] P.-D. Coureux, A. L. Wells, J. Ménétrey, C. M. Yengo, C. A. Morris, H. L. Sweeney, and A. Houdusse, "A structural state of the myosin V motor without bound nucleotide," *Nature*, vol. 425, pp. 419–423, Sept. 2003.

[141] C. Batters, H. Ellrich, C. Helbig, K. A. Woodall, C. Hundschell, D. Brack, and C. Veigel, "Calmodulin regulates dimerization, motility, and lipid binding of Leishmania myosin XXI," *Proceedings of the National Academy of Sciences*, vol. 111, pp. E227–E236, Dec. 2013.

[142] C. Batters, H. Ellrich, C. Helbig, K. Woodall, C. Hundschell, D. Saczko-Brack, and C. Veigel, "Calmodulin and Lipid Binding Regulate Dimerisation and Motility of Myosin-XXI in Leishmania," *Biophysical Journal*, vol. 106, p. 180a, Jan. 2014.

[143] D. Saczko-Brack, H. Ellrich, C. Werner, C. Batters, and C. Veigel, "Regulation of Myosin VI Studied by Electron Microscopy," *Biophysical Journal*, vol. 108, p. 618a, Jan. 2015.

List of publications

- C. Batters, H. Ellrich, C. Helbig, K. A. Woodall, C. Hundschell, D. Brack, and C. Veigel, "Calmodulin regulates dimerization, motility, and lipid binding of Leishmania myosin XXI," *Proceedings of the National Academy of Sciences*, vol. 111, pp. E227–E236, Dec. 2013

- C. Batters, D. Brack, H. Ellrich, B. Averbeck, and C. Veigel, "Calcium can mobilize and activate myosin-VI," *Proceedings of the National Academy of Sciences*, vol. 113, pp. E1162–E1169, Jan. 2016

- D. Saczko-Brack, E. Warchol, B. Rogez, M. Kröss, S. M. Heissler, J. R. Sellers, C. Batters, and C. Veigel, "Self-organization of actin networks by a monomeric myosin," *Proceedings of the National Academy of Sciences*, vol. 113, pp. E8387–E8395, Dec. 2016

Published abstracts

- C. Batters, H. Ellrich, C. Helbig, K. Woodall, C. Hundschell, D. Saczko-Brack, and C. Veigel, "Calmodulin and Lipid Binding Regulate Dimerisation and Motility of Myosin-XXI in Leishmania," *Biophysical Journal*, vol. 106, p. 180a, Jan. 2014

- D. Saczko-Brack, H. Ellrich, C. Werner, C. Batters, and C. Veigel, "Regulation of Myosin VI Studied by Electron Microscopy," *Biophysical Journal*, vol. 108, p. 618a, Jan. 2015

Scientific activity

- 2017 Talk DPG-Frühjahrstagung Dresden, Germany
- 2016 Organization CeNS Workshop: SPIDER & Scipion Workshop (Image Processing) with Dr. Tanvir Shaikh, Jose Miguel de la Rosa, Dr. Daniel Wilson, Prof. Rasmus Schröder and Prof. Jürgen Plitzko
- 2016 Poster CeNS Conference Venice
- 2016 Talk CeNS Science Rocks!
- 2016 Poster 15th Alpbach Meeting, Austria
- 2015 Talk SFB863 Conference Ringberg, Germany
- 2015 Poster BPS 59th Annual Meeting, Baltimore
- 2014 Talk SFB863 Conference Hohenkammern, Germany
- 2014 Poster SFB863 Conference Hohenkammern, Germany
- 2013 Organisation CeNS Workshop: SPIDER (Image Processing) with Dr. Tanvir Shaikh & Jose Miguel de la Rosa
- 2013 Poster CeNS Conference Venice, Italy

List of Figures

1.1 Introduction: Structure of F-actin 7
1.2 Introduction: Spectrum of the absorbance measurement of G-actin in two different dilution 9
1.3 Introduction: SDS PAGE gel of three samples taken during the G-actin preparation process 10
1.4 Introduction: Phylogenetic tree of myosin classes 11
1.5 Introduction: Structure of myosin molecules 12
1.6 Introduction: Crystal structure of calcium-free calmodulin 12
1.7 Introduction: Structural composition of the myosin motor domain 14

2.1 Introduction: First electron based microscope built by Ruska and Knoll . 18
2.2 Introduction: First practical electron microscope developed by Dr. James Hiller 18
2.3 Introduction: Scheme and real setup of a TEM 21
2.4 Introduction: Condenser lenses C1 and C2 23
2.5 Introduction: Effect of the second condenser lens C2 23
2.6 Introduction: Single grid holder for the TEM 25
2.7 Introduction: Origins of lens aberrations 26
2.8 Introduction: Undamped and damped contrast transfer functions (CTF) . 28
2.9 Introduction: Contrast transfer functions (CTF) for different defocus values 29
2.10 Introduction: Representation of the CTF by diffraction pattern 30
2.11 Introduction: EM micrographs of nucleotide-free FL myosin XXI at different foci 31
2.12 Introduction: Comparison of positive and negative staining 32
2.13 Introduction: Stain migration into the gap of the tweezer and under the TEM grid 33
2.14 Introduction: EM micrographs of myosin molecules with different stain depths 33
2.15 Introduction: EM micrographs of actin filaments with different stains . . 35

3.1 Introduction: Demonstration of the cross correlation between two images 39
3.2 Introduction: Reference free vs. reference based alignment 41
3.3 Introduction: Demonstration of variance images 43
3.4 Introduction: Demonstration of correspondence analysis with a synthetic image data set 47
3.5 Introduction: Histogram showing the percentage of interimage variances of the synthetic image data set 48
3.6 Introduction: Factormap of the synthetic image data set for the first two eigenvectors 49
3.7 Introduction: Factormap of the synthetic image data set for the third and fourth eigenvector 49
3.8 Introduction: Reconstruction of the noisy image data set by a linear combination of the first three eigenvectors 50
3.9 Introduction: Dendogram from the classification of the synthetic image data set 51
3.10 Introduction: Binary tree showing classes produced from a dendogram at different cut off values 52
3.11 Introduction: Illustration of the classification by K-means 53
3.12 Introduction: Classification of the synthetic image data set by K-means . 53
3.13 Introduction: Principle of resolution determination of EM reconstructions using FRC 55
3.14 Introduction: Terminal based SPIDER software 57
3.15 Introduction: SPIRE, the GUI for SPIDER: Dialogs 58
3.16 Introduction: SPIRE, the GUI for SPIDER: Dialogs 58
3.17 Introduction: SPIRE, the GUI for SPIDER: Dialogs 59

4.1 Introduction: Crystal structure projection matching, Step 1 63
4.2 Introduction: Crystal structure projection matching, Step 2 65

5.1 Myosin VI: Cellular functions of myosin VI 71
5.2 Myosin VI: Sequence of FL myosin VI 72
5.3 Myosin VI: Crystal structures of FL myosin VI and V 72
5.4 Myosin VI: number of recorded, picked and used images of myosin VI in four different conditions 75
5.5 Myosin VI: Preferred orientations of full-length myosin VI on the EM grid 76
5.6 Myosin VI: Crystal structure projection matching of the bent conformation of myosin VI using a nucleotide-free myosin VI structure 77

5.7 Myosin VI: Crystal structure projection matching of the straight conformation of myosin VI using a nucleotide-free myosin VI structure 78
5.8 Myosin VI: Negative stain EM of FL myosin VI by Song et. al 79
5.9 Myosin VI: Crystal structure projection matching of the straight conformation of myosin VI using an ADP model of myosin VI 80
5.10 Myosin VI: Analysis of the lever arm of the straight conformation: EM . 82
5.11 Myosin VI: Analysis of the lever arm of the straight conformation: Interpretation 83
5.12 Myosin VI: Eight discovered conformations of nucleotide-free myosin VI 84
5.13 Myosin VI: Statistics on the calcium and calmodulin dependency of the shape of myosin VI 85
5.14 Myosin VI: Piechart: Statistics on the calcium and calmodulin dependency of the shape of myosin VI 85
5.15 Myosin VI: Piechart: Statistics on the shapes found in all four calcium / CaM conditions conditions, reduced to two parameter 87
5.16 Myosin VI: Maximum resolution class averages of molecules in the bent conformation at low and high calcium 89
5.17 Myosin VI: Cross correlation comparison of the myosin VI bent conformation at low and high calcium 90
5.18 Myosin VI: Intensity profile plots of myosin VI at low and hight calcium 92
5.19 Myosin VI: Superposition of crystal structure model and EM average . . 93
5.20 Myosin VI: Binding studies reveal bipartite binding site of myosin VI . . 95
5.21 Myosin VI: Lever arm stiffness of myosin VI investigated using EM-movies 97
5.22 Myosin VI: Summary of the single particle analysis on FL myosin VI . . . 101
5.23 Myosin VI: Possible calcium-induced unfolding of the 3HB 104
5.24 Myosin VI: Intensity profile plot of the 3HB at low and high-calcium . . 104
5.25 Myosin VI: Analysis of flexibility of the neck and tail region of myosin VI dependence of calcium and calmodulin 106
5.26 Myosin VI: Calcium and calmodulin dependent distribution of long and short myosin VI molecules 107
5.27 Myosin VI: Possible pathways for CaM changing from the standard to the bipartite binding site in myosin VI 108
5.28 Myosin VI: Model of two-step mechanism for calcium regulation of the myosin VI molecular motor 111
5.29 Myosin VI: Application of a round mask with blurry edges to the data set prior to the alignment 118
5.30 Myosin VI: SPA on the low CaM-low Ca^{2+}-condition: Step 1 119

5.31 Myosin VI: SPA on the low CaM-low Ca^{2+}-condition: Step 2 120
5.32 Myosin VI: SPA on the low CaM-low Ca^{2+}-condition: Step 2 121
5.33 Myosin VI: Realignment and reclassification of the initial data set by multireference-based alignment 122
5.34 Myosin VI: SPA on the low CaM-low Ca^{2+}-condition: Step 3 123
5.35 Myosin VI: Creation of a myosin VI crystal structure model, PDB: 2BKI+3GN4 124
5.36 Myosin VI: Crystal structure of ADP-myosin VI, PDB: 4ANJ 125
5.37 Myosin VI: Detailed analysis of EM averages using a crystal structure model 126
5.38 Myosin VI: Workflow for creating outlines of EM averages and color coded versions 127
5.39 Myosin VI: Workflow: Analysis of flexible parts within EM classes by creating a movie 128

6.1 Myosin IX: Actin filaments are involved in a variety of structures in the cell 153
6.2 Myosin IX: Sequence of myosin IX 154
6.3 Myosin IX: Actin crosslinking molecules 156
6.4 Myosin IX: Formation of actin crosslinks induced by myosin IX 157
6.5 Myosin IX: Negative stain electron micrographs of single myosin IX molecules crosslinking actin 159
6.6 Myosin IX: Global average of 981 images of myosin IX - actin - crosslinks 160
6.7 Myosin IX: Negative stain electron micrographs of single myosin IX molecules crosslinking actin 161
6.8 Myosin IX: Illustration of parallel, antiparallel, in phase and out of phase crosslinks 163
6.9 Myosin IX: Prealigned data set of myosin - actin - crosslink images . . . 163
6.10 Myosin IX: SPA of myosin IX- actin - crosslink images: Step 1, Consistency check 165
6.11 Myosin IX: SPA of myosin IX- actin - crosslink images: Step 2, Polarity analysis 166
6.12 Myosin IX: Polarity and phase assignment to crosslink images 167
6.13 Myosin IX: Polar features seen in the EM global crosslink average 169
6.14 Myosin IX: Resolution of EM global crosslink average 169
6.15 Myosin IX: Reclassified polarity aligned data set with respect to the three binding conformations: diagonal, bent and inchworm 170
6.16 Myosin IX: Variance analysis of EM global crosslink average 171

6.17 Myosin IX: Crystal structure projection matching of a myosin IE model onto the myosin - actin - crosslink EM average 173
6.18 Myosin IX: Superposition of a calmodulin-myosin IE structure with the myosin - actin - crosslink EM average 174
6.19 Myosin IX: Raw micrographs of myosin IXa and actin in the presence of calcium 176
6.20 Myosin IX: general shape of myosin IXa- actin - crosslinks 177
6.21 Myosin IX: Prealigned data set of myosin - actin - crosslink images in the presence of calcium 178
6.22 Myosin IX: SPA of myosin IXa- actin - crosslink images in the presence of calcium: Step 1, Consistency check 179
6.23 Myosin IX: SPA of myosin IXa- actin - crosslink images in the presence of calcium: Step 2, Polarity analysis 180
6.24 Myosin IX: Polarity and phase assignment to crosslink images in the presence of calcium 182
6.25 Myosin IX: Three shapes adopted by actin-crosslinking myosin IX in the presence of calcium 183
6.26 Myosin IX: Variance analysis of myosin IX crosslinks in the presence of calcium 184
6.27 Myosin IX: Myosin IE - actin crystal structure overlaid onto calcium-crosslink global average 184
6.28 Myosin IX: Detailed analysis of three myosin conformations using crystal structure projection matching 185
6.29 Myosin IX: Summary of myosin IX - actin - crosslinks in the presence of calcium in the form of a cartoon 186
6.30 Myosin IX: Cartoon of myosin IX bound to actin 189
6.31 Myosin IX: Summary of SPA on myosin IXain the presence of actin and absence of calcium 190
6.32 Myosin IX: How a monomeric myosin self-organizes actin networks . . . 192
6.33 Myosin IX: Discussion: Comparison of myosin IX - actin - crosslinks in the absence and presence of calcium 194
6.34 Myosin IX: Statistical distribution of calcium dependent F-actin polarity and phase 195
6.35 Myosin IX: Statistical distribution of calcium dependent myosin IX-conformations 195
6.36 Myosin IX: Comparison of EM averages of the myosin VI and IX catalytic domain 198

6.37 Myosin IX: Possible three dimensional arrangement of how myosin IX molecules crosslink actin filaments 199
6.38 Myosin IX: Possible simplified three dimensional arrangement of myosin crosslink in a perspective, side and top view 199
6.39 Myosin IX: Magnification-series of one actin-myosin crosslink EM micrographs 202
6.40 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks 205
6.41 Myosin IX: Difference between alignment of small and large crosslink images 207
6.42 Myosin IX: Crystal structure of filamentous actin, PDB:3b63 208
6.43 Myosin IX: Creation of a first crosslink average 209
6.44 Myosin IX: Classification of an aligned and windowed data set with respect to binding to one of the actin monomers 211
6.45 Myosin IX: Possible arrangements of two actin filaments within one crosslink with respect to their polarity and phase 212
6.46 Myosin IX: Possible orientations of each crosslink-conformation 213
6.47 Myosin IX: Simulated crosslinks and real crosslink images from raw micrographs 214
6.48 Myosin IX: Creation of prealigned crosslink data set: step 1 215
6.49 Myosin IX: Why aligning data set with respect to the myosin was legit . 216
6.50 Myosin IX: Creation of prealigned crosslink data set: step 2 217
6.51 Myosin IX: Polarity analysis on generic data set, step 1 219
6.52 Myosin IX: Polarity analysis on generic data set, step 2: consistency check 221
6.53 Myosin IX: Polarity analysis on generic data set, step 3 223
6.54 Myosin IX: Polarity analysis on generic data set, step 4 225
6.55 Myosin IX: Variance analysis of myosin IX - actin - crosslinks 227
6.56 Myosin IX: Variance analysis of myosin IX - actin - crosslinks 228
6.57 Myosin IX: Variance analysis of myosin IX - actin - crosslinks in the presence of calcium 230
6.58 Myosin IX: Suitable crystal structure to analyze myosin IX - actin - crosslinks 231
6.59 Myosin IX: Low-pass filtered 2D projections of a myosin IE - actin crystal structure 232
6.60 Myosin IX: Crystal structure projection matching of myosin IE - actin - model to EM averages 234
6.61 Myosin IX: Crystal structure projection matching of myosin IE - actin - model to EM averages 235
6.62 Myosin IX: Modeling calmodulin onto loop2 of myosin IE 237

6.63 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Overview 1 240
6.64 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Overview 2 241
6.65 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Overview 3 242
6.66 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Overview 4 243
6.67 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Two actin filaments involved 1 244
6.68 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Two actin filaments involved 2 245
6.69 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Three actin filaments involved 1 246
6.70 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Three actin filaments involved 2 247
6.71 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: Three actin filaments involved 3 248
6.72 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: More then three actin filaments involved 1 249
6.73 Myosin IX: Raw micrograph of myosin IX - actin - crosslinks: More then three actin filaments involved 2 250
6.74 Conclusion: Model for myosin lipid interaction using nanodiscs 254
6.75 Conclusion: First cryo-EM images of myosin IX crosslinking actin filaments 255
6.76 Conclusion: Myosin XXI can crosslink actin filaments in the form of a dimer. Imaged by negative stain EM 256
6.77 Conclusion: Special insert in loop 2 of myosin IXa 257

List of Tables

1.1 Introduction: Concentration measurement and the Warburg-Christian method . . . 9

5.1 Myosin VI: Absolute numbers of images . . . 86
5.2 Myosin VI: Statistics on the myosin VI at different calcium and calmodulin concentrations . . . 86
5.3 Myosin VI: Cross correlation coefficients of EM averages at low and high calcium . . . 88
5.4 Myosin VI: Calcium dependent maximum intensities and area under the peaks in graph 5.24 . . . 103
5.5 Myosin VI: Buffer for negative stain EM . . . 116

6.1 Myosin IX: Absolute values for graphs **C** in Figure 6.11 . . . 164
6.2 Myosin IX: Absolute values for graphs **D** in Figure 6.11 . . . 164
6.3 Myosin IX: Absolute values for graphs **E** in Figure 6.11 . . . 167
6.4 Myosin IX: Absolute numbers of the statistical distribution of the diagonal, bent and inchworm conformatio . . . 171
6.5 Myosin IX: Absolute values for graphs **C** in Figure 6.23 . . . 178
6.6 Myosin IX: Absolute values for graphs **D** in Figure 6.23 . . . 181
6.7 Myosin IX: Absolute values for graphs **E** in Figure 6.23 . . . 181
6.8 Myosin IX: Statistics on shape-appearance . . . 183
6.9 Myosin IX: Buffer for negative stain EM . . . 203
6.10 Myosin IX: Values for graphs in Figure 6.53 **C** . . . 222
6.11 Myosin IX: Values for graphs in Figure 6.53 **D** . . . 222
6.12 Myosin IX: Relation between polarity / phase and applied rotations / mirroring during the alignment . . . 224
6.13 Myosin IX: Values for graphs in Figure 6.53 **E** . . . 226

Acknowledgments

Most of the results presented in this thesis would have been impossible to achieve without the help, support and ideas of a number of people, who i would like to acknowledge in the following:

In the first place I thank my supervisor **Claudia Veigel**, who made it possible for me to research in this fascination field. Besides her amazing job supervising me during my PhD, her extreme broad range of knowledge often led to profound discussions and not seldom resulted in approaches from a completely different point of view, which was the foundation for our research. Apart from that, she let me always work independently and freely experiment with my own ideas and at the same time provided a guided framework to stay efficient and focused
Second, I would like to thank **Hermann Gaub** for taking his time to discuss the topic of my work, reviewing my thesis and supervising during this time. I further thank him for setting the course to write my dissertation in physics.

I would like to give credit to a number of more people that made the results of dissertation possible:
Chris Batters for doing a perfect job at Myosin VI sample purification. Anyone who has ever used the electron microscope to study small biological objects knows how unforgiving this technique is when it comes to sample purity. The experiments would have been much more challenging, if not impossible with samples of less quality.
Ewa Warchoł and **Markus Kröss** for their work on expressing and purifying Myosin IX, which resulted in ideal sample conditions for the electron microscopy.
Tanvir Shaikh for not only introducing me into SPIDER, but also for being almost constantly available whenever I had questions about the software and SPA in general. Together with Jose Miguel he also instructed two hands-on SPIDER / Scipion Workshops for which I am very thankful. Without Tapu, I would have get lost and gone wild in the EM data-jungle.
Stan Burgess for giving me first hand insights into single particle analysis, being avail-

able for any of my questions and for providing me with tips&tricks in this field. I am in particular thankful for the time he took to discuss my many questions, usually writing in detailed answers that exceeded my expectations by far.
Hendrick Dietz for making his electron microscopes available to me and allowing me to operate them by myself.
Evi Stahl, Klaus Wagenbauer and Fabian Schneider from the Dietz lab for instructing me into the proper usage of the Philips CM100 EM, helping out whenever something did not go as it was supposed to and for giving me the possibility to make first images of my samples in cryo EM.
All students, PhD's, post-docs, technicians and workers from our group who helped and supported me and with their discussions and insights from different perspectives made this an unforgettable time, in particular: **Heike Ellrich**, **Christian Hundschell**, **Nam Le**, **Ana Petrovac**, **Michaela Pollauf**, **Susanne Schickle**, and **Robert Waberer**.
The **CeNS** and **SFB863** for their financial and organizational support and for organizing great conferences and meetings.

My wife **Justyna** and my daughter **Zofia** for everything.