DISSERTATION

submitted to the
Combined Faculties for the Natural Sciences and for
Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Submitted by

Dipl. Phys. Sarah Elise Freyja Wulf
born in Hamfelde Kr. Hzgt. Lauenburg, Germany

Oral examination: 16. April 2014

Application of new reconstruction strategies to enhance the interpretable resolution in 3D Electron Microscopy

–

Studies of Rigor and ADP binding complexes of Actin and Myosin V

Referees: Prof. Dr. Rasmus R. Schröder
Prof. Dr. Ulrich Schwarz

Abstract

Subject of these studies was the strong binding complex of actin and myosin V. How the molecular motor protein myosin powers muscles by generating mechanical force out of chemical energy at high efficiency fascinates researchers already for a long time. Myosin V offers the unique possibility to study not only the rigor state but also its preceding strong binding ADP state, which is only transiently present in other myosins. Since the actomyosin complex could not be studied by protein crystallography so far, transmission electron microscopy (TEM) of vitrified protein complexes is the method of choice. New reconstruction strategies were applied, to enhance the interpretable resolution in 3D electron microscopy, together with subsequent molecular dynamics (MD) simulations of protein crystal structures. Such a combined approach allows for almost the same resolution as X-ray diffraction facilitates. As a new – and in principle superior – reconstruction scheme a filtered least squares reconstruction algorithm was applied in order to obtain a better defined density localization. Thorough analysis combined with electron tomography studies revealed flexible specimen properties that limit the applicability of the current implementation of the least squares method due to necessary ad hoc assumptions. As a first step to reduce these assumptions iterative helical reconstruction was applied and yielded densities with reliable 8 Å resolution. Following MD simulations were able to resolve functional differences of myosin between the rigor and the strong binding ADP state.

Zusammenfassung

Gegenstand der Untersuchungen war der stark gebundene Komplex aus Aktin und Myosin V. Die Fragestellung, wie das molekulare Motorprotein Myosin Muskelaktivität hervorbringt, indem es chemische Energie in mechanische Arbeit transformiert, fasziniert Forscher bereits seit langem. Myosin V bietet die einzigartige Möglichkeit nicht nur den Rigor-Zustand, sondern auch den ihm vorausgehenden stark gebundenen ADP-Zustand zu untersuchen, der in anderen Myosinen nur transient ist. Da der Actomyosin-Komplex bisher nicht mit Proteinkristallographie untersucht werden konnte, ist die beste Methode die Transmissionselektronenmikroskopie (TEM) anhand vitrifizierter Proteinkomplexe. Es wurden neue Rekonstruktionsstrategien zur Erhöhung der interpretierbaren Auflösung der 3D Elektronenmikroskopie angewendet. In Kombination mit molekulardynamischen (MD) Simulationen der Proteinkristallstrukturen erlaubt dies, vergleichbare Auflösungen wie in der Röntgenbeugung zu erreichen. Als ein neues – und im Prinzip überlegenes – Rekonstruktionsschema wurde eine algebraische *Least-Squares*-Methode angewendet in der Absicht, eine besser definierte Lokalisierung der Objektdichte zu erhalten. Eine gründliche Analyse und weitere Untersuchungen mittels Elektronentomographie zeigten flexible Probeneigenschaften auf, die die Anwendbarkeit der *Least-Squares*-Methode in ihrer derzeitigen Implementierung wegen notwendiger Ad-hoc-Annahmen einschränken. Als ersten Schritt diese Annahmen zu reduzieren, wurde eine iterative, helikale Rekonstruktionsmethode verwendet, die es ermöglichte Dichten mit einer Auflösung von 8 Å zu rekonstruieren. Nachfolgende MD-Simulationen erlaubten, funktionelle Unterschiede zwischen dem Rigor- und dem stark gebundenen ADP-Zustand festzustellen.

Contents

Nomenclature

2D	two-dimensional
3D	three-dimensional
Å	Ångström
µl	microliter, 1 µl=$1 \cdot 10^{-6}$ l
µM	micromolar, 1 µM=$1 \cdot 10^{-6}$ mol/dm^3=$1 \cdot 10^{-3}$ mol/m^3
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate, the universal energy source of living organisms
CCD	charge-coupled device
Chimera	Program for interactive visualization and analysis of molecular structures, University of California, San Francisco, USA
CTF	contrast transfer function
EDTA	Ethylene-diamine-tetraacetic acid, used as protein buffer
EM	Electron microscopy
EMAN	Suite of image processing tools aimed primarily at the transmission electron microscopy community, available versions EMAN1 and EMAN2, National Center for Macromolecular Imaging, Baylor College of Medicine, Houston, Texas, USA
eV	electron volt, 1 eV=$1.602 \cdot 10^{-19}$ kg m^2/s^2
f-actin	filamentous actin
fs	femtosecond, 1 fs=$1 \cdot 10^{-15}$ s

FSC	Fourier shell correlation
g-actin	globular actin
IHRSR	iterative helical real space reconstruction
IHRSR++	User-friendly extension to the initial IHRSR software, Timothy Baker Lab, University of California, San Diego, USA
IMOD	Electron tomography software, Boulder Laboratory for 3D Electron Microscopy, University of Colorado, USA
JWeb	Java implementation of the graphics interface of SPIDER
kDa	kilo-Dalton, 1 Da=1 u=1.66$\cdot 10^{-27}$ kg, u is the unified atomic mass unit
MD	molecular dynamics
MDFF	molecular dynamics flexible fitting
mM	millimolar, 1 mM=1$\cdot 10^{-3}$ mol/dm^3=1$\cdot 10^{0}$ mol/m^3
NAD	nonlinear anisotropic diffusion
NAMD	Nano-scalable molecular dynamics software from the University of Illinois at Urbana-Champaign, USA
nm	nanometer, 1 nm=1$\cdot 10^{-9}$ m
NMR	nuclear magnetic resonance
ns	nanosecond, 1 ns=1$\cdot 10^{-9}$ s
rmsd	root mean square deviation
SPARX	Single particle processing environment, distributed together with EMAN2
SPIDER	Image processing system for electron microscopy, Wadsworth Center, New York State Department of Health, USA
TEM	transmission electron microscopy
Tris	tris-hydroxymethyl-aminomethane, used as protein buffer

U	enzyme unit, defined as amount of enzyme that catalyzes the conversion of 1 micro mole of substrate per minute
VMD	Visual Molecular Dynamics, molecular graphics software from the University of Illinois at Urbana-Champaign, USA

Chapter 1

Introduction

Prof. Dr. Kenneth C. Holmes (Max Planck Institute for Medical Research, Heidelberg), who is known for his outstanding contribution to muscle research, once characterized muscle as "machine for turning chemical energy into mechanical work at high efficiency and at constant temperature". Even to date, "no nanotechnology comes close to emulating this process. Therefore we can learn much by studying the molecular mechanism of muscle" (Holmes, 2004). Despite its long history and many great achievements the field of muscle research still has to offer interesting open questions. Especially details of structural and functional properties of the so-called chemo-mechanical transduction, the process whereby myosin transforms chemical energy into mechanical work, are difficult to investigate because of its highly dynamic and transient character. Structural data can be acquired only of stable intermediate or end states. X-ray diffraction using protein crystals has the greatest resolving power so far and yielded stand-alone structures for both actin and myosin with about 2 Å resolution. But the crystallization of the functional complex of filamentous actin and myosin has not succeeded yet. Hence, state-of-the-art methods and techniques have been exploited that can be used instead of X-ray crystallography.

Transmission electron microscopy of vitrified protein complexes offers the possibility to study the functional complex. 3D reconstruction combined with subsequent molecular dynamics simulations of protein crystal structures into the TEM densities allows to gain insight into the functional states with almost the same resolution as X-ray diffraction facilitates. In order to analyze not only some structural conformations but also to learn about the evolution of states in their functional cycle two subsequent states of the myosin crossbridge attached to actin have been investigated. The rigor state, known from *rigor mortis*, is a stable conformation in the absence of high-energy nucleotides, namely adenosine triphosphate (ATP). This state can be biochemically prepared for all myosins and has already been studied previously. For this work the non-muscle myosin V (from chicken) has been chosen to investigate due to the fact that its unique kinetics offers a stable structural state that precedes the rigor

state, the strong binding ADP state, that is only transiently present in other myosins as it represents an intermediate of the force generating process. Both the rigor and the strong binding ADP state of myosin V have been investigated in the complex with filamentous actin. The molecular motor mechanism of myosins is highly conserved by evolution thus insights from one myosin are also valid for other myosins.

The first part of this thesis gives an introduction into the field of research. Chapter 2 outlines the background of muscle research and deals with the properties of actin and myosin, that are the main components of the force producing machinery. Helical symmetry is introduced as well as cellular functions of unconventional myosins. Likewise kinetics and structural states of the force generating cycle, known as Lymn-Taylor- or crossbridge cycle, are discussed. While chapter 3 describes available methods for protein structure determination including its physical concepts and its applicability, chapter 4 gives a detailed delineation of how a protein structure can be solved by transmission electron cryo microscopy. Image formation in a transmission electron microscope (TEM) and the contrast transfer to the detector are described as well as the necessary correction for the unwanted effects of the so-called contrast transfer function (CTF). Cryo sample preparation and data acquisition schemes are addressed both for single particle analysis and for electron tomography including restrictions due to radiation damage. Besides standard 3D reconstruction methods the algebraic least squares reconstruction algorithm is introduced that was first investigated for model data in the preceding diploma thesis (Wulf, 2009) and now should have been applied to real micrographs. How molecular dynamics simulations can combine results from 3D reconstructions of TEM images with structures obtained by protein crystallography is outlined in chapter 5.

The second part presents the results. Chapter 6 deals with the application of the algebraic least squares method to an actomyosin data set. Its property of better density localization in the reconstructed density was believed to optimize information retrieval in order to enhance the interpretable resolution. Thorough analysis combined with electron tomography studies (chapter 7) revealed flexible specimen properties that limit the applicability of the current implementation of the least squares method with its ad hoc symmetry assumptions. Therefore, iterative helical reconstructions were applied as first step to reduce ad hoc assumptions (chapter 8). Difficulties are explained to correct the contrast transfer function due to a high noise level in the data. However, finally two reconstructions are reported with resolutions for the actin of about 7 Å and the myosin motor domain of reliable 8 Å. Subsequent molecular dynamics flexible fitting of crystal structures into the TEM densities (chapter 9) was able to resolve functional differences of myosin. Thus, the present investigation was able to visualize for the first time that the central β-sheet, myosin's *transducer*, occupies different states in the actomyosin complex. The third part gives a final discussion of the achieved results and presents an outlook.

Part I

Prerequisites

Chapter 2

Actomyosin complex

The actomyosin complex is formed by the interaction of the two proteins actin and myosin. The cyclic interaction of both under consumption of adenosine triphosphate (ATP) powers muscle contraction. ATP is the cellular energy unit, the universal energy source of living organisms. Today myosin is known to be a molecular motor with enzymatic activity. Studying myosins started long ago with the question: how does muscle contract and generate force. Decorated actin proved to be a very successful minimal model system to study the strong interaction of actin and myosin. Actin filaments (f-actin) incubated with only the motor domain of myosin yield the so-called decorated actin. In this non-physiological system every binding side of f-actin is occupied by a myosin motor domain. In the meantime knowledge is gained about a whole family of different myosin isoforms. However, as the core mechanism of all myosins seem to be conserved by evolution, information from any myosin isoform can help to gain insight into the molecular motor mechanism (Holmes, 2004). Especially the non-muscle myosin V, as studied in this work, has provided important additional understanding. This chapter gives a short introduction into the beginnings of muscle research and then deals with the structural and functional properties of the two proteins actin and myosin.

2.1 Muscle

As reviewed by Ken Holmes the muscle research has always been accompanied by the development of analytic methods and innovation of new techniques. Actually in many cases techniques have been pioneered for understanding muscle, which are now in general use (Holmes, 2004). In the early development of microscopy it was van Leeuwenhoek in 1674 who used his own innovation, a simple light microscope, to discover the myofibrils and cross-striations in muscle fibers. A short time after that it was already suggested, that the units delineated by the cross-striations, today known as sarcomeres,

may serve as contraction units (reviewed by Holmes, 2004). After Wilhelm Kühne discovered in 1864 the muscle's main component, the myosin protein (Kühne, 1864), it were Engelhardt and Ljubimova, who could show in 1939 that myosin has the enzymatic activity of an ATPase (Engelhardt and Ljubimova, 1939), namely to hydrolyse ATP into ADP (Adenosine diphosphate) and a free phoshate ion. The free energy liberated by the decomposition can then in turn be used to catalyze another chemical reaction that otherwise would not happen. In the case of muscle the energy is transformed into mechanical work. This has been termed chemo-mechanical transduction. Szent-Györgyi (1942) manages a little later to show that fibers of actin and myosin contract on adding ATP. The work of Huxley and Niedergerke (1954) and Huxley and Hanson (1954) was a major breakthrough in understanding sarcomere function. They postulated the so-called sliding filament model. Against the common belief that during muscle contraction and relaxation there is a shorting or stretching of some kind of continuous filamentous structure, they showed that during muscle shortening myosin and actin filaments slide past each other without much change in length (Squire et al., 2005). More recently it was directly proven by Kron and Spudich (1986) that the filaments do really slide by fixing myosins to a glass surface and showing the sliding movement of fluorescently-labelled actin filaments across them in the presence of ATP. Today this method is commonly called in-vitro motility assay. A detailed review about the molecular architecture in muscle contractile assemblies is given by Squire et al. (2005). For studies of the underlying basis of force generation often the reduced actomyosin system has been used. As it has become clear that actin seems to be the passive partner, providing binding sites for myosin, the focus has been on studying myosin. The next two sections focus on the properties of actin and myosin.

2.2 Actin and helical symmetry

Actin is not only present in the so-called thin filaments of muscle, as part of the cell cytoskeleton it has also a central importance in cell motility (Dominguez and Holmes, 2011). It can build up stress fibers or cell protrusions on the mobile edge of a cell (lamellipodia and filopodia) and is also involved in cell division, since the contractile ring is powered as well by the interaction of myosins with actin filaments. Actin exists in two forms, a monomeric or globular (g-actin) and a filamentous one (f-actin). Because of its tendency to polymerize it had been particularly difficult to obtain a high-resolution structure (Cooke, 2004). Kabsch et al. (1990) succeeded to solve the structure of g-actin by X-ray crystallography (cf. Chapter 3 for a brief description of the method) in a complex with a small molecule which prevented polymerization. G-actin consists of 375 amino acids and has a molecular weight of 42 kilo-Dalton (kDa). As shown in Fig. 2.1 g-actin consists of two major domains each

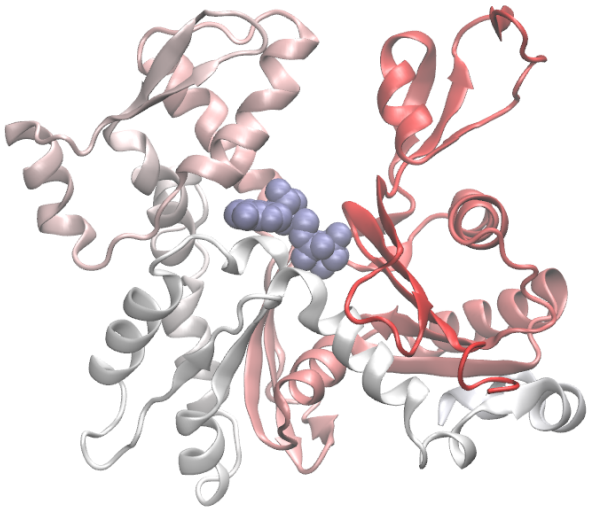


Figure 2.1: Structure of the actin monomer g-actin. The coordinates are taken from PDB entry 1ATN (Kabsch et al., 1990) and have been visualized using the program VMD (Humphrey et al., 1996). In this orientation the major domains are positioned on the left and right side with the subdomains 1 (lower right), 2 (upper right), 3 (lower left) and 4 (upper left). The upper subdomains (2 and 4) are separated by the so-called nucleotide cleft. The color is coded from N- to C-terminus as transition from red to white. The nucleotide ATP is shown in iceblue.

having two subdomains. The four subdomains form a rather flat globular actin monomer that fits into a rectangular prism with dimensions of (55 x 55 x 35) Ångström (Å) (Dominguez and Holmes, 2011). Between the major domains actin binds a nucleotide. G-actin is normally bound to ATP, whereas upon polymerization actin hydrolyzes ATP to ADP which is kept bound in the filamentous arrangement. Till now many more structures of the actin monomer have been obtained, all of them in a complex with molecules that prevent polymerization (Cooke, 2004).

The structure of actin is highly conserved. Therefore, in order to build the actomyosin complex it is actually common that the actin and the myosin being investigated must not originate from the same organism. For example the actomyosin complex that is analyzed in this work is build up by rabbit actin and chicken myosin.

The actin polymerization is a highly dynamic process. At one end the filament may grow whereas at the other end depolymerization may dominate. Therefore, inside cells there are a lot of capping, cross-linking or anchoring proteins to stabilize the filaments, if needed. Isolated filaments can be stabilized with phalloidin. The toxin phalloidin can be isolated from poisonous fungi and prevents actin depolymerization by locking adjacent f-actin subunits together (Cooper, 1987). The first model of the f-actin

structure was obtained by X-ray fiber diffraction (cf. Chapter 3 for a brief description of the method) and the use of the g-actin crystal structure (Holmes et al., 1990). More recent Oda et al. (2009) could show that upon g-actin to f-actin transition the two major domains undergo a twisting that flattens the actin monomer. This has not been considered in the initial model. Recently electron cryomicroscopy by Fujii et al. (2010) yielded a 6.6 Å resolution density map of f-actin (cf. Chapter 4 for a description of the method). The helical structure obtained from this study is illustrated in Fig 2.2. The helical filament has a diameter of 10 nm and a structural polarity due to the fact that all monomers have the same direction. Decorated actin filaments show an arrowhead pattern originated in the angle the myosin motor domain binds the actin and thereby visualize both the polarity and the symmetry of the filaments. The pointed end of the arrowheads is also referred to as minus end and the barbed end as plus end due to the polarity of polymerization. The filaments appear as two strings of beads twisting gradually around each other (Squire et al., 2005). The spacing between crossovers is about 36 nm. This distance depends on the amount of twisting. Fujii et al. (2010) reported an azimuthal rotation between successive subunits along the helix of $-166.6°$, also referred to as the symmetry angle. The subunit axial translation is 27.5 Å. The generic helix that runs through all monomers is left-handed and has a pitch of 59 Å (Squire and Knupp, 2005). The structure can approximately be described by a helix with 13 monomeric subunits in 6 full turns (a 13/6 helix) (Squire et al., 2005). As the symmetry angle is not far form 180° the impression of two slowly twisting strings appears. The twist of this so-called long-pitch helix is right-handed (Squire et al., 2005).

A 13/6 helix has a symmetry angle of $-166.15°$, but not

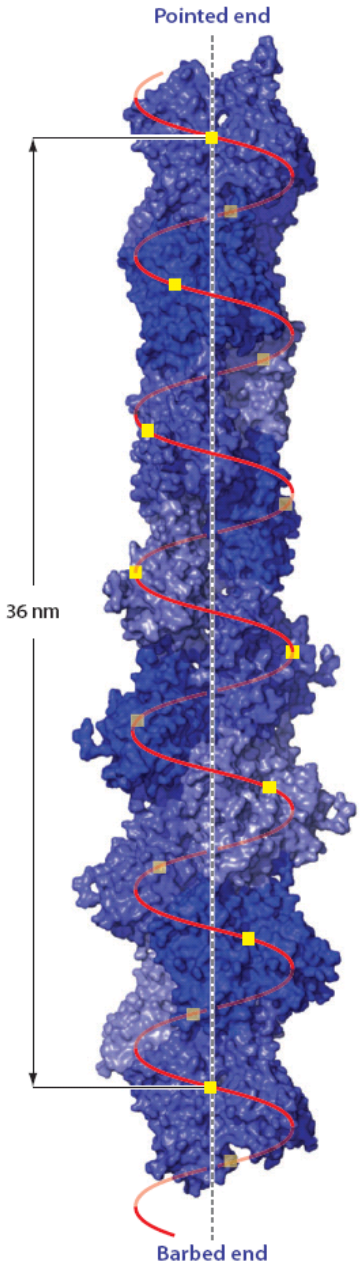


Figure 2.2: Helical structure of f-actin. The image is adopted from Dominguez and Holmes (2011) and has been constructed using the structure from Fujii et al. (2010) with 13 g-actin monomers in 6 left-handed turns (13/6 helix), resulting in a 36 nm crossover spacing of the right-handed long-pitch helix.

all filaments show this symmetry. Some actin filaments have also been found to have a 28/13 helical symmetry, which corresponds to an azimuthal rotation of $-167.14°$ and can be obtained from a 13/6 configuration by a slight additional twist. The subunit axial translation stays the same but the rotation may vary and also depend on the actin binding proteins, since they are known to be able to alter the twisting slightly. Although there are works reporting nearly perfect symmetry, actin filaments seem to have no absolute symmetry. Instead, they have shown to be rather flexible and are even described by Egelman et al. (1982) as having a random variable twist around an average symmetry. This implies that also the crossover spacing of 36 nm is an average value. Local azimuthal twisting and untwisting may give rise to a rather short so-called coherence length along the filament (Egelman et al., 1982).

2.3 Myosin

Myosin was originally extracted from muscle and was subsequently named according to that by Wilhelm Kühne. Myosin not only constitutes the thick filaments in muscle, today knowledge is gained of many other non-muscle myosins, which are hence called unconventional myosin isoforms. In 1972 the discovery of different isoforms began with the work of Pollard and Korn (1972) and it was shown that the isoforms can have widely varying structures and properties (reviewed by Cooke, 2004). According to Sweeney and Houdusse (2010) there are today more than 35 classes, with 13 of which are present in humans. Fig. 2.3 shows the general organization of the different myosin isoforms. All members of

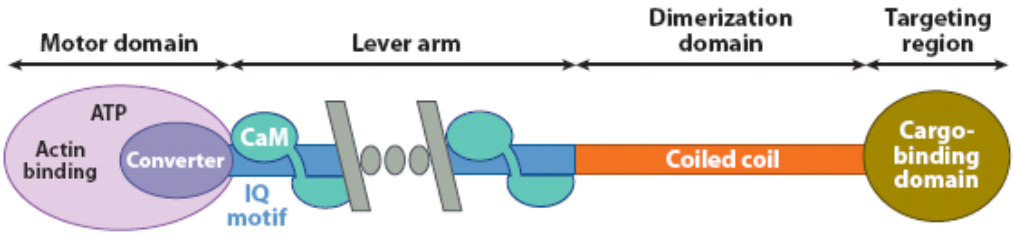


Figure 2.3: General organization of the myosin isoforms. See text for a description. The image is adopted from Sweeney and Houdusse (2010).

the so-called myosin superfamily share a highly conserved motor domain. It contains both the actin binding side and the nucleotide binding site and is thus the catalytic active part of the myosin. It is followed by a neck region, which is known to act as a lever arm (swinging lever arm hypothesis (Holmes, 1996)). It can amplify small changes inside the motor domain and is composed of a long α-helix which binds a variable number of light chains. A light chain has low molecular weight and is either calmodulin or homologous to calmodulin which bind to the amino acid consensus sequence known as

IQ-motif (reviewed by Cooke, 2004). The light chains stabilize the α-helix to make it stiff (Holmes, 2004) and have mostly regulatory function. The neck is followed by the C-terminal tail region. It can be very diverse depending on the myosin class. It can contain coiled-coil sequences for dimerization, if the molecule is dimeric, and a cargo-binding domain for the targeting and attachment of the myosin to its cargo. The size of a myosin greatly depends on the nature of its tail region. The motor domain is composed of about 800 amino acids and has a molecular weight of about 110 kDa.

2.3.1 Cellular functions

The skeletal muscle myosin II has two light chains and is a dimer. Its tails build up the thick filaments in muscle. In a well-orchestrated collaboration they are able to move whole organisms or generate large forces during isometric contraction. Smooth muscle myosins together with actin filaments build up the contractile tissue of many hollow organs, blood vessels and lymph vessels. For the diverse functions inside the cell the various myosins are required to have different kinetic properties and structural adaptations (Hammer III and Sellers, 2012). Indeed they have optimized for performing their distinct biological functions. They participate in functions like cargo transport during endocytosis and exocytosis (intracellular trafficking), cell adhesion, cell division (contractile ring), movement of pigment granules and cell motility (Hammer III and Sellers, 2012). A class I myosin for example is involved in hearing as tension sensor in the inner ear (for a review cf. Gillespie (2004)). All myosins work together with actin filaments and move generally towards the barbed / plus-end of the filament. From all studied myosins so far only class VI myosins move in the opposite direction, towards the minus-end (Wells et al., 1999).

The unconventional myosin V is known to be a single molecule cargo transporter inside the cell. Mehta et al. (1999) reported for the first time that myosin V is a processive motor that shows long runs along actin filaments making many steps without detachment. Myosin V has a long neck region with six light chains and builds dimers. The long lever arm enables the myosin to take 36 nm steps. In comparison for the muscle myosin II a sliding movement of about 10 nm has been reported. It has been shown that the length of the displacement depends on the length of the lever arm. A step size equal to the symmetry of the actin filament allows it to move straight along one side of the filament without need to spiral in order to find new binding sides (Fig. 2.4). It was demonstrated that myosin V walks by a hand-over-hand mechanism so that at any time at least one head is strongly bound to actin (Yildiz et al., 2003). Using high-speed atomic force microscopy it was actually possible to visualize the processive hand-over-hand movement per video imaging (Kodera et al., 2010). For a detailed review about the roles of myosin V as cargo transporter the reader is referred to Hammer III and Sellers (2012). They point out that class V myosins seem to tether organelles to actin as well as serve

Figure 2.4: Processive movement of myosin V. The actin filament is illustrated in gray with its pointed end toward the top. Symmetry related binding sites a shown in green. The two myosin V motor domains are in different binding states (blue) and depict the critical two-head-bound intermediate for processive movement. The long neck is stabilized by six light chains each (yellow and red). On the right the beginning of the dimerization domain is shown in gray with an arrow pointing downwards, indication the direction of movement towards the barbed end. The diameter of actin (10 nm) sets the scale. The image is adopted from Vale and Milligan (2000). In this thesis here only the blue drawn motor domain followed by one light chain is analyzed.

as short-range point-to-point organelle transporters. The actin cytoskeleton is organized in a way the plus-end of actin filaments pointing towards the cell cortex, thus myosin V moves its cargo bound to its tail domain toward the cell periphery, usually following long-range kinesin and microtubule dependent transport (Hammer III and Sellers, 2012). A comparison of the motility of myosin and kinesin, which processively walks along microtubules, and some similarities due to evolutionary conservation can be found in Vale and Milligan (2000). Brawley and Rock (2009) address the problem how a specific cargo may reach its destination inside a crowded cell. They report on unconventional myosin traffic using a selective actin cytoskeleton. Rock (2012) even compares the distinct actin tracks being chosen during intracellular trafficking with the complex network of subway lines.

2.3.2 Kinetics and structural states

The cyclic interactions of myosin with f-actin can be described by the crossbridge cycle proposed by Lymn and Taylor (1971). The cycle is illustrated with later on discussed structural states in Fig. 2.7. The myosin motor domain, sometimes referred to as crossbridge, binds ATP, hydrolyses it and keeps the products bound. This primes the motor domain featuring the lever arm in the “up” position and is followed by attachment to a nearby actin binding site. Binding to actin leads to the release of the products. At first the phosphate is released followed by the actual power stroke. Structural changes in the motor domain cause the lever arm to be positioned downwards, which leads to a few nanometer displacement or if constrained to the generation of a few piconewton of force and to the release of

ADP. Binding of a new ATP molecule causes a rapid detachment of myosin and completes the cycle (Holmes, 2004). In the absence of ATP (rigor mortis) myosin cannot detach and remains strongly bound to actin. According to this the nucleotide free strong binding complex of actin and myosin is called rigor state. Kinetic analysis showed a strong negative coupling between ATP binding affinity and actin binding affinity. Once the other one is bound the affinity of the previous bound one is greatly reduced. Myosin is thus a product-inhibited ATPase (Geeves and Holmes, 2005). Without actin it prefers to keep the hydrolysis products and without ATP it remains bound to actin. A combination of enzymatic and structural studies has greatly helped to understand the motor mechanism. Today there is evidence of many more biochemical states during the ATPase cycle than in the original kinetic scheme of Lymn and Taylor (1971) (Sweeney and Houdusse, 2010). However, only the stable end states of a dynamic process can yield structural data (Geeves and Holmes, 2005). Due to the fact that the unconventional myosin V is kinetically tuned to enable its processive movement, it offers different dominant conformations. Not only the rate it proceeds through the cycle is altered but also the relative amount of the cycle it spends strongly bound to actin (Sweeney and Houdusse, 2004). Therefore – and due to the strong conservation of the motor domain – studies of the non-muscle myosin V offer new insights into the muscle contraction mechanism.

The first crystallographic myosin structure solved was the motor domain from chicken skeletal muscle (Rayment et al., 1993b). It represents the so-called post-rigor state, when myosin is detached after ATP binding and before repriming of the lever arm (cf. Fig. 2.7). Figure 2.5 and 2.6 show the essential elements of the myosin motor domain. As described by Holmes et al. (2004) the myosin motor domain consists of four subdomains (U50K, L50K, N-terminal, converter) with a central seven-stranded β-sheet. A deep cleft, known as the actin binding cleft, runs from close to the nucleotide binding side to the actin binding side. It separates the subdomains named upper and lower 50K domain (U50K, L50K). The nucleotide binding side consists of three important, flexible elements named P-loop, switch 1 and switch 2, like they have been found very similar in G-proteins before. Switch 2 is connected to the α-helix named relay helix. Another conformationally flexible element, the converter domain, is attached to the relay helix. The converter transmits rearrangements within the motor domain to the lever arm, which is anchored in the converter (cf. Fig. 2.6). The discovered structural homologies with G-protein suggested directly mechanisms for the function of the myosin (Cooke, 2004). The first actomyosin model (Rayment et al., 1993a, Schröder et al., 1993) was build using the post-rigor myosin structure. It was suggested that the actin binding cleft being open in post-rigor probably closes upon strong binding to actin. So far crystal structures of myosin have been obtained showing three different conformations. The structural states assigned to the crossbridge cycle are shown in Fig. 2.7. The post-rigor state exhibits an open cleft, a straight relay helix and a lever arm

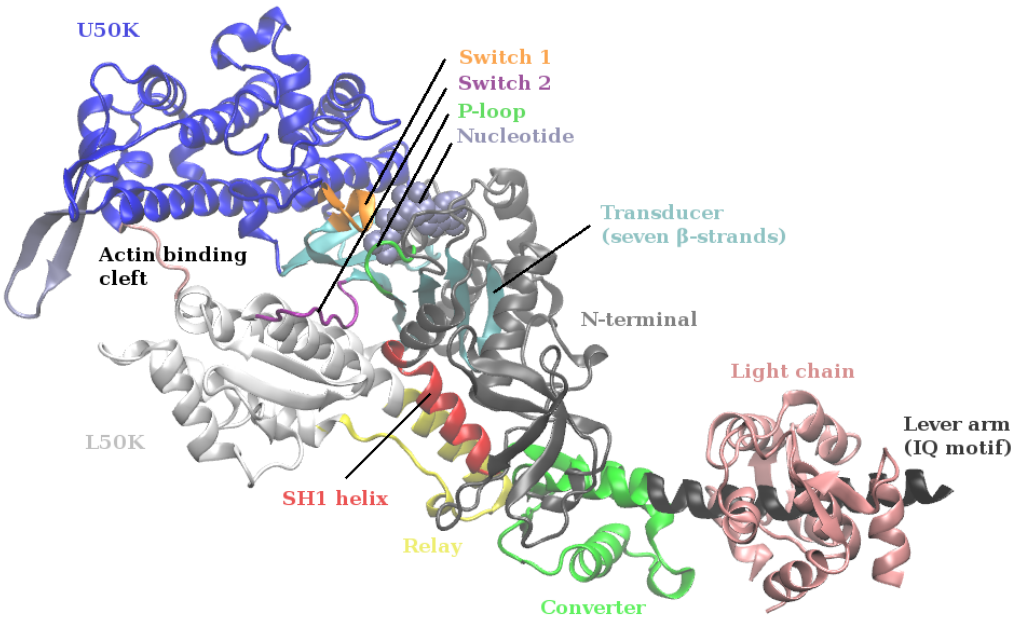


Figure 2.5: Structural elements of the myosin motor domain. The coordinates are taken from PDB entry 1W7J and represent myosin V in the post-rigor conformation with an open actin binding cleft. It is visualized using the program VMD.

"down" position. The pre-powerstroke state could be solved using ADP-P_i analogs. It shows as well an open cleft but the relay helix has a kink at its middle point and the lever arm is in an "up" position. A structure of the nucleotide-free myosin V in the absence of actin shows a closed actin binding cleft and a straight relay helix (Coureux et al., 2003). A conformation that could not be obtained with muscle myosin II. As it may closely represent the actin bound rigor state, it is referred to as rigor-like state (Sweeney and Houdusse, 2004). Following the force-generating cycle from pre- to post-powerstroke conformation the kink in the relay helix is removed causing the converter and lever arm to rotate (Holmes, 2004). Therefore, "the mechanism by which actin binding brings about the straightening of the kinked relay helix becomes a central issue in understanding muscle contraction" (Holmes, 2004). Energy-filter electron cryo-microscopy has yielded a 14 Å resolution density map of rabbit skeletal actin decorated with chicken skeletal myosin motor domains (Holmes et al., 2003). They could confirm the prediction that the actin binding cleft is closed on strong binding to actin and that the rigor-like myosin V crystal structure really is close to rigor, as it fits without deformation the myosin II rigor density (Holmes et al., 2004). The cleft closure seems to induce the opening of the nucleotide-binding pocket via changes in the position of switch 1 and the P-loop. Additionally, it appears to involve a twist of the central β-sheet (Holmes, 2004). This allows Holmes (2004) to suggest a

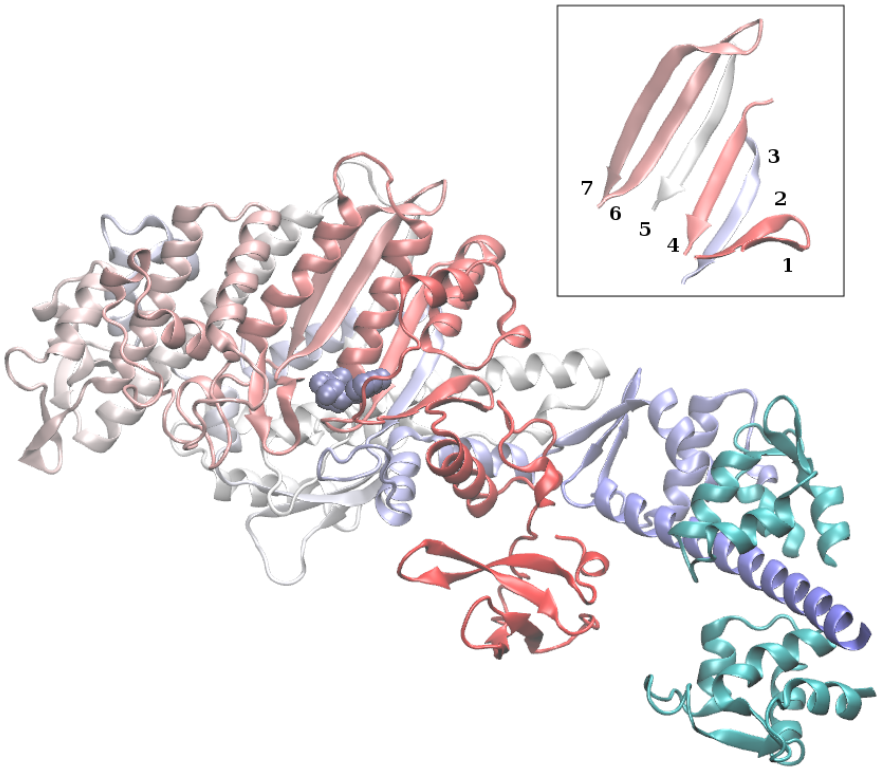


Figure 2.6: Myosin motor domain colored by index from N- to C-terminus. The coordinates are the same like in Fig. 2.5. The orientation is chosen to allow a view onto the nucleotide binding side with the seven-stranded β-sheet in the back. The color changes from red (N-terminal β-barrel) to blue (truncated lever arm). The light chain is shown in cyan and the nucleotide in iceblue. The insert shows how the seven-stranded β-sheet is composed of different parts of the amino acid sequence and thus allows to integrate structural rearrangements of the different subdomains. Below the insert the white relay helix can be seen how it interacts with the light blue three-stranded β-sheet of the converter which anchors the lever arm. (Image rendered by the program VMD.)

mechanism whereby "straightening of the relay helix comes about through the twisting of the β-sheet. If the cleft closure at the actin-binding side causes the twisting of the β-sheet then one has a mechanism for actin binding driving the powerstroke". Up to now there are no high-resolution structural details of the central β-sheet and the surrounding structural elements obtained from the strong binding actomyosin complex either in rigor state or strong ADP-binding state. To date crystal structures exist only of the unbound myosin motor domain. Coureux et al. (2004) refer to the central β-sheet as the *transducer* since it seems to be the central integrator of the motor. A detailed description of the different structural states and their differences can be found in Geeves and Holmes (1999), Geeves and Holmes (2005), Holmes et al. (2004), Coureux et al. (2004), Sweeney and Houdusse (2004) and Sweeney and Houdusse (2010).

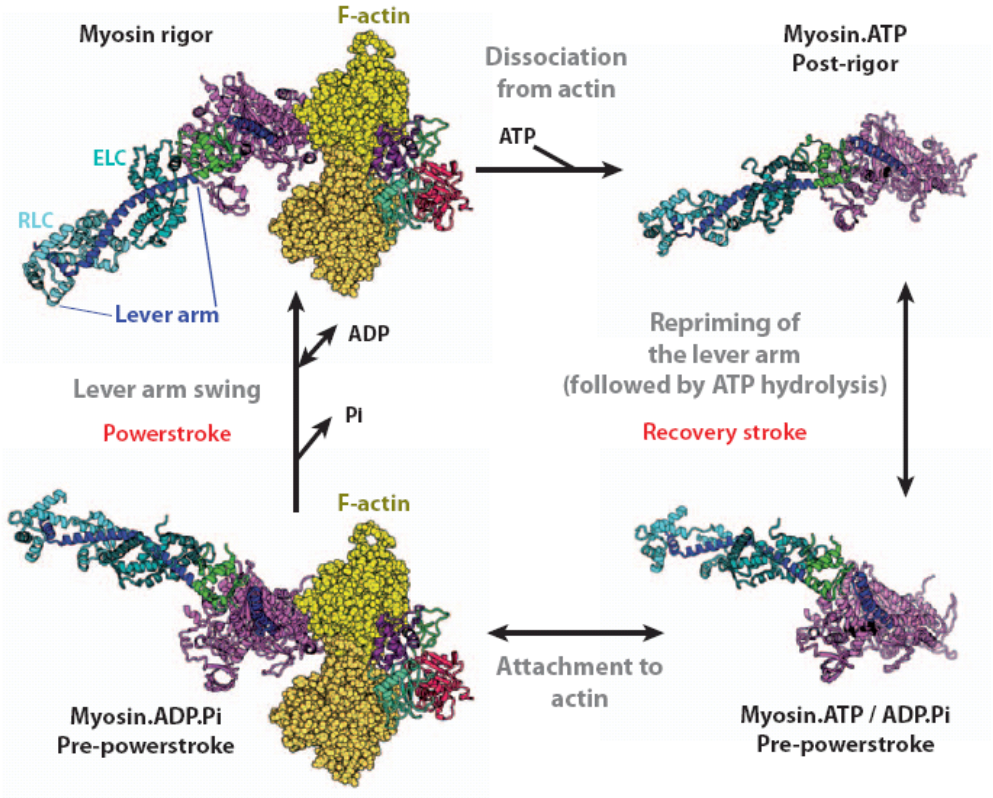


Figure 2.7: Structural states of the crossbridge cycle. The myosin is shown with two light chains like in skeletal muscle (ELC: essential light chain, RLC: regulatory light chain). The converter is shown in green, the relay helix in blue and the remaining motor domain in purple. Two actin monomers are depicted in yellow, the third one is shown with its four subdomains. The three structural states, as described in the text, are assigned to the corresponding states of the cycle as proposed by Lymn and Taylor (1971) (adopted from Sweeney and Houdusse, 2010).

Chapter 3

Determining protein structure

Chapter 2 showed that besides biochemical and cell biological studies the three-dimensional (3D) structure determination of a protein can provide important insights into its function and how it accomplishes its tasks. There are several methods that can be used for structure determination. Most of the published protein structures (over 80% (RCSB Protein Data Bank, 2013)) are obtained by X-ray crystallography, about 10 % by solution nuclear magnetic resonance (NMR) spectroscopy. This chapter gives a brief overview of the basic principles and the limitations of the common methods and offers the connection to the analysis of the actomyosin complex by means of transmission electron microscopy and its combination with crystallography derived molecular structures, the so-called "hybrid microscopy".

3.1 X-ray crystallography

In order to visualize small objects the wavelength of the electromagnetic radiation has to be chosen small enough according to the Abbe diffraction limit for the object to diffract the wave. The typical distances of a protein range from its size in the order of 10 nm to its intramolecular architecture with a covalent carbon-carbon bond length of typically 1.5 Å. To analyze such distances with electromagnetic radiation the wavelength of X-rays is needed. As described for instance by Rhodes (2006) it is not possible to obtain focused images of single proteins with X-rays. On the one hand there are no lenses to focus X-rays. Therefore, diffraction patterns are acquired. On the other hand the interactions of X-rays, emanated by a conventional source, with a protein are so weak that no diffraction signal can be detected. This drawback can be circumvented by analyzing a regular array of identical proteins, a protein crystal. Both problems lead to further implications. First the nature of interaction is considered. As explained by Warren (1990) the approximation of classical Thomson scattering, the elastic X-ray

scattering by free charged particles, is adequate for most crystallographic applications. Disregarding the correct wave mechanical description the interaction can be thought of charged particles inside the molecule being accelerated by the oscillating electric field component of the incident X-rays. As the Thomson cross section decreases with the inverse square of the mass of the charged particle, the interaction can be described as scattering by the electrons inside the protein only. The atomic nuclei yield no significant contribution. As the electrons oscillate they in turn emit electromagnetic dipole radiation of the same wavelength. When the protein and hence its electrons are arranged in a crystal the emitted radiation from the oscillating electrons can interfere. In most directions the interference will be destructive. It is constructive for those directions given by Bragg's law, when the diffraction can be regarded as reflection by a set of parallel planes through the unit cells of the crystal (Branden and Tooze, 1998). The direction and intensity of the diffracted radiation can be detected. It is the diffraction pattern of the protein crystal which consists of thousands of spots. In order to obtain a representation of the molecule in real space the spatial interpretation of each reflection has to be obtained. In addition to the position and intensity there needs to be information about the relative phases, which is lost during detection. It is a major task to determine the phases, to solve the so-called phase problem. There are different attempts including multiple isomorphous replacement using heavy atoms, anomalous scattering or multiple-wavelength anomalous dispersion. A short introduction is given by Branden and Tooze (1998) for instance. The phases are of great importance because they contribute to the positional accuracy in the electron density to be determined. Whereas the amplitudes affect the relative weights of the density (Chiu, 1993). Together with the estimated phases a computer can be used to simulate an image-reconstructing lens (Rhodes, 2006) and iteratively refine the phases and the density. As result a 3D representation is obtained of the electron clouds of the molecule, the electron density map. Depending on the quality of the crystal and of the diffraction data, resolutions in the order of 1 Å are obtainable. Finally the map has to be interpreted by fitting into it a sequence of amino acids yielding a molecular model of the protein, that often exhibits some uncertainties due to flexible parts of the peptide chain, disorder or thermal motion.

The structure of proteins is described by a hierarchical composition: the amino acid sequence of the protein's polypeptid chain is called primary structure, the regions of the sequence which arrange in regular structures like alpha helices or beta sheets are called secondary structure. The packing of the secondary structure into compact domains is described as the tertiary structure and the arrangement of the tertiary structure of more than one polypeptid chain is named the quarternary structure (Branden and Tooze, 1998). Analyzing a protein by means of X-ray crystallography the primary structure has to be known by other means. For the determination of the secondary and tertiary structure X-ray

crystallography is a very powerful method. According to the statistics of the protein data base it is the main method so far. The major limitation of this method is to obtain crystals of the desired protein or protein complex. The chance of crystallization for globular proteins is much higher than for those having highly hydrophobic or flexible, unstructured portions. Crystals of proteins with distinct quarternary structure or even protein complexes of binding partners are rare. While several myosins and single globular actin molecules each for itself have been solved by X-ray crystallography as described in Chapter 2, the crystallization of the complex of actin and myosin has not been successful so far.

3.2 X-ray fiber diffraction

Proteins which build fibrous structures, like actin which polymerizes to build helical filaments, are difficult to study by crystallography, since filaments do not form perfect crystals. Instead, the helices can be studied packed as highly ordered fibers. As the helices become aligned with each other and show a high degree of regularity in their side-by-side packing they form structures described as crystallites within the fiber (Branden and Tooze, 1998). Fiber diffraction analysis has already been used very successfully for instance to study the structure of Deoxyribose Nucleic Acid (DNA) by Watson and Crick (1953) and led to their double helical DNA model or more recent to study f-actin with diffraction intensities to 3.3 Å and 5.6 Å depending on the orientation with respect to the fiber axis by Oda et al. (2009). In the 2D diffraction pattern of such oriented fibers the reflections are grouped into layer-lines due to the repeating structure along the filaments. These layer lines allow to deduce the helical symmetry and together with model building in some cases also the structure can be determined (Holmes, 1998). The method has been extensively used for time-resolved studies of intact muscle fibers. Thereby, changes in the diffraction pattern during muscle contraction can be detected. A detailed review on this topic is given by Squire and Knupp (2005). Up to now there is no high resolution structure of the actomyosin complex solved by fiber diffraction.

3.3 Nuclear magnetic resonance spectroscopy

Besides X-ray diffraction NMR spectroscopy can be used to obtain the secondary and tertiary structure of proteins. NMR makes use of the quantum mechanical properties of those atomic nuclei, which have a resulting magnetic moment, a nuclear spin. The nuclei of many elements, including the most abundant carbon and oxygen isotopes ^{12}C and ^{16}O, have no resulting nuclear spin. Proteins naturally offer only the magnetic moment of protons (^{1}H) for NMR analysis. Sometimes during protein synthesis other isotopes with nuclear spin, such as ^{13}C, ^{15}N or ^{31}P are introduced. Quantum mechanical a strong external magnetic field leads to a splitting of the energy levels of the different spin states, that

can be populated (the effect goes back to the findings by Zeeman, 1897). The distance of the energy levels corresponds to the Larmor frequency. The nuclear spins of the hydrogen atoms of the protein will equilibrate and align in the presence of the external magnetic field. Upon irradiation by a radio frequency pulse, close to the Larmor frequency, by resonance the spins can be exited to populate states of higher energy levels. During relaxation back to equilibrium state they emit in turn radio frequency radiation, that can be detected by a coil. Since the Larmor frequency depends on the magnetic field the nuclear spin experiences, the frequency of the emitted radiation varies slightly due to changes in local molecular environment, e.g. due to induced magnetic fields by the surrounding electron distribution. The detected frequencies are measured relative to a reference signal to have them independent of the magnetic field strength and are called chemical shifts (Keeler, 2010). Depending on the applied pulse sequence different molecular properties of the sample can be probed. Two-dimensional (2D) spectroscopy methods are used to yield information about interacting atoms and their distances. Spacial proximity allows magnetization transfer from one to another nucleus. Correlation spectroscopy (COSY) measures peaks when magnetization is transferred by scalar coupling (also called J-coupling) between hydrogen atoms, that are connected covalently through not more than one or two other atoms (reviewed by Branden and Tooze, 1998). The coupling of the spins is mediated through bonds as it can be explained by hyperfine interactions between the nuclear spins and local electrons (Hahn and Maxwell, 1952). The coupling signal provides information about connectivity inside the protein including bond distances and angles. Nuclear Overhauser effect spectroscopy (NOESY) gives peaks when magnetization is transferred via cross-relaxation by dipolar coupling (Schirra, 1996). It does not depend on how distant the hydrogen atoms are in the primary structure, they only need to be close enough in space, not through bonds. Normally a signal is only observed if the proton distance is smaller than 5 Å (Schirra, 1996). Therefore, NOESY gives valuable information about secondary and tertiary structure.

The result of NMR spectroscopy is a set of estimated distances between specific pairs of atoms. A COSY spectrum allows the assignment of a hydrogen atom to the kind of amino acid it belongs to and a NOESY spectrum gives distance constraints from hydrogen atoms in one residue (amino acid in the sequence) to hydrogen atoms in another residue (reviewed by Branden and Tooze, 1998). Having a sufficient number of bonded and non-bonded distance constraints, the number of possible configurations of a model becomes finite. Via sequential assignment an ensemble of structural models can be derived that is consistent with the data. As for crystallography the primary structure has to be known. However there is no need of crystals, as the proteins can be studied in high concentrated solution. A unique feature of NMR is the ability to probe dynamics of molecular conformations (Chiu, 1993).

The main limitation of this method is, that it is restricted in size to only small proteins (about 30 kDa), since the interpretation of complex spectra is difficult and the peaks may overlap despite of using a high magnetic field strength. Therefore, NMR is the method of choice only for small proteins that have not yet formed nice crystals. Some studies of the same protein by X-ray crystallography and NMR spectroscopy showed results that nicely agree with each other (reviewed by Branden and Tooze, 1998).

3.4 Transmission electron microscopy

Transmission electron microscopy (TEM) is a complementary method to X-ray diffraction and NMR methods, as it is only in rare cases useful to determine the secondary or tertiary structure of a protein. In turn it can be used to determine the quaternary structure of large proteins or complexes of more than one protein, an area of application difficult to facilitate with X-ray or NMR methods. How a protein structure can be solved by means of TEM will be explained in detail in the next chapter. The achievable resolution ranges nowadays in many cases in the order of 5-10 Å. In cases of stable structures with high symmetries, like icosahedral symmetries of viruses or 2D crystals of membrane proteins, resolutions down to about 3 Å have been achieved (Zhang et al., 2008, 2010, Henderson et al., 1990). As the resolution becomes comparable with X-ray crystallography secondary structure elements can be identified. α-helices are described to become visible in the resolution range from 6-10 Å, loop densities and the protein backbone at about 5-6 Å and strands of the β-sheets may be traced at resolutions higher than 4.5 Å (Fernández et al., 2008). In all those cases with resolutions not reaching this regime it is of great use to already have information about the secondary and tertiary structure by one of the other methods in order to interpret the protein density map yielded by TEM. As in the case of the complex studied in this work, where structures of all binding partners by themselves have already be determined, one can combine the results by a hybrid approach, sometimes called "hybrid microscopy", using molecular dynamics methods as described in Chapter 5. The resulting protein structure in the functional state as studied with TEM can then often be interpreted close to crystallographic resolution.

Chapter 4

Transmission electron cryo-microscopy

Transmission electron microscopy has its beginnings in the early 1930s, when Ernst Ruska and Max Knoll invented the first electron microscope which uses electrons instead of photons. Two fundamental discoveries preceded the invention (as reviewed by Rose, 2008). After J. J. Thomson had discovered the electron in 1897, L. de Broglie and H. Busch set the stage for the era of electron microscopy. Theoretical considerations led L. de Broglie in 1925 to the understanding that every elementary particle can be described as a wave as well as a particle. H. Busch discovered in 1926 how to use magnetic fields to focus electrons like a lens, built of glass, is focussing light. Since the early transmission electron microscopes the technique has been refined. Together with the development of a wave-optical imaging theory and the rise of the digital age, it is nowadays not only possible to acquire almost fully computer-controlled digitized TEM images but also to calculate a 3D representation of the analyzed specimen using linear imaging theory and subsequently emerged 3D reconstruction techniques. This chapter deals first with an outline of the image formation theory and the contrast transfer by the optical system. Afterwards it is discussed how to prepare a biological sample to use it as a specimen for TEM and at the same time preserve its structure. Topics like radiation damage and strategies to optimize information retrieval both during imaging and 3D reconstruction are outlined as well to give an introduction how to assess a protein structure by means of transmission electron microscopy.

4.1 Image formation and contrast transfer

For the interpretation and usage of TEM images it is important to know what kind of information they contain. In contrast to X-ray methods, where no focussing lenses are available, electron microscopy has the advantage to be able to record images which contain both amplitude and phase information being essential for the 3D reconstruction of an object (Glaeser, 1982). The weak-phase object ap-

proximation of thin biological specimens builds the base for image interpretation, as it relates the object's density function linearly with its detectable images (Unwin and Henderson, 1975). Before going into detail with this approximation, the nature of interaction shall be considered first. Electrons are charged particles and are scattered in the coulomb potential of the atoms of the specimen. They experience the specimen as potential distribution. Since the interaction cross-section is much higher than for the interaction of X-rays (cf. Henderson, 1995), the specimen has to be quite thin. Otherwise it will be electron-opaque and no signal can be detected in transmission. For image formation the most important interaction is the elastic scattering. Elastically scattered electrons have the same energy as the unscattered ones and hence can coherently interfere in the detection plane to form an image (Williams and Carter, 2009). The elastic scattering has a wide angular distribution and gives rise to high-resolution information (Frank, 2006a). Inelastic scattering involves energy transfer from the electron to the specimen, e.g. by some kind of electronic excitation, or energy loss due to Bremsstrahlung. Its angular scattering distribution is narrow and implies a lower resolution (Frank, 2006a). Furthermore, the inelastically scattered electron wave cannot interfere coherently with the unscattered wave. There are attempts to use only the elastically scattered electrons for image formation (cf. Section 4.3.3). In the following the wave-optical description of image formation upon elastic scattering is outlined. The TEM is operated in such a way that a coherent, parallel electron beam irradiates a thin specimen. In the wave-optical description the illuminating beam is a plane wave ψ_0. The scattered wave function behind the specimen can be written as (Reimer and Kohl, 2008)

$$\psi_s = \psi_0 a_s(r) e^{i\varphi_s(r)} \,. \tag{4.1}$$

r is the radius in the specimen plane with its origin coinciding with the optical axis. $a_s(r) \leq 1$ is the local decrease of amplitude due to absorption or obstruction by an objective aperture. $\varphi_s(r)$ is the phase shift caused by the elastic scattering. Thus, the phase of the wave carries the high-resolution imprint of the potential distribution of the specimen. In the framework of the phase-grating-approximation (Hawkes, 2006) the phase shift is given by

$$\varphi_s(x,y) \approx -\frac{\pi}{\lambda e U} \int V(x,y,z) dz \,. \tag{4.2}$$

λ is the relativistic de Broglie wavelength of the electron, e the elementary charge, U the acceleration voltage and $V(x,y,z)$ the specimen's potential distribution. The integration is carried out along the optical axis in thin slices through the specimen. The thinner the specimen is, the higher the acceleration voltage is and the lighter the atoms of the specimen are, the better this high energy approximation holds true. Then the phase shift imprinted information in the scattered wave function can be described as the projection of the specimen. The possibility to interpret TEM images as projections of the specimen offers the ability for information retrieval in the form of 3D reconstructions. However,

the optical system is not perfect. Its aberrations affect scattered and unscattered wave and thus the image formation in the detection plane. On its way to the detector the exit wave propagates through the objective lens's field and further down through the magnifying lenses of the projective system. Thereby, it accumulates geometric and phase distortions. The aberrations can be summed up to a point-spread function as in light microscopy, describing the blurring of a point to point mapping. The image wave function in the detection plane is given by the convolution of the scattering wave function directly behind the specimen (the so-called exit wave) with the point-spread function $h(r)$ of the microscope:

$$\psi_i = \frac{1}{M}\psi_s * h(r)\,, \tag{4.3}$$

where M is the magnification of the microscope and $*$ symbolizes the convolution operation. By applying the convolution theorem, the convolution is equivalent a multiplication in frequency space:

$$\tilde{\psi}_i = \frac{1}{M}\tilde{\psi}_s \cdot H(q)\,, \tag{4.4}$$

where $\tilde{\psi}_i$, $\tilde{\psi}_s$ and $H(q)$ denote the Fourier transforms of ψ_i, ψ_s and $h(r)$. q is the reciprocal distance, the spatial frequency in frequency space. The Fourier transform $H(q)$ of the point-spread function is called contrast transfer function (CTF) and is given by a phase factor composing the wave aberrations $W(q)$ and an envelope function $E(q)$ restricting and damping the information transfer at high spatial frequencies q:

$$H(q) = e^{-iW(q)}E(q)\,. \tag{4.5}$$

The wave aberrations cause an additional phase shift. Considering only the lower order aberrations, which have the greatest impact, while neglecting higher orders, the aberration function is given by (Reimer and Kohl, 2008):

$$W(q) = \frac{\pi}{2}(C_s\lambda^3 q^4 - 2\Delta z\lambda q^2)\,. \tag{4.6}$$

C_s is the spherical aberration of the objective lens and results in shorter focal lengths for electrons passing at high radial distances from the optical axis through the lens. The defocus Δz represents a slight decrease or increase of the focal length of the objective lens due to changes in the lens's current about its nominal value. The envelope function can contain the masking due to a small objective aperture and the influence of a partially coherent illumination due to a finite source size and finite energy spread. The detectable signal is the expectation value of the image wave function: $\psi_i\psi_i^*$. For a simplified description of the image wave function the weak-phase object approximation is introduced. Biological specimens consist mostly of atoms with low atomic number Z and they give rise to only small changes of the amplitude and phase of the illuminating plane wave. Thereby, they produce only weak contrast. In approximation they can be described as weak-phase objects with $\varphi_s(r) << 1$ and

as weak-amplitude objects with local amplitude variations $a_s(r) \approx 1$. $a_s(r)$ can be expressed by $a_s(r) = 1 - \varepsilon_s(r)$ with $\varepsilon_s(r) << 1$. As a consequence thereof the exponential function of equation 4.1 can be approximated by only the linear terms of its Taylor expansion. Additionally, all cross-terms are extremely small, so they are negligible too. The scattering wave function (Eq. 4.1) becomes by using the weak-phase object approximation (Reimer and Kohl, 2008)

$$\psi_s \simeq \psi_0 \left(1 - \varepsilon_s(r) + i\varphi_s(r)\right) . \tag{4.7}$$

With the Euler identity the exponential function of the wave aberration function in Eq. 4.5 can be written as $\cos(W(q)) - i\sin(W(q))$. Considering the approximations and accounting for only linear terms in amplitude and phase shift the intensity of the detector signal becomes (Reimer and Kohl, 2008)

$$I = \psi_i \psi_i^* = 1 - 2 \cdot FT^{-1}\left\{E(q)\tilde{\varepsilon}_s \cos(W(q))\right\} + 2 \cdot FT^{-1}\left\{E(q)\tilde{\varphi}_s \sin(W(q))\right\} , \tag{4.8}$$

with $\tilde{\varepsilon}_s$ and $\tilde{\varphi}_s$ being the Fourier transforms of ε_s and φ_s. FT^{-1} denotes the inverse Fourier transform. ψ_0 has been set to unity for simplification. The function $\sin(W(q))$ is called phase contrast transfer function (pCTF) and $\cos(W(q))$ amplitude contrast transfer function (aCTF). The resulting equation shows how the phase and amplitude changes induced by a weak scattering object are transferred into the image. Images of biological objects following a cryo preparation (cf. Section 4.2) are assumed to have less than 10 % amplitude contrast. They exhibit mainly phase information.

4.1.1 Corrections for the contrast transfer function

The previous section has shown that both, the amplitude and phase information, are modulated by a sinusoidal function influenced by the spherical aberration and the defocus. Nevertheless, the images contain projection information of the object density distribution. Dealing in a corrective way with the contrast transfer the high resolution information can be recovered in a subsequent 3D reconstruction. The contrast transfer function (CTF) leaves an imprint in the power spectrum of a recorded image. By computational analysis it can be estimated. The absolute square of the Fourier transform of a single image is also called periodogram (Jenkins and Watts, 1968). It shows only very weakly the CTF imprint. A power spectrum also calculates the expectation value of the absolute-squared Fourier transform, but it includes the averaging of a large number of images with the same statistical properties (Frank, 2006a). A power spectrum of a single image can be calculated by extracting smaller subimages with a certain degree of overlap and forming the average. The incoherent averaging of the amplitudes yields a better estimation of the CTF imprint, which appears as concentric rings. The so-called Thon rings denote the radially symmetric regions of no information transfer (Thon, 1971). Their inspection

serves as quality control, since they can reveal specimen drift or strong astigmatism. Both types of images should be disregarded. The rotational average of the power spectrum yields a one-dimensional curve that builds the basis for computational CTF parameter estimation. Especially the positions of the zeros are important. The more zeros are present in the data, the more reliable the estimation will be. The simplest correction of the CTF is done by a so-called phase flipping operation to account for the contrast reversals between pairs of CTF zeros. The phase-corrected Fourier transform of the image is given by (Frank, 2006a):

$$\tilde{I}'(q) = \begin{cases} -\tilde{I}(q) & \text{for } H(q) < 0, \\ \tilde{I}(q) & \text{for } H(q) > 0 \end{cases} \tag{4.9}$$

$\tilde{I}$ denotes the Fourier transform of the image. A complete correction would be given by a deconvolution. It can be described by a division of the image transform in Fourier space by the transfer function. Unfortunately, this cannot be done directly due to the CTF zeros and noise. For the spatial frequencies of the zero positions the Fourier components are not transferred into the image. Close to the zeros the information is weighted down. A direct deconvolution will lead to singularities and an amplification of noise. The method called Wiener filtering represents a feasible deconvolution approach. The phase- and amplitude-corrected image transform is then given by (cf. Frank, 2006a):

$$\tilde{I}'(q) = \frac{\tilde{I}(q)H^*(q)}{|H(q)|^2 + f} \tag{4.10}$$

The additive term f has to be chosen with regard to the spectral signal-to-noise ratio of the data and prevents excessive noise amplification close to the CTF zero positions. Often there are not enough zero positions resolved to estimate all parameters of the CTF. Therefore, often the envelope functions are summed up and are treated as an approximate overall amplitude decay towards high spatial frequencies. Saad et al. (2001) showed that a simple Gaussian function with width B adequately describes a cumulative envelope function:

$$E(q) = e^{-Bq^2} \tag{4.11}$$

Zhou and Chiu (2003) pointed out that although the dampening function is mathematically analogous to the crystallographic temperature factor, the physical causes are quite different. This cumulative B-factor is composed of experimental and computational causes. By applying a negative B-factor to the final 3D reconstruction the high-resolution contrast can be restored in order to reveal the structural features concealed in the map (Rosenthal and Henderson, 2003, Fernández et al., 2008).

4.2 Cryo microscopy

Transmission electron cryo microscopy (abbreviated cryo-TEM or cryo-EM) denotes the kind of TEM work with a special specimen preparation in order to preserve the biological object in its natural environment without the need of chemical fixation. As the specimen has to be transferred into the high vacuum system of the TEM, it cannot be just proteins in solution. Chemical fixation techniques lead to artifacts and prohibit obtaining high resolution information of the object (cf. Frank, 2006a). As reviewed by Henderson (2004) the birth of cryo-EM goes back to the work of Dubochet et al. (1982) and Adrian et al. (1984). They pioneered the specimen preparation to preserve biological objects in rapidly frozen amorphous ice, also known as sample vitrification. Purified proteins or other small biological objects can be prepared in a buffer, suiting their needs to be stable. A copper or molybdenum grid of 3 mm diameter serves as support. Typically it has a thin holey carbon film on top of it with a hole diameter ranging from 0.6 to 3.5 µm. A few µl of the sample solution are applied to the grid, which has often been treated in a plasma cleaner to become hydrophilic. Excess buffer is removed by blotting for a few seconds. The grid with only a thin liquid layer is rapidly plunged into a cryogen, usually liquid ethane. The high cooling rate of ethane, supercooled by liquid nitrogen, prevents the water to form crystalline ice (Stewart, 1990). Later on the grid can be stored in liquid nitrogen and be transferred into the microscope keeping the temperature well below -160°C to -175°C. At higher temperatures crystalline ice will appear.

4.3 Data acquisition schemes

In order to optimize the information retrieval from 2D images for a subsequent recovering of its original 3D information it is required to adapt the data acquisition with regard to radiation damage. The scattering density of proteins is only slightly higher than of vitreous ice ($\sim$1.36 : 0.97 g/cm^3 according to Penczek, 2010), thus the contrast formed by the proteins surrounded by amorphous ice is quite low. Reduction of shot noise can be achieved by increasing the number of applied electrons and will yield a better contrast, but at the same time a high electron exposure gives rise to artifacts owing to radiation damage and leads to loss of resolution. For light atoms like carbon, the main component of proteins, unfortunately the inelastic scattering cross-section for electrons is about three times higher than the elastic one (cf. Henderson, 1995, Reimer and Kohl, 2008). As reviewed by Baker and Rubinstein (2010) most beam damage occurs due to electrons that deposit an energy between $\sim$5 to $\sim$100 eV (on average $\sim$20 eV) in the specimen (Langmore and Smith, 1992). The energy suffices to excite or ionize valence electrons and thereby covalent bonds, which are of the order of a few eV, are broken and free radicals produced. In turn this causes the emission of secondary electrons. The cooling of

the specimen inside the microscope has no effect on the primary radiation damage. The secondary damage, caused by the free diffusion of secondary electrons that break more bonds and free radicals that begin to catalyze chemical reactions, can be reduced by cooling. The ice matrix is thought of to restrict their mobility (Henderson, 1990). The electron exposure has to be chosen in such a way to optimize the signal-to-noise ratio in the images at the desired resolution. It cannot be optimal for all resolutions at the same time. Depending on the kind of specimen two data acquisition schemes have arised.

4.3.1 Single particle approach

Protein structures are usually solved following a so-called single particle approach. It is based on the assumption that the object is available in many identical copies. Then like in X-ray crystallography, where the dose is partitioned throughout the crystal compound, the dose can be distributed over many images each containing one view of the object. Each subsequent image is taken from an identical, but not the same copy of the object. To recover a high-resolution protein structure it is recommended to use not more than 10 to 15 $e^-/Å^2$ per image (per particle) for an electron energy of 200 keV (Baker and Rubinstein, 2010). Data acquisition is done under so-called low-dose conditions, meaning all imaging parameters especially the focus are adjusted nearby the exposure position. Thereby, it is avoided to pre-irradiate the particle. Depending on the spectral signal-to-noise ratio present in the data set, many ten thousand single particle images may be needed in order to recover the object density properly. To achieve an optimized sampling of the object's structure factors two things have to be accounted for: There should be sufficient projection views in different orientations and the defocus has to be varied in order to have the suppressed information transfer due to the CTF zeros at varying positions in frequency space. To allow in subsequent data processing the determination of the particle's orientation they are supposed to have at least a molecular weight of 200 kDa and are sampled in a moderate defocus range. The contrast depends strongly on the adjusted defocus (cf. Eq. 4.5). Too close to focus the particles can barely be distinguished from the background, because the low spatial frequencies are strongly CTF suppressed and thus the alignment will probably fail. A very high defocus leads to a strong oscillating CTF that in general cannot be corrected properly and thus prohibits likewise a high-resolution recovery. The applied defocus is chosen always in the so-called under-focus.

4.3.2 Electron tomography

Biological objects that rarely exist in copies with identical structure exclude the use of any averaging technique (Frank, 2006b). Cell components with an entirely different size range like organelles or other subcellular structures typically belong to this category. The lack of redundancy allows a 3D

visualization only by means of electron tomography. Different projections are obtained by tilting the specimen inside the microscope with respect to the optical axis. To yield a sufficient orientation sampling of one and the same object a higher electron exposure has to be tolerated than following the single particle approach. Therefore, resolutions obtainable with electron tomography range between 20–60 Å, which is generally lower than for single particle EM (Baker and Rubinstein, 2010). The recommended exposures range between 20 and 100 $e^-/Å^2$ at 200 keV to optimize the signal-to-noise ratio between 25 and 100 Å (Baker and Rubinstein, 2010). Furthermore, it is suggested to have rather a slight over-irradiation than to underexpose the specimen. Electron tomography of small objects is difficult due to the resolution restriction caused by the unavoidable radiation damage.

4.3.3 Zero-loss energy filtering

The image formation theory as described in Section 4.1 considers only the elastic scattering of the electrons within the specimen. However, there is a significant amount of inelastic scattering (cf. Section 4.3). Simulations of inelastic images by Angert et al. (2000) show that for low spatial frequencies there can actually be a contrast reversal compared to the elastic scattering contrast. Furthermore, based on chromatic aberration the inelastic information is blurred. Zero-loss energy filtering can overcome the difficulties because it allows to obstruct the inelastically scattered electrons and thus to prevent them from adding unfavorable image contrast (Schröder et al., 1990). For this purpose the TEM has to be equipped with an electromagnetic energy filter which allows to separate the electrons according to their energies. By placing a slit into the so-called energy-dispersive plane one can select the elastically scattered electrons and mask out the inelastically scattered ones. In practice normally a slit width of about 10-15 eV around the zero-loss peak is chosen. It has been shown that the removal of inelastically scattered electrons leads to higher specimen contrast and this can be described in the CTF by an additional filter contrast (Angert et al., 2000).

4.4 3D reconstruction

Today there are many different software packages and 3D reconstruction algorithms that can be used to recover the original object density distribution. Reconstructions of three-dimensional structures from electron micrographs go back to the work of DeRosier and Klug (1968). Their reconstruction method is based on the Fourier-slice theorem. The theorem implies that the Fourier transform of a 2D projection equals a central slice of the 3D Fourier transform of the object. The 3D Fourier transform can be constituted slice by slice from all projection images. Fourier inversion yields the reconstructed object density. This method allows a direct consideration of the sampling of the Fourier space and

allowed Crowther et al. (1970) to make a statement of how many projections are required for a given final resolution. Obviously, the object's Fourier space is nonuniformly sampled. Close to the origin, where all slices intersect, the spatial frequencies are overrepresented, while high spatial frequencies are underrepresented due to a larger separation of the central slices. When data acquisition gives rise to a missing volume in Fourier space, that could not be sampled, e.g. in electron tomography where a 90° tilt is impossible, the resolution will be inferior in that particular direction and gives rise to artifacts in real space. In the following reconstruction algorithms that are used in this work are introduced. They are based on the approximation of the linear image formation theory.

4.4.1 Weighted backprojection

The weighted backprojection is an empirical method that reverses the linearly assumed process of projection formation. It has become a standard reconstruction method and can handle huge data sets as well. A backprojection is the inverse operation of a projection. The 2D image is transformed into a volume by a translation or dilation of the image perpendicular to the image plane, forming a so-called backprojection body p_j^b. Mathematically it can be expressed by a convolution of the projection p_j in the plane (x_j, y_j) with the 3D spread function l_j (Rademacher, 2006):

$$l_j = \delta(x_j, y_j)c(z_j) \quad \text{with} \tag{4.12}$$

$$c(z_j) = \begin{cases} 1 & \text{for } -a \leq z_j \leq +a \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

$$p_j^b(x_j, y_j, z_j) = p_j(x_j, y_j) * l_j(x_j, y_j, z_j) \tag{4.14}$$

δ is the Dirac's delta distribution and a denotes the width of the spread. A simple backprojection b is formed by the summation of all backprojection bodies:

$$b(x, y, z) = \sum_j p_j^b(x_j, y_j, z_j) \tag{4.15}$$

The algorithm of the simple backprojection is linear and shift-invariant and allows a description of its performance with the concept of a point-spread function or a transfer function in frequency space (Rademacher, 2006). Following that $b(x, y, z)$ is not yet the original object density distribution $o(x, y, z)$, instead it is the object density convolved with the point-spread function $h(x, y, z)$ of the algorithm, that contains all errors the simple summation brings about:

$$b(x, y, z) = o(x, y, z) * h(x, y, z) \tag{4.16}$$

The point-spread function can be assessed by considering the backprojection of an ideal point object $\delta(x, y, z)$. In the ideal case its Fourier transform, the transfer function $H(q_x, q_y, q_z)$, can be expressed

analytically and fulfills the relation $H \neq 0$. Then the original object can be recovered by a deconvolution:

$$o(x, y, z) = FT^{-1} \left\{ \frac{B(q_x, q_y, q_z)}{H(q_x, q_y, q_z)} \right\} \tag{4.17}$$

B denotes the Fourier transform of the backprojection b and $o(x, y, z)$ is the weighted backprojection of the object density distribution. In the simplest case, for an axial projection geometry without missing orientations the transfer function becomes by approximation (Rademacher, 2006)

$$H(R) = \frac{1}{R} \qquad \text{and} \qquad W(R) = R \tag{4.18}$$

The inverse transfer function is the corresponding weighting function $W(R)$. In this case it is the radius R in Fourier space perpendicular to the tilt axis. In a more general case the weighting can be applied by Fourier filtering of the simple backprojection by an appropriate weighting function.

4.4.2 Filtered least squares

The algebraic least squares method is based on a strict mathematical derivation within linear theory. Reconstruction of the three-dimensional structure density from projections is a case of general inverse problems. Crowther et al. (1970) proposed already in 1970 a direct, non-iterative algebraic reconstruction method, which is based on a least squares solution of the set of projection equations. At that time it turned out to be computationally impracticable, because of the large computing power the application of this method requires. In 2003 Holmes et al. (2003) showed the nowadays applicability to 3D data reconstruction for an axial projection geometry. The diploma thesis (Wulf, 2009), that precedes this work, presents a thorough analysis of the filtered least squares algorithm in the implementation of K.C. Holmes (Holmes and Schröder). It was investigated for model data and now should be applied to real micrographs (cf. Chapter 6). In order to analyze the reconstruction problem in real space one has to define the sought-after object density on a grid and to formulate a set of linear projection equations, which relate the object volume and the projection image data with the sum of projected gridpoints as observations. Since in electron microscopy one typically deals with ill-conditioned linear systems an analogous proceeding to a Moore-Penrose generalized inverse can be used to solve the least squares approach of the projection equations. The method of inverting ill-defined matrices was described by Diamond (1969). 3D reconstruction as an eigenvalue problem was first analyzed by Klug and Crowther (1972).

Like iterative algebraic reconstruction methods the filtered least squares algorithm starts with setting up the projection equations, which are then to be solved to determine the structure density. The continuous object density is represented by the values x_j, which it takes on a regular grid. Since each projection of the object should be integrated with the same uniform spacing, in order to maintain

equal weighting of the observed data, some form of interpolating function is needed to represent the density x'_j between lattice points (Crowther et al., 1970). In the present implementation lattice points convolved with a Gaussian function of suitable width are used for this purpose (cf. Fig. 4.1). Some analysis during the diploma thesis (Wulf, 2009) has shown that a standard deviation of 0.5 is an optimal value. The observables σ_i (pixels of the projection images) are given by the sum of a subset

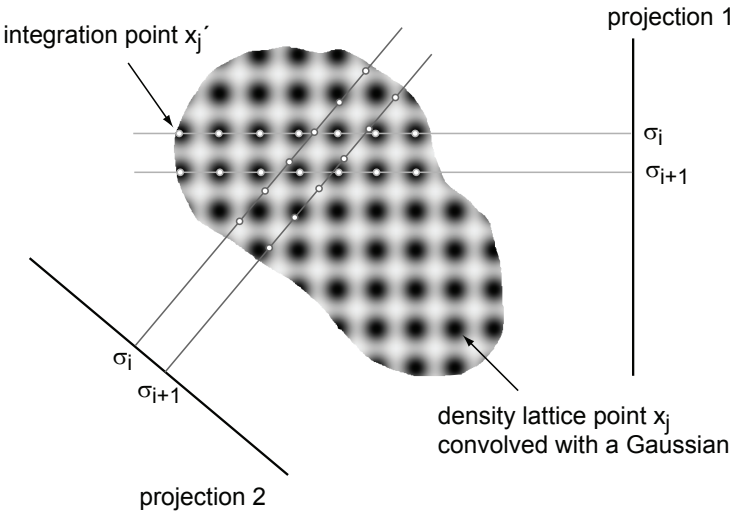


Figure 4.1: 2D illustration of how the projection equations are set up. σ_i are the observables (pixels of the projection images), each σ_i leads to one projection equation.

of density integration points x'_j, which then have to be expressed by interpolation in terms of nearby lattice points x_j. The projection equations can be written as

$$\sum_j h_{ij} x_j = \sigma_i \tag{4.19}$$

where h_{ij} are the weightings, which are determined by the direction of projection and by the Gaussian interpolation. Writing them in form of a matrix equation and more correctly according to experimental data with observables containing errors ($\sigma_i \to b_i + \varepsilon_i$) the projection equations are given by

$$Hx = b + \varepsilon \tag{4.20}$$

where H is the projector matrix, x the vector containing all unknown density points and $b + \varepsilon$ the vector containing all observables. The classical least squares solution that minimizes the sum of the squares of errors $\varepsilon^T \varepsilon$ (cf. Penrose, 1956) is given by

$$H^T H x = H^T b \tag{4.21}$$

$H^T H$ is the normal matrix. If its inverse $(H^T H)^{-1}$ exists, it would directly provide the solution x. Instead, the normal matrix typically will be multiple singular and cannot be inverted directly. As was suggested by Crowther et al. (1970) one can diagonalize the normal matrix and subsequently invert it in the subspace of significant eigenfunctions.

In this thesis a completely new mathematical matrix notation could be derived. The sought object density can be expressed by the following matrix equation: Let λ_i denote the eigenvalues of the symmetric normal matrix, sorted in decreasing order, and let v_i be a set of orthonormal eigenvectors of $H^T H$ (such that $H^T H v_i = \lambda_i v_i$ for all i), then the least squares solution of Eq. 4.20 is given by

$$x = \sum_i \frac{1}{\lambda_i} \left(v_i^T \cdot H^T b \right) v_i \tag{4.22}$$

The inclusion of very small eigenvalues λ_i into the inversion of the normal matrix leads to subsequent amplification of any errors present in the data resulting in an inaccurate determination of the density. Therefore, in practice there is a lower limit for the minimum eigenvalue for any matrix inversion and thus one has to truncate the series in Eq. 4.22. The numerically posed truncation of the inversion acts as a filter. Filtering by truncation of a series to obtain a filtered least squares solution was proposed in the context of Fredholm integral equations (cf. Lee and Prenter, 1978, Luecke and Hickey, 1984). Let $\delta > 0$ and $\lambda_i > \delta$ then

$$x = \sum_{i \in \{i:\lambda_i > \delta\}} \frac{1}{\lambda_i} \left(v_i^T \cdot H^T b \right) v_i \tag{4.23}$$

is the filtered least squares solution of Eq. 4.20 with respect to the size of eigenvalues. Another possibility is filtering according to the spatial frequency content of eigenfunctions. The projection equations (Eq. 4.20) are then solved in the subspace of significant eigenvectors. This subspace can be determined by considering the Crowther criterion (Crowther et al., 1970). It gives the maximum spatial frequency $\mathrm{R_{Crowther}}$ that can be achieved in the reconstruction for a given projection geometry. Thus, the relevant subspace can be selected by analyzing which eigenfunctions contribute significant frequencies inside the Crowther frequency R_{Crow} (and inside Nyquist frequency* R_{Nyq} to avoid artifacts). The superposition of these relevant eigenfunctions leads to the filtered least squares solution with respect to spatial frequencies. The derivation and implementation of this solution with respect to spatial frequencies was achieved during both the preceding diploma thesis and this thesis.

To point out how the object density x is being determined one has to look at the different components of the newly obtained Eq. 4.23. Since H is the projector matrix $H^T b$ is the simple backprojection

*The Nyquist frequency is the highest spatial frequency contained in the data at a given sampling distance, defined by the pixel size p (Sampling theorem). $R_{Nyq} = 1/(2p)$

(Holmes and Schröder). The inner product in brackets is the projection of the two vectors onto each other and thus it is the test of how much of each eigenvector is contained in the simple backprojection. It gives scalar values telling how much of each eigenfunction v_i to put into the reconstruction. Thus, the new matrix notation illustrates the reconstruction process in a quite descriptive manner. Division by the appropriate eigenvalue λ_i acts as the geometrical correct weighting (Holmes and Schröder). The range of the index i defines which eigenvectors v_i are superposed for the reconstruction or in other words how the least squares solution is being filtered. A mask is used to define the area within which the density of the object is sought. If there is no knowledge about the shape of the object, one can use a circular mask. Whereas if one has some knowledge about it, one can use a geometrical suited mask. The use of the mask will reduce the size of the normal matrix, which has two effects. The calculation of the eigenvalues and eigenfunctions is faster and the resolution obtainable from the given data is improved since one seeks to solve for a smaller number of variables.

There are close parallels to the weighted backprojection. Thorough analysis during this work and the preceding diploma thesis showed that eigenvalues corresponding to low order eigenfunctions are large and fall off monotonically with higher orders. Thus, the filtering of the simple backprojection, being present in both methods, is analogous to the R-weighting described in Section 4.4.1 (Gilbert, 1972). The significant difference is pointed out by Holmes and Schröder: "The least squares method uses custom-made orthogonal functions for computing the transform of the back projection" and a geometrical correct weighting. Thus, it "yields the best possible solution for the given geometry without any ad hoc assumptions" (Holmes and Schröder). The present implementation is restricted to a single axis tilt geometry because it makes use of the dimensional reduction from 3D to 2D. A strict single axis tilt geometry allows to build up a volume by independently reconstructed slices. Even so it is restricted to small projection images because the required memory to set up the normal matrix for the 2D problem scales with the power of four with the image dimension. For instance, the attempt to reconstruct images with 513 x 513 pixels will require about 130 GB random access memory. An image size of 161 x 161 pixels, that has been used for the application of the method in this thesis, already gives rise to a normal matrix of size 17645 x 17645. The diagonalization is calculated using the linear algebra package LAPACK (Anderson et al., 1999). Because the matrix is determined by the projection geometry and not by the object, a previously calculated diagonalization can be reused if only the object changes. The preceding diploma thesis showed that the least squares method has the potential to optimize information retrieval from projection images compared to the weighted backprojection method (Wulf, 2009). For simulated data sets especially the density localization could be recovered with a higher accuracy, leaving the neighboring pixels / voxels less affected.

4.4.3 Iterative helical reconstruction

The studies of helical disorder (Egelman et al., 1982), which showed that for some isolated helical polymers there are no forces or factors to maintain a long-range order and that disorder can arise over rather short distances, led to the development of the iterative helical real space reconstruction (IHRSR) algorithm (Egelman, 2000, 2007). The method is based on the weighted backprojection combined with an iterative refinement of the helical symmetry. Small segments of the projection images of the helical filaments are treated as individual single particles. Projections of a reference structure are used for a multi-reference alignment of the segments. The initial reference can be for instance a featureless cylinder with the diameter of the object to be reconstructed. A first estimation of the object density is obtained by backprojection. The helical symmetry of the volume can be determined by means of least squares fitting. Those symmetry that minimizes the density variations along the helix is subsequently imposed to the volume. The symmetrized volume in turn can be used as reference for the next iteration. Until convergence the orientation parameters of each segment and the helical parameters of the volume are refined. Recently a revision of the IHRSR algorithm has been published (Behrmann et al., 2012b). It is implemented in the SPARX single particle image processing environment (Hohn et al., 2007) and enables parallelized execution.

4.5 Fourier shell correlation

The Fourier shell correlation (FSC) is a method to compare two reconstructed volumes in Fourier space. When f and g denote two volumes and F and G its Fourier transforms the mathematical formula is given by (Penczek and Frank, 2006)

$$\mathrm{FSC}(f,g,R) = \frac{\sum_{|\|X_n\|-R|\leq\varepsilon} F(X_n)G^*(X_n)}{\left[\left(\sum_{|\|X_n\|-R|\leq\varepsilon} |F(X_n)|^2\right)\left(\sum_{|\|x_n\|-R|\leq\varepsilon} |G(X_n)|^2\right)\right]^{\frac{1}{2}}} \tag{4.24}$$

The width of the Fourier shells is 2ε, the coordinates X_n form a regular grid in reciprocal space and $R = \|X_n\|$ denotes the size of the spatial frequency. The FSC yields a one-dimensional curve of correlation coefficients as function of the Fourier radius R. Coefficients close to one depict a high resemblance of the two volumes, close to zero they denote a lack of any similarity at all. One possibility to calculate the FSC is to split the data set into halves, reconstruct them independently and compare them in Fourier space. A direct relation between the FSC and the spectral signal-to-noise ratio (SSNR) exists, which is given in the case of a data-halves test by (Unser et al., 1987)

$$\mathrm{SSNR}(R) = 2 \cdot \frac{\mathrm{FSC}(R)}{1-\mathrm{FSC}(R)} \tag{4.25}$$

This relation allows to interpret the FSC in the sense of a resolution criterion. The spatial frequency for which the spectral signal-to-noise ratio drops below a pre-defined value is determined (Penczek and Frank, 2006). There is more than one value used in the electron microscopy community. Often the resolution of a reconstructed volume is stated with the reciprocal spatial frequency $1/R$ in Ångström using the FSC(R)=0.5 criterion. It corresponds to a SSNR=2 at the cut-off resolution. As Ludtke (2011) points out it is often used due to is reliability. The FSC curve at lower values can be very sensitive to artifacts that can be introduced during the processing (e.g. by overfitting). Such artifacts will have only minimal impact on the resolution based on the 0.5 criterion. A FSC(R)=0.33 criterion corresponds to interpreting the data up to the spatial frequency of equal signal and noise strengths (SSNR=1). Another often used criterion is FSC(R)=0.143 which corresponds to the threshold value used in X-ray crystallography (Rosenthal and Henderson, 2003). Fernández et al. (2013) emphasize that in the absence of artifacts (e.g. overfitting) the 0.5 criterion can severely underestimate the resolution. For the application to electron microscopic data the FSC reflects mostly phase errors. A name like measuring phase consistency would therefore be more appropriate than calling it resolution, since the context it is used in is completely different from the one e.g. optics uses it (Penczek and Frank, 2006).

Chapter 5

Molecular dynamics simulation

This chapter gives an introduction to molecular dynamics simulations and how they can be used to combine the results obtained by cryo-TEM and X-ray crystallography. Recently Karplus, Levitt and Warshel have been awarded for the "development of multiscale models for complex chemical systems" with the Nobel prize in Chemistry (Royal Swedish Academy of Sciences, 2013). It emphasizes the great impact computational modeling has. They combined classical physics for simulating large proteins with quantum theoretical calculations to account for chemical reactions in the catalytic center. Molecular dynamics (MD) methods have been developed in the theoretical physics community during the late 1950s and with increasing computing power and theoretical and methodological developments over the years they are now widely used in the fields of material science, (bio-)chemistry and biophysics. For the first time McCammon, Gerlin, and Karplus (1977) applied MD simulations to proteins. Since then it has been used frequently to study biomolecular phenomena like conformational properties of macromolecular complexes, transport of small molecules, ligand docking, homology modeling, drug design and even ab initio prediction by simulating protein folding. Structures yielded by X-ray or NMR methods are often subject to MD refinement (Adcock and McCammon, 2006). For the structural analysis of macromolecules quantum effects, relevant on the timescale of chemical reactions, can be neglected and classical physics can be used to describe the relevant biomolecular force fields.

5.1 Classical molecular dynamics

Neglecting electronic motion and quantum effects in general the classical approach of molecular dynamics describes the motion of the atomic nuclei in a many-body system by solving the Newtonian equations of motion. After Newton described 1687 in his *Philosophiae Naturalis Principia Mathe-*

matica the basic concepts about force and motion it was Laplace (1820) mentioning the idea of linking structural states of the past and future by understanding and animating all the nature's forces. Following this Schlick (2010) "termed MD as Laplace's vision of Newtonian mechanics on supercomputers". Modern programs numerically integrate the Newtonian equations of motion. To keep the numerical algorithm stable the integration time step "must be an order of magnitude smaller than the fastest motions of the system" (Scheraga et al., 2007). This are normally bond vibrations in the order of 10 femto-seconds (fs), so that the integration time should be chosen of the order of 1 fs. Simulation systems can contain up to the order of 10^6 atoms and in order to cover relevant molecule dynamics and reach equilibration of the system the simulation should cover a time span in the order of nanoseconds. To limit the required computing time per time step typically simple mathematical functions (e.g. harmonic oscillators) are used to describe the molecular forces, while the atoms and their bonds are modeled by spheres connected by springs (Adcock and McCammon, 2006). However the description must faithfully represent the system, therefore "the potential energy, represented through the MD "force field", is the most crucial part of the simulation" (Phillips et al., 2005). The Newtonian equations of motion can be written with the force F expressed by the total potential energy U_{total}

$$\vec{F_i} = m_i \ddot{\vec{r}_i} = -\frac{\partial}{\partial \vec{r_i}} U_{total}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N), \qquad i = 1, 2, \ldots, N, \tag{5.1}$$

where m_i is the mass of atom i and $\vec{r}_i$ is its position. Based on the forces calculated from the potential energy the position and velocity of every atom are propagated in time, forming a trajectory and providing a dynamical description of the system. As trade-off between accuracy and low computational costs or ease of parameterization the formulation of the potential energy function is composed of additive energy functions and the non-bonded interactions of atoms are calculated as sum of contributions of two interacting atoms, usually called pair potentials, the simplest way to describe many interacting particles. Electrostatic interactions are modeled as interactions between point charges. Normally only the dipole moment is considered. Higher moments and polarization effects are neglected (Scheraga et al., 2007). NAMD (Phillips et al., 2005), the MD simulation program used in this work, sets up the total potential energy as described by Phillips et al. (2005) using following contributions:

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{Coulomb} \tag{5.2}$$

The first three terms describe the bonded interactions. Harmonical or sinusoidal functions are used for the bond stretching, bending and torsion,

$$U_{bond} = \sum_{bonds\, i} k_i^{bond}(r_i - r_{0i})^2, \tag{5.3}$$

$$U_{angle} = \sum_{angles\, i} k_i^{angle}(\theta_i - \theta_{0i})^2, \tag{5.4}$$

$$U_{dihedral} = \sum_{dihedral\, i} \begin{cases} k_i^{dihe}[1 + \cos(n_i\phi_i - \gamma_i)], & n_i \neq 0 \\ k_i^{dihe}(0_i - \gamma_i)^2, & n = 0 \end{cases} \tag{5.5}$$

with the covalent bond length r and the backbone angles θ and dihedral ϕ as depicted in Fig. 5.1. The

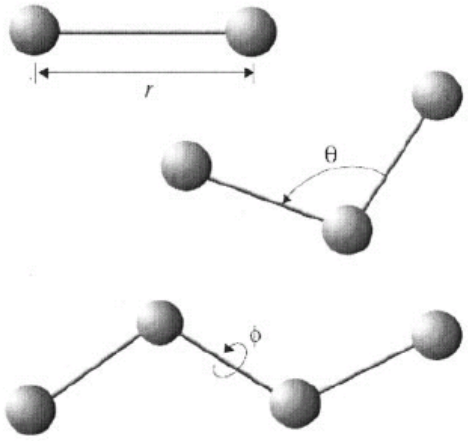


Figure 5.1: Coordinates for the description of bonded interactions in MD simulations: r is the length of a covalent bond, θ represents the bond angle and ϕ governs the dihedral angle (adapted from Phillips et al. (2005)).

non-bonded interactions are accounted for in Eq. 5.2 by a Lennert-Jones potential approximating the van der Waal's interaction U_{vdW} and a Coulomb potential $U_{Coulomb}$ for the electrostatic interaction, both are pair potentials:

$$U_{vdW} = \sum_i \sum_{j>i} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{5.6}$$

$$U_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}. \tag{5.7}$$

Via proper adjustment of the parameters the functions can describe biological molecules with high accuracy (Mackerell, 2004). For the ease of calculation force field parameter files are used for the bonded interactions. A proper determination of these parameters is quite important and is generally

accomplished through quantum mechanical calculations as well as empirical techniques (Phillips et al., 2005). The non-bonded interactions have a non-local nature, are long-range interactions and hence exist between every pair of non-bonded atoms in the system. Therefore, its calculation limits the speed of MD simulations. There are different approaches to handle this problem. One possibility is to use multiple time stepping. The non-bonded interactions are calculated then less often than the bonded ones, e.g. every 2 fs instead of each 1 fs. Another possibility is to use a cut-off radius, e.g. 10 Å, to spatially truncate the long-range interactions. This speeds up the calculations enormously and is often used in simulations of molecules in vacuo, in which additional to the cut-off the natural environment of the molecules is omitted intentionally so that the number of atoms is decreased enormously as well. A more accurate model of the biological system includes the molecule's surrounding, in most cases water with ions. Explicit solvent simulations use periodic boundary conditions to avoid surface effects at the boundary of the simulated system and to make use of Fourier methods. Periodic boundary conditions are achieved by replicating the system to infinity via periodic translations. The full, non-truncated non-bonded interactions can then be calculated using Ewald summation (Ewald, 1921). It "is considered more reliable than a cut-off scheme" (Phillips et al., 2005). A fast numerical approach to compute the Ewald sum is the particle-mesh Ewald algorithm (Darden et al., 1993). Nevertheless, it is much more time-consuming and requires a better computational infrastructure. A compromise in accuracy and computational costs are implicit solvent methods which treat the solvent as continuum. The molecule is immersed in a dielectric and collisions of the solute molecule with its environment can be mimicked by introducing friction and collision forces as described by Scheraga et al. (2007). The Newtonian dynamics turn into Langevin or Brownian dynamics and Eq. 5.1 becomes a stochastic differential equation:

$$\vec{F}_i = -\nabla_{\vec{r}_i} U(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) - m_i \gamma_i \dot{\vec{r}}_i + \vec{R}_i(t), \qquad i = 1, 2, \ldots, N, \tag{5.8}$$

with the potential force $-\nabla_{\vec{r}_i} U$, the velocity $\dot{\vec{r}}_i$ of atom i, the friction coefficient γ_i and random forces $\vec{R}_i(t)$ arising from collisions. Xia et al. (2002) describe that compared to simulations in vacuo using implicit solvent methods a higher percentage of backbone angles will be in the most favorable regions and the structures will be more similar to those from explicit solvent treatment.

5.2 MDFF – combining cryo-TEM and X-ray crystallography – hybrid microscopy

In the context of this work MD methods can be used to combine cryo-TEM and X-ray crystallography results to yield a conformation of X-ray structures representing the functional state captured in the

TEM. Known stand-alone structures can be used as initial structures of macromolecular complexes to simulate a trajectory converging to the functional state represented by the less detailed EM density of the complex. It is worth to explore this hybrid approach when the functional complex cannot be crystallized but all the contributing molecules or cofactors have been solved by themselves by means of crystallography or other methods with high resolution. The development of molecular dynamics flexible fitting (MDFF) by Trabuco et al. (2009) addresses this problem. They extended the NAMD software by two additional terms. NAMD has a *gridForce* feature that allows to add to any simulation an arbitrary potential defined on a three-dimensional (3D) grid. The EM density $\Phi(\vec{r})$ is converted into a potential energy function U_{EM}:

$$U_{EM} = \sum_i w_i V_{EM}(\vec{r}_i) \qquad \text{with} \tag{5.9}$$

$$V_{EM}(\vec{r}) = \begin{cases} \xi \left[1 - \frac{\Phi(\vec{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right] & \text{if } \Phi(\vec{r}) \geq \Phi_{thr}, \\ \xi & \text{if } \Phi(\vec{r}) < \Phi_{thr}. \end{cases} \tag{5.10}$$

The solvent contribution of the map is removed by introducing a threshold value Φ_{thr}, yielding a flat potential in the solvent region. The scaling factor ξ uniformly adjusts the strength of the new EM *gridForce* during the MD simulation. The weighting factor w_i can be defined per atom and is generally set to the atomic mass. It "avoids strong differences in the acceleration of atoms due to mass disparities, ensuring stability of the simulation" (Trabuco et al., 2009). The non-physical potential energy U_{EM} drives the atoms in the cryo-EM density by applying forces proportional to the gradient of the EM density. High density regions of the map correspond to low energy regions of the potential. The potential function U_{total} (Eq. 5.2) preserves the correct physical parameters of the model during simulation. Due to the flat distribution of U_{EM} outside the EM density of the molecular complex and the fact that the initial model structure might have structural elements inside as well as outside the EM density there is the need to ensure integrity of the secondary structure elements such as α-helices and β-sheets. Otherwise they might be teared apart. Therefore, Trabuco et al. (2009) introduced another potential function U_{ss} to keep the integrity of the secondary structure, whose strength can likewise be tuned for different MD runs.

For the interpretation of MDFF results one should keep in mind some limitations of the method: Based on the non-physical potential functions only rearrangements of structural elements can be simulated – the class of structural folds, i.e. opening up of α-helices, cannot be simulated. In most cases such rearrangements are expected for different functional states, as has been described for the known myosin structural states in Section 2.3.2. Already existing secondary structure might change only by turning off U_{ss}. More general, MD methods as described in Section 5.1 and used in this work optimize

potential energy. Considerations on free energy of the system and its differences are not included in the underlying classical description. Properties known to contribute to structural favorable configurations such as conformational entropy of the polypeptide chain or hydrophobic effects are not accounted for.

Part II

Optimizing information retrieval from 2D TEM images to enhance the interpretable resolution of the reconstructed 3D object

Chapter 6

Reconstruction by filtered least squares

Analysis of the filtered least squares reconstruction algorithm suggested that the density localization in the reconstructed object will have a higher accuracy compared to applying weighted backprojection (Wulf, 2009). Thus, it offers the potential to optimize the density recovery from the original object. This chapter describes the attempts to use the method to reconstruct the actomyosin complex from TEM images. The present implementation of the filtered least squares reconstruction algorithm allows to reconstruct the symmetry unit of a helical filament, from which a helix can be build up considering its screw symmetry. A filament with helical symmetry lying straight and plane in the image plane resembles actually an axial tomogram as it offers with each symmetry unit a new, symmetry related view of its structure. The implementation makes use of this. Out of one single image a corresponding volume can be reconstructed. First of all, the sample preparation and data acquisition are explained which have been used to obtain the data set, which was used to implement the data processing pathway. It is followed by a description of how to prepare the data prior to the reconstruction and how to reconstruct a volume with the newly implemented algorithm. The main part of the chapter deals with the implementation and testing of the alignment procedure of the individual volumes to obtain the final result. Contrary to the weighted backprojection, where one volume is been reconstructed from all aligned images, not the images are subject to alignment. Instead the reconstruction of single volumes precedes the alignment, which is then carried out in 3D for this method. Concluding the applicability of the filtered least squares reconstruction in the present implementation is discussed.

6.1 Materials and methods

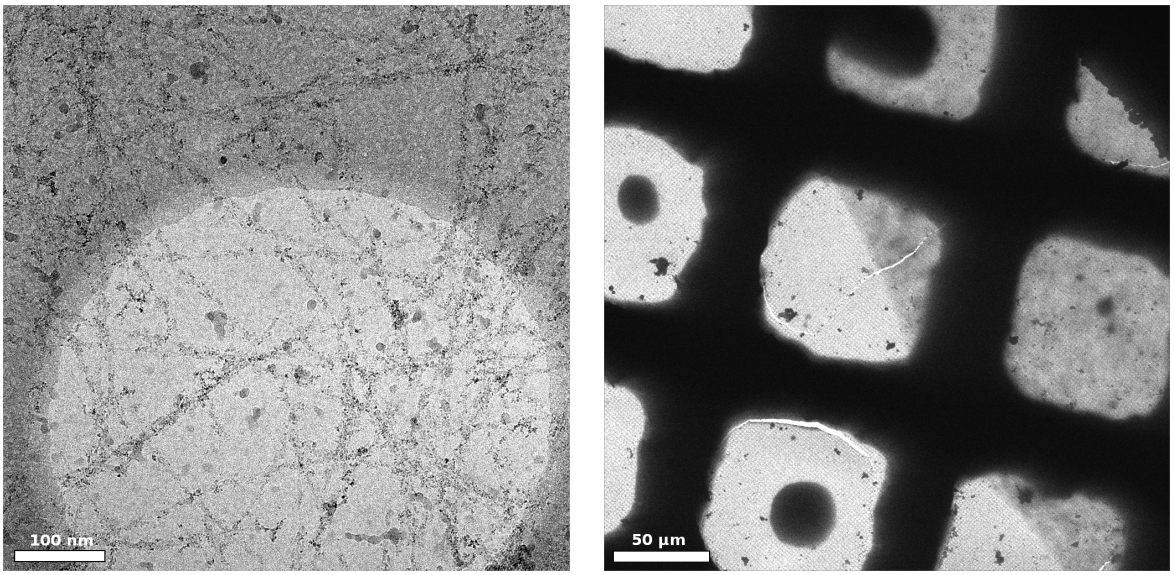
6.1.1 Sample preparation

Actin was purified from rabbit skeletal muscle. The polymerized f-actin has been kept in 5 mM Tris buffer (organic compound tris(hydroxymethyl)aminomethane with the formula $(HO\text{-}CH_2)_3C\text{-}NH_2$) with pH-value of 7.5 containing 60 mM KCl, 2 mM $MgCl_2$, 0.2 mM $CaCl_2$, 1 mM NaN_3 and 0.2 mM ATP. The filaments have been stabilized with 200 µM phalloidin to prevent depolymerization. The solution can be stored on ice at 4 °C for about two weeks. Myosin V was obtained by recombinant DNA technology. It originates from chicken (*Gallus gallus*). The construct is the same as analyzed by X-ray crystallography (nucleotide-free myosin V, PDB entry 1OE9 (Coureux et al., 2003)) and was kindly provided by Prof. Dr. H. Lee Sweeney (University of Pennsylvania, Philadelphia, USA). The construction, expression in insect cells and purification was done by the Sweeney Lab. From the messenger ribonucleic acid (RNA), which encodes the protein's amino acid sequence, a complementary DNA can be obtained. It was truncated at the codon corresponding to the residue 792 to obtain a construct containing only the motor domain and the first IQ motif (also referred to as myosin V-1IQ). The fragment consists of the residues 1-792. A DNA sequence of a polypeptide protein tag, a so-called FLAG-tag, was added at the truncated C-terminus. The FLAG-tag consists of seven positively and negatively charged amino acids plus one with neutral charge and is used for chromatographic purification of the protein after overexpression. The one light chain bound to the truncated lever arm is the human essential light chain LC-1sa (a fragment of 150 amino acids with the residues 59-208) and has been co-expressed with the myosin V-1IQ in insect cells (De La Cruz et al., 1999). The protein has been shipped in May 2010 with a high concentration of 9.8 mg/ml = 106 µM. Since the solution contains 50 % glycerol it can be stored at −20 °C in the freezer for very long times. The aqueous dilution buffer used to obtain arbitrary actin and myosin concentrations for the sample preparation contains 5 mM Tris with pH-value 8.0, 10 mM KCl and 0.5 mM $MgCl_2$.

The preparation of TEM grids with acceptable quality in protein concentration, ice thickness and full decoration of f-actin with the myosin fragment turned out to be quite difficult. Often the quality was far from acceptable. Some of the problems are summarized in the following. There is a general connection between protein concentrations, ice thickness and quality of decoration. Two extremes exist. Low filament concentrations leads to thin ice layers, that are necessary for transmission imaging, but often the ice layer contains no filaments at all or only very few undecorated f-actin filaments. Sometimes there even is no ice layer at all. High filament concentrations often give rise to bundles of filaments resulting in a thick ice layer crowded with far too many filaments. In this case the decoration

often is very good. Both cases cannot be used for data acquisition. In order to find a compromise, resulting in non-overlapping and fully decorated actin filaments contained in an amorphous ice layer with moderate thickness, many parameters have extensively been tested. The following parameters can be varied: the protein concentrations, the protein ratio, the decoration time, the grid type, the time and power of the glow discharge of the grid, the applied volume of protein solution to the grid, the blotting paper to remove excess buffer, the blotting time, the blotting force and the temperature and humidity inside the blotting chamber. The proteins are diluted to the desired concentrations separately. They can be mixed in a small reaction tube prior to the application onto a TEM grid (pre-mixed) or they can be applied one after another to the grid (on-grid decoration). The decoration time for the pre-mixed decoration can be up to half an hour or one hour, for the on-grid decoration it is within the range of 2–10 minutes. In both cases the myosin is pipetted to the actin. The ratio of actin and myosin should exceed the theoretical ratio of 1:1 and sometimes has been chosen with a high excess of myosin (up to 1:20). Different TEM grid types have been tested. Commercial grids with a thin holey carbon film with different hole sizes (0.6 to 3.5 µm hole diameter) have been used as well as self-made holey carbon films with irregular forms and spacings (so-called lacey films) on top of a copper grid. The mesh size of the copper grid is offered with values of 200, 300 and 400. It affects the blotting behavior likewise. Prior to the application of the protein solutions to the grid, the grid can be made hydrophilic by means of glow discharge / plasma treatment in low pressure nitrogen-oxygen atmosphere. The power, duration and pressure of glow discharge control the degree of hydrophilicity of the holey carbon film. Without glow discharge the film can hardly be wet and subsequent blotting will take away almost all applied proteins and buffer. Strong glow discharge leads to larger remaining volume after blotting and subsequent to thicker ice layers. It turned out also that myosin tends to stick to the hydrophilic carbon film and thereby does not bind to f-actin. The charged purification tag (FLAG-tag) may have an impact on the bundling of decorated filaments which has been seen often at higher filament concentrations. Unfortunately, it cannot be cut easily after purification in order to test its impact. The applied volume of total protein solution ranges from 2-4 µl. The blotting parameters are device-dependent. Both a FEI Vitrobot Mark III and the newer Mark IV have been used to vitrify the samples. Only the Mark IV allows to set a blotting force. Blotting paper has been tested from high absorbent to quite hard and low absorbent. The temperature inside the blotting chamber has been adjusted both to 4 °C and to room temperature (about 22 °C). The humidity is always set to 100 %. As some of the parameters are correlated and not independent in the way they affect the specimen conditions it is a difficult undertaking. Additionally, the whole procedure is far from reproducible. Gradually it became apparent that the age of the grids must not to be underestimated. With time the grids become more hydrophobic. The wetting properties change in a way not to be compensated satisfactorily by glow discharge. For

instance, grids out of a freshly delivered package have been prepared. About two months later four times the protein concentration was needed to have approximately the same filament amount inside the ice layer in the holes of the carbon film. A combination of a small applied volume with higher blotting force and shorter blotting time has been tested as well as a larger applied volume combined with lower blotting force and longer blotting time. Figures 6.1 and 6.2 show a few images of various attempts of sample preparations, that have not yielded the desired quality and thus demonstrate some of the discussed problems. The samples that have been used for data acquisition have been prepared

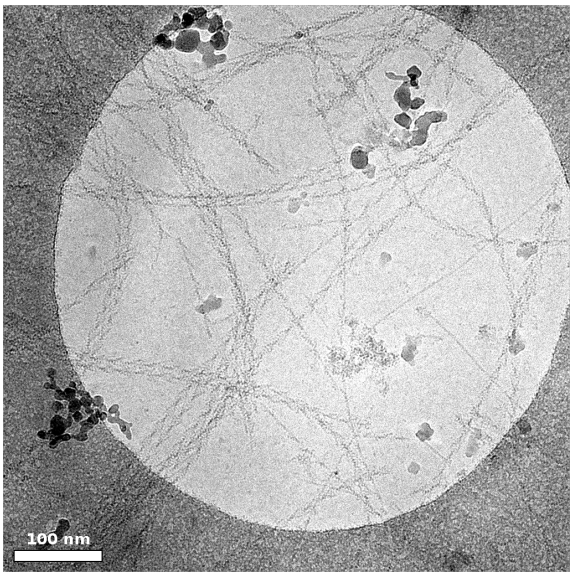


(a) quite bad protein quality

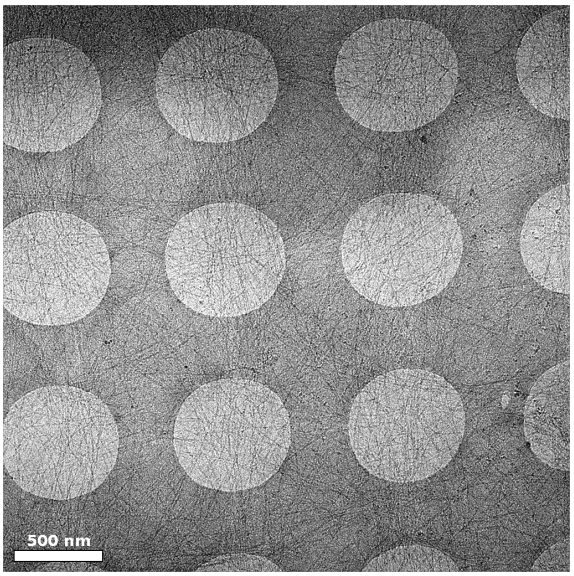
(b) crowded proteins in darker squares

Figure 6.1: The images depict some of the specimen's quality problems due to sample preparation. The image in (a) is taken of an unusable preparation. It shows quite bad protein quality. (b) is an overview image of the grid. In black the grid bars can be seen. The fringed edge is due to thicker ice close to the bars. Some squares show ruptures in the carbon film, others a frozen, residual drop of protein solution. A clear boundary separates ice without any filaments at all from a cloudy area with crowded filaments. Both areas cannot be used.

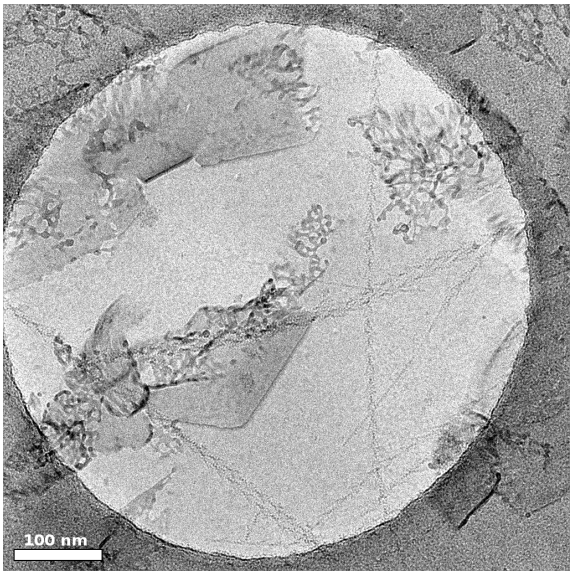
in cooperation with the group of Prof. Dr. Keiichi Namba (Osaka University, Osaka, Japan). The preparation was done as follows: Quantifoil® holey carbon molybdenum grids with 0.6 µm hole diameter (Quantifoil R0.6/1, Mo 200 mesh) have been used. They have successively been washed with chloroform, toluol and water to remove any residual plastic, which is needed for the production of the holey carbon film. The grids have been glow discharged in a low pressure nitrogen-oxygen atmosphere for 20 s directly before usage. On-grid decoration proved to result in less bundling. 1.25 µl of 1.92 µM concentrated phalloidin-f-actin solution has been applied to the grid clamped by a lying tweezer. Two minutes later, 1.25 µl of 9.93 µM concentrated myosin V solution and 0.2 µl apyrase have been added.



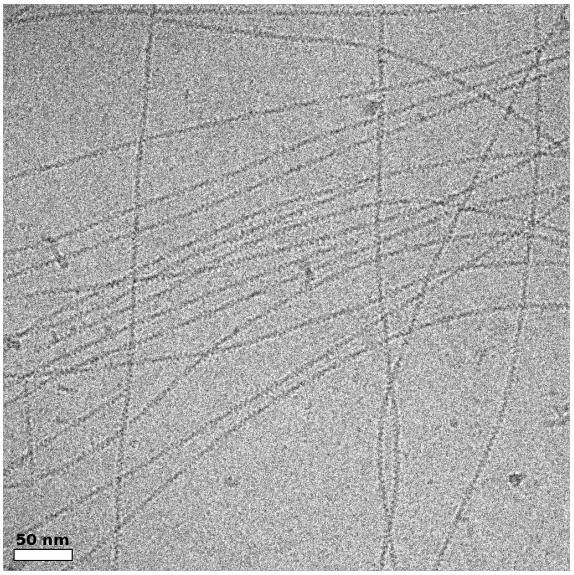
(a) bundling of filaments



(b) filament concentration too high



(c) contamination and bundles



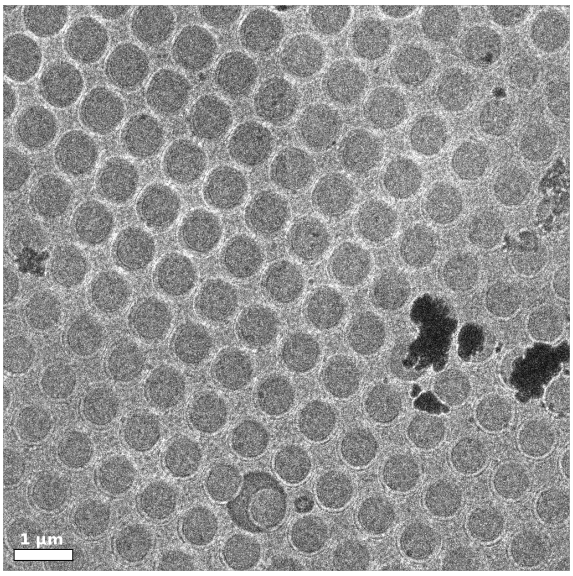
(d) undecorated f-actin in the presence of myosin

Figure 6.2: The images depict some quality drawbacks of the specimen due to problematic sample preparation. (a) shows the tendency of decorated f-actin to bundle. Some filaments are fully decorated, others are undecorated or only partially decorated. The high filament amount leads also to filament bending. The darker blobs are contaminations. (b) is an overview image of the holey carbon film with a very high amount of filaments in the ice layer. (c) depicts other forms of contamination and (d) shows the case that even in the presence of myosin the actin filaments can be completely undecorated.

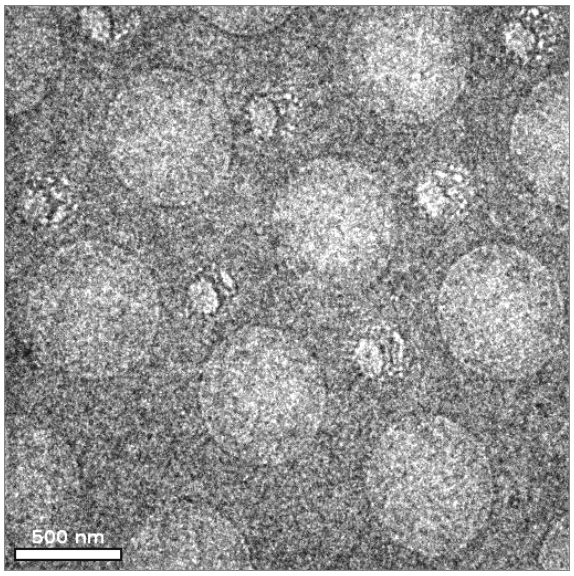
Apyrase is an adenosine phosphatase enzyme that catalyzes the hydrolysis of ATP to yield adenosine di- and monophosphate (ADP / AMP) and inorganic phosphate P_i (ATP $\rightarrow$ ADP+P_i $\rightarrow$ AMP+2P_i). Due to the actin polymerization a residual final concentration of about 5.5 µM ATP is left inside the sample solution, that should be digested by the apyrase. The final protein concentrations are 0.89 µM f-actin, 4.6 µM myosin V and 18.5 µU/µl apyrase. Thus, the actin to myosin ratio is about 1:5. After two minutes incubation / decoration time the sample was plunge-frozen into liquid ethane using an automated vitrification device from FEI, the Vitrobot Mark III. The Vitrobot chamber has been kept at 100 % relative humidity and at room temperature (about 22 °C). Excess buffer has been blotted with Whatman paper No. 1 from both sides for 3.5 s. This preparation was intended to yield the rigor state of the actin and myosin interaction. A second state has been prepared to give the strong-ADP-binding state that the unique kinetics of myosin V offers. Instead of apyrase 0.2 µl of 250 µM concentrated ADP solution have been added, resulting in a final ADP concentration of 18.5 µM. The volumes and concentrations of actin and myosin have been the same as for the rigor state preparation.

6.1.2 Data acquisition

The prepared TEM grids have been observed using a JEOL JEM3200FSC electron microscope, which is situated in Osaka in the lab of Prof. Dr. Keiichi Namba (Osaka University, Japan). It is equipped with an Ω-type energy filter and a liquid-helium cooled specimen stage. The field-emission electron gun was operated at 200 kV. The autoloader of the TEM allows to store one grid over night without refilling of liquid nitrogen. Thus, it allows to work many days with the same specimen. The data has to be acquired manually, because the stage cannot be controlled by a computer. Figure 6.3 shows two overview images. A lower magnification with a very low electron dose has been used to locate suitable areas for data acquisition. The focussing was done close to the acquisition area with a strongly decreased illumination area, in order to avoid pre-exposing the acquisition area. The focus areas can be seen in Figure 6.3 by the beam damage occurring during focussing with a magnification of approx. 180000x. The illumination area for the actual image has been slightly bigger than the holes in the carbon film. Figures 6.3 and 6.4 depict the workflow of data acquisition. Zero-loss energy filtered images with an energy slit width of 10-15 eV have been recorded with a 4096 x 4096 pixel slow-scan CCD camera from TVIPS (TemCam-F415MP) with 15 µm sized pixels. The magnification, including post-magnification, is around 112000x, resulting in an image pixel size of 1.35 Å. Images have been acquired in the temperature range of 50-60 K of the specimen stage without occurring specimen drift. The nominal defocus values have been chosen in the range of 1.0-2.5 µm. The electron dose is estimated by $\sim$15 $e^-/Å^2$ and has been distributed over 2 s exposure time. Into some specimens tobacco mosaic virus was mixed to calibrate the magnification. Its 23 Å layer-line spacing in the Fourier trans-



(a) overview image with darkened illumination area



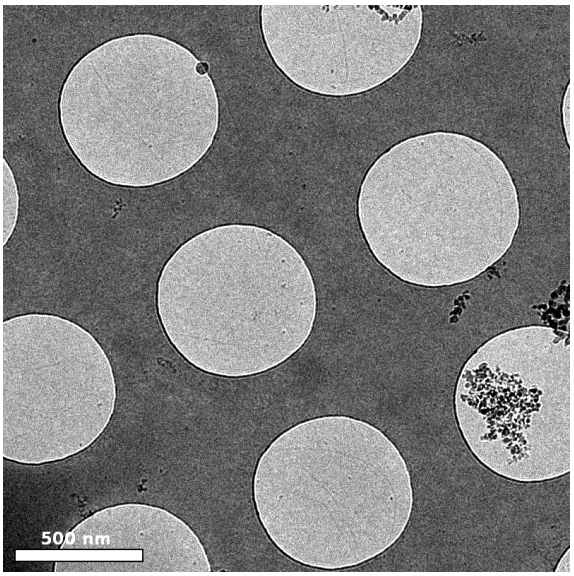
(b) overview with beam damage at the focussing positions

Figure 6.3: Two overview images at lower magnification show the 0.6 µm sized holes containing the vitrified proteins inside the ice layer. Both the size of the illumination area for the acquisition and for the focus spot can be seen. Image (b) exhibits a low signal-to-noise ratio due to a very low electron dose.

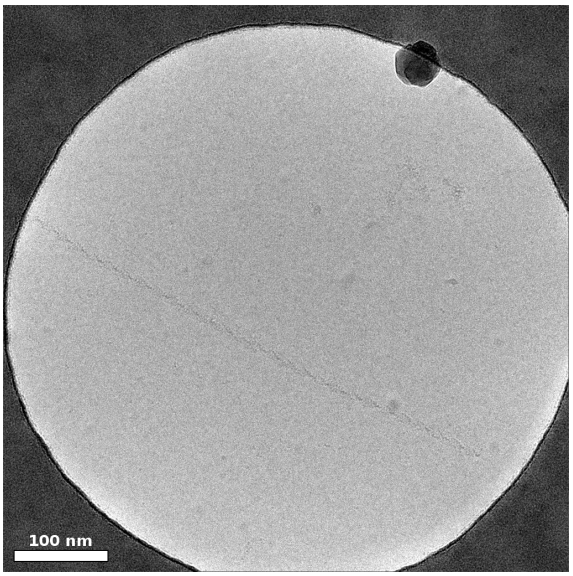
form of the images can be used, since the tobacco mosaic virus has a very rigid symmetry compared to the actin. In total about 4500 images have been collected manually. Out of it only 1000 images of the rigor state show decorated filaments and only 70 images of the second preparation could be selected for the image analysis. Some of the selected images exhibit a high background signal probably due to the myosin excess, some have filament cross-overs or the quality of decoration decreases along the course of the filaments. Nevertheless, it was the best compromise to yield decorated and sparsely distributed filaments. The main focus of data acquisition has been on the rigor state due to time restrictions to obtain at least one complete data set. The strong-ADP-binding state preparation showed worse decoration compared to the rigor state preparation. The analysis shown in the next section is done with the main data set of the 1000 images. The second, very small data set is only discussed later.

6.2 Results

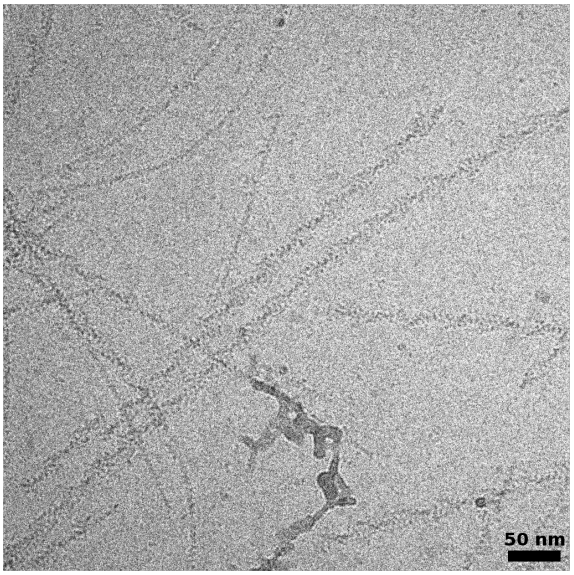
The data set of TEM images contains filaments of varying lengths and orientations. Prior to reconstruction the filaments have to be selected manually and prepared in a way the algorithm can read it as input data. The image processing pipeline is built up using the SPIDER image processing system



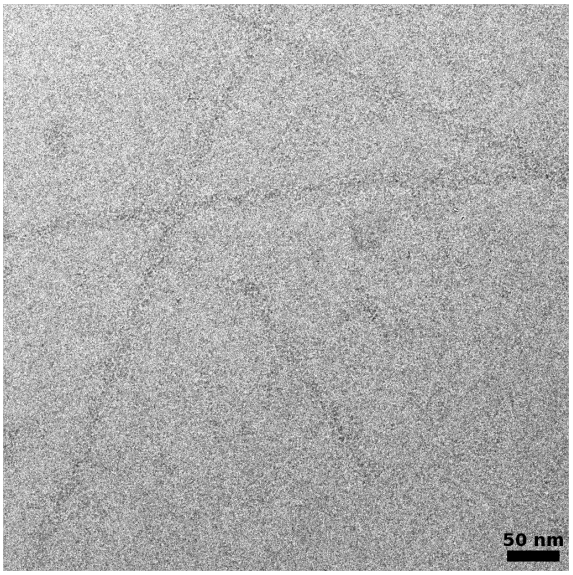
(a) strong defocus allows to locate filaments



(b) fully decorated straight actin filament



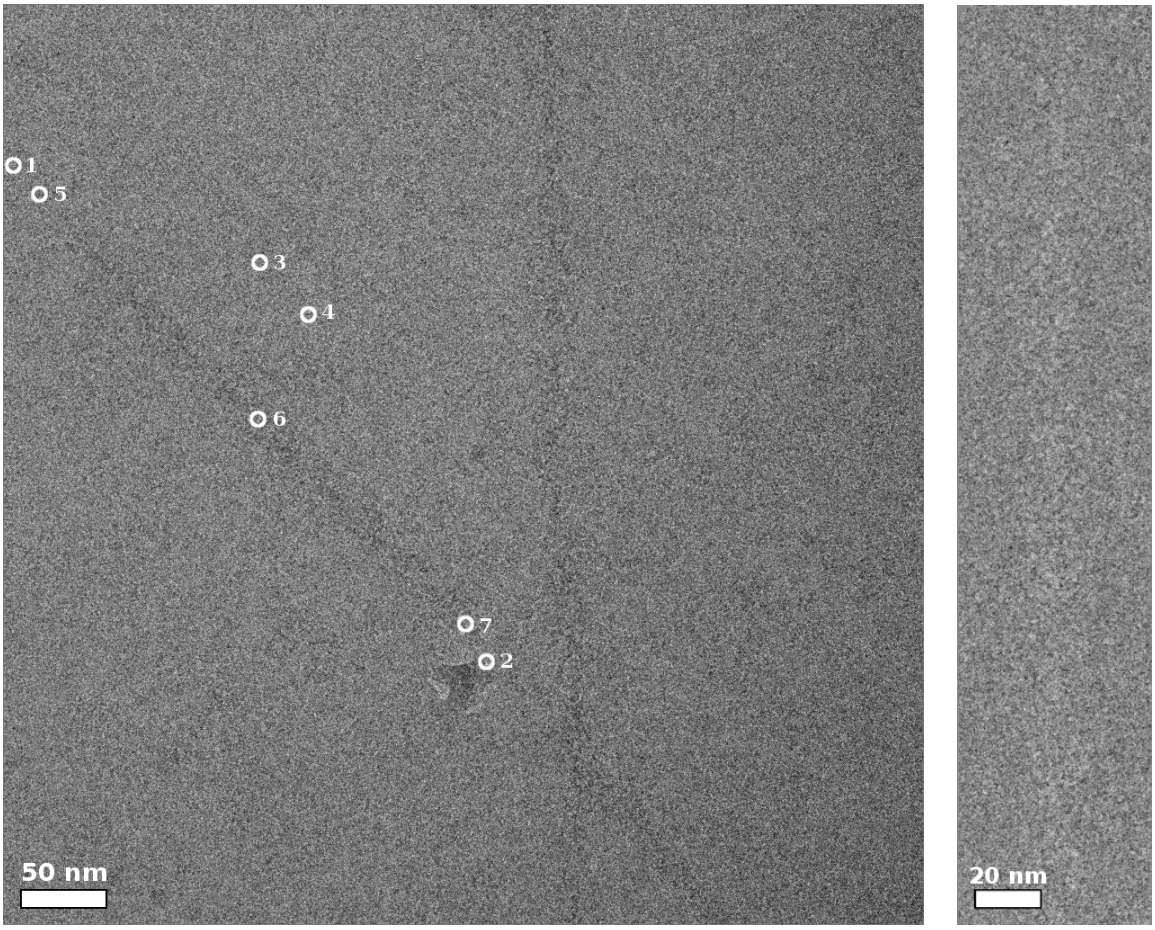
(c) strongly defocused image with good and bad decoration



(d) about 1.3 µm defocused TEM image

Figure 6.4: Images of the actomyosin complex in amorphous ice. (a) A strong defocus of up to 50 or even 100 µm allows to locate the filaments at lower magnifications and low electron doses. (b) shows a single actin filament that is fully decorated with myosins V motor domains. (c) and (d) are both acquired at a magnification of around 112000x using an electron dose of ∼15 electrons per square Ångström. For the image in (c) a strong defocus of about 4 µm has been used to inspect the quality of decoration. (d) shows a typical image that is subject to further processing.

for electron microscopy (Frank et al., 1996). The filament selection is done by a SPIDER script in combination with JWeb, which is the Java implementation of the graphics interface of SPIDER. Image by image has to be opened in JWeb in order set marker points for the course of the filament or filaments present in the image. Additional marker points close to the filament have to be set to select a flat background region for background correction of the image (cf. Fig. 6.5). The manual selec-



(a) Filament selection (b)

Figure 6.5: The images depict the process of filament selection with JWeb (Frank et al., 1996). (a) Two markers are used to define the area the filament is lying in, the next two markers define a nearby, flat background region for background correction of the image and three markers are used to appoint the course of the filament. By linear regression to the filament marker points the angle is calculated to rotate the filament in vertical direction. (b) shows the rotated, background corrected and inverted filament that is subject to further processing.

tion follows general selection criteria: the filament should be almost straight, should hence exhibit no curvature, preferably have no cross-overs with other filaments or any contamination and should be excluded if it is apparently undecorated or partially decorated. The coordinates of the marker points

can be used to cut out the filament and subsequently rotate it into vertical direction. In a first attempt the filament image has been mirrored along the vertical axis. A rotational orientation search should give the doubled angle of the filament with the vertical axis. It turned out to be not reliable. A rotational orientation search with an oriented reference seems to determine likewise not every angular orientation accurately, especially for filaments near the edge. It proved to be most reliable to directly use the marker coordinates to calculate the in-plane rotation angle. As it might be influenced by an imprecise manual marker setting, it is checked later if it can be improved. In order to calculate the rotation angle a script, written in the Python programming language, has been used to fit a regression line to the filament's marker points. The gradient of the regression line directly gives the angle. The Python script can be called during the SPIDER run. File statistics of the selected background area allow to subtract the average background signal form the filament image. For normalization it is further divided by the background standard deviation. The contrast is inverted to have a positive signal for the protein density. Depending on the desired image pixel size the images can be binned or interpolated also at this early stage in processing. Normally, twofold binned images increase the signal-to-noise ratio for the image analysis. Unfortunately, interpolation is also needed to suit the requirements of the least squares algorithm to reach a symmetry related pixel size. The input data for the read-in has to be an integer divisor of the symmetry unit size of 27.5 Å. For this data set it is therefore either 2.75 Å or 1.375 Å. The resulting set of images is a mixture of pointing up and down filaments. About 3000 filaments of different length have been selected. Correlations with pointing up and down reference filaments are due to the low signal-to-noise ratio not sufficient to certainly determine which direction the filament is pointing to. Visual inspection allows no decision either.

The individual filament images are cut further into quadratic filament segments using a high oversampling. For an image pixel size of 2.75 Å every 10 pixels along the filament a new segment has been cut, resulting in one segment image per symmetry unit of the filament. The segments are centered precisely and their in-plane rotation is checked again. To center them first it has been tried to mirror them and find per cross correlation the doubled shift towards the center position. Again, like for the rotation it turned out not to be precise. Correlation with a centered reference works better. It turned out to be best using a one-dimensional (1D) projection of the segment and of a centered reference to align each segment to the center position. The integer value of the determined cross correlation peak is applied as x-shift to the segment. The in-plane rotation is checked via rotational correlation with a centered, vertical reference in a restricted angle range close to the vertical direction (plus / minus a few degrees). In order to suppress noise inside the determined parameters the continuity of the segments belonging to one filament has been used. Big variations along one filament cannot be possible. Hence, a low pass filter has been applied to the list of x-shifts and in-plane rotation angles to get rid

of outliers prior to the application. The filtered, integer values have been applied to the segments. Finally, the images are converted to the opposite endianness, since unfortunately the image processing system SPIDER and the least squares reconstruction program use a different endianness (SPIDER: big endian, least squares algorithm: little endian). Figure 6.6 shows an example of the resulting oriented filament segments. There are about 200000 segments in total. Each resembles one single particle in the framework of single particle reconstructions.

In the framework of the least squares reconstruction and its image processing workflow the alignment is not done with the normalized and vertical oriented filament images. Instead the reconstruction of single volumes of each segment precedes the alignment, which is then carried out in 3D for this method. The reconstructed 3D volumes of each filament segment are then aligned to each other. Averaging of the aligned volumes gives the final reconstructed density representation of all particles of the data set. The image size of the segments has been chosen to have at least two arrowheads inside the image. Some power spectra calculated along the helix of strongly defocused images show up layer lines that suggest a helical symmetry close to the 28/13 crystallographic symmetry. Therefore, the 28/13 symmetry has been used in the following. It describes the helix in repeats of 28 symmetry units that form exactly two arrowheads. The least squares reconstruction algorithm has been used as described in the preceding diploma thesis (Wulf, 2009). To reduce the computational costs and let the calculations become feasible the algorithm uses the axial geometry to reduce the 3D problem to a 2D one. The present implementation allows to reconstruct the symmetry unit of a helical filament slice by slice, from which a helix can be built up considering its symmetry. Each segment is treated as one axial tomogram. It offers with each symmetry unit a new, symmetry related view of its structure. The implementation makes use of it and therefore out of one image a corresponding volume can be reconstructed. In order to read in the segment data and to set up the projection equations that are subsequently solved by the least squares algorithm some assumptions of the symmetry are necessary. The pixel size has been scaled to 2.75 Å in a first attempt, leading to 10 slices along the helical symmetry axis that compose the symmetry unit. Then, every tenth pixel row of the scaled segment image corresponds to the same slice of the symmetry unit. Each row has a new symmetry related projection angle which is a multiple of $-13 \cdot 360°/28 \approx -167.143°$. Figure 6.7 illustrates the data read-in. Pixel rows from two arrowheads are read in, further rows will give no new data as the projection orientations are the same. For each filament segment the image slices composing the helical symmetry unit are reconstructed. The necessary filtering of the least squares solution is done with respect to spatial frequencies and additionally by the criterion eigenvalue $\lambda_i > 1$ which includes for the given symmetry and number of projections 2019 eigenvalues and eigenfunctions into the reconstruction. The time-consuming setting up of the projection matrix and its inversion in the subspace of significant eigenfunctions can

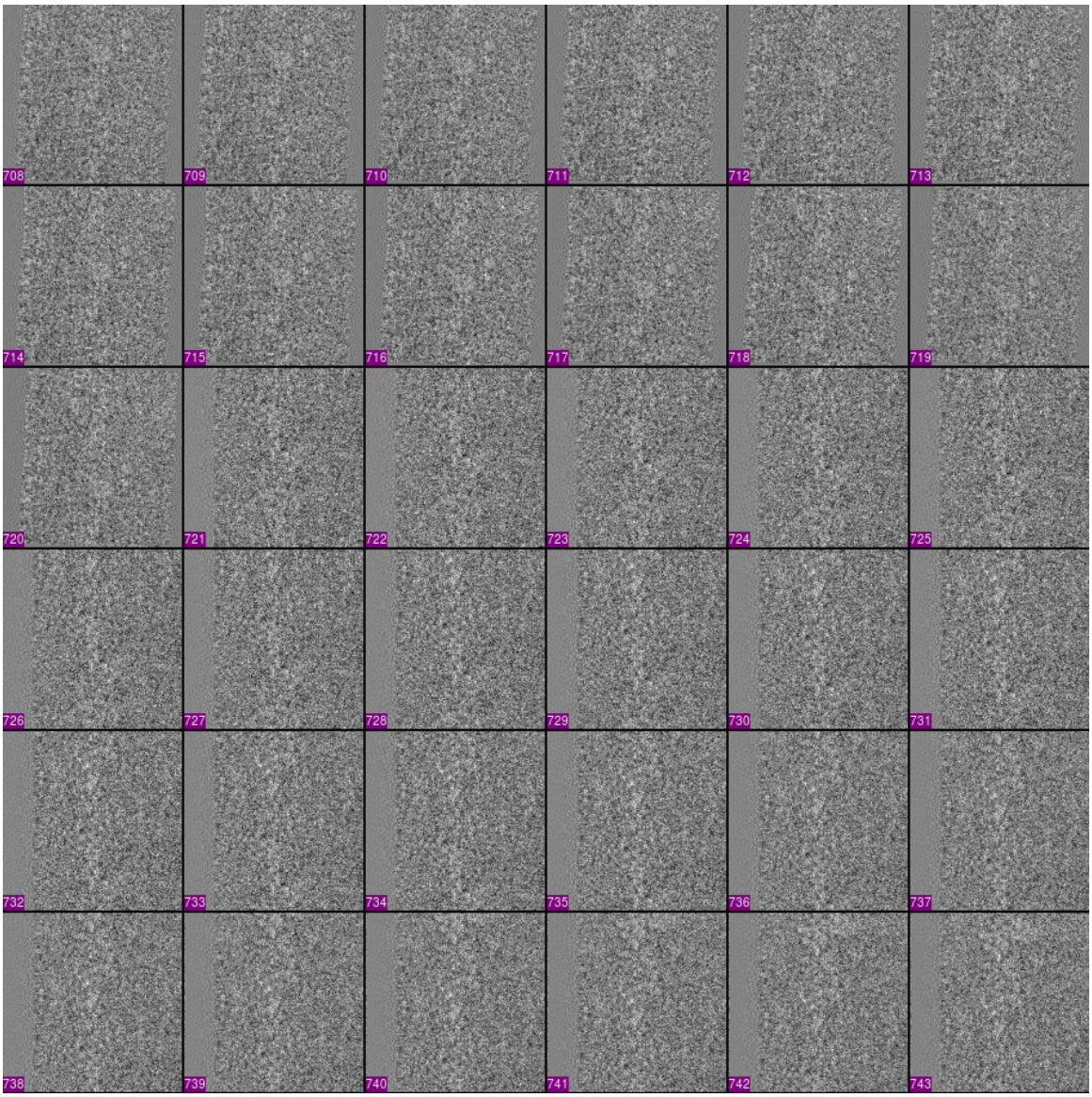


Figure 6.6: Excerpt of the filament segments that have been cut from the filaments of variable length. The images have a size of 83 nm x 83 nm. The segment stack is visualized by the v2 program. It belongs to EMAN, version 1, a suite of image processing tools aimed primarily at the transmission electron microscopy community (Ludtke et al., 1999).

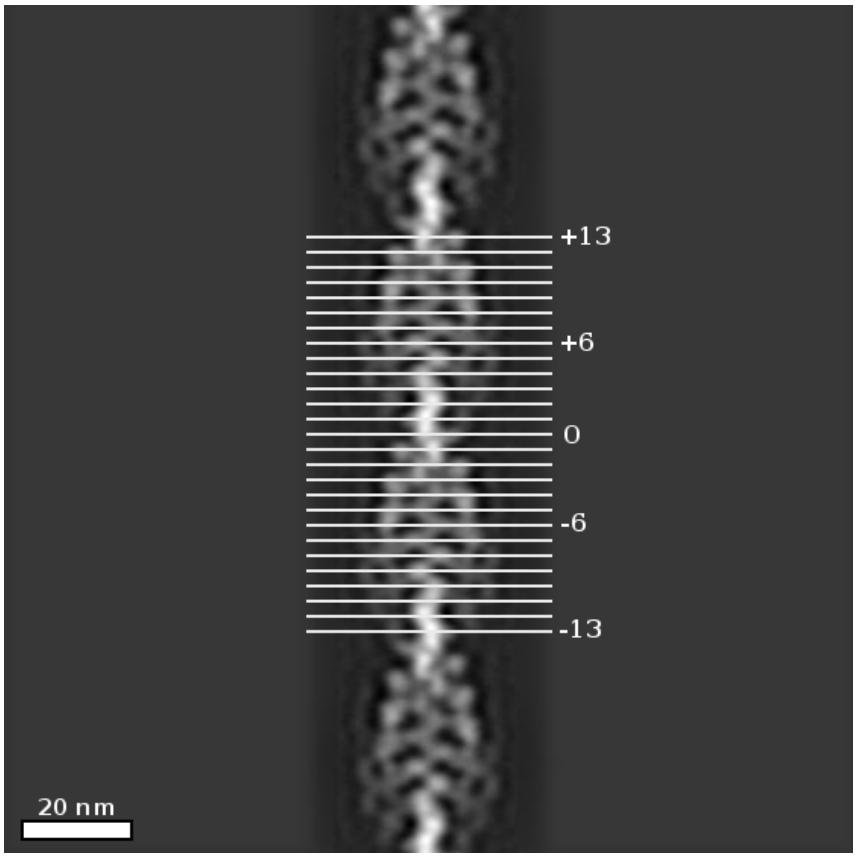
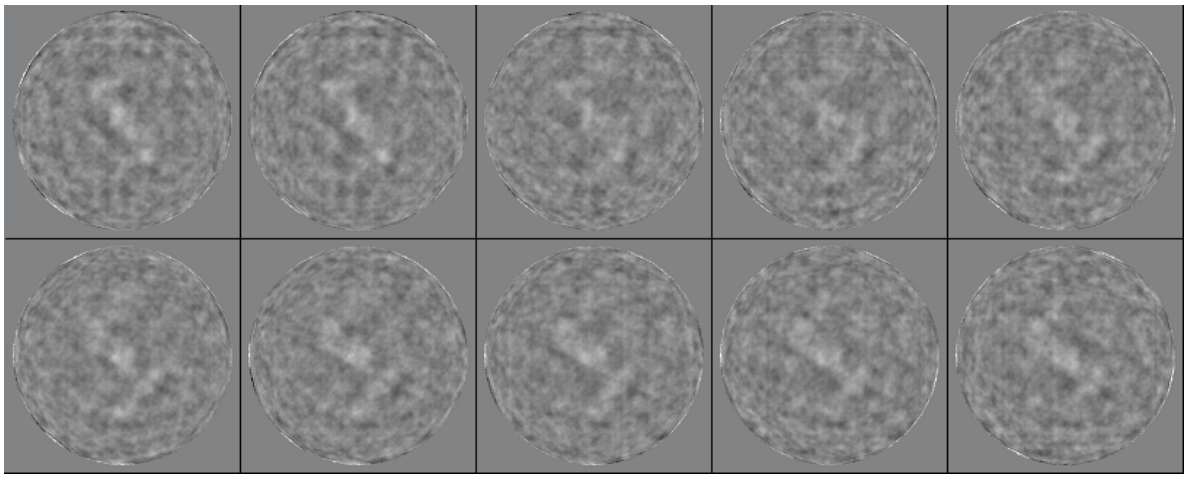


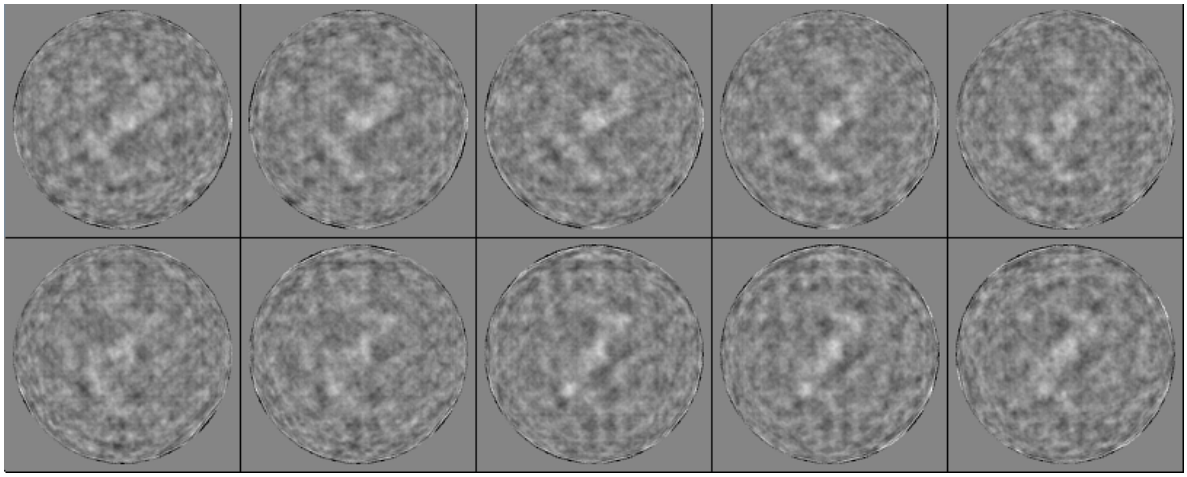
Figure 6.7: The image illustrates the read-in of the filament segment data. From the center of each image two arrowheads are read in, one above the center, one below the center. The symmetry unit in this case is composed of ten image slices. The symmetry unit related pixel size allows to assign each tenth pixel row a new projection angle given by the symmetry. A crystallographic 28/13 symmetry gives a symmetry angle of $-167.143°$ between two adjacent symmetry units. Therefore, the axial tomogram of each filament segment provides projection data from -13 to $+13$ times the symmetry angle. For each pixel one projection equation is set up. Shifting the reading frame successively the projection data for each of the ten slices of the symmetry unit can be read in.

be precomputed and be stored. Then it can be reused for all other segments. Static compilation of the program with all its library dependencies even allows its use on different machines of a computer cluster. As the directions of the filaments are unknown so far, both directions have been reconstructed. To avoid the need of a second projection matrix with a positive symmetry angle ($+167.143°$ instead of $-167.143°$ between adjacent symmetry units), the filaments have been rotated by 180°. A subsequent reconstruction with the same, negative symmetry angle gives the reconstructed symmetry unit of the filament pointing in the opposite direction. Only one of the directions assigned to the angle is correct. Subsequent analysis of the volumes allows to distinguish which of the assignments was correct.

A main part of the data processing workflow has been the implementation of the 3D alignment of the single segment volumes to each other. In a first attempt it was supposed that the coherent length of the filament is longer than the portion of the filament contained in one projection image. It implies that there are no symmetry changes during the imaged length of the filament. Therefore, the continuity of the filament has been used to average all segment volumes extracted from one filament. The segment volumes have been rotated back by the amount of the multiple symmetry angle corresponding to their position in the filament and then have been averaged. Figure 6.8 shows the resulting reconstructed slices of the symmetry units for one filament. The reconstructions for the different directions exhibit a density that seems to be mirrored compared to the other direction. Building up a helix with the wrong assignment of the filament direction with respect to the sign of the symmetry angle will thus lead to a helix as well, but with a wrong handedness. Nevertheless, the directional imprint in the reconstructed volumes is stronger than in the filament images. The correct handedness can be distinguished using a reference. The initial reference has been the reconstructed volume of one filament, whose direction was known. The handedness is determined as follows: The two reconstructed volume units per filament are projected to obtain only one image per volume. This image has a better signal-to-noise ratio than the volume. The same is done with the reference. A rotational alignment of the projected volumes with the reference gives the best rotation to match the reference. The images are rotated by 360° in 1° steps. Each rotated image is correlated with the reference. The cross correlation value of the central pixel is stored for each rotation step and is visualized for one filament in Figure 6.9 as line profile. The projected image is rotated by the angle with the highest cross correlation value. It is done for both filament directions. A subsequent cross correlation of the rotated, projected volumes with the reference gives the correct direction for the filament. The direction which correlates best is assumed to be true. Figure 6.10 shows for one filament the projected volume units, that are subject to determine the correct handedness. They show a clear directional imprint. As quality control only those filaments have been used, for which one direction has correlated clearly better than the other, i.e. filaments with such bad signal to give no clear directional signal have been disregarded. Once the handedness is de-



(a) Reconstructed slices of the symmetry unit of one filament.



(b) Reconstructed slices of the symmetry unit of the same filament, but rotated by 180° prior to reconstruction.

Figure 6.8: The reconstructed slices of one symmetry unit are shown for one filament. Stacked on top of each other they compose the volume unit that builds up the helix. All segment volumes belonging to that filament have been averaged using the symmetry relation between them according to their position in the filament. (a) and (b) show the reconstruction of one of the two potentially correct directions the filament is pointing to. Both are reconstructed using the same symmetry angle, which can be correct only for one of the directions. The image size is 44.3 nm x 44.3 nm.

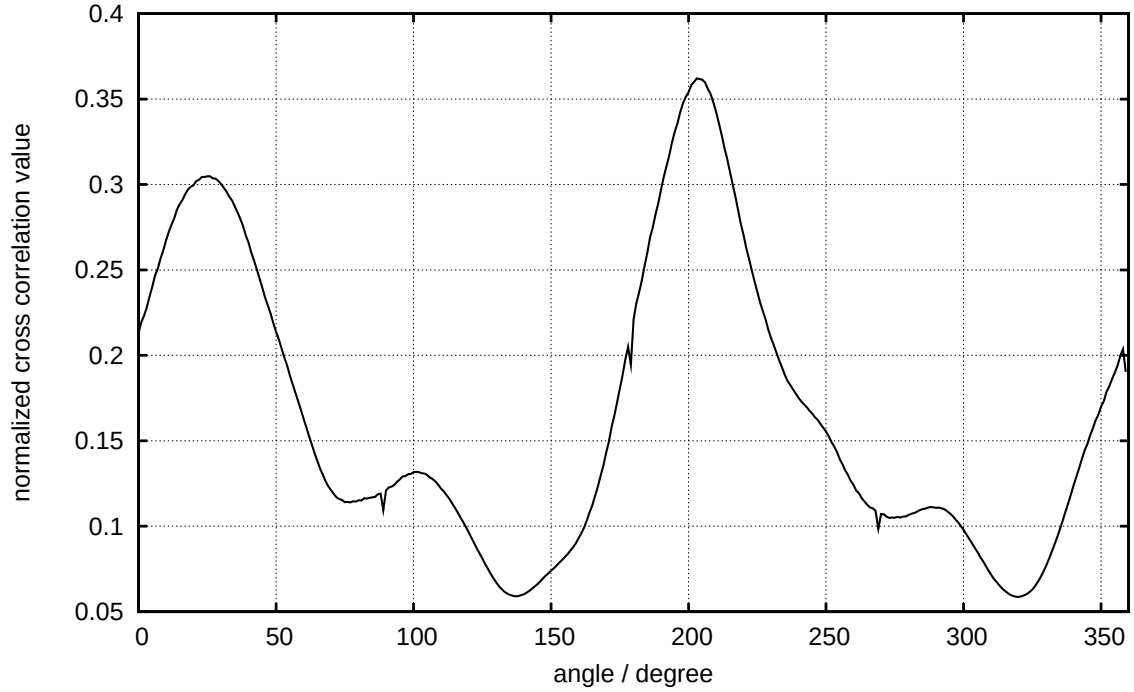
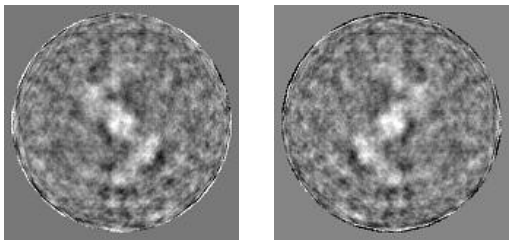


Figure 6.9: Line profile of the rotational cross correlation to obtain the correct handedness of the filaments. The profile shows a clear peak that gives the rotation angle to best match the reference.



(a) (b)

Figure 6.10: The projected images of the reconstructed symmetry units of one filament are shown. (a) and (b) show the two reconstructed filament directions. They have a clear directional imprint and together with a reference the correct handedness of the filament can be distinguished (44.3 nm x 44.3 nm image size).

termined the one symmetry volume that describes the density of the filament needs to be aligned with respect to the other volumes. It is accomplished by a simultaneous determination of the height and azimuth angle, which describe the volume's orientation. Also for this purpose a reference is used. The reconstruction of an exceptionally well resolved filament has been used as initial reference. In order to test each cyclic permutation of possible matching height a 30 slices reference is used. The central ten slices are one reference symmetry unit. The unit is once rotated by the positive symmetry angle and once by the negative symmetry angle and inserted before and after the central unit to build a small helix of three symmetry units. Thereby, all possible relative heights of the volume, to be aligned with the reference, can be accounted for. The possible heights range from shifting down the volume by half the symmetry unit to shifting up of half the symmetry unit. To account for all possible combinations of height and azimuth angle the entire 2D parameter space is sampled as illustrated in Figure 6.11. For each of the volume height positions (x-axis in Fig. 6.11) a rotational cross correlation (y-axis in

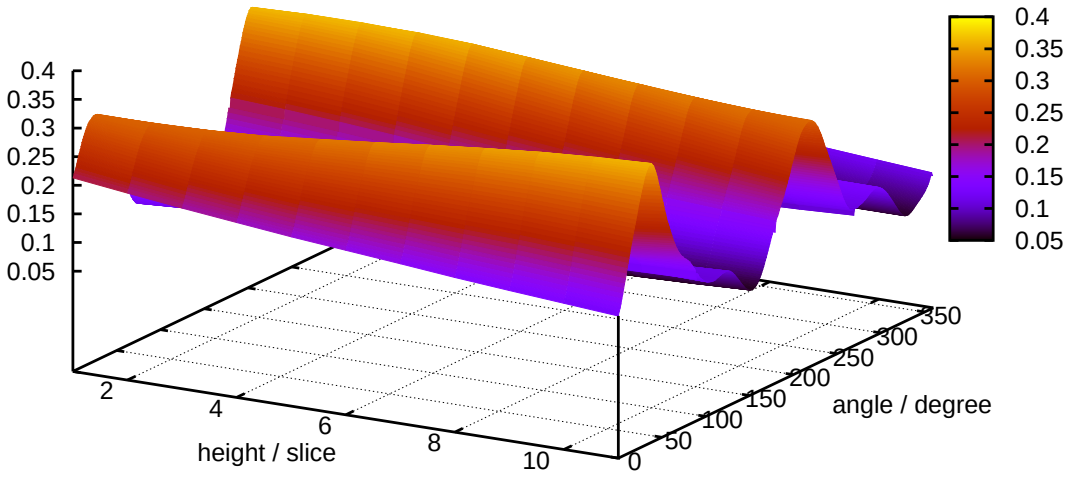


Figure 6.11: For one volume the 2D parameter estimation is been illustrated. The 2D parameter space of possible heights and azimuthal rotations is sampled in order to find the best matching parameters. The z-axis shows the normalized cross correlation values between the volume and the reference. The highest peak gives the parameters that align the volume with the reference.

Fig. 6.11) of each slice with its possible matching one of the reference is done (the same as shown in Fig. 6.9). The cross correlation values from all slices are averaged, since the slices cannot have a different orientation as they form one symmetry unit. The overall best rotational correlation value

is stored for that possible height. All other heights are tested in the same way. Figure 6.12 shows a profile of the height parameter estimation. It is composed of the cross correlation values of the best azimuthal rotation for each height (the maxima along the y-axis in Fig. 6.11 give the values for each height). A peak search will give the height with the best correlation. The corresponding azimuth angle

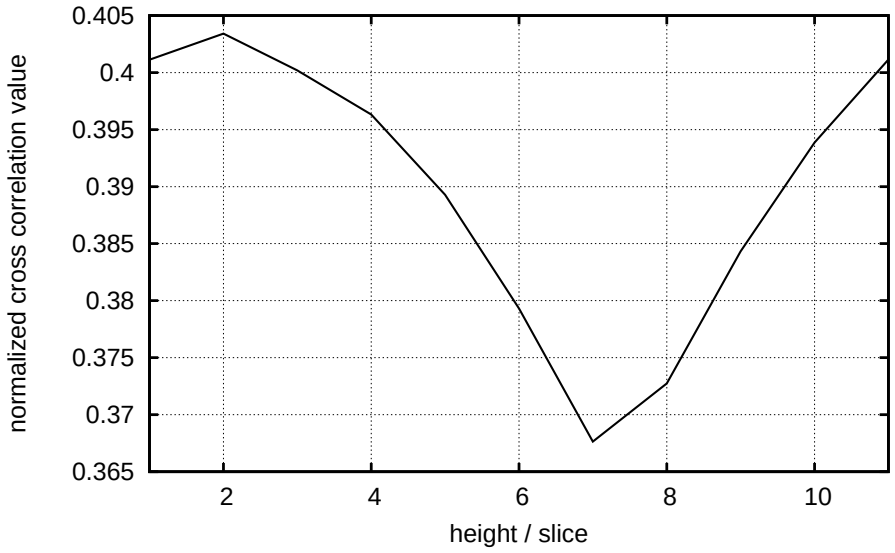


Figure 6.12: The figure shows a line profile illustrating the parameter estimation of the height of one volume. Each entry belongs to the overall best rotational correlation for the respective height (the maxima along the y-axis in Fig. 6.11). The highest correlation peak for the combination of the height and azimuth angle gives the alignment parameters of the volume. Height 1 and 11 are the same, they only belong to different azimuth angles due to helical continuation.

combined with helical continuation and cyclic permutation of the slices is used to create the aligned volume of the filament. The alignment is done for all averaged filament volumes. It can be distributed as single jobs to the computer cluster. A submit script restricts the maximal number of submitted jobs to the cluster to avoid problems with the cluster scheduler.

For CTF parameter estimation, whole CCD frames are used, not the small segment images. This has been implemented in SPIDER as well. A phase flipping has been done to correct the phases of the aligned volumes in order to have in different defocused volumes the same sign for the phases in frequency space. The contrast transfer function seems not to be the limiting factor for this reconstruction method due to other issues discussed below. Therefore, the CTF correction is explained in detail only in Chapter 8. The phase flipped and aligned volumes of all filaments are averaged to give the final reconstructed density. In order to improve the reconstruction a refinement has been done as illustrated

in Figure 6.13. In a first refinement the reconstructed volume has been used as new reference for the

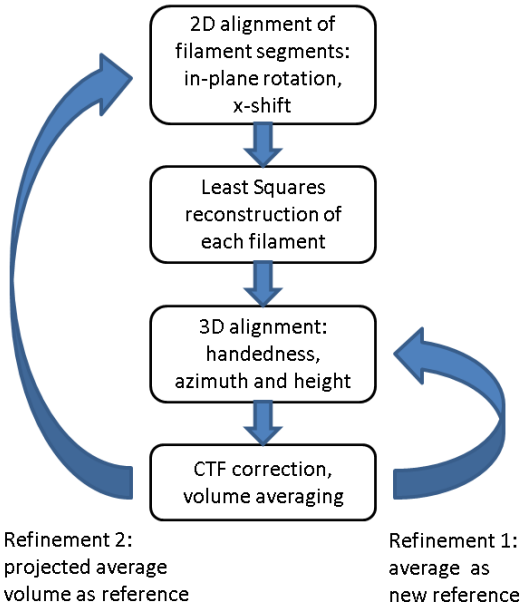


Figure 6.13: Illustration of the refinement of the alignment. The reconstructed volume is used as new reference. In a first refinement only for the 3D alignment of the volumes, in a second step also for the input alignment before the reconstruction.

3D alignment of the volumes. In a second refinement the projection of the helix, built up from the reconstructed symmetry unit, is used for the alignment of the images prior to the reconstruction. Figure 6.14 shows the refined, reconstructed slices of the helical symmetry unit for the supposed 28/13 symmetry. Outside the density some circular density prolongations can be seen that hint at something that might be incorrect. Figure 6.15 shows the corresponding helical arrangement. The symmetry angle has been used to build up the helix, stacking the rotated symmetry units on top of each other. Because the reconstructed density seems to have some shortcomings and a poor resolution further analysis has been done.

A quality control step has been implemented. The quality of each filament segment has been tested by cross correlation with the projection of the reconstructed helix (from Fig. 6.15). Some filaments show an overall uniform distribution of correlation values of their segments with the reference, some show areas of inferior quality or even areas which do not correlate. Two examples have been chosen and are shown in Fig. 6.16. Obviously, the first half of the filament in Fig. 6.16 (b) has poor quality and may

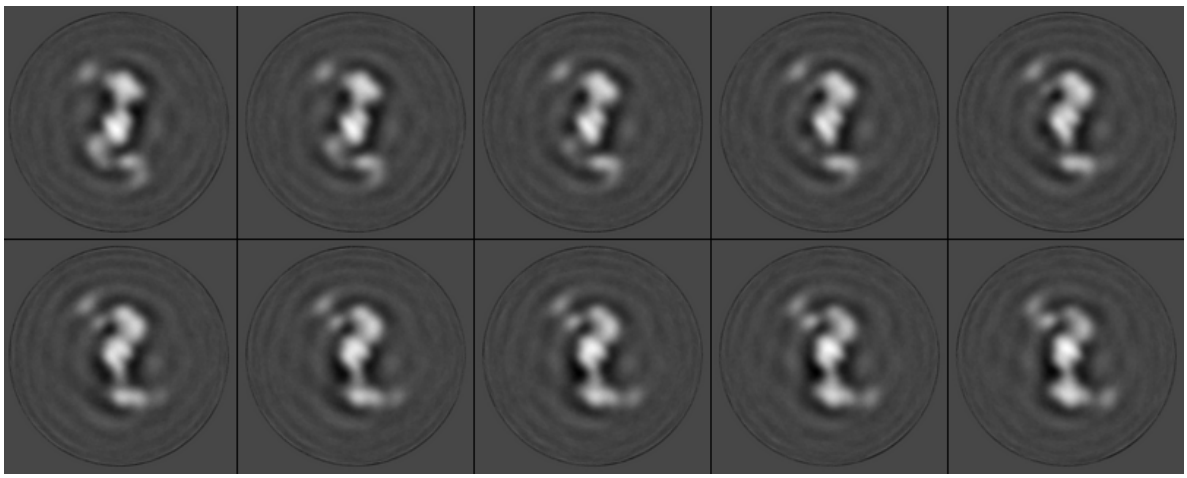


Figure 6.14: Shown are the refined, reconstructed slices of the helical symmetry unit for the supposed 28/13 helical symmetry. It is the averaged density for all filaments. Composed of roughly 3000 filaments of variable length which correspond to about 200k individual particles / monomers. Each particle (filament segment) has been reconstructed separately by the filtered least squares algorithm using all eigenfunctions with eigenvalues $\lambda_i > 1$. The averaging and 3D alignment has been done as described in the text. The image size is 161 x 161 pixels with 2.75 Å per pixel (44.3 nm x 44.3 nm).

better be excluded from the final reconstruction. Therefore, the individual cross correlation results have been used together with an arbitrary set threshold to select only the better filament segments. The same averaging and 3D alignment process has been done with the subset of chosen segments. Nevertheless, the final reconstruction shows the same features as the one without quality control as shown in Fig. 6.14. The questionable quality of some filament segments therefore seems not to be the immediate problem present in the reconstruction of Fig. 6.14.

Furthermore, a different reconstruction area has been tested. The form of the mask, which defines the reconstruction area, has been fitted to resemble the myosin appearance. A smaller reconstruction area theoretically improves the reconstruction, because one seeks to solve the equation system for a smaller number of variables with the same number of projection data. Figure 6.17 illustrates the different reconstruction areas both for the 2.75 Å pixel size and for 1.375 Å that is discussed later on. The number of density points that have to be accounted for in the linear equation system can greatly be reduced by the geometrically adapted form. Thereby, it greatly reduces the size of the normal matrix, which has as many entries as the square of the number of density points. Technically interesting is what happens to the eigenfunctions by the use of the geometrically adapted mask. Figure 6.18 shows an excerpt of the first eigenfunctions belonging to the highest eigenvalues. The perfect cylinder sym-

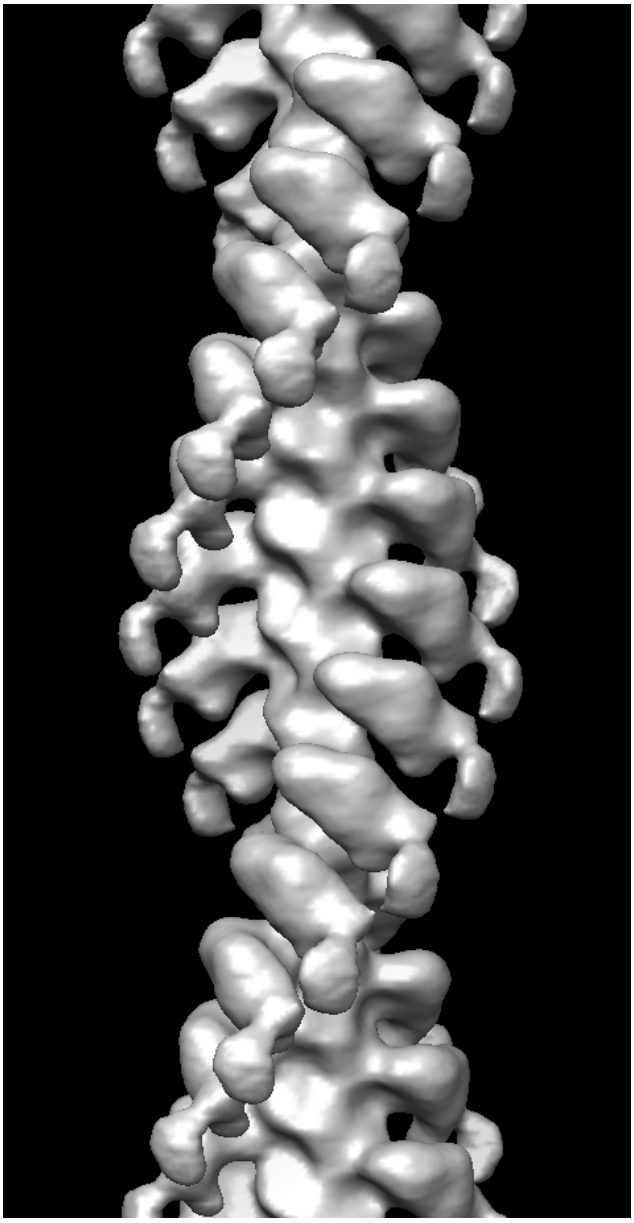
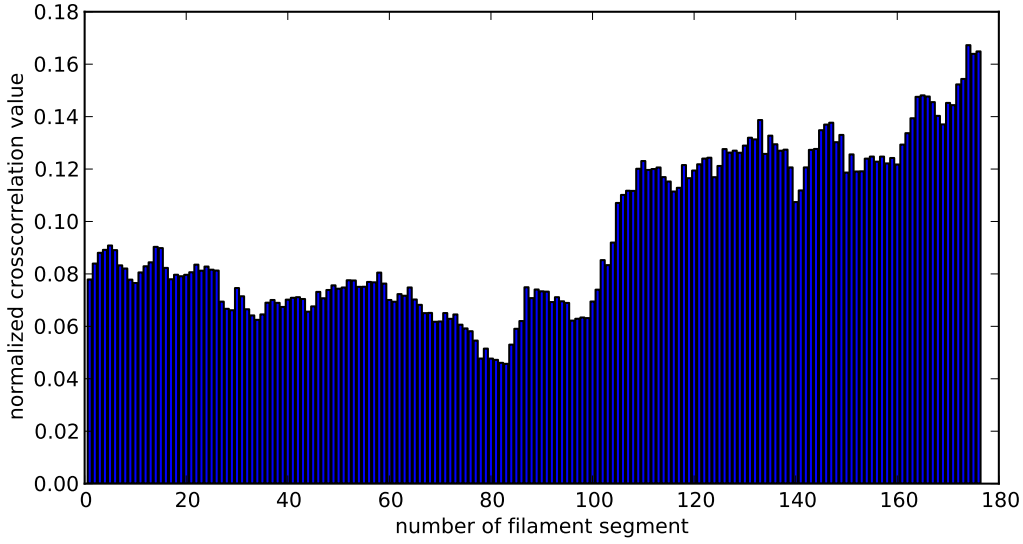
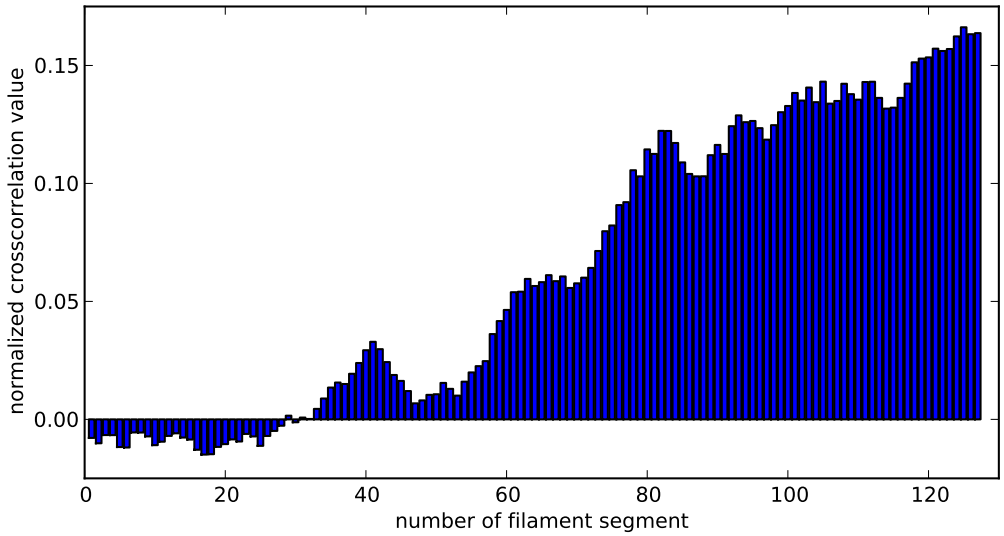


Figure 6.15: The surface representation shows the helix that has been built up from the symmetry unit (shown in Fig. 6.14) together with symmetry related rotations of the unit that have been stacked on top of each other. One arrowhead equates to 38.5 nm. The surface rendering was done in Chimera, an interactive molecular graphics program (Pettersen et al., 2004).

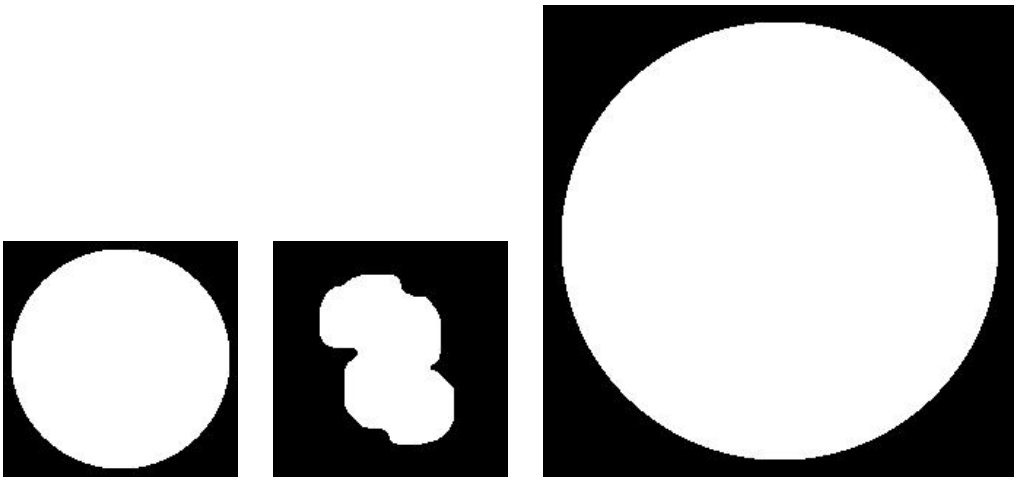


(a)



(b)

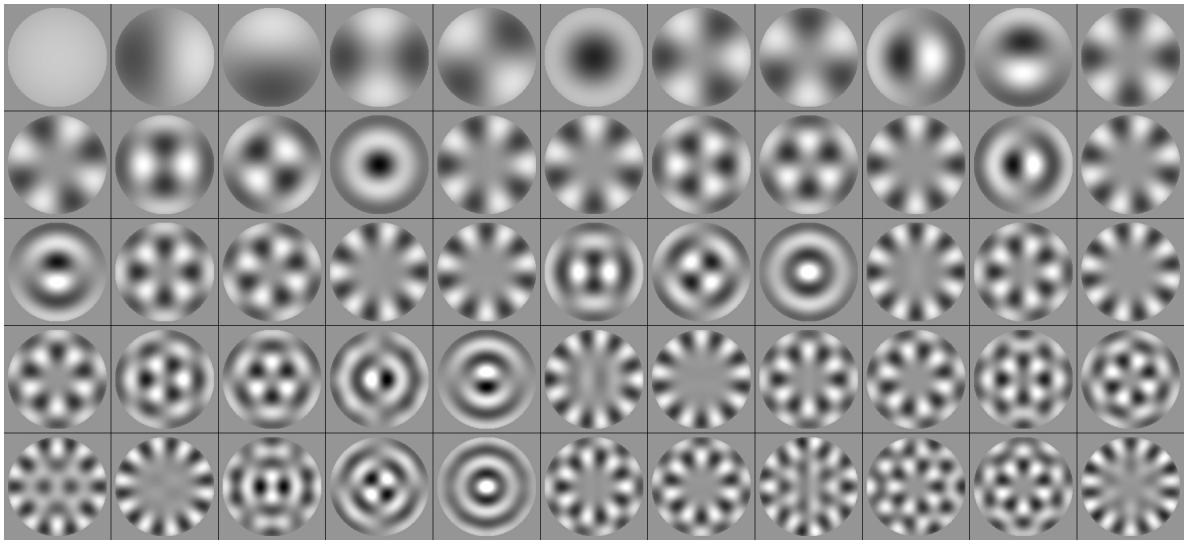
Figure 6.16: The bar plots show for two example filaments the quality distribution of the individual fiament segments along the filaments. The quality of the filament segments has been tested via cross correlation between the segment and the projection of the reconstructed helix. While one half of the filament in (a) correlates better than the other half, the beginning of the filament in (b) seems to have nothing in common with the reference.



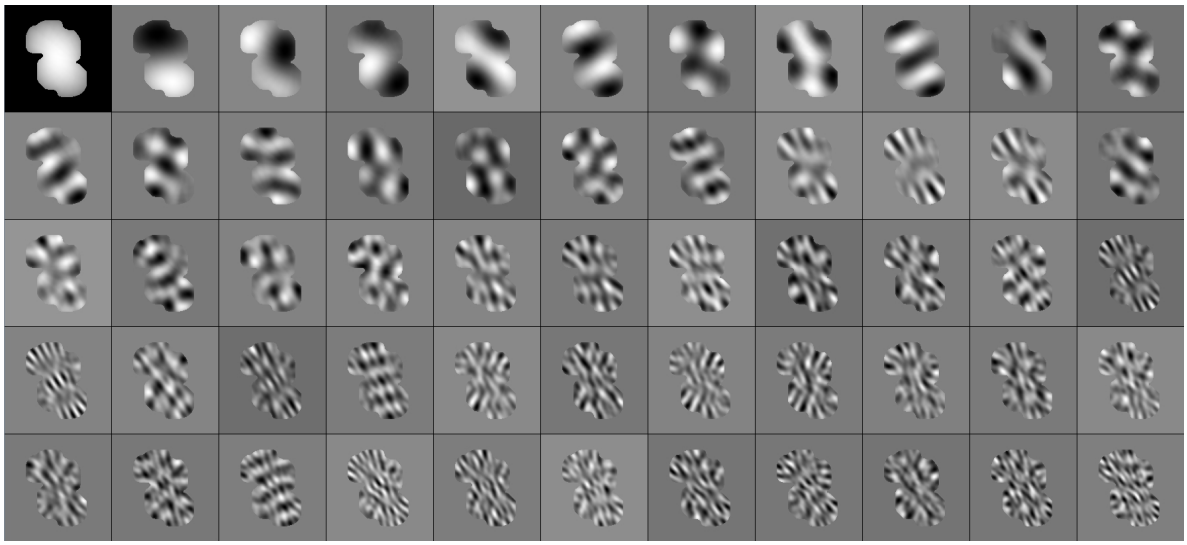
(a) about 18000 d. p. (b) about 8000 d. p. (c) about 71000 density points

Figure 6.17: Different forms and sizes of the reconstruction areas. The areas are labelled with the number of the density points (d. p.) they comprise. The round mask defines a reconstruction area without information about the object. The myosin-like shaped mask uses information of the object to define the reconstruction area. (a) and (b) are suited for the object with a pixel size of 2.75 Å and (c) is for reconstructions with half the size per pixel (1.375 Å).

metric shape of the eigenfunctions that can be seen in (a) changes to adapt to the myosin shaped form in (b). The spatial frequency content of the eigenfunctions increases in (b) faster than in (a). It corresponds to a lower number of significant eigenfunctions in total. Using a filtering of the least squares solution following the criterion $\lambda_i > \delta$ and $\delta > 1$ the number of significant eigenfunctions for the round reconstruction area is about 2000, for the myosin shaped area there are about 1400 significant eigenfunctions. Their spatial frequency composition has further been analyzed. For all significant eigenfunctions the absolute squares of their Fourier transforms (periodograms) have been calculated and averaged. The result can be seen in Fig. 6.19 (a) and (c) for the round and myosin-like shaped areas, respectively. The star-shaped appearance of the averaged periodogram is due to the limited projection sampling. The Fourier-slice theorem illustrates this. Each projection, symmetry related every $-167.14°$, corresponds to a central slice in Fourier space, resulting in a sampling of the Fourier space as depicted by the two images. The profiles along one of the arms of the star reveal the sampling of the different spatial frequencies. For the round shape the sampling is isotropic. The two profiles are within the accuracy of reading the same. However, the myosin shape exhibit a clear directional difference. For the long axis (blue) there is an area, directly following the steep shoulder in the center, that is sampled with a higher intensity than in the direction of the short axis (red) of the myosin shape. It might give rise to an anisotropic resolution of the reconstruction. It follows a brief description of

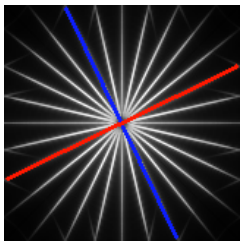


(a) round reconstruction area

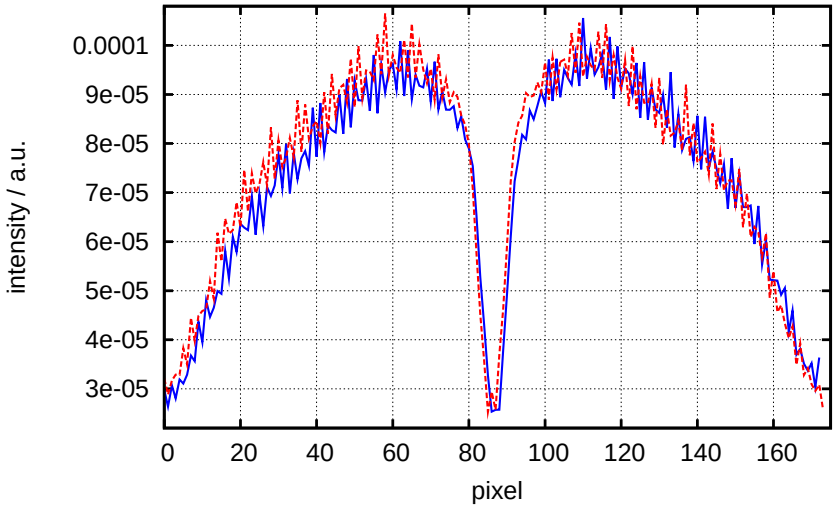


(b) geometrically adapted reconstruction area

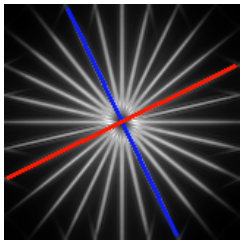
Figure 6.18: Illustrated are some of the eigenfunctions that are used during reconstruction to establish the inverse of the normal matrix. (a) and (b) show the first eigenfunctions for the different reconstruction areas. The eigenfunctions are sorted by their corresponding eigenvalue in decreasing order. They are composed of low spatial frequencies that gradually increase with the decreasing of the eigenvalue. Due to the reduction to a 2D reconstruction problem by the axial symmetry the eigenfunctions are two-dimensional.



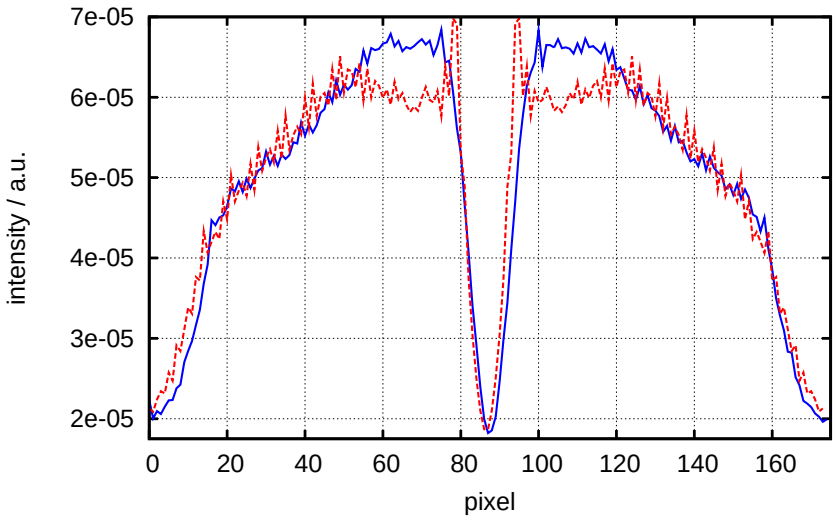
(a) round shape



(b) line profiles for the round shape



(c) myosin shape



(d) line profiles for the myosin shape

Figure 6.19: The figure illustrates the analysis of the spatial frequency sampling for both reconstruction areas. (a) and (c) show the periodograms averaged over all significant eigenfunctions. (b) and (d) show for two opposing directions line profiles through the center as indicated in (a) and (c) with the same color. The dip in the middle marks the origin in frequency space. In both directions from the center the spatial frequencies increase till the Nyquist frequency.

how the reconstruction with the myosin shaped area has been accomplished and what results from it. The mask has been built using a small helix of the symmetry unit reconstructed with the round area. The helix has been composed of three symmetry units and has been projected along its helical axis resulting in a broadened projection image of the density distribution of all slices. Together with an edge filter and broadening of the central area a new binary mask has been created. In order to combine the geometrically restricted area with the data processing workflow described above the input preparation of the images prior to the reconstruction had to be changed. The same assumptions about the filament coherent length, the symmetry and pixel size has been made as stated before. Because the filaments are not aligned in 2D normally each filament may need a different rotation of the mask so that its density is not going to be obstructed or hidden by the mask. In order to avoid the need to invert many different projection matrices each for one rotation of the mask, the 3D alignment results of the round reconstruction area have been used to adjust the filament to fit rotationally inside the mask. The determined azimuth angle and height have been used together with the position of each filament segment inside the filament to find the same phase origin of the helix for all segments. To accomplish that, the slices of the segment images have been shifted cyclically according to their phase origin to suit the readframe for the algorithm. Since the readframe is chosen that is closest to the previously estimated azimuth angle the rotation of the reconstructed density might not be exactly the same for all volumes. Therefore, the same processing after the reconstruction has been done as described above. The result can be seen further below in Fig. 6.24 (a). Shown are only the first five slices of the symmetry unit. They show even a stronger circular density prolongation around the central actin density. The 3D alignment, intended to result in small rotational adjustments, has resulted obviously also in rotations opposing the direction of the long density axis. Due to these severe shortcomings, originated probably by an incorrect data preparation for the read-in before the reconstruction, the effect of solving the equation system for a smaller number of variables cannot be seen in the result. The analysis discussed further below reveals why the data preparation could not succeed. This attempt points out more clearly than the one with the round area that something went wrong.

Another technical note relating to the reconstruction algorithm is the achievement to solve the equation system even for half the size per pixel (1.375 Å). The size of the reconstruction area is shown in comparison with the smaller one in Fig. 6.17. The image dimension could be increased from 161 pixels to 323 pixels. The area comprises about 71000 density points. The corresponding normal matrix that has to be diagonalized thus has 71000 x 71000 entries. For a long time it seemed to be computationally impossible. The entries of the matrix could not be described by a 32-bit index parameter space. The default integer type of the necessary libraries has been 32-bit. 71000^2 exceeds 2^{32}. A recompilation of the linear algebra libraries LAPACK and BLAS with default 64-bit integer type solved the problem.

After about 15 days of computing, using a SunFire 4600 8x Quad-Core AMD Opteron 8384 equipped with 256 GB memory, the matrix has been diagonalized. The reconstruction of each filament segment and its 3D alignment account for another two to three months computing time. The resulting reconstructed symmetry unit, now composed of 20 slices, can be seen in Fig. 6.20. Except for the doubled

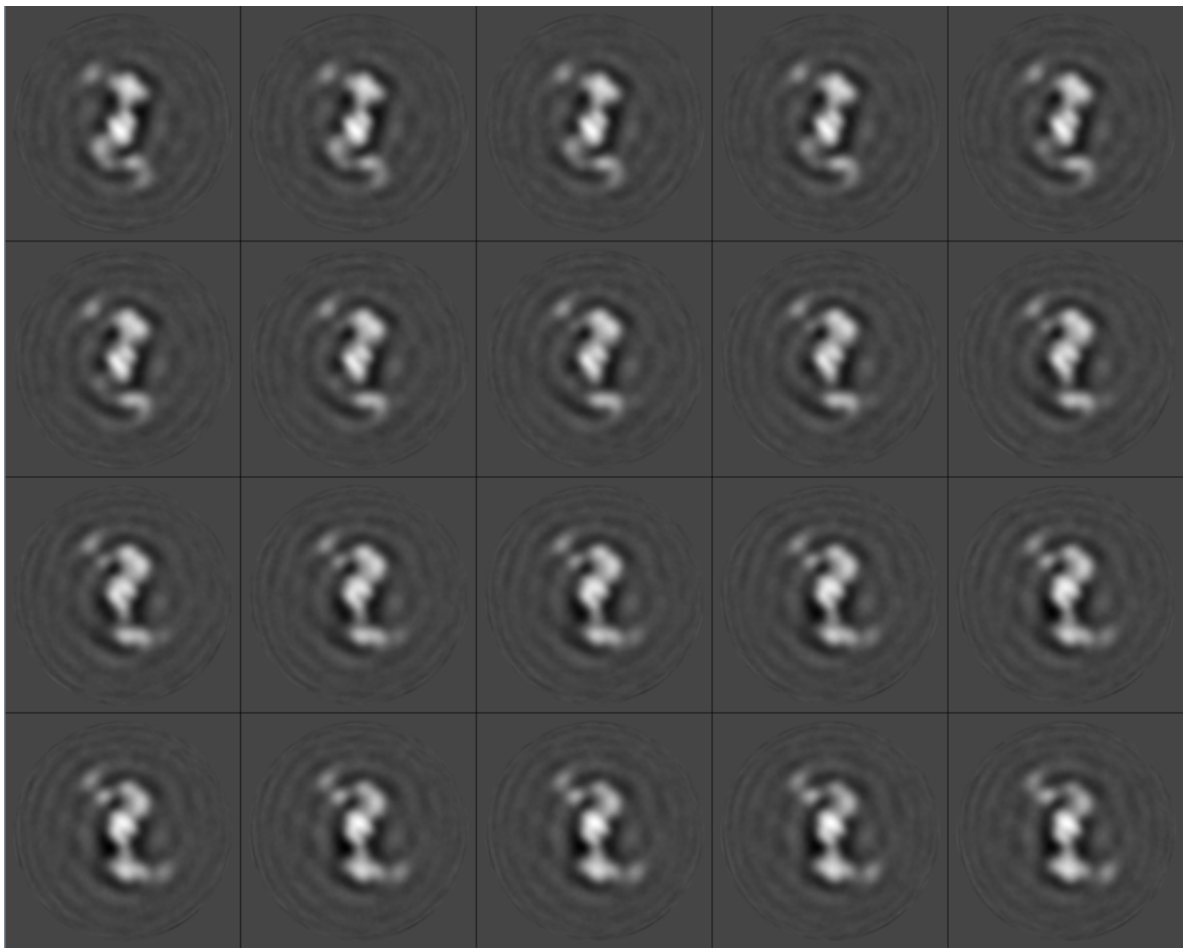
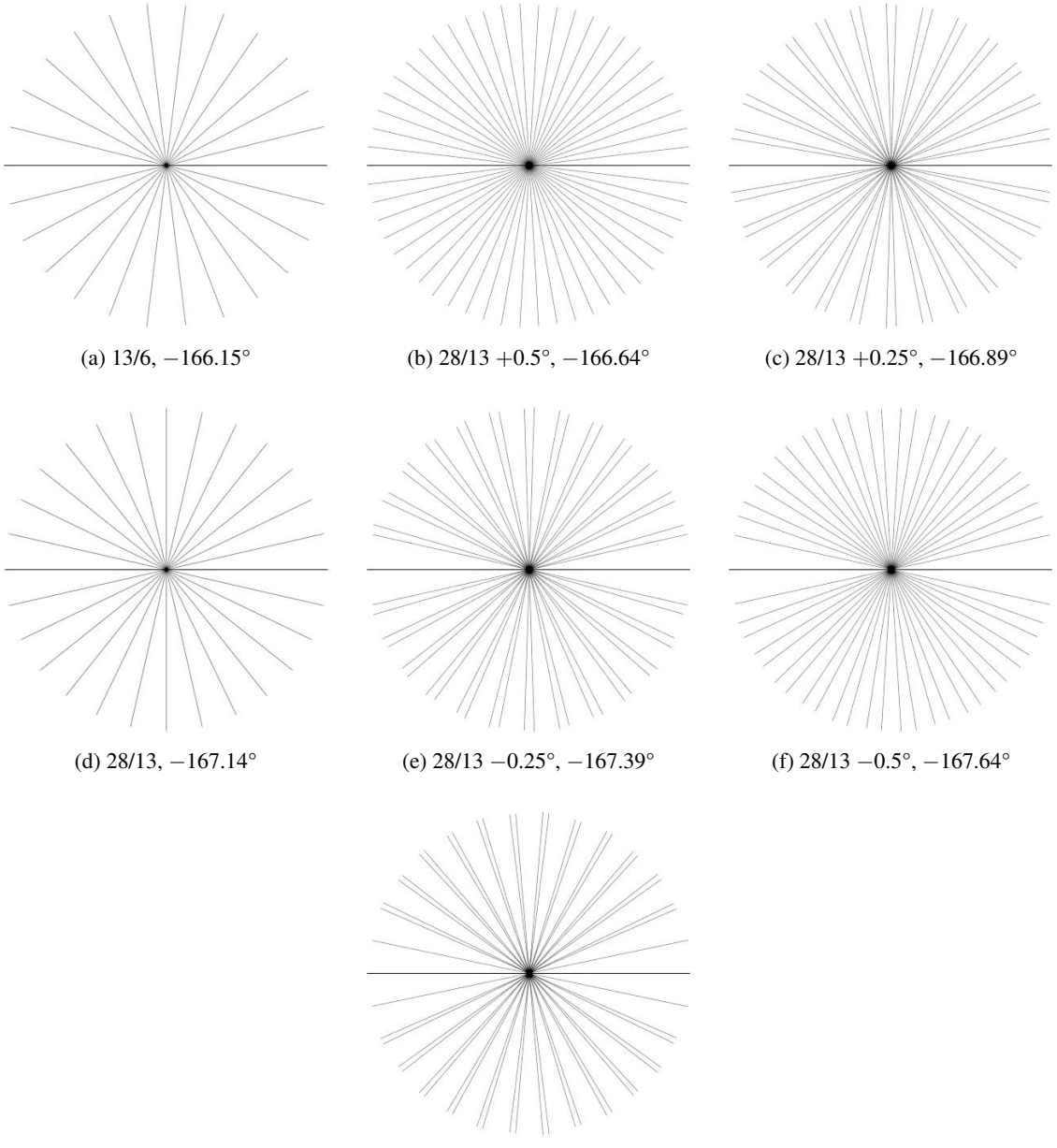


Figure 6.20: Shown are the reconstructed slices of the helical symmetry unit for the supposed 28/13 helical symmetry using half the size per pixel than in Fig. 6.14. The pixel size of 1.375 Å divides the symmetry unit of 27.5 Å in 20 slices that have been reconstructed separately. The image size is 323 x 323 pixels instead of 161 x 161 pixels and as before 44 nm x 44 nm. The reconstruction and 3D alignment have been the same as before.

number of slices it shows very much the same features as the reconstruction with the binned pixel size. Considering the additional amount of computing time, it clearly reassures to use a binned pixel size as far as the Nyquist frequency becomes not too small compared to the features one seeks to resolve within the density.



(a) 13/6, $-166.15°$ (b) 28/13 $+0.5°$, $-166.64°$ (c) 28/13 $+0.25°$, $-166.89°$

(d) 28/13, $-167.14°$ (e) 28/13 $-0.25°$, $-167.39°$ (f) 28/13 $-0.5°$, $-167.64°$

(g) 28/13 $-1°$, $-168.14°$

Figure 6.21: The projection sampling of the Fourier space is shown for different, assumed symmetry angles. The symmetry angles have been chosen close to the 28/13 symmetry angle, since it seems to be close to the true angle. The images have been created using the Fourier-slice theorem and the projection angles.

The further analysis aimed to reduce some of the former stated assumptions as far as the method allows to do so. In the following the testing of different helical symmetries is reported. Unfortunately, it cannot be treated as a free parameter that can be estimated during the reconstruction. Therefore, some other angles are assumed to fit the data. Because initial analysis suggested a symmetry close to the crystallographic symmetry described as 28/13, small deviations from that symmetry angle have been tested. For each symmetry angle a matrix has to be set up and be solved. Upon read-in of the projection data into the reconstruction algorithm each successive projection is assigned a multiple of the new symmetry angle, consequently the Fourier space is sampled in a different way. Figure 6.21 illustrates the sampling for the different symmetry angles. Symmetry angles that have been considered are the 28/13 symmetry angle $\pm 0.25°$, $\pm 0.5°$ and $-1°$. The angle $+1°$ has been replaced by the close by angle of the crystallographic symmetry 13/6. The sampling images have been created considering again the Fourier-slice theorem. For each projection that is considered in the reconstruction a central line with its corresponding projection angle has been placed inside the sampling image. Some symmetry angles result in angular equally spaced projection distributions. Others show missing areas and unequally spaced projections. A 28/13 helix is built up by repeats of two arrowheads (arrowhead A and B: ABABAB...). A changed symmetry angle by about plus 1° results in a 13/6 helix that is composed of repeats of only one arrowhead (AAAAAA...). A symmetry angle in between or close by results in a longer filament fragment that can be used for the read-in, because there are many arrowheads before one equals exactly another. If these cases can be approximated by a crystallographic symmetry u/t the values of the subunits u and the angular full turns t must be very high (ABCDEF...). The missing angular area in some projection samplings of the Fourier space shows up that more projections of a longer filament segment can be used to fill up the Fourier space.

Nevertheless, for all symmetry angles projection data from two arrowheads has been used, except for the 13/6 symmetry. Normally the symmetry allows to use only one arrowhead. The use of exactly two arrowheads leads mathematically to the situation that the projection equation system, that is greatly under-determinated, is over-determinated in all equations at the same time. All projection directions are described by two equations. In the sense of a least squares solution it still can be solved. One has to be careful if the oversampling is not the same for all equations. When the read-in is adopted from the 28/13 symmetry directly without changes to the 13/6 symmetry, stripes appear in all eigenfunctions. 26/12 is the same as 13/6 in the sense of the symmetry it describes. The read-in of 28 instead of 26 symmetry units as projection data results in an unequal over-determination of some projection equations. Subsequently, this leads to stripes that appear in the solution of the equation system. It shows how sensitive the algorithm can be to changes in the projection geometry. Figure 6.22 shows

the spectra of the eigenvalues for the different symmetries. The eigenvalues are sorted in descending

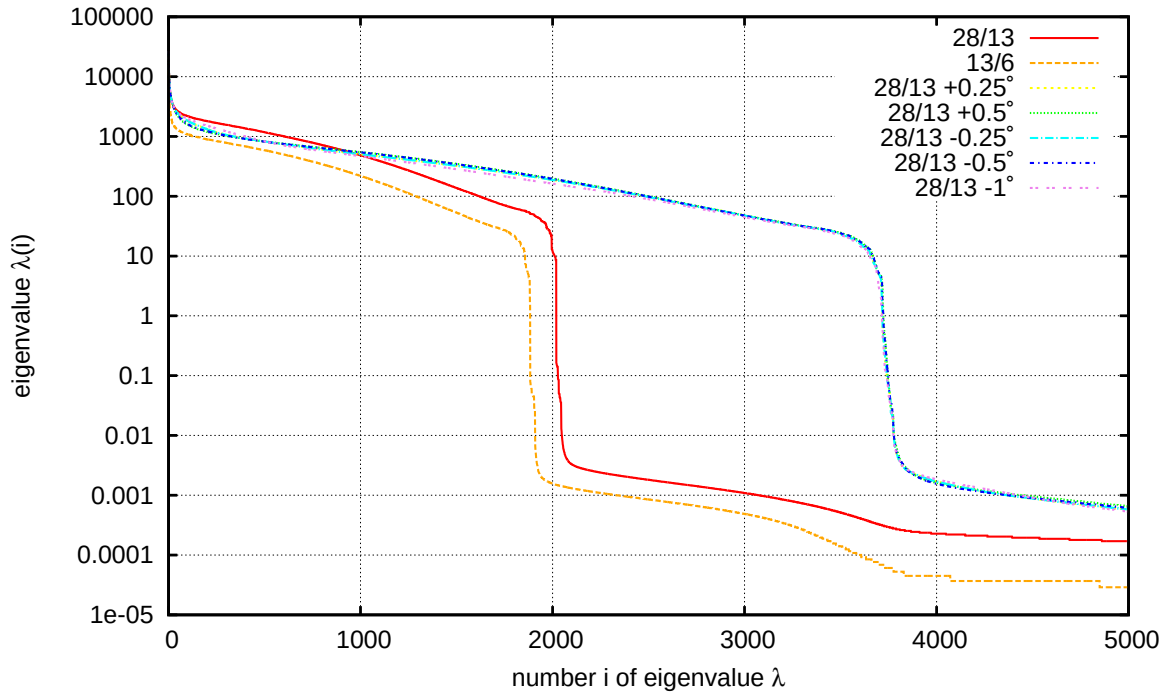
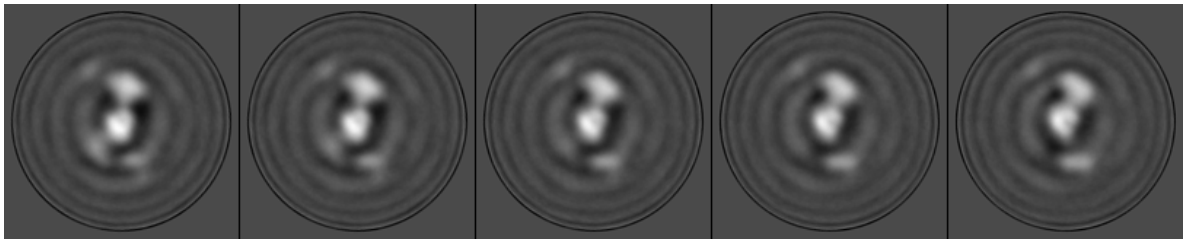
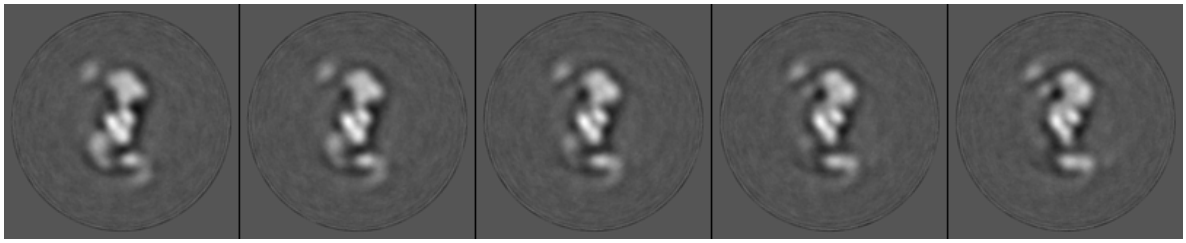


Figure 6.22: The spectra of the eigenvalues are shown for the different symmetries. The eigenvalues are sorted in descending order. Note the logarithmic representation of the ordinate. Non-crystallographic symmetries offer a higher number of significant eigenvalues for the reconstruction due to a finer sampling of the Fourier space.

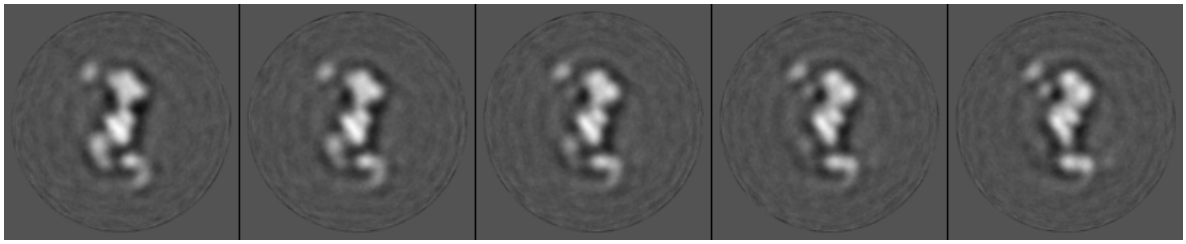
order. It can be seen that for the non-crystallographic symmetries the number of eigenvalues with $\lambda_i > 1$ is almost doubled. For those symmetries a subspace of about 3700 eigenfunctions can be used for the reconstructions due to a finer, albeit unequal sampling of the Fourier space. The reconstructions have been done in the same way as for the 28/13 symmetry except for the different symmetry angle and an adapted read-in for the 13/6 symmetry and the use of more significant eigenfunctions for the non-crystallographic symmetries. The filtering criterion has been the same using all $\lambda_i > \delta$ and $\delta > 1$. The reconstruction of each filament segment is followed by a 3D alignment of the same kind as described above. The final reconstructions are shown in Fig. 6.23 and Fig. 6.24. They are represented each by the first five slices of the symmetry unit and are shown together with the reconstruction using the myosin shaped reconstruction area and a 28/13 symmetry. First of all the form of the density stayed the same for all reconstructions. Only the quality seems to be affected. The symmetry angles with small deviations from the 28/13 symmetry seem to improve the results greatly. However, the different projection samplings of the Fourier space lead to unequal conditions for a comparison of the results. It is difficult to judge what has been the effect of the sampling and what is due to the



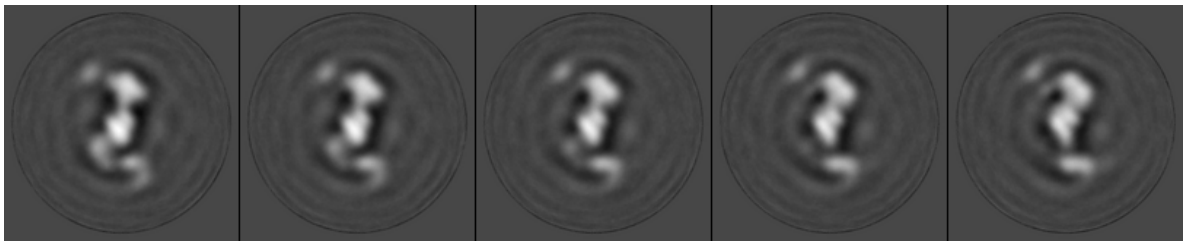
(a) 13/6, $-166.15°$



(b) 28/13 $+0.5°$, $-166.64°$

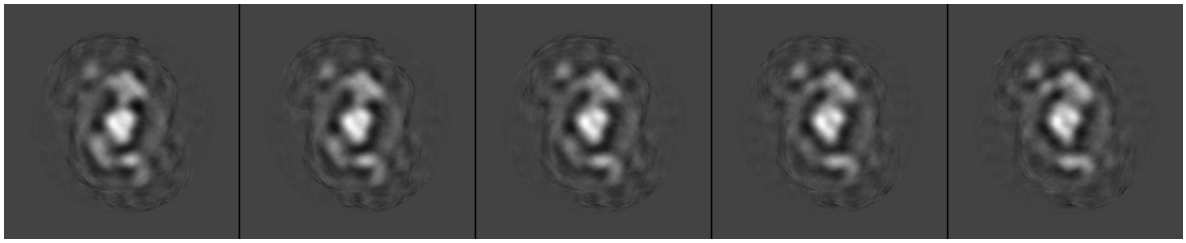


(c) 28/13 $+0.25°$, $-166.89°$

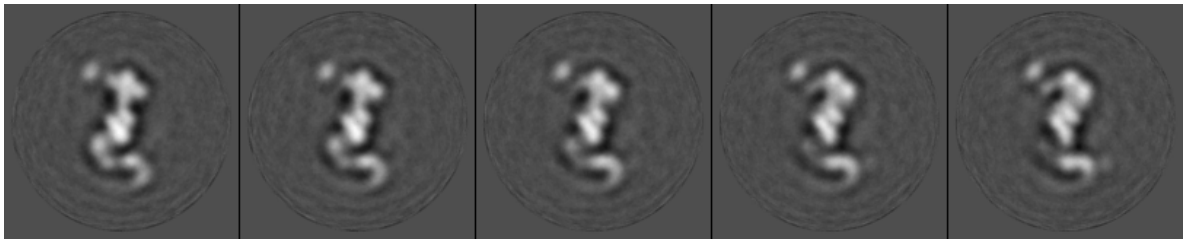


(d) 28/13, $-167.14°$

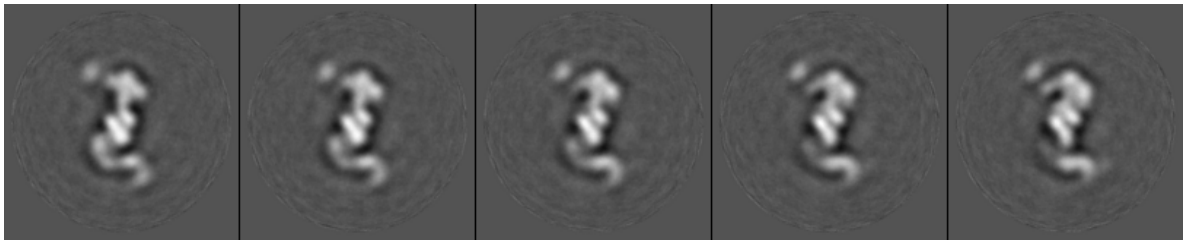
Figure 6.23: Shown are the first five slices out of ten of the final reconstructions for the different, assumed symmetry angles. Each of them is an average over all separately reconstructed filament segments. The reconstruction was done using the filtered least squares algorithm followed by a 3D alignment as described before in the text. The reconstructions are labelled by their symmetry angle. The image size is 161 x 161 pixels with 2.75 Å per pixel (44.3 nm x 44.3 nm).



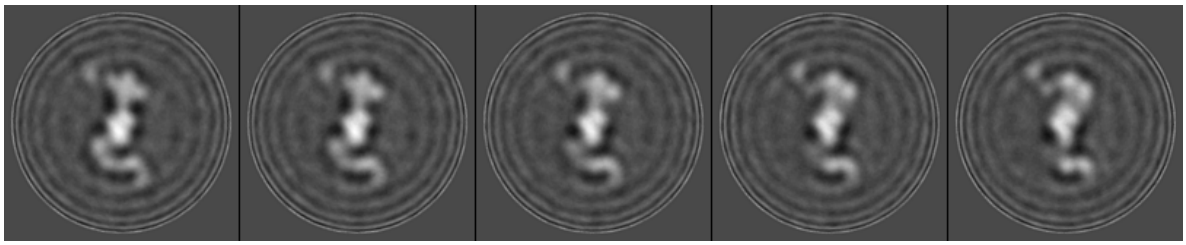
(a) 28/13 myosin mask



(b) 28/13 $-0.25°$, $-167.39°$



(c) 28/13 $-0.5°$, $-167.64°$



(d) 28/13 $-1°$, $-168.14°$

Figure 6.24: Shown are the first five slices out of ten of the final reconstructions for the second part of the different, assumed symmetry angles. Each of them is an average over all separately reconstructed filament segments. The reconstruction was done using the filtered least squares algorithm followed by a 3D alignment as described before in the text. The reconstructions are labelled by their symmetry angle. The image size is 161 x 161 pixels with 2.75 Å per pixel (44.3 nm x 44.3 nm).

changed symmetry angle that may describe the data set better or worse. The 13/6 symmetry has the worst projection sampling due to its short repeat of one arrowhead. The quality of the reconstruction therefore is inferior to those of the 28/13 symmetry. All reconstructions with symmetry angle deviations from 28/13 by ± 0.25 and ± 0.5 exhibit a higher quality than the 28/13 reconstruction. The reconstruction with the symmetry angle of $-166.64°$ shows an almost flat background without density prolongation. Furthermore, it is the symmetry with the best projections sampling. The sampling is fine, equally spaced and without a missing area. The fact that the reconstructions become better with a symmetry angle deviating in both directions based on the 28/13 symmetry suggest the assumptions that the sampling has a dominating influence on the quality of the reconstruction. This makes it even harder to discern which angle may describe the data best. The reconstruction with $-168.14°$ symmetry angle shows fringes close to the border of the reconstruction area. They seem not to be related with the reconstructed object. Therefore, they might emanate from the eigenfunctions that are based on a projection sampling that is unequal to a higher extent than for the other symmetries. In order to find the appropriate angle for the data set a closer look has been taken at the myosin density at higher radius, following the concept that a correct symmetry angle should give a maximal density. Radial density profiles have been calculated. Since there might be a different amount of negative density in the results, a rotational average has not been used. Instead, the symmetry units have been projected to obtain one image and within thin rings the maximum intensity has been estimated. The result is shown in Fig. 6.25. Both reconstructions with 28/12 $\pm 0.5°$ exhibit higher intensity for the central actin density. They belong to the projection samplings with the finest and thereby as much as possible equal distribution of projection directions. However, the one with the best sampling ($-166.64°$) has not the highest intensity for the myosin density at higher radius. It reassures that this analysis indeed might be sensible to the true symmetry angle present inside the data set. The highest intensities for the myosin density belong to the reconstructions with symmetry angles of $-167.39°$ and $-167.64°$. Both are deviations to the same direction of 28/13. The intensities for the $-168.14°$ symmetry are rising again towards the outer border. It is due to the fringes present in those reconstruction. Further, the integrated area below the radial density profiles has been estimated and the results are shown in Fig. 6.26. The areas have been calculated by trapezoidal numerical integration using MATLAB, from the MathWorks Inc., version R2013a. It can be seen, that the $-167.39°$ symmetry gains the highest amount of intensity for the myosin density. It catches up even with the higher density the $-166.64°$ symmetry has for the actin. It might be interpreted that the $-167.39°$ symmetry represents the data set best out of those symmetries that have been tested.

Next follows the description of the attempt to reduce some more ad hoc assumptions. In the frame-

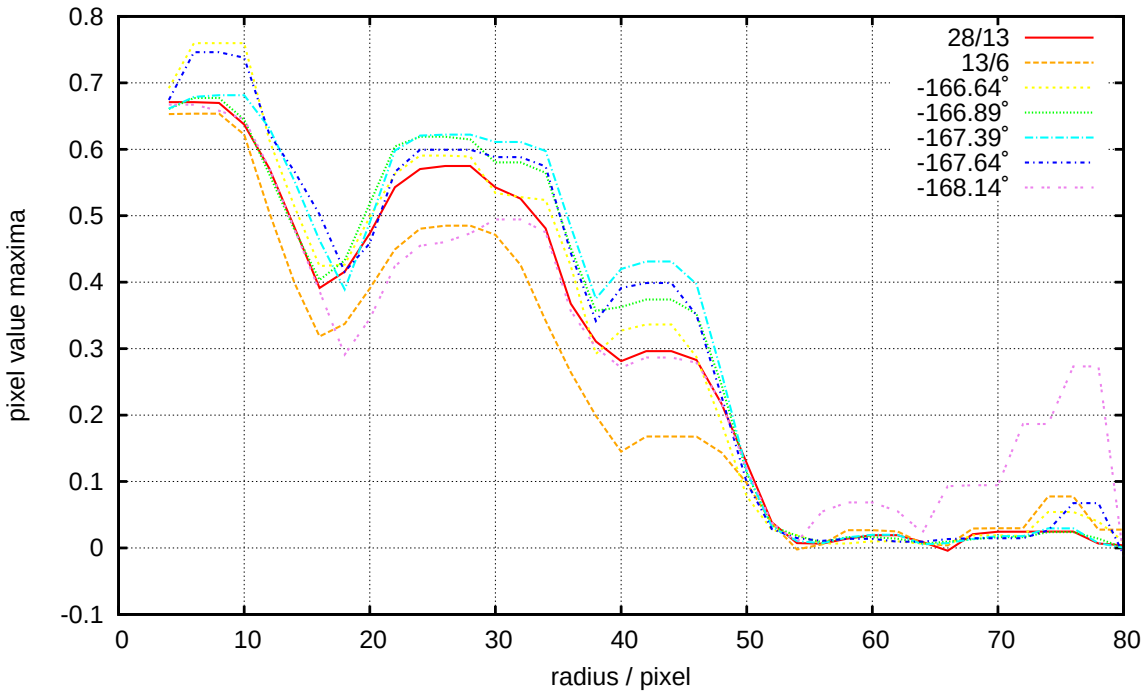


Figure 6.25: Radial profiles have been obtained for the projected symmetry units of the different reconstructions shown in Fig. 6.23 and 6.24. The profiles reflect the maximal values within thin rings of the projected volumes.

work of the current implementation of the filtered least squares method the reconstruction as axial tomogram cannot be changed, but the alignment of each filament segment volume can be altered. The coherent length of the filaments has been assumed previously to be longer than the imaged length of the filament. It may not be justified. A long coherent length still has to serve for the read-in of the projection data, but during alignment the length can be assumed to be shorter. Unfortunately, it is only a semi-reduction of the assumptions. For a start the segment volumes still have been averaged for the filaments they belong to, but the azimuth angle between them has been estimated with an additional alignment and was no longer set to the symmetry angle. Each symmetry unit volume of a filament segment has been projected to one image prior to the alignment in order to have a higher signal-to-noise ratio and less computational costs. A reference-free rotational alignment of each image series belonging to one filament has been done with SPIDER. The difference angle between adjacent segments has been plotted as histogram in Fig. 6.27. The angle distribution shows a nice peak without a broad distribution close to it. The peak is located within the accuracy of the alignment quite close to the angles discussed above. The resolution of the rotational alignment is not high enough (cf. Fig. 6.27 (b)) to determine the angle as precise as used before. The clear peak assures that the symmetry

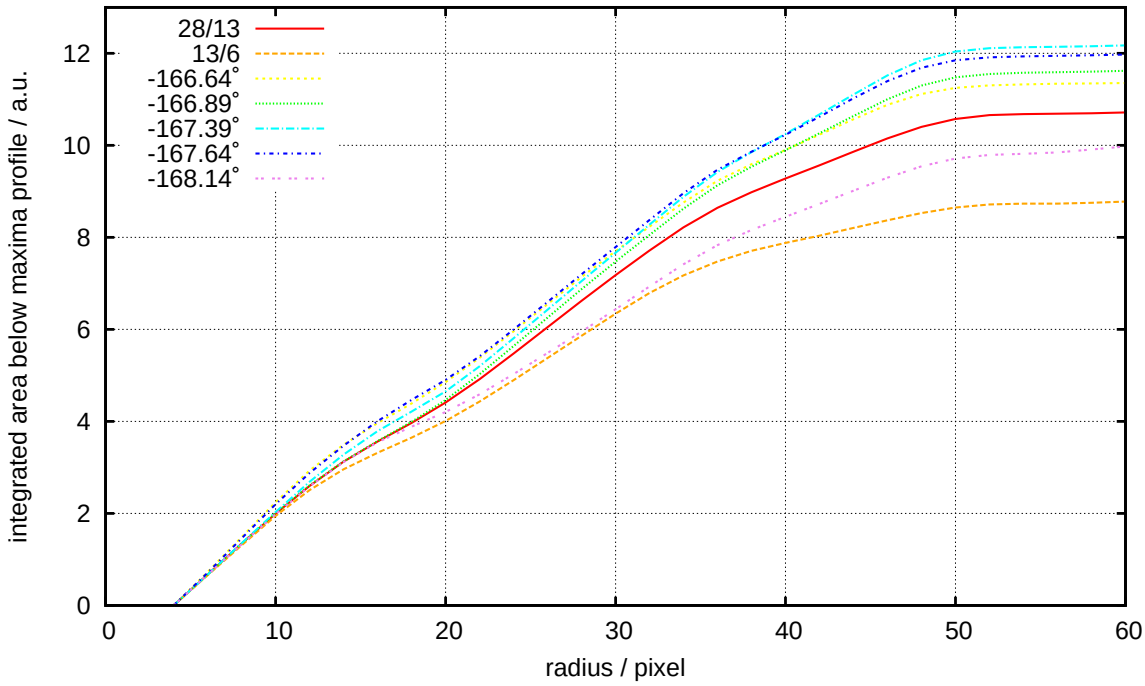
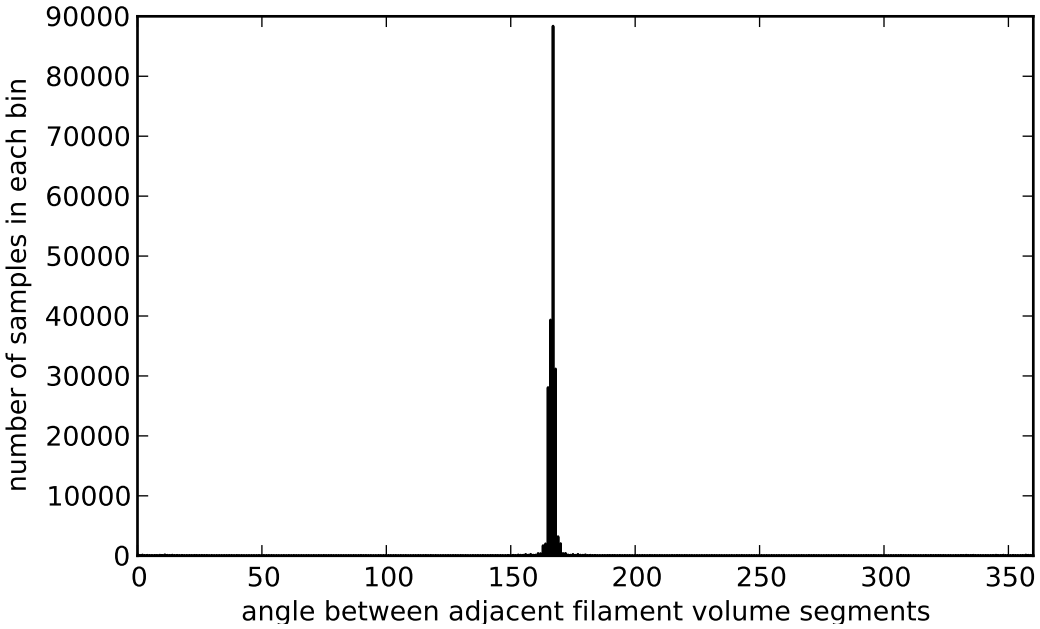


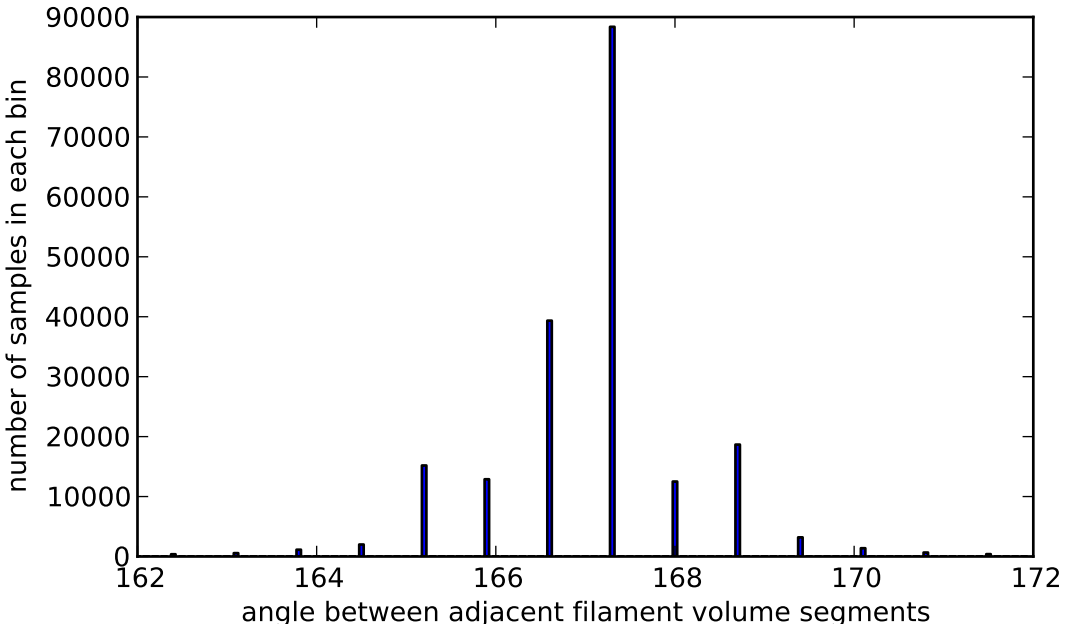
Figure 6.26: The figure shows the integrated areas below the radial density profiles for the different symmetries. The area has been integrated numerically as function of the radius and gives an estimate of which symmetry reconstructs the highest myosin density.

angle really might be one of the above tested ones. At the same time it shows that there is enough signal inside the single segment volumes for a further alignment. All segment volumes of one filament have been averaged after rotation by the estimated angle. The subsequent 3D alignment of all filament volumes has been the same as stated above. Figure 6.28 shows the comparison between the result of the former alignment and the new result for one filament volume. The comparison shows that using a rotational alignment to average the segment volumes of one filament more density of the object is coming up. It indicates that the coherent length of the filaments indeed may not be as long as assumed earlier. At the same time the analysis shows with its narrow angular distribution that the symmetry angle is quite constant.

The certainty to have enough signal inside the single segment volumes for further alignments and the insight that the coherent length might be rather short allows to dare the attempt to have a complete 3D alignment for every segment volume without averaging along the filaments. It corresponds to the attempt to account even for a changing height of the symmetry units along the filament and thereby it reduces once more a former ad hoc assumption. Although it is again only a semi-reduction of the

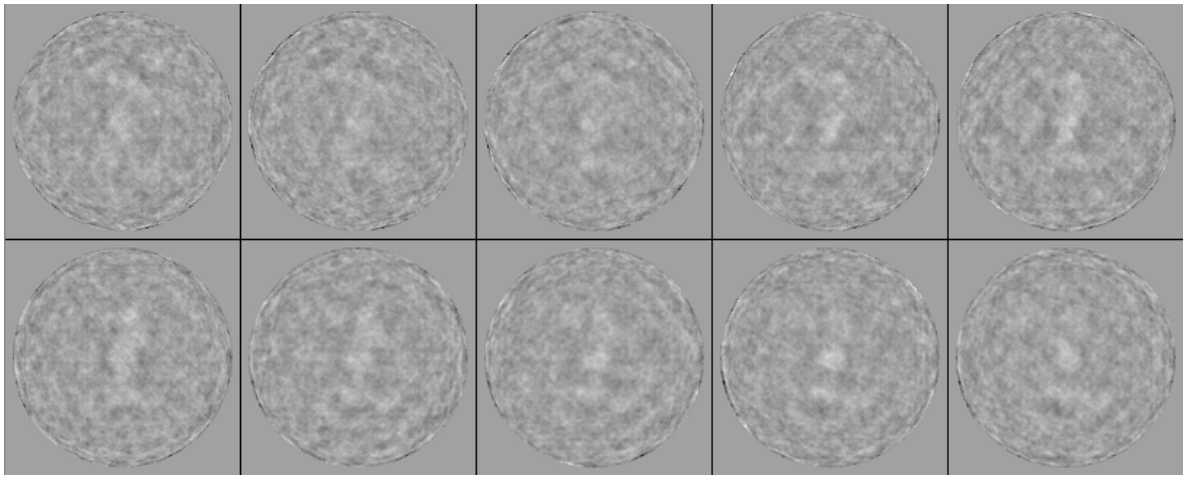


(a)

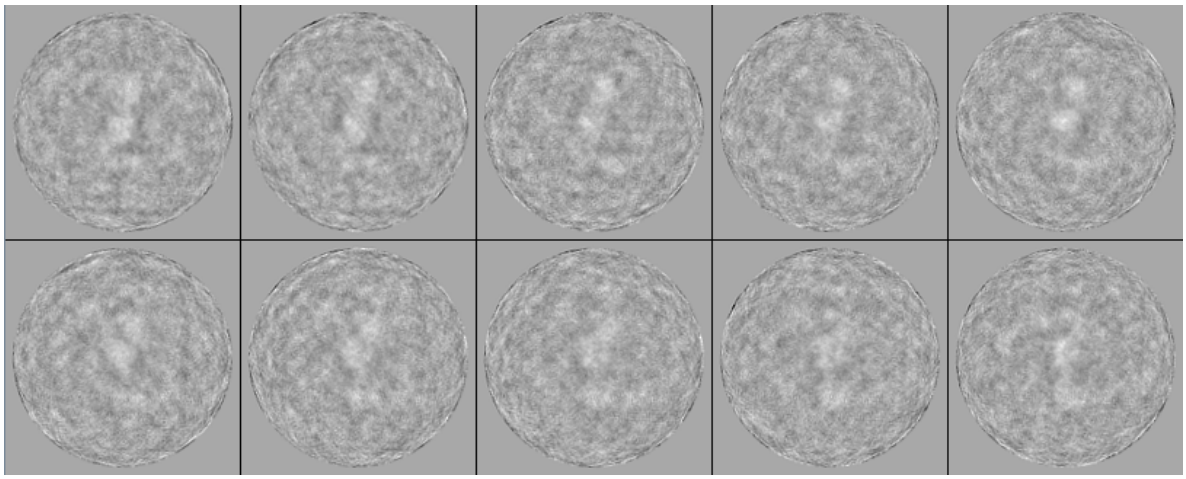


(b)

Figure 6.27: Histogram of the distribution of the angle between adjacent segment volumes. All segments, in total about 200000, have been considered. (a) shows the whole angular range and (b) is zoomed in to the interesting range. The distance of the bins in (b) reflects the limited resolution of the rotational alignment. The difference angles are shown with positive sign.



(a) averaging along the filament due to the symmetry angle



(b) averaging along the filament with rotational alignment

Figure 6.28: Reconstructed slices of the symmetry unit for one filament. (a) shows the result of the former alignment using the fixed symmetry angle to rotate the segment volumes. It corresponds to an assumed coherent length of the filament longer than the imaged part of it. (b) presents the result using a rotational alignment of the segment volumes prior to averaging. It represents the attempt to account for a possibly shorter coherent length or imperfect symmetry along the filament. Both reconstructions have been done using the filtered least squares reconstruction with $\lambda_i > 1$ and 28/13 symmetry. The size per pixel is 2.75 Å and the image size is 161 x 161 pixels as before (44.3 nm x 44.3 nm).

assumption, since the read-in of the projection data still assumes a read frame of 10 slices. A 3D alignment of each segment volume allows to determine the azimuth and height separately for each. Thereby, the computational costs increase dramatically. Instead of about 3000 alignments 200000 alignments are needed. In order to compare the result with the former obtained ones the independently 3D aligned segment volumes have been averaged for one filament after the alignment. The result can be seen in Fig. 6.29. The reconstructed slices of the symmetry unit belong to the same fil-

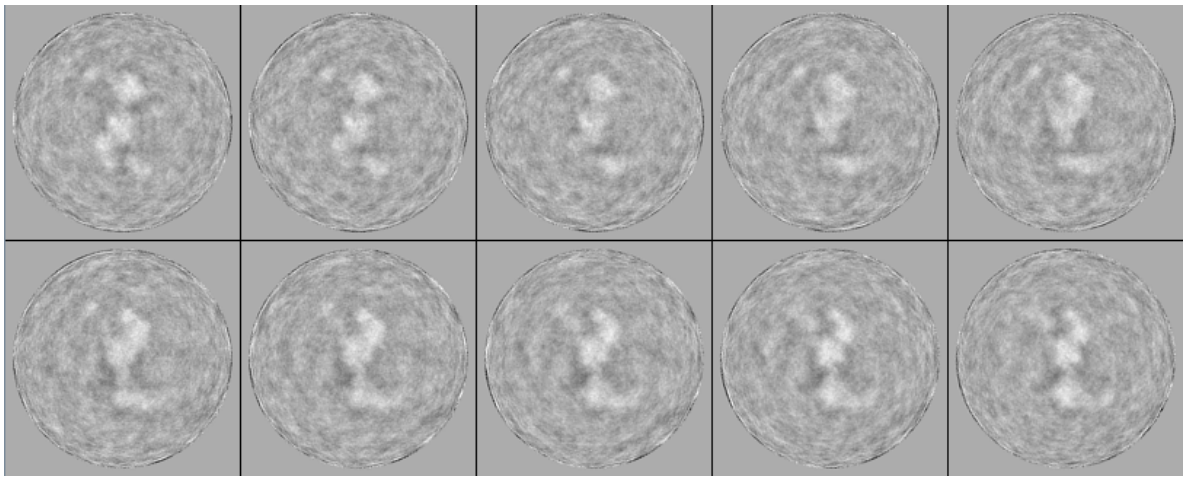


Figure 6.29: Reconstructed slices of the symmetry unit for the same filament as in Fig. 6.28. Each segment volume has been subject to an individual 3D alignment in order to determine its azimuth and height with respect to a reference. The segment volumes have been averaged afterwards to be comparable with the former results. The averaging before the alignment has been omitted to account in a better way for symmetry changes along the filament. The alignment thus has been done without any ad hoc assumptions. The reconstruction has been done as before using the filtered least squares reconstruction with $\lambda_i > 1$ and 28/13 symmetry. The image size is the same as before (44.3 nm x 44.3 nm).

ament as in Fig. 6.28. The quality of the result could be improved dramatically using no assumptions for the 3D alignment. Only the necessary assumptions have been made in order to use the reconstruction algorithm in its current implementation for axial tomography. It reveals how far the acquired data set is from being a set of axial tomograms of ideal helical objects with a long coherent length. Such an increase in reconstruction quality can only be explained by the fact that the alignment parameters differ from those that are theoretically given by a perfect helical symmetry. The inspection of the alignment parameters reveals why the former attempts had to fail. Figure 6.30 and 6.31 show the azimuth angle and height parameter with one data point per segment volume for two different filaments. Figure 6.30 greatly helps to explain the small improvement between Figure 6.28 (a) and (b) compared to the huge improvement of Fig. 6.29. The azimuth angle shows again a quite narrow distribution

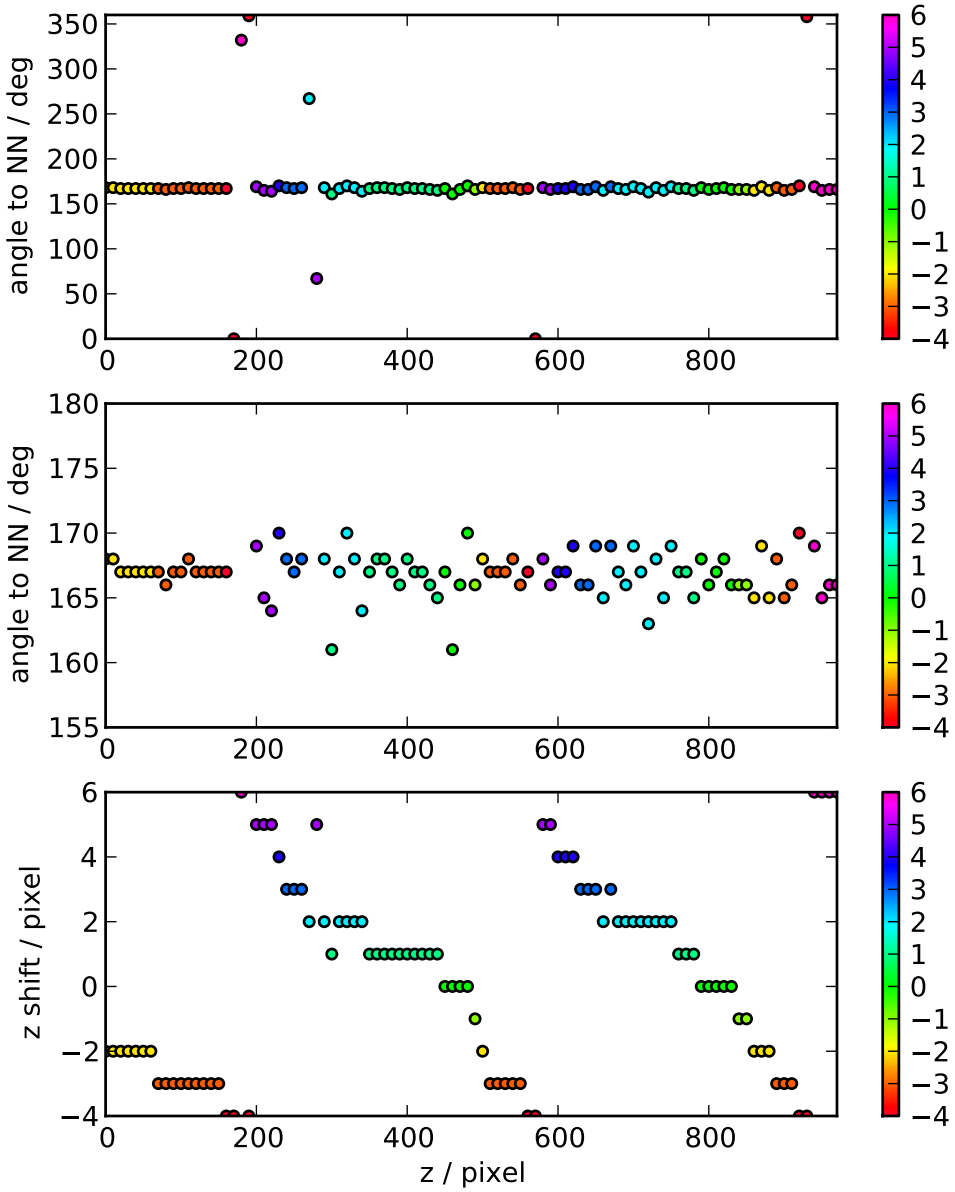


Figure 6.30: Alignment parameters for all segments of one filament. The 3D alignment has been done without any ad hoc assumptions. The filament is the same as for the reconstructed slices shown in Fig. 6.28 and 6.29. Plotted are the azimuth angle to the nearest neighbor (NN) segments and the determined height (color coded) as function of the filament length z. Height -4 and 6 are the same but correspond to different azimuth angles. The upper plot shows the whole azimuth range of 360°. The plot below is zoomed in. The lowermost plot illustrates the height shift along the filament not only by the color code but also as data of the ordinate.

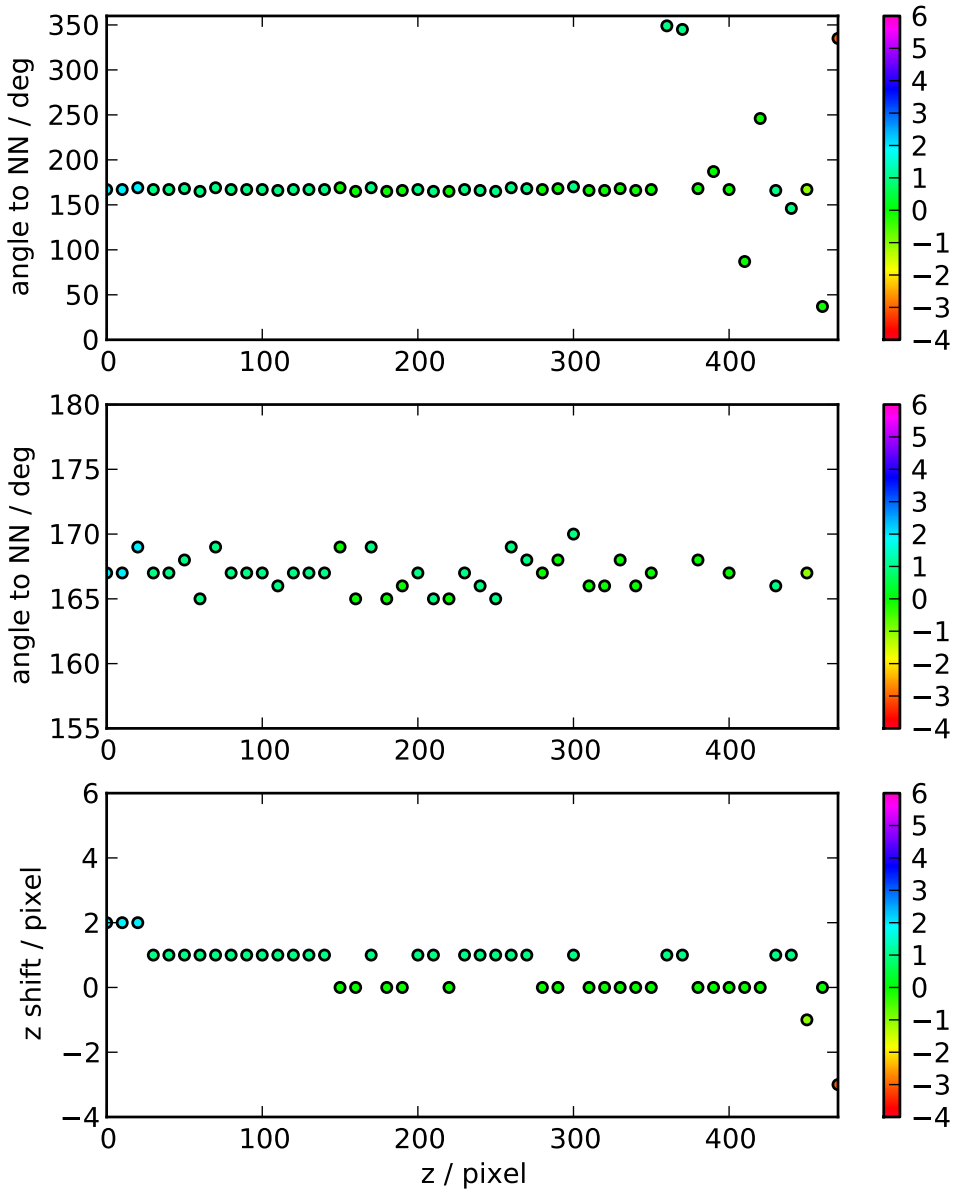


Figure 6.31: Alignment parameters for all segments of a second example filament. Again the azimuth angle to the nearest neighbor (NN) segments and the determined height (color coded) as function of the filament length z are plotted. Each segment is represented by one data point. The upper plot shows the whole azimuth range of 360°. The plot below is zoomed in to the symmetry angle. The lowermost plot illustrates the height parameter of the segments not only by the color code but also as data of the ordinate. Again the difference angles are shown with positive sign.

around the symmetry angle. Hence, the rotational alignment alone yields only a small improvement, since the former assumed fixed symmetry angle has been not as bad as it might have been. The huge problem of the data set is exposed by the continuous shift of the height from one color to the next. Also depicted by the steep gradient in the lowermost plot. Therefore, only the alignment with the height as free parameter for all segments could give the improved result. The symmetry unit obviously does not correspond to ten slices of the projection data. This could have two reasons. One could be an incorrect estimated pixel size. Then the gradient for the height shift has to be the same for all filaments. This is not the case as can be seen in Fig. 6.31. The other reason is that the filaments possess a tilt out of the projection plane and therefore appear shortened in the projection image. Depending on the amount of the out-of-plane tilt all kinds of gradients for the height shift may appear. Only filaments without a gradient in the lowermost plot are lying within the projection plane. The second filament (cf. Fig, 6.31) shows only a minor height shift. So far it cannot be excluded that additionally to an out-of-plane tilt there is an inaccuracy in the pixel size. It might be a combination of both. The consequences are severe with respect to the applicability of the filtered least squares reconstruction algorithm in its current implementation. There are many filaments that show a height shift in the parameter analysis. The one of Fig. 6.30 exhibits a shortening of about one symmetry unit within 400 pixels. It could even be used to estimate the out-of-plane tilt. But it does not help much. The current implementation of the reconstruction algorithm needs to read the projection data as axial tomogram. Only for an axial tomogram the dimensional reduction from 3D to 2D can be done. The necessity of an axial tomogram implies that the data set has to be composed of filaments without any out-of-plane tilt. Obviously, the present data set does not meet this requirement. Upon read-in subsequently wrong projection angles have been assigned to the image slices using the reading frame of ten pixels. A data set which resembles axial tomograms might even have produced acceptable results with the first attempt to use a perfect symmetry both during the reconstruction and the alignment. The coherent length might not be that short. The biggest problem is given by the out-of-plane tilt. The last-mentioned alignment without any assumptions probably could achieve such improvement as it tries to correct the errors that have been made during the read-in. It is amazing that, despite a wrong projection assignment for the reconstruction to some of the image slices or rather the read-in of wrong images slices for the next projection angles, the result can be like that in Fig. 6.29. Due to the fact that the data set cannot be reconstructed to its true resolution by a treatment as axial tomogram, no final result will be shown of a 3D alignment of all 200000 segment volumes. Certainly it will be better than the one shown in Fig. 6.15 but it cannot reflect all the details present inside the data set.

6.3 Discussion

A data processing workflow has been implemented from the filament selection to the final 3D alignment of the reconstructed volumes to be able to use the filtered least squares algorithm for the reconstruction. The reconstruction is calculated using centered, small segments of the filaments. Later on the single segment volumes are aligned to each other in 3D. Different alignment strategies have been followed. Unfortunately, the results have not been as good as expected. Thus, the applicability of the filtered least squares reconstruction algorithm needs to be discussed: It has been intended to be used due to its potential to reconstruct the object's density with a better defined density localization than the widely used weighted backprojection. Thereby, it should optimize the information retrieval from the projection images and enhance the quality of the reconstructed density. Unfortunately, the positive properties that have been shown in the analysis in the preceding diploma thesis (Wulf, 2009) could not be utilized here. The necessary requirements for the current implementation of the algorithm do not meet the flexible properties of the filaments contained in the acquired data set. Due to the dimensional reduction of the reconstruction problem from 3D to 2D the algorithm requires properties of the data that have to be fulfilled exactly. The filaments have to lie exactly within the projection plane to justify the treatment as axial tomogram followed by the 2D reconstruction. Next the pixel size has to be exactly determined to meet the projection data read-in for the reconstruction. If the size in pixel of the symmetry unit does not meet the reading frame wrong image slices will be read for the assignment of the subsequent projection directions. The analysis presented in this chapter finally ended up with the finding that both requirements are not fulfilled within the data set that has been acquired.

The data set may have some shortcomings. The filament amount has been such that there are still filament cross-overs present in the data. It may have promoted the high out-of-plane tilt distribution that could be revealed by the final 3D alignment step. Compared to the out-of-plane tilts the search for the accurate symmetry angle seems to be solvable. Unfortunately, it cannot be treated as a free parameter for the reconstruction as it can be done for the alignment. However, the symmetry angle could be located quite accurately although the projection sampling of the Fourier space has a strong influence on the quality of the reconstruction and may overshadow the actual effect of the angle. With a better resolution for the rotational alignment of the azimuth determination the mean angle between nearest neighbor segments can be used to determine the exact symmetry angle being present in the data set. Later on it could be used to set up a new matrix for a refinement step. The present out-of-plane tilt distribution of the filaments cannot be accounted for in the current implementation. A completely new implementation of the least squares algorithm without the simplification for axial tomograms could solve the problem. But it might still be questionable if the computational complexity due to a huge 3D

projection matrix will be manageable. So far the program can deal best with rigid filaments that are lying flat inside the amorphous ice layer and feature a long coherent length without symmetry variations. How far these attributes can be fulfilled by the actomyosin complex was part of the analysis described in the next chapter.

Chapter 7

Electron tomography reveals flexible specimen properties

Electron cryo tomography of the actomyosin complex was performed parallel to the data analysis using the least squares algorithm. The intent has been to investigate single actin filaments with myosin decoration without the need to average over different samples in order to obtain 3D information about the preferred way the filaments distribute themselves inside the ice layer, their straightness, their tilts out of the projection plane and additionally to gain insight into the quality of the decoration and its symmetry. The work has been done in the framework of the bachelor thesis of Sebastian Bierbaum (Bierbaum, 2012). The supervision has been part of this dissertation. Electron tomography of specimens with a size that normally is going to be analyzed using single particle methods and high magnifications brings along some technical challenges. Normally larger objects are the subject to tomography and often in a first instance chemically fixed samples are used. The need of a sufficiently high magnification to reveal the details and at the same time the need to use a very low electron dose in order not to burn the cryo specimen posed the challenge to visualize single filaments with higher signal-to-noise ratio without any need of averaging techniques.

7.1 Material and methods

7.1.1 Sample preparation

The sample preparation again has been quite difficult. The aim has been to prepare grids that resemble the specimen properties that have been seen within the data set of the 3000 filaments. The same circumstances appeared as stated in Section 6.1.1. The samples that could be used for data acquisition have been prepared as follows: The actin buffer was composed as before of 5 mM Tris, pH-value of

7.5, 60 mM KCl, 2 mM $MgCl_2$, 0.2 mM $CaCl_2$, 1 mM NaN_3 and 0.2 mM ATP. An aqueous dilution of fresh actin was prepared with an actin concentration of 14.2 µM. The emerging actin filaments were stabilized using 200 µM phalloidin. The myosin had been the same as for the sample preparations before, the myosin V provided by Prof. Dr. H. Lee Sweeney (University of Pennsylvania, Philadelphia, USA) in May 2010 and stored in the freezer at -20 °C. The dilution buffer to obtain arbitrary protein concentrations consisted of 5 mM Tris with pH-value 8.0, 10 mM KCl and 0.5 mM $MgCl_2$. Tomography needs additionally fiducial markers inside the ice layer to be able to align later on the different tilted views to each other. For this purpose gold nanoparticles of 10 nm diameter from Aurion have been used. They have been saturated with the protein bovine serum albumin (BSA) to prevent them from clustering that has been observed before. Excessive BSA has been removed by centrifugation. The original solution of the nanoparticles has been exchanged with the previously described dilution buffer to be able to mix them with the proteins without changing their buffer conditions. The nanoparticle concentration has been increased ten times compared to the original solution. The proteins were pre-mixed in a small reaction tube prior to the application to the grid. 17 µl of 2 µM concentrated myosin solution was added to 17 µl of 0.5 µM concentrated actin solution. Additionally, 3.5 µl gold nanoparticle solution was added. The final actin to myosin ratio has been about 1 : 4 with the final actin concentration of 0.22 µM and final myosin concentration of 0.9 µM. Again Quantifoil® molybdenum grids with a mesh size of 200 and 0.6 µm holes within the carbon support film (R0.6/1) have been used. This time they have not been washed with organic solvents. Directly before usage they have been glow discharged using the advanced plasma cleaning system Solarus Model 950 from GATAN. The parameters have been set to the power of 50 W for a duration of 3 s in a low pressure nitrogen-oxygen atmosphere. Thereby, the grids become quite hydrophilic and any hydrocarbon contamination is being removed. The grid has been placed into the humidity chamber of the vitrification device, a Vitrobot Mark IV from FEI. Inside the chamber 4 µl of the pre-mixed sample solution have been applied to the grid. The vitrifying conditions were set to room temperature (22 °C), 100 % relative humidity within the chamber, 7 s blotting time and -25 for the blot force parameter. Whatman paper No.1 was used for two-sided blotting. The sample was rapidly frozen following the automated procedure of the Vitrobot.

7.1.2 Data acquisition and 3D reconstruction

The tomographic tilt-series have been acquired under low dose conditions using a FEI Titan Krios electron microscope at an acceleration voltage of 200 kV. The microscope is equipped with a GIF Quantum post-column energy filter and an UltraScan 1000 CCD camera with 2048 x 2048 pixels, both devices are from GATAN. The energy filter has been used with a slit width of 15 eV to obtain zero-loss

filtered images without a huge inelastic contribution. A 70 µm sized objective aperture was inserted. The image pixel size has been estimated to 4.2 Å using a nominal magnification of 33000x. The tilt-series ranges from -65° to +65° and followed a Saxton scheme (Saxton et al., 1984) with the starting point at 0°. The increment at zero degree $\Delta\alpha_0$ has been 3°. Instead of using a fixed tilt increment the Saxton scheme uses variable tilt increments $\Delta\alpha_i$ according to the formula $\Delta\alpha_i = \Delta\alpha_0 \cdot \cos(\alpha_i)$. Because the tilt increment depends on the cosine of the tilt angle, the angular sampling becomes finer at higher tilts. It is described as an optimized dose distribution with the need of a minimized number of images for a given resolution (Saxton et al., 1984). The Explore 3D software from FEI has been used for automated data acquisition. The automation has not been reliable in all attempts. In order to have the same features centered in the field of view in each subsequent acquired image, tracking has been done at the nearby focus position, that has to be situated on the tilt axis in order to have the same height as the exposure position. The tracking has been done before each subsequent image. Nevertheless, in many attempts the position drifted away from image to image. It might be promoted by the small field of view of the camera. It had to be checked prior to the acquisition if at high stage tilts the view is going to be obstructed by the grid bars or the edge of the grid. The eucentric height was adjusted mechanically at the beginning of the series and for the tomogram discussed here the defocus was set to 3 µm. The total electron dose does not exceed 100 $e^-/Å^2$. It has been distributed by adjusting the exposure time with an intensity ratio of 2.1 for the image intensity at 0° compared to the intensity at 60° tilt. The initial exposure time was 0.61 s. The series consists of 57 images.

The 3D reconstruction was accomplished using the IMOD software from the Boulder Laboratory for 3D Electron Microscopy, University of Colorado, USA (Kremer et al., 1996). It is developed especially for electron tomography and is based on a weighted backprojection reconstruction. The alignment of the image stack was done using 54 fiducial markers that could be tracked within the field of view in all tilt images. To account for local distortions the images have been divided into 5 x 5 local image patches. The final reconstruction was low pass filtered by a Gaussian filter to reduce high frequent noise. Furthermore, nonlinear anisotropic diffusion (NAD) filtering within IMOD has been applied to the tomogram to achieve better contrast (Frangakis and Hegerl, 2001). NAD is a technique to reduce image noise while preserving much of the significant image details like edges. Thereby, it optimizes the contrast without a significant loss of resolution. The parameters for the NAD filtering, that control the filtering strength (K) and the number of iterations (I), have been tested on a small subvolume. Finally a value of 20 was used for both parameters K and I. The final tomogram thickness (z-direction) comprises 139 nm in 330 image slices. Its x and y dimensions are about 860 nm. The z-direction corresponds to the direction of the incident electron beam.

7.2 Results

The amount of filaments within the ice layer of the prepared TEM grids is higher than for the data set discussed before. Because for lower filament concentrations a good decoration could not be achieved, the high amount of filaments has been accepted. Three good tilt-series could be acquired. The one that is going to be discussed here is characterized by the nicest resolved helical pattern of the filaments and is at the same time the thinnest tomogram. Figure 7.1 shows a three-dimensional representation

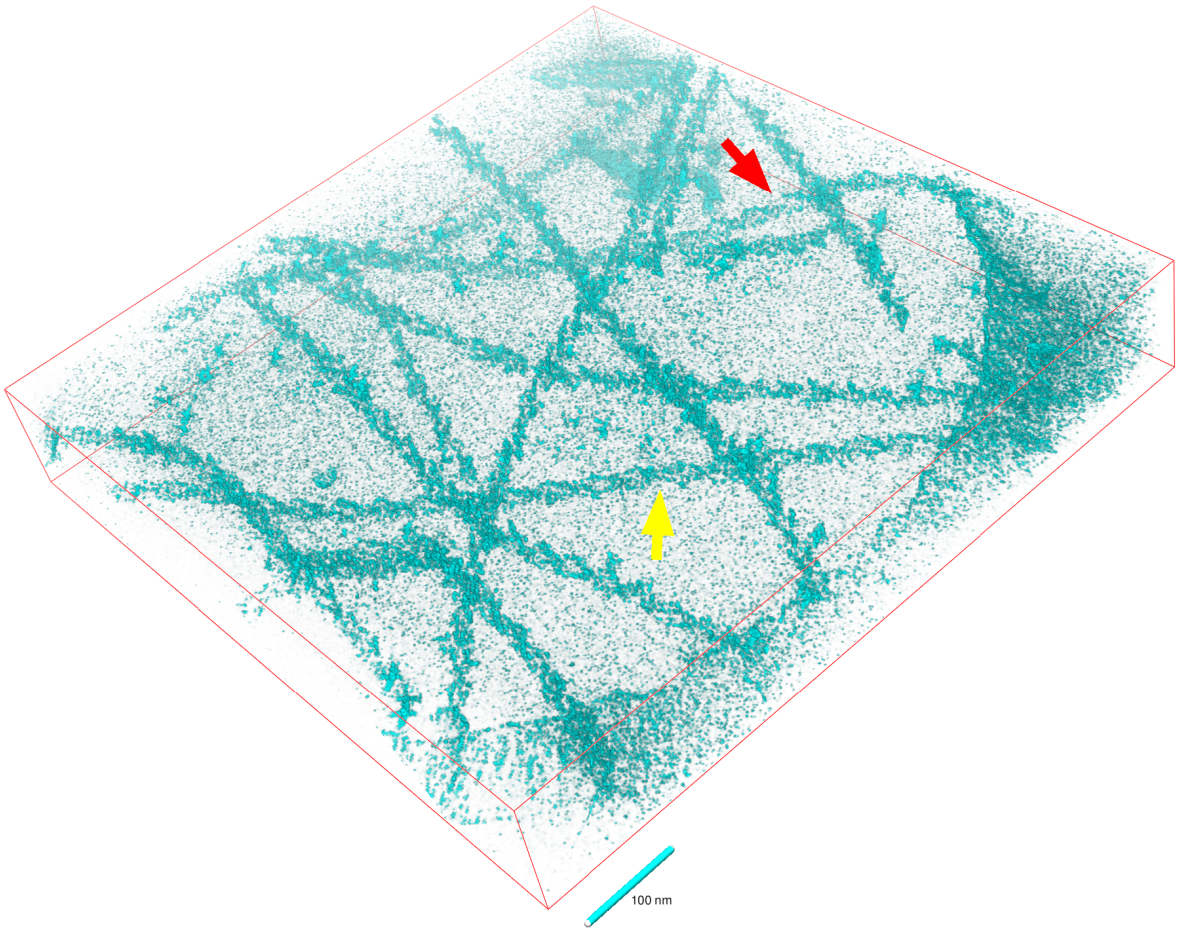


Figure 7.1: Three-dimensional representation of the noise-reduced tomogram of actin filaments, that are decorated with myosin V motor domains (Myosin V-1IQ fragment). The edge of a hole of the holey carbon film can be seen. The arrows mark a strongly bent (red) and a rather straight (yellow) region of a filament. The maximum density representation has been generated in Chimera (Pettersen et al., 2004). The x-,y-,z-dimensions of the tomogram are 2048 x 2048 x 330 pixels with the pixel size of 4.2 Å resulting in about 860 x 860 x 140 nm.

of the tomogram. A thicker ice layer in the other tomograms may have promoted to visualize the filaments only with less well resolved myosin repeats. Also the decoration might be less homogeneous. The high amount of filaments within the shown tomogram led to many filament contacts. The filaments are attached to the upper surface of the holey carbon film and arrange themselves rather flexible with a lot of filament-filament contacts. Some filaments appear rather straight and only slightly bent, others are strongly bent around each other. Inside the volume of the 4 µl droplet, applied to the grid before vitrification, certainly they do not have to have such a plane orientation. Upon removing most of the buffer volume during blotting their network seems to have been compressed and some filaments have been wrapped around others. A closer look at the yellow labelled filament is given by Figure 7.2. A series of individual images has been extracted form the full volume with a slice separation of 10 slices in z-direction. The repetitive arrowhead pattern of the diagonal filament is well resolved. Going from slice to slice first one half of the arrowheads appears, then the appearance is almost like the projection view of the filaments and moving further the other half of the arrowhead pattern shows up. The quality of the second filament is not the same. A repetitive pattern can hardly be distinguished. It shows that although all filaments in the field of view are decorated with the myosin motor domains its quality may very well vary. Strain and friction due to filament contacts may lead to flexible adaptations that result in less symmetrical arrangements along the filaments.

The red labelled filament has been used to estimate its out-of-plane tilt. It is clamped both by adhesion to the carbon film and by several other filaments. Its coordinates have been extracted from the tomogram in order to determine its maximum slope by using geometric mathematics and Gaussian error propagation. A tilt of $\theta = 18° \pm 3°$ was estimated. It may correspond to the maximal occurring out-of-plane tilt. Lower density of filaments will certainly achieve two improvements: the number of filament cross-overs will decrease and the ice thickness has shown to decrease as well. Both improvements will lead to smaller tilts out of the projection plane. Inside the tomogram the strong bending can clearly be seen. It is important whether it can be detected also in the projection images. Figure 7.3 shows the untilted projection image from the acquired tilt-series. The arrows indicate the same filaments as before. Both filaments obviously look alike. Hence, a tilt out of the projection plane accompanied by bendings and distortions cannot be seen within the projections. It is impossible to distinguish between slightly and strongly bent filaments. The shortening of the helical repeat owing to the out-of-plane tilt alone is not sufficient to detect them within the projection images. Furthermore, the helical periodicity of the single filaments in the tomogram has been analyzed. The analysis displays inhomogeneities in the estimated arrowhead length. 25 complete arrowheads from rather straight filaments could be included into the analysis. The corresponding coordinates have been determined in the volume. The resulting length distribution is quite broad (cf. Bierbaum, 2012). Although not many

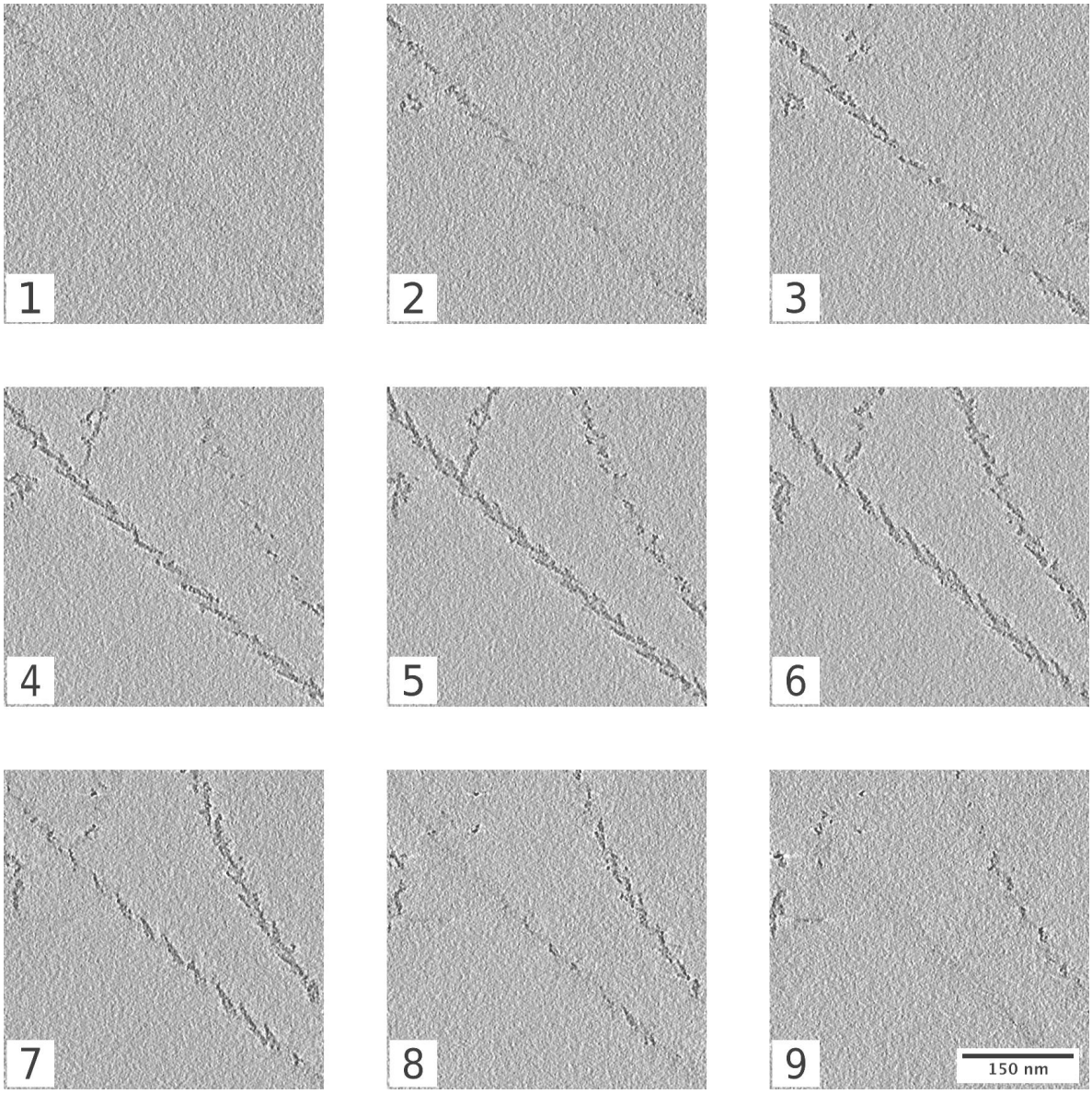


Figure 7.2: Slices through part of the tomogram. The images show every tenth slice of the volume, hence the separation between each of them is 4.2 nm. The diagonal filament lies almost straight within the ice layer. Its repetitive helical structure is nicely resolved. In contrast the slightly bent filament on the right does not show the same quality.

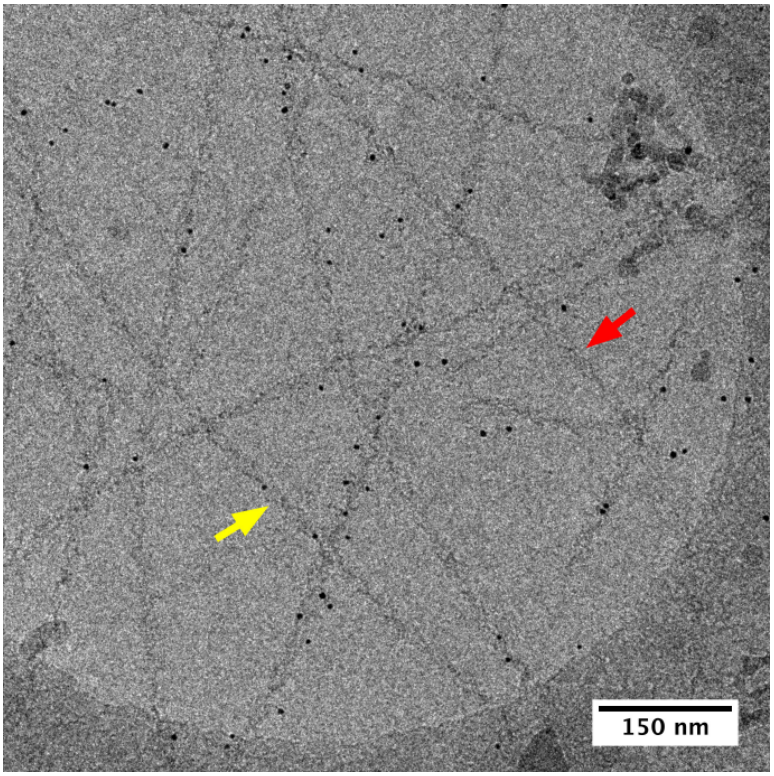


Figure 7.3: Zero degree tilt image from the tilt-series. The arrows label the same filaments as in Fig. 7.1: the red arrow the strongly bent filament and the yellow arrow the rather straight filament. The small black dots are the fiducial markers. They show a high affinity to be close to the filaments and are not distributed completely randomly.

arrowheads could be measured and thus the distribution has no high significance, a tendency can be seen. Filaments with disturbances no longer seem to exhibit a clear symmetry. The disturbances are followed by some kind of relaxation to reduce the intrinsic strain. Thereby, obviously also the length of the arrowheads seems to be affected.

7.3 Discussion

Electron tomography has been applied successfully to the actomyosin complex under cryo conditions. Thereby, the arrowhead structure of the myosin decorated actin filaments could be visualized. The sample preparation once again has been a bottleneck. It would have been nice to see the filament properties both for a high and a low filament density. A low density could not be prepared for this purpose, since the decoration could not be maintained in those preparations. Nevertheless, the impact of a high filament density could be analyzed. The filaments showed to be rather flexible and thereby

produced large out-of-plane tilts. Further analysis revealed that filaments behave like polymer bands rather than rigid filaments with fixed crystallographic symmetry. Filament-filament contacts together with adhesion to the support carbon film lead to a flexible adjustment of the filaments. The high filament concentration promotes a thicker ice layer and thereby allows for higher tilt angles out of the projection plane. The measured angle of $\theta = 18° \pm 3°$ may represent the upper limit of out-of-plane tilts. It corresponds to a shortening within the projection image by about 5 %. For the other filaments without strong external influences (hindrance by other filaments) no such high tilts and projection shortenings occurred. The broad distribution of the arrowhead length, measured in rather straight filaments, indicates that the filament cross-overs can have a severe impact on the helical symmetry. The disturbances lead to quite flexible adjustments, probably to reduce intrinsic strain. Instead of helical order the filaments then can feature structural disorder over long ranges.

Our findings imply consequences for the single particle reconstruction. The best way to deal with it, is to prepare better samples to exclude all disordering influences from the data set to be reconstructed. Filaments without cross-overs and not more than one adhesion contact to the carbon support film in principle should exhibit their intrinsic symmetry. Especially for the use of the least squares method, that treats filaments as axial tomograms, the sample preparation is critical. Certainly the method can deal with an ideal data set very well. If unfortunately no such sample with good decoration at the same time could be prepared, one has to be careful in dealing with filaments that show cross-overs. Most likely a filament cross-over is accompanied by a filament tilt out of the projection plane. The associated projection shortening probably cannot be detected due to the low signal-to-noise ratio within the images that hardly allows to see arrowheads. All those filaments cannot be used in a productive manner within an axial tomography approach. So far the filaments have been used till a cross-over occurs. Filaments that exhibit bends within the x-y plane have been excluded from the filament selection. However, the results show that a bending cannot be distinguished if its direction does not lie predominantly within the projection plane. Hence, the flexible adjustments and structural disorders that follow a bend in z-direction are most likely included into the reconstruction. Thereby, the reconstruction represents indeed only the average symmetry. In case the symmetry deviations are symmetrically distributed around the intrinsic symmetry it might be all right. If they can shift the mean with respect to the intrinsic symmetry, the result might be misleading. The disturbed length along the filament starting with the disturbance could not be estimated by this analysis. If it is rather short resulting in a tilted but otherwise straight filament it might be accurate to use such filaments in the context of a reconstruction method that can deal with these tilts.

Chapter 8

Iterative helical reconstruction as first step to reduce ad hoc assumptions

The iterative helical real space reconstruction (IHRSR) method after Egelman (2000) has been applied to reduce the ad hoc assumptions that are still necessary for the least squares reconstruction method. It suites in a better way the flexible properties of the actomyosin complex as the symmetry angle and the helical rise are free parameters which are subject to an iterative determination. Both parameters have to be assumed in advance for the current implementation of the least squares method. The pixel size does not need to have a specific value for the this method and out-of-plane tilts can be treated also. The images are aligned in 2D for this method and a weighted backprojection is used to obtain the 3D reconstruction of the whole data set. Because a better density localization of the least squares reconstruction could not be obtained for this data, a normal backprojection that needs less assumptions, is the best way so far to deal with the flexible properties of the data as long as a new implementation of the least squares method is not available. The sample preparation and data acquisition is illustrated again because of the evidence that the former preparation might have led to a kinetically impure mixture of structural states. The old data* set is also discussed in the context of this method in order to assess the findings of the least squares method. Complete data sets† have been acquired of both the rigor state and the strong binding ADP state. In comparison of these states the former data set can be included also to get additional information about the biochemical state of the sample. This chapter here points also to different approaches to correct the unwanted effects of the contrast transfer function (CTF) and reports on much improved results compared to chapter 6 that finally allow for a biological interpretation combined with a further processing using molecular dynamics (cf. chapters 9 and 11).

*"old" refers to the data acquired at the JEOL microscope in Japan

†here is referred to the new data acquired at the FEI microscope in Heidelberg

8.1 Material and methods

8.1.1 Sample preparation

Discussions with Prof. Dr. H. Lee Sweeney (University of Pennsylvania, Philadelphia, USA) led to the evidence that possibly the former buffer conditions did not suit the myosin in the intended way. Both the temperature prior to vitrification and the $MgCl_2$ concentration might have favored an ADP bound state instead of the rigor state. Therefore, a new preparation protocol has been designed together to obtain a better definition of the structural state that is captured within the reconstructed density. Myosin V has again been provided by Prof. Dr. H. Lee Sweeney in April 2013, this time in two different preparations. For the rigor state the myosin was treated already with apyrase to remove ADP and it was concentrated to 91 µM in an aqueous buffer of 4 mM EDTA, 10 mM KCl and 10 mM imidazole with pH value of 7.0, containing no magnesium ions or any nucleotide. The former used Tris was replaced by imidazole, an aromatic heterocycle with the chemical formula $(CH)_2N(NH)CH$. Ethylene-diamine-tetraacetic acid (EDTA, chemical formula $C_{10}H_{16}N_2O_8$) has the ability to sequester residual magnesium ions, that are present in the final solution due to the actin preparation, avoiding a non-intended magnesium influence on the myosin kinetics. The actin was freshly polymerized directly before usage. G-actin (of 58.5 µM concentration contained in a buffer of 5 mM Hepes, 2 mM $MgCl_2$, 1 mM NaN_3 and 0.2 mM ATP) was polymerized at room temperature for four hours using 5 mM Hepes (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer at pH value 7.5 and a second high-salt-concentration buffer containing 2 mM $MgCl_2$, 100 mM KCl and 10 mM imidazole at pH value 7.0. The resulting f-actin solution with a concentration of 5 µM and 50 mM KCl was further diluted with 4 mM EDTA and 5 mM imidazole at pH value 7.0 to a concentration of 0.25 µM f-actin. It was sonicated for 5 min to shorten the filaments and phalloidin was used to stabilize them. The myosin V was diluted to a concentration of about 5 µM using a buffer of 4 mM EDTA, 10 mM KCl, 5 mM imidazole at pH value 7.0.

New Quantifoil® holey carbon molybdenum grids of type R0.6/1 with mesh size 200 were used. They were washed in organic solvents (chloroform, toluol and afterwards in water) and glow discharged using the advanced plasma cleaning system Solarus Model 950 from GATAN with a low power program of only 5 watt for 3 s in a low pressure nitrogen-oxygen atmosphere. To ensure the plasma ignition under low power conditions a higher watt program had to be run beforehand without the grids. The grid preparation was achieved using an on-grid-decoration procedure. 2 µl of actin solution were applied onto a lying grid clamped by a tweezer, after two minutes followed by 2 µl myosin solution. The tweezer has carefully been mounted onto the Vitrobot Mark IV from FEI. The vitrification was done

two minutes after adding the myosin following the automated procedure of the Vitrobot. Vitrifying parameters have been chosen to room temperature (about 22°C), 100 % relative humidity within the chamber, 3 s blotting time with Whatman paper No. 1 and blotting force parameter −25. The final concentrations for the rigor preparation are 0.125 µM f-actin, 2.5 µM myosin V, 10 mM KCl, 0.43 µM residual ATP from the actin stock solution and 0.042 mM $MgCl_2$ also from the actin stock solution. The free magnesium ions should be chelated by EDTA. The actin to myosin ratio is 1 : 20.

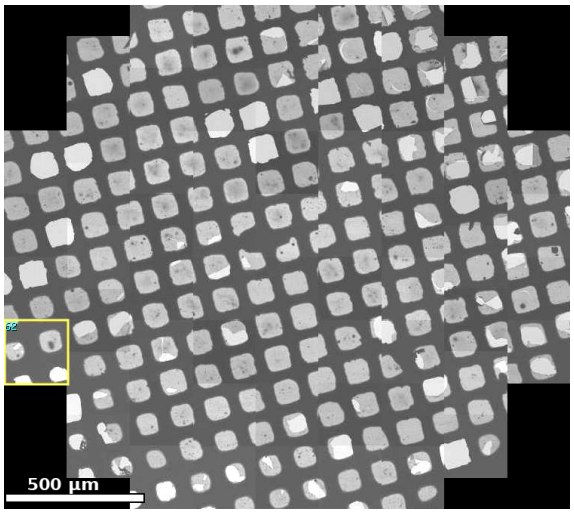
For the strong ADP binding state a 106 µM concentrated myosin V preparation has been provided contained in a buffer system with suitable magnesium and ADP concentrations and less KCl (12 mM $MgCl_2$, 50 µM MgADP, 5 mM KCl and 5 mM imidazole at pH value 7.0). The actin was polymerized also in a buffer containing magnesium and ADP (using both a 5 mM Hepes buffer at pH value 7.5 and a buffer system of 12 mM $MgCl_2$, 100 mM KCl and 5 mM imidazole at pH value 7.0) for about four hours with a final f-actin concentration of 5 µM and 50 mM KCl. The filaments have been stabilized by adding phalloidin. Afterwards the f-actin was diluted with a buffer system of 12 mM $MgCl_2$, 5 mM imidazole at pH value 7.0 and 50 µM ADP to obtain a final actin concentration of 0.25 µM. Again the f-actin was sonicated to shorten the filament lengths. The myosin was diluted in the buffer system of 12 mM $MgCl_2$, 5 mM KCl, 5 mM imidazole at pH value 7.0 and 50 µM ADP to a final myosin concentration of 3 µM. The grid preparation and vitrification was the same as for the rigor state. The final concentrations for the ADP state are 0.125 µM f-actin, 1.5 µM myosin V, 5 mM KCl, 0.43 µM ATP from the actin stock solution, 48.75 µM MgADP and 11.85 mM $MgCl_2$. The actin to myosin ratio is thus 1 : 12.

8.1.2 Semi-automated data acquisition

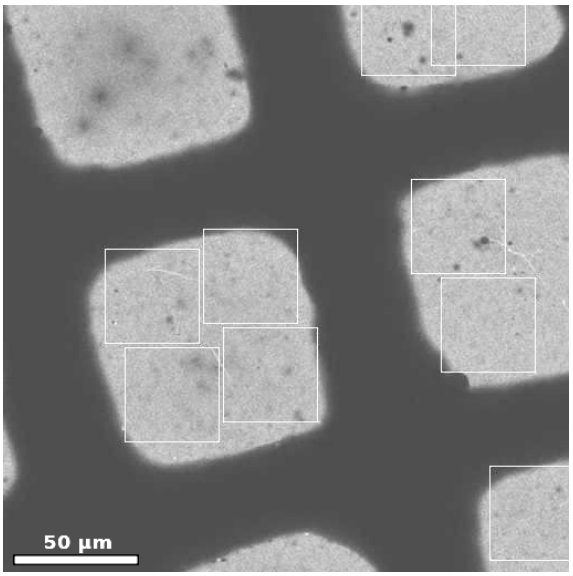
The new data sets have been acquired using the same microscope as for the tomographic tilt-series: a FEI Titan Krios electron microscope with a GATAN GIF Quantum post-column energy filter and a GATAN UltraScan 1000 CCD camera with 2048 x 2048 pixels. The microscope offers the possibility to load twelve grids at the same time into the so-called autoloader and store them at liquid nitrogen temperature for a long time. The related dewar vessel is refilled automatically. It allows for an automated data acquisition. Following parameters have been used: an electron acceleration voltage of 200 kV, an energy filter slit width of 10 eV, an electron dose of approx. 15 $e^-/Å^2$, a 100 µm sized objective aperture and a 50 µm sized condenser aperture. The emission current of the field emission gun of the TEM has not been stable over time. Therefore, the spot size and the exposure time have been readjusted several times. Each time the exposure was adjusted to have an electron dose of approx. 15 $e^-/Å^2$, measured with the specimen removed from the field of view. In the beginning a spot size of

7 was used, later on spot size 6 which is more intense. The exposure time was varied between 600 ms to 1200 ms. A quick tune of the energy filter using the Digital Micrograph software was done before each new session and a zero-loss peak alignment also in between the sessions, when a shadow of the filter slit showed up in the images. In the beginning the energy filter has been quite unstable and many readjustments were needed. In the meantime the energy slit width was enlarged to 15 eV due to the instabilities. For most of the images a width of 10 eV could be used, since later on with an optimized alignment the energy filter has been more stable. A nominal magnification of 105000x was used that led to an image pixel size of about 1.36 Å/pixel. In order to restrict the illumination to a small area to be able to use as much neighboring locations on the grid as possible the so-called nano-probe mode was used. It allows to illuminate a small area while maintaining approximately parallel illumination. In the beginning a beam diameter of 900 nm in the image plane was used. As the outer beam boundary moved too often into the field of view, the illumination area was enlarged to 1100 nm.

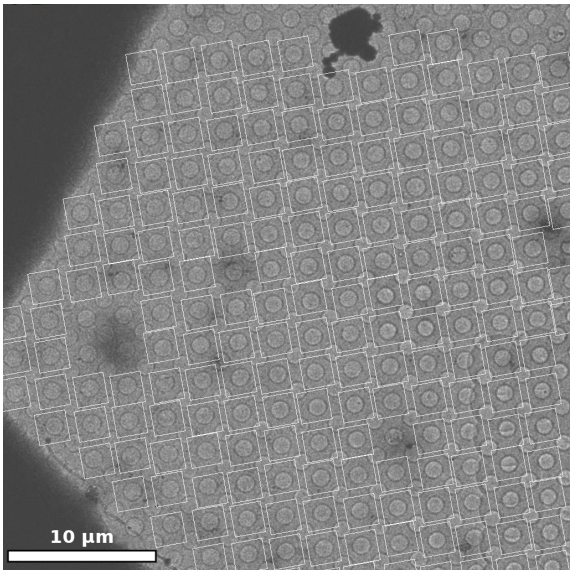
The Leginon software (Carragher et al., 2000) was used for automated data acquisition. It is a system designed for automated collection of images from a transmission electron microscope. The automation was reduced into a semi-automated procedure in order to allow for a manual choice, where to acquire the images. This step led to a higher amount of images that show filaments in the field of view. Again strongly defocused images at lower magnifications were used to locate the filaments inside the ice layer. The workflow of the multi-scale imaging using Leginon is outlined in the following and in addition illustrated in Fig. 8.1 and 8.2. First almost the whole grid is imaged using a low magnification of 135x to obtain a so-called grid map. A single grid image is shown in Fig. 8.1 (b). The grid map in Fig. 8.1 (a) is built up by 62 such images, indicated by the yellow frame. The overview allows to choose squares, defined by the grid bars, that show an intact holey carbon film with an ice layer of moderate thickness. The square images were taken at a magnification of 740x. Four adjacent images could be acquired within one square. Fig. 8.1 (c) shows an image that is outlined by the white frames in (b). At the beginning a magnification of 360x was used for the square images. Then only one image fits into a square. As the z-height of the grid is mechanically adjusted once per square it turned out, that images of 740x magnification have the advantage to allow for four different height adjustments, that were useful, because some grids turned out to have a quite uneven carbon film. Leginon was configured to allow to queue its targets. Then each acquired image waits for a manual interaction. The square images were queued for the so-called hole targeting. Inside each square image all the holes have been selected manually where an acquisition with higher magnification was desired (indicated by the white frames in (c)). Automated procedures have not been used at this step. Fig. 8.1 (d) shows a hole image acquired with 15000x magnification. A focus target has to be set per square image. Then



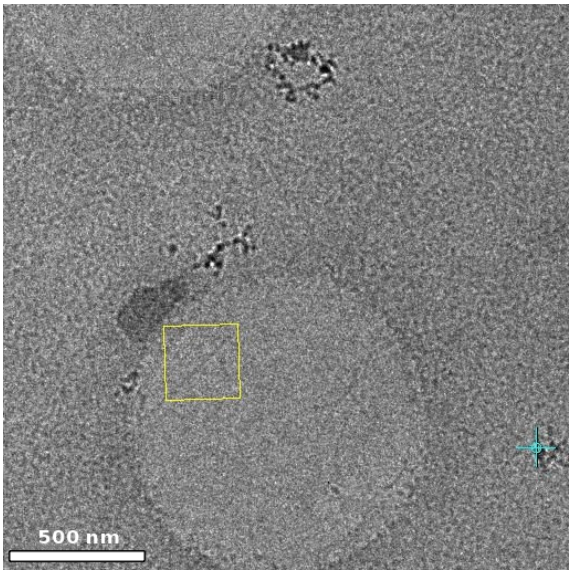
(a) grid map



(b) grid image

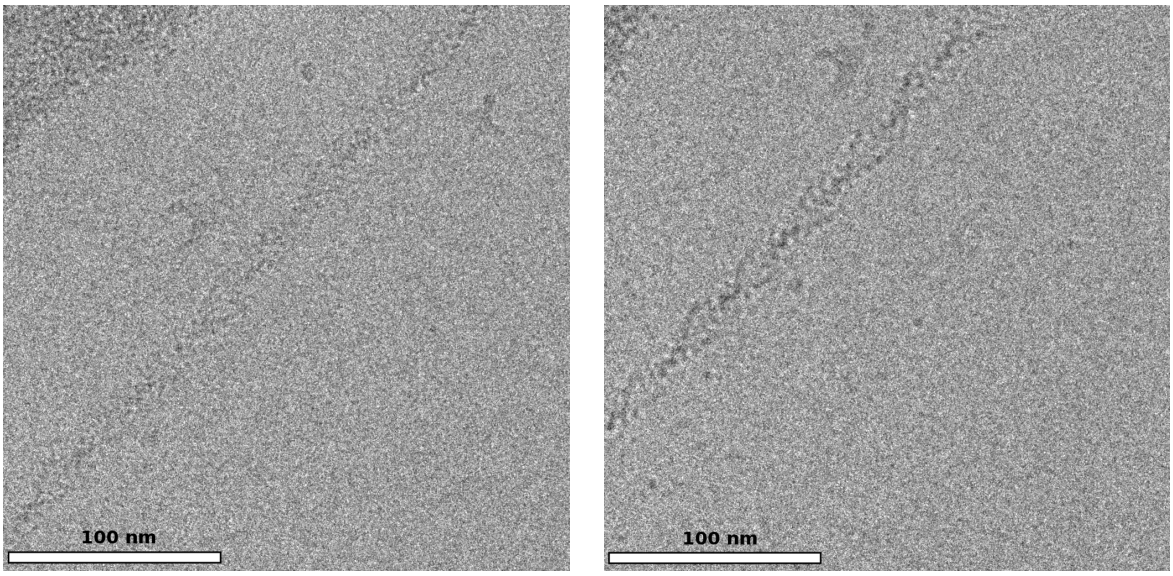


(c) square image



(d) hole image

Figure 8.1: Leginon multi-scale imaging. Each subsequent image is acquired with a higher magnification and its size is indicated in the previous image by a white or yellow frame. The grid map is composed of 62 grid images and shows the metal bars of the grid forming regular squares. The carbon support film in some squares is broken. The grey level indicates the ice thickness. A huge set of hole images is acquired and manually examined to find the filaments inside the ice layer. Both a focus target (cyan cross) and an exposure target (yellow frame) are set per filament.



(a) 1st exposure, 1.7 µm defocus

(b) 2nd exposure, 5 µm defocus

Figure 8.2: The final images were acquired with a magnification of 105000x and an electron dose of 15 $e^-/Å^2$. The defocus was randomly chosen by Leginon in the range of 1.0–2.5 µm. A second image at higher defocus was acquired in order to visually check the quality of the decoration. In (b) the arrowheads can clearly be seen. Only the image in (a) is going to be used for the processing and 3D reconstruction.

first the eucentric height is adjusted mechanically, followed by the acquisition of all hole targets. The hole settings have to be configured such that Leginon uses one ancestor (the z-focus in this case), otherwise the height adjustment of the grid will be ignored for the hole acquisitions. The mechanical focussing was done at 15000x magnification and was inspected manually, because sometimes upon a huge height difference of subsequently chosen squares, the automatic focussing had failed. If this is not checked and corrected, all hole images are useless as they are acquired not in eucentric height. Although filaments might be seen in the hole images the Leginon presets of the different magnifications are not going to match each other using a wrong height, so that the field of view will contain most probably a different area of the specimen in the higher magnification images. If the height is completely off, all exposure images will have an useless strong defocussing. Each hole image is acquired with a very low electron dose of about 0.03 $e^-/Å^2$. Spot size 10 and a strong defocus of 100 µm were used combined with four-fold binning to obtain an acceptable signal under low-dose conditions. The strong defocus allows the filaments to appear as thin black lines within the hole images. Leginon moves from one targeted hole position to the next without going back to a lower magnification. The coordinates of the specimen stage for the hole image are calculated from the marked position within the square image. If Leginon's calibrations are quite accurate, the chosen position will be found blindly. Leginon

has been configured such that all hole images are waiting for the manual interaction for the exposure targeting. As illustrated in Fig. 8.1 (d) filaments, that are detected within the hole images, are marked by an exposure target and a nearby focus target. The exposure targets are queued as well and then are acquired one after the other. Leginon can calculate the shortest path between them and then goes blindly back to the stage coordinates of each subsequent target using the hole preset. Using Leginon's navigator under low-dose conditions at 15000x magnification it has been checked via cross-correlation with the former hole image if the desired position has been reached. In this manner specimen drift can be accounted for, that might have occurred during refilling of the nitrogen dewar vessel when Leginon is processing a long queue of many hundreds of exposure positions. If a position cannot be reached in several attempts the target is going to be skipped.

Several configurations have been tested to reach the target position quite accurate. Because the field of view of the exposure position is only about 270 nm x 270 nm the stage has to be quite accurate to image the intended filament. First the navigator with a lower requested accuracy has been used combined with a final image shift. Then the stage has only to come close to the position and the rest is done by image shifting. It has the advantage that less images and cross-correlations of the navigator are needed, that speeds up the whole process and needs less electron dose. Unfortunately, the final image shift is not going to be used for the strongly defocused second exposure, that is quite useful to assess the quality of the filament by visual inspection. Approaching the target position only by image shift, without any cross-correlation checks if it is really reached, is the fastest method and at the same time requires the lowest additional electron dose. In short Leginon runs it proved to work quite well, but during long Leginon runs of many hours it cannot account for any occurring drift and becomes less accurate in approaching the intended positions. Therefore, finally the specimen stage alone has been used to mechanically approach the position. In the exposure settings an accuracy of 100 nm has been requested. Once a position is reached Leginon changes to the focus preset that was configured to use the same magnification as the exposure (105000x) and a slightly smaller illumination area of 900 nm in order not to pre-irradiate the exposure position. The focussing is done at the previously targeted focus position nearby the exposure position. The defocus of the final image of the filament (cf. Fig. 8.2 (a)) is chosen randomly out of a defined defocus range between 1.0 µm to 2.5 µm. A second image was acquired with 5 µm defocus (cf. Fig. 8.2 (b)). The stronger defocus reveals details of the arrowheads more clearly. Only the first images have been processed further. Leginon is combined with a htlm-based observer interface that allows to check from anywhere what the microscope is doing. Combined with a Virtual Network Computing (VNC) software it allows to interact with Leginon even without being present at the microscope. Appion (Lander et al., 2009) is a pipeline for processing and analysis of EM images and is integrated within Leginon. They use the same database.

A web-based image assessor from Appion was used to select all those images that exhibit filaments inside the field of view. Bad images were deselected. The second, stronger defocused images were used for this step to have a better estimate of the quality of the decoration. In total about 14500 exposure images have been acquired. All positions, where to acquire the images, have been set by hand. For the rigor state there are 8887 images out of which 3725 have been selected. 5618 images were acquired for the strong binding ADP state with 3602 images selected to be good. So far automated attempts to determine whether a good filament lies inside the image or not, proved by far not to be as good as visual inspection. A higher number of images compared to the data set as discussed with the least squares method was necessary due to the smaller detection area of the camera by a factor of four. The grids that were used for the data acquisition showed many separately distributed filaments and less filament cross-overs as before. Despite the sonication of the f-actin solution the filaments were rather long. Most filaments exhibit adhesion contacts to the carbon support film.

8.1.3 Image processing and 3D reconstruction

All selected images were extracted from the Leginon database. With the help of the program *e2proc2d* from the image processing suite EMAN2 (Tang et al., 2007) the images were converted into SPIDER format. The filament selection was done as before using JWeb from SPIDER, but this time the strongly defocused image was displayed in parallel using EMAN2 (*e2display*), to find out more clearly which filaments are worth to be selected. Each filament was marked using seven marker points. In total 4852 filaments have been manually selected for the rigor state and 4211 for the ADP state. The image processing prior to the reconstruction was adopted from that of the least squares method. The filaments were rotated into vertical direction due to the calculated angle determined by the marker points, the background was corrected, the contrast inverted and the images were binned twofold. No interpolation to a specific pixel size has been done. Filament segments were windowed this time every 30 pixels along the filament instead of every 10 pixels leading to an overlap of 90 %. The segment size has again been chosen to be 301 x 301 pixels. The segments have been checked for in-plane rotation and have been centered like before using only integer pixel shifts.

Because this methods needs aligned images in 2D in order to reconstruct only one volume from all aligned filament segments, the correction for the unwanted effects of the contrast transfer function (CTF) is done prior to the reconstruction within the images. For this purpose the CTF parameters have been determined for each image. The whole CCD images have been used to calculate the power spectra, not the small segment images, to have a better signal to noise ratio. The CTF parameters have been determined using both SPIDER, in a first instance, and the program CTFFIND3 (Mindell and

Grigorieff, 2003) later on. The CTF correction has been implemented in SPIDER. All CTF corrected filament segments have been combined to build one image stack per data set or two data halves stacks for the resolution estimation. The data halves stacks have been built using all segments of the odd or even numbered filaments to have all overlapping segments from one filament within one data half.

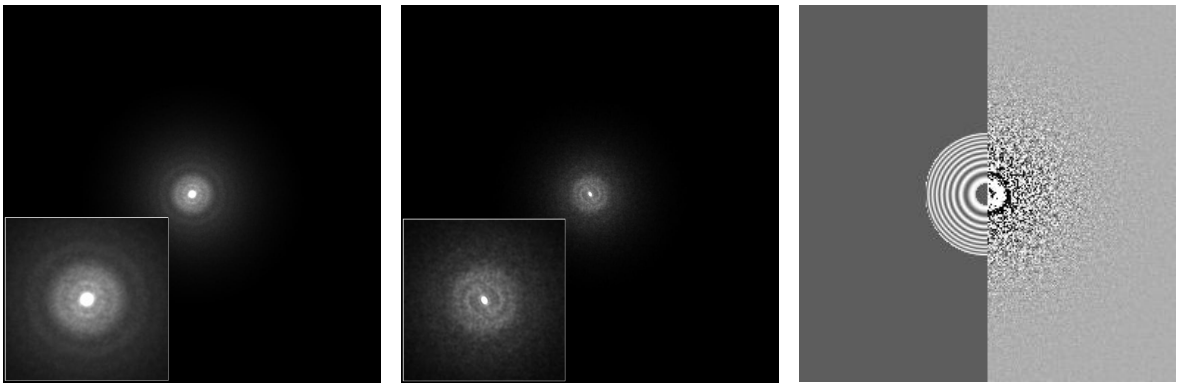
The 3D reconstruction has been accomplished in the beginning with the program IHRSR++. It is a user-friendly extension to the initial Iterative Helical Real Space Reconstruction (IHRSR) software. IHRSR++ is provided by the Timothy Baker Lab, University of California, San Diego and is a SPIDER program. However, the program uses old SPIDER conventions and an alignment operation that is no longer supported in current SPIDER versions, therefore it has been adapted to be usable on the computer cluster. The SPIDER alignment operation AP NQ had become obsolete and was replaced by the AP SH procedure. SPIDER version 21.00 was used for this reconstruction method. Most parts of the image processing have been done with the SPIDER version 18.15. The attempts to run the IHRSR++ program using a Message Passing Interface (MPI) for parallel computing have not been successful. Thus, later on the revision of the IHRSR algorithm called *sxihrsr* that is implemented in the SPARX single particle processing environment has been used (Hohn et al., 2007). It fully supports MPI usage. For this program the image stacks were converted using *e2proc2d* (EMAN2) from SPIDER format to the primary EMAN2 and SPARX format HDF5. It supports arbitrary metadata like header information for every image inside the stack. The SPARX program *sxheader* was used to prepare the header parameters prior to the reconstruction. The alignment parameters are stored as normal text file and can be imported into the header for a subsequent iteration. In this manner the image interpolations are kept to a minimal number. Filtering of the 3D reconstructions was done in EMAN2 using the program *e2proc3d*, while the b-factor estimation and correction was accomplished using the stand-alone package EM-BFACTOR (Rosenthal and Henderson, 2003, Fernández et al., 2008).

8.2 Results

8.2.1 Rigor state

Different approaches have been made to correct the unwanted CTF effects. The 3D reconstruction method requires no particular pixel size, but the CTF estimation and correction need to have this parameter. Instead of using simply the calibrated pixel size from Leginon it has been checked by calculating power spectra along some strongly defocused helices. The layer lines indicated a pixel size of rather 1.32 Å instead of the former calibrated value of 1.36 Å. The new value has been used in a subsequent CTF parameter determination using the SPIDER operation TF ED. In the first approach

the phases of each filament image have been flipped using the binned pixel size of 2.64 Å to have the correct phases in all images of different defocus values. It has been implemented using SPIDER's operation TF CT to calculate binary CTF Fourier images containing only the values $+1$ and -1 instead of the sinusoidal CTF function. An amplitude contrast ratio of 0.07 has been assumed. The microscope's spherical aberration parameter C_s is 2.7 mm. The additional filter contrast induced by the zero-loss filtering has been tried to be accounted for. It has been implemented following the descriptions of Angert et al. (2000) within the program from Yang et al. (2009) that is based on MATLAB from the MathWorks Inc. Unfortunately, the parameter estimation now with two additional parameters was not reliable. It might be due to the low signal-to-noise ratio and to the fact that not enough CTF zero positions are present in the data. An unsupervised parameter estimation within many thousands of images was inconceivable. The first results of the 3D reconstruction using the IHRSR method indicated a repeat height of 27.72 Å for the symmetry unit. This size, known to be 27.5 Å, has been used to recalibrate the pixel size to a new value of 1.31 Å or 2.62 Å for the twofold binned data. Further analysis showed also that many defocus values close to the focus have not been estimated correctly. Fig. 8.3 shows the very weak CTF imprint in the data. While the power spectrum from the old data



(a) JEOL data, 4k camera (b) FEI data, 2k camera (c) CTFFIND3 estimation

Figure 8.3: Power spectra calculated from the old data set (a) and from the new one (b). Both images have a defocus of about 1.6 µm and almost the same pixel size. They show a clear difference in the strength of the CTF imprint. The inserts show the enlarged center. In (c) the result of the parameter estimation of the program CTFFIND3 is shown. The left image half depicts the calculated CTF, the right half the power spectrum. The continuous grey areas inside and outside the half rings on the left are due to the defined frequency range for the parameter estimation.

set acquired at the JEOL microscope equipped with a 4k x 4k camera shows a lighter, second ring, that clearly defines the zero position in between, the power spectrum from the newer data set acquired at the FEI microscope equipped with a 2k x 2k camera shows no further rings and thereby makes the CTF

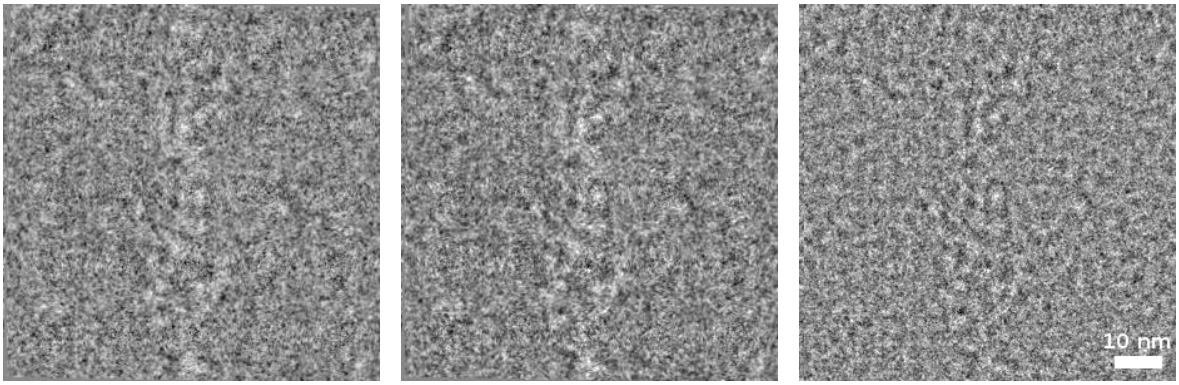
parameter estimation less reliable. The differently strong CTF imprints might be due to the quality of the cameras and their different detector sizes. Different envelope functions of the CTF due to the different microscopes and different ice thicknesses might also play a role.

In a second approach the recalibrated pixel size has been used to re-estimate the CTF parameters. CTFFIND3 proved to yield a better result than the SPIDER function TF ED. CTFFIND is also implemented in the new SPIDER versions, but in many cases a huge astigmatism is found which is not present in the data. Due to the low CTF imprint this parameter estimation has been unstable too. The original CTFFIND3 program offers an additional restriction as it allows to define the expected amount of astigmatism and thereby is much more reliable with really low signal-to-noise data. Images close to the focus had to be identified in a second approach also with CTFFIND3. Using a general parameter set with a wide defocus range from 300 nm to 6.5 µm they have not been estimated correctly with a defocus in the order of 10 nm. Only those have been subjected to a second estimation with a narrower search range. The general parameter set has been chosen to be 256 pixels for the box size, 45 Å minimal resolution, 8 Å maximal resolution, 300 nm minimal defocus, 6.5 µm maximal defocus, 500 nm defocus steps and 50 nm expected astigmatism. After some tests the second parameter set has been chosen to be 512 pixels for the box size, 30 Å minimal resolution, 8 Å maximal resolution, 300 nm minimal defocus, 3.0 µm maximal defocus, 100 nm defocus steps and 10 nm expected astigmatism. The newly estimated CTF parameters have been applied by phase flipping using SPIDER. The height of the symmetry unit converged now to 27.51 Å. The quality improvement of the reconstruction is discussed later by means of the Fourier Shell Correlation (FSC) that has been calculated for each approach using two independently reconstructed data halves (cf. Fig 8.14).

In a third CTF correction approach not only the phases have been corrected but also the amplitudes following the description of Sachse et al. (2007). The Fourier transform of the images $\tilde{I}(q)$ has been multiplied by the simulated contrast transfer function $H(q)$ without any dampening envelope functions: $\tilde{I}'(q) = \tilde{I}(q) \cdot H(q)$. The dampening functions can be omitted in SPIDER by setting the source size and defocus spread to zero and the gaussian envelope half-width in Fourier space to a high value, like 100 Å^{-1}. The multiplication of the images by the CTF has been implemented in SPIDER using the operation TF C followed by a multiplication in Fourier space. The resulting phase-corrected and amplitude-weighted segment images have again been stacked into an image stack. The corrected image stack is used for the alignment and 3D reconstruction by the IHRSR method. The multiplication by the CTF suppresses the noise around the CTF zero positions, that according to literature allows for a better alignment. The reconstructed volume V has then been subject to an amplitude

correction in 3D: for all filaments the squared 3D CTFs have been calculated using SPIDER's operation TF C3 (simulated 3D power spectra without envelopes) and have been summed. The final 3D reconstruction V is deconvoluted by this sum to yield the fully CTF corrected 3D reconstruction V': $\tilde{V}'(q) = \tilde{V}(q)/(\sum CTF^2_{3D,i} + f)$. $\tilde{V}(q)$ denotes the Fourier transform of the final volume. The deconvolution is represented in Fourier space by a simple division. The parameter f prevents noise amplification close to the CTF zero positions and was chosen to be 1% of the maximal value of the summed 3D CTF volume. This two-step correction approach results summed up in the Wiener filtering described in Eq. 4.10 on page 24.

Figure 8.4 shows the visual effect of the CTF correction on the filament images. The phase flipped



(a) no CTF correction (b) phases flipped (c) multiplied by CTF

Figure 8.4: Filament segment images without CTF correction, with phase flipping and after multiplication by the simulated contrast transfer function (CTF). The image size is 301 x 301 pixels with 2.62 Å/pixel.

image equals the uncorrected image visually quite well, whereas the image multiplied by the CTF, simulated according to the estimated CTF parameters using CTFFIND3, has a slightly different visual appearance. The IHRSR++ program allows to create an initial reference volume out of the image stack by random angular assignments. The resulting cylindrical volume of suitable diameter can be seen in Fig. 8.5. The scale can be estimated by the scale bar in Fig. 8.6. Fig 8.6 shows image slices perpendicular to the filament axis. It can be seen how the density converges in a few iterations from the initial volume to the final density distribution. The reconstruction was done using the *sxihrsr* program from SPARX in parallel execution. Following parameters have been used: 2.62 Å pixel size, 1° angular step to generate reference projections, initial helical parameters of 27.5 Å for the helical rise and $-167.1°$ for the helical angle, $\pm 15°$ in-plane rotation angle (how far the filament may deviate from vertical direction), a fraction of the volume used for helical search of 0.67 and a padding size of 2 for the 3D reconstruction. The first ten iterations have been done without accounting for out-of-plane tilt angles

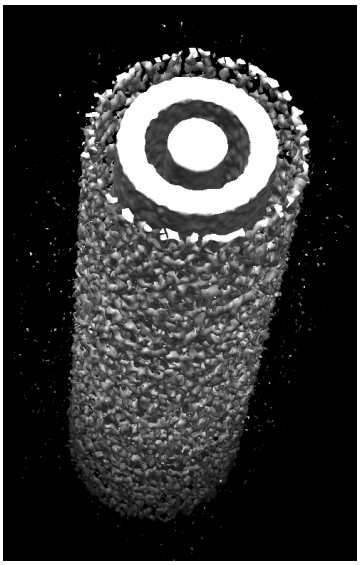
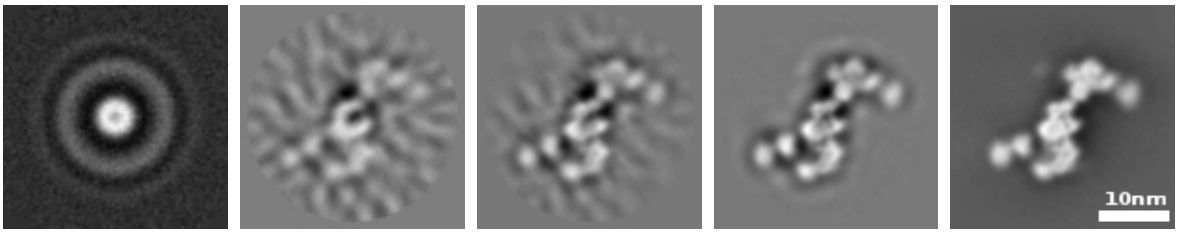


Figure 8.5: The initial volume of the iterative helical real space reconstruction (IHRSR) method can be created by random angular assignments to the individual filament segments. The program IHRSR++ offers this option. The surface rendering was done with the Chimera software (Pettersen et al., 2004). The scale can be estimated by the one from Fig. 8.6.



(a) iteration 1 (b) iteration 3 (c) iteration 6 (d) iteration 17 (e) 3D CTF corr.

Figure 8.6: Slices perpendicular to the filament axis from different IHRSR iterations. (a) shows the initial distribution, that belongs to the volume shown in Fig. 8.5. The first ten iterations were done without accounting for out-of-plane tilt angles. Four iterations followed with $\pm 18°$ out-of-plane tilts using an angular spacing of 3°. The last three iterations sampled the out-of-plane tilts with 1° spacing. (e) shows the 3D CTF corrected density from (d). Especially in (b)–(d) a large negative (black) density can be seen close to the central actin density.

using 360 reference projections. After the first two iterations the refinement of the helical symmetry was started. For the iterations 11 to 14 reference projections have been generated also for out-of-plane tilt angles from $-18°$ to $+18°$ using an angular spacing of 3°. In the last three iterations the angular sampling of the out-of-plane tilts was increased to 1°. The performance dropped considerably in speed due to the high number of reference projections that have to be considered in the projection-matching step. The final slices of almost one symmetry unit can be seen in Fig. 8.7. Using the whole rigor data set the symmetry parameters converged to the helical rise of 27.51 Å and the symmetry angle of $-167.405°$. The volume is amplitude- and phase-corrected, including the 3D amplitude correction after the reconstruction. A clear improvement can be seen compared to the least squares reconstruction approaches. To account for the structure factor decay in Fourier space the B-factor has been

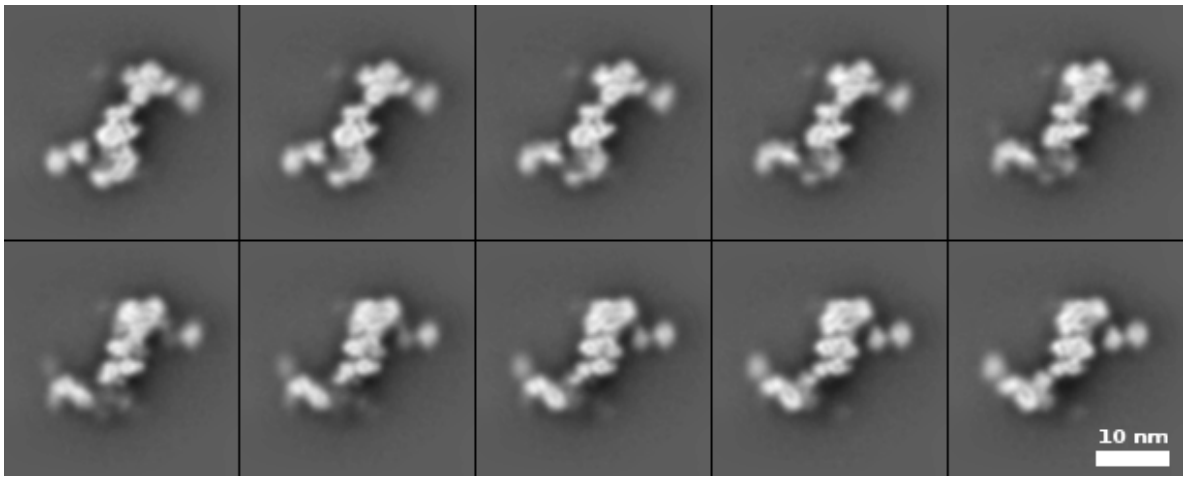


Figure 8.7: Slices of the final amplitude- and phase-corrected 3D reconstruction by the IHRSR method of the new rigor data set. A 3D amplitude correction was performed, but the B-factor, that describes the structure factor decay towards higher frequencies in Fourier space, has not been corrected so far. Likewise no other filtering has been done.

estimated using the program EM-BFACTOR. Figure 8.8 shows a so-called Guinier plot that illustrates the way how to compensate the amplitude decay towards higher spatial frequencies. The Guinier plot shows the natural logarithm of the spherically average structure factors F as a function of the square spatial frequencies d^{-2}, where d is the resolution in Ångström. The spherically averaged structure factor amplitudes are computed and named experimental amplitudes. To represent the signal better in the reconstructed density a noise-weighting function is applied in the form of $(2FSC/(1+FSC))^{\frac{1}{2}}$, where FSC denotes the Fourier Shell Correlation curve estimated using two data halves. In the Guinier representation a large frequency range is approximately linear. By least squares fitting of the function $y = a + b \cdot x$ to the noise-weighted amplitudes in the resolution range from 18 Å to 9 Å following parameters have been found: $a = -14.1$ and $b = -346.7$. Thus, the B-factor that affects the data is estimated to be 346.7 $Å^2$ according to the notation of Eq. 4.11 on page 24. The amplitudes are restored by applying the negative B-factor of -346.7 $Å^2$ till a maximum resolution of 9 Å (sharpened amplitudes in Fig. 8.8). The Wilson statistics, normally used in protein crystallography, can be used to set the density on an absolute scale as the average amplitude in the Wilson regime is described to equal the square root of the number of atoms N_{atoms} and the zero angle scattering (indicated in the plot by a red cross) equals $0.28 \cdot N_{atoms}$. The scaling has been omitted here, because the molecular weight of the helix in the field of view cannot be stated exactly. The final phases-, amplitudes- and B-factor corrected density of the whole rigor data set is shown as surface representation in Fig. 8.9. A clear separation of the actin monomers and a nice structuring of some secondary structure elements can be

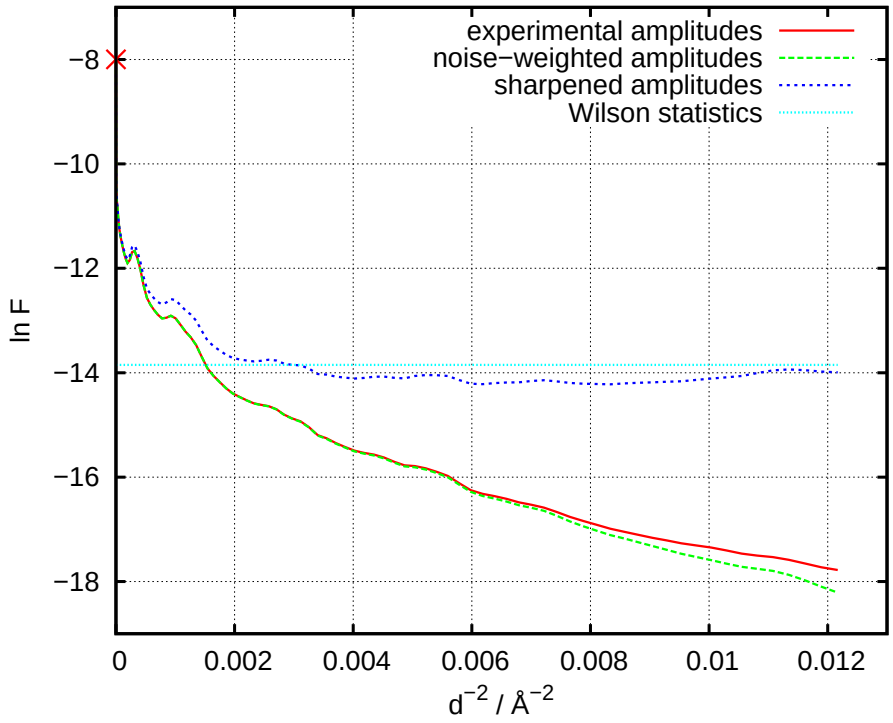


Figure 8.8: The Guinier plot shows the natural logarithm of the spherically average structure factors F as a function of the square spatial frequencies, where d is the resolution in Ångström. It illustrates the restoring of the amplitudes by applying a B-factor of $-346.7\,\text{Å}^2$ according to the formula e^{-Bq^2}. The process is also denoted as sharpening of the density. The Wilson statistics can be used to put the density on an absolute scale.

seen. The quality of the reconstruction is going to be investigated and estimated in the following. First the distribution of defocus values and out-of-plane tilts are analyzed in Fig. 8.10 and Fig. 8.11. Most of the estimated defocus values are within the range of 1.0–2.5 µm, that was set during data acquisition. Leginon used random values out of this range. Therefore, an equal distribution could be expected. But there are several influences that may have changed an equal distribution. The focussing might have been incorrect in some case, this could account for the quite low and high defocuses. However, a large effect might come from the selection to keep only images with good filaments. Since out of 8887 images only 3725 have been selected, the random distribution might have been changed into one that virtually would allow to fit a Gaussian distribution to it. The distribution of the out-of-plane tilts (cf. Fig. 8.11) also could be described by a Gaussian distribution. Its full width at half maximum is quite broad. Possible out-of-plane tilts of $\pm 18°$ have been considered in the reconstruction, thus 72° and 108° are the largest tilts reference projections have been calculated for. The high scores for both largest

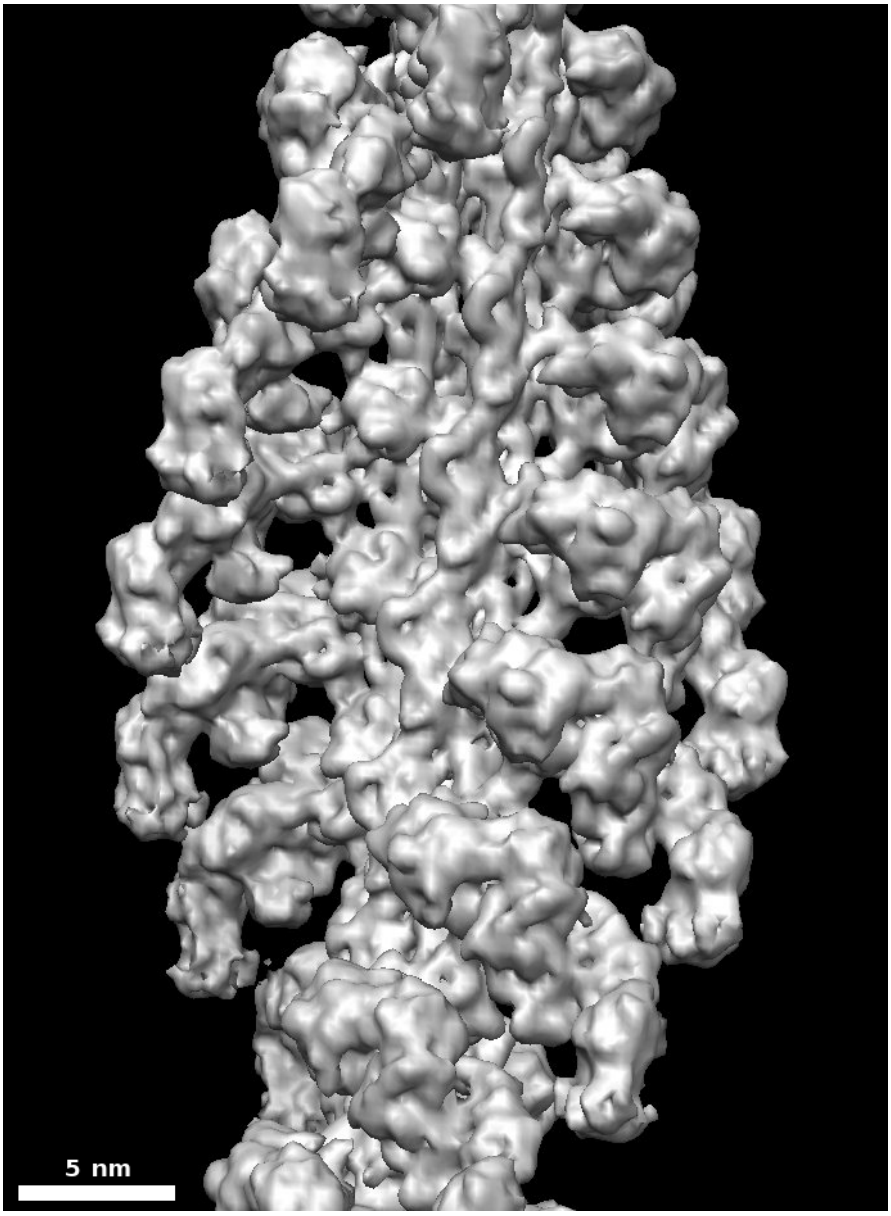


Figure 8.9: Surface representation of the final, fully CTF corrected reconstruction of the rigor data set. Using the IHRSR method the symmetry parameters converged to helical rise of 27.51 Å and a symmetry angle of $-167.405°$. The density was sharpened by a B-factor of $-346.7\,\text{Å}^2$. In total the density is reconstructed by using approx. 259000 symmetry units / single particles / actin and myosin monomers. The surface rendering was done in Chimera.

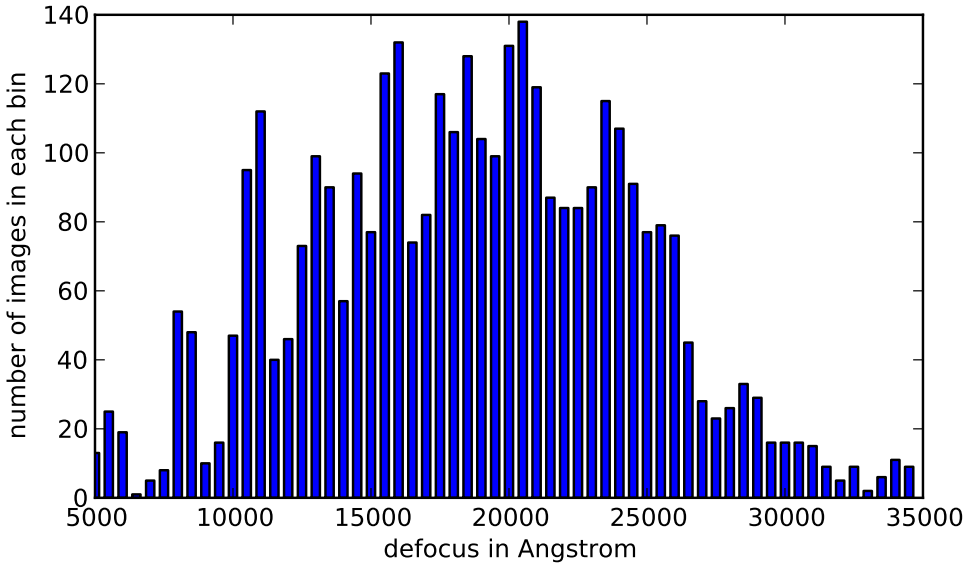


Figure 8.10: Distribution of estimated defocus values of the 3725 rigor images that have been acquired with random defocus assignments by Leginon. The defined defocus range was 10000–25000 Å. All values were applied as so-called under-focus.

values may indicate that 18° out-of-plane tilt, that has been estimated for strongly bending filaments using electron tomography, might not be the maximum occurring tilt. Thus, considering a broader distribution during the reconstruction might improve the result. In order to detect heterogeneities within the data set in a first instance six data subsets have been formed following intentionally the order of data acquisition, because it may reveal systematic deviations. Figure 8.12 shows bar plots of the resulting symmetry parameters for the six subsets that have been reconstructed independently of each other. The smaller helical rise of subset 2 and 3 is suspicious as the other four subsets converged to the very same value of 27.51 Å. If a systematic deviation of the magnification is the reason for that, it might have led during the IHRSR reconstruction to projection assignments of a higher out-of-plane tilt due to the shortening of the filament that cannot be distinguished from a projection shortening due to an out-of-plane tilt. It might have broadened the distribution discussed above. Subset 2 and 3 are also those subsets, that show a slightly smaller symmetry angle. It might come along with the different height. Except for that the parameters are quite stable within the subsets. Furthermore, the radial density distribution has been analyzed for the reconstruction of the whole data set. Figure 8.13 shows three different radial profiles. Two profiles have been calculated for the reconstructed density and one profile has been simulated using crystal structures that have been rigid-body docked into the density with the program Situs (Wriggers et al., 1999). The position of the structures and their atoms has been refined using molecular dynamics simulation combined with the extension of the method by

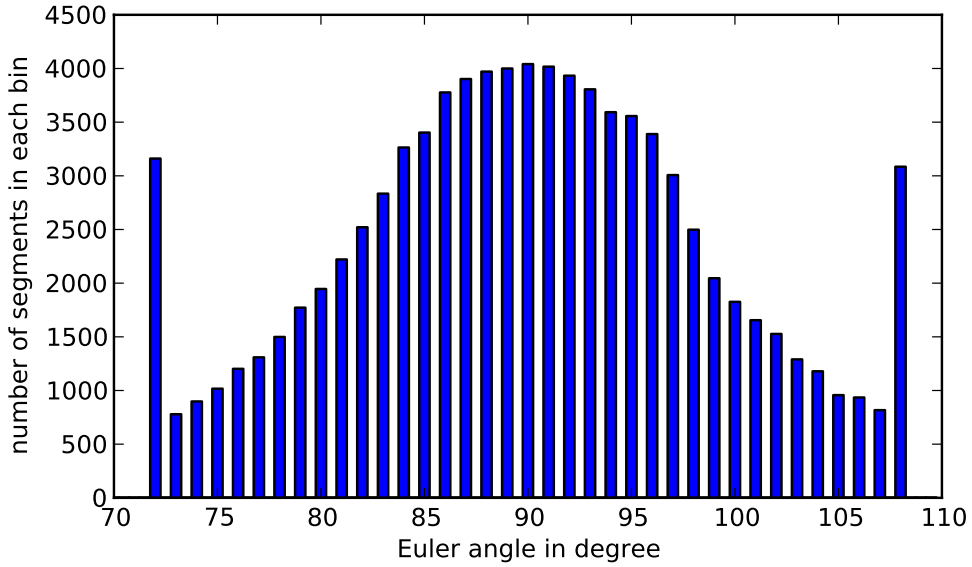
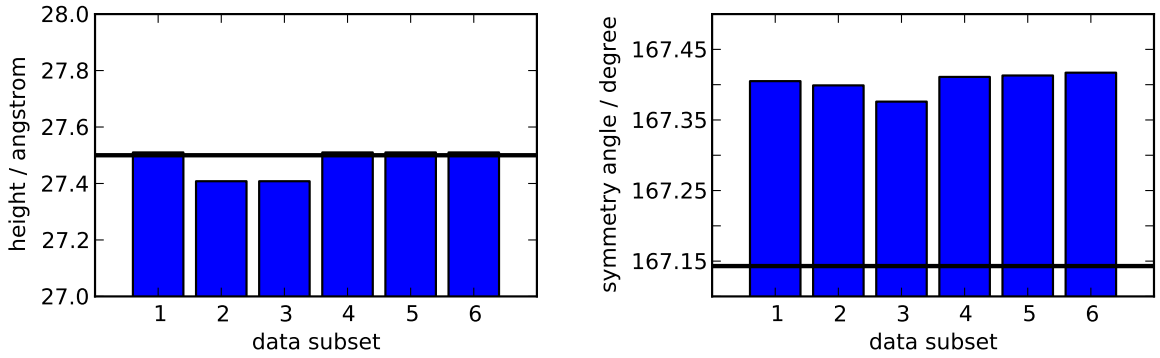


Figure 8.11: Distribution of estimated out-of-plane tilts considering all 90634 segment images, that have been used from the rigor data set. The Euler angle ϑ describes the tilt out of the projection plane. A value of 90° corresponds to segments that lie within the projection plane.

the molecular dynamics flexible fitting (MDFF) (Trabuco et al., 2009) (cf. Chapter 9). From the fitted crystal structures a density map of 4 Å has been calculated using the program VMD (Humphrey et al., 1996). This simulated density has been used to determine the ideal radial density distribution. The profiles have been scaled to have the same maximum value. It can be seen clearly that the minima are quite different. For the phase flipped volume the density becomes actually negative after the central actin peak. It is not much better for the fully corrected volume as it drops down to almost zero. The second peak of the profile belongs to the myosin main density, the third peak or shoulder corresponds to the light chain. The full CTF correction compared to the phase flipping corrects the position of the first peak, but also lets the other peaks fade out. Due to the strong negative density the positions of the second and third peak are shifted a bit towards higher radius. Both experimental profiles end up with negative density at high radius. The phase flipped volume offers for the myosin peak almost the same intensity as the simulated data. The light chain density is difficult to compare. A weaker intensity can result from negative density background derived from alignment errors that are at higher radius more severe or it might even be possible, that at some myosins no light chain has been present, since it is not covalently bound and might be lost during preparation in some cases.

Finally the resolution is going to be assessed by means of Fourier Shell correlations (FSC). For this purpose the whole rigor data set has been divided into two data halves, one containing only the even numbered filaments, the other the odd numbered ones. The segments from one filament thus have



(a) helical rise

(b) symmetry angle

Figure 8.12: Symmetry parameters for six data subsets, that have been reconstructed independently of each other. The data is chronologically divided into subsets to enable systematic errors to show up. The height of the symmetry unit converged to a mean value of (27.476 ± 0.048) Å and the symmetry angle to a mean value $(-167.403 \pm 0.014)°$. Lines indicate the symmetry parameters for a 28/13 crystallographic symmetry (27.5 Å and $-167.14°$). The symmetry angle is shown with positive sign.

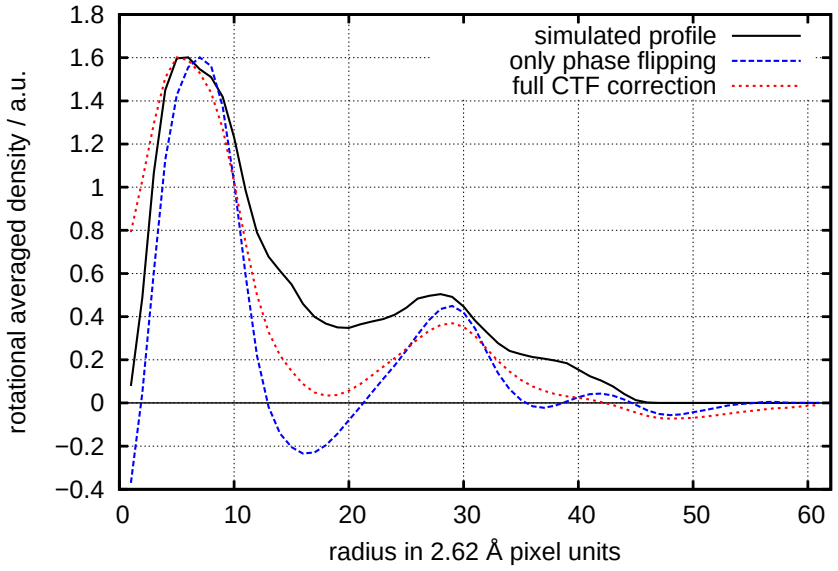


Figure 8.13: Radial density distribution of the reconstructed rigor density. The helices have been projected down their filament axis and the resulting image has been rotationally averaged to yield a one-dimensional profile. Profiles are shown both for the phase flipped reconstruction and for the fully CTF corrected density. For assessment of the quality a profile has been simulated.

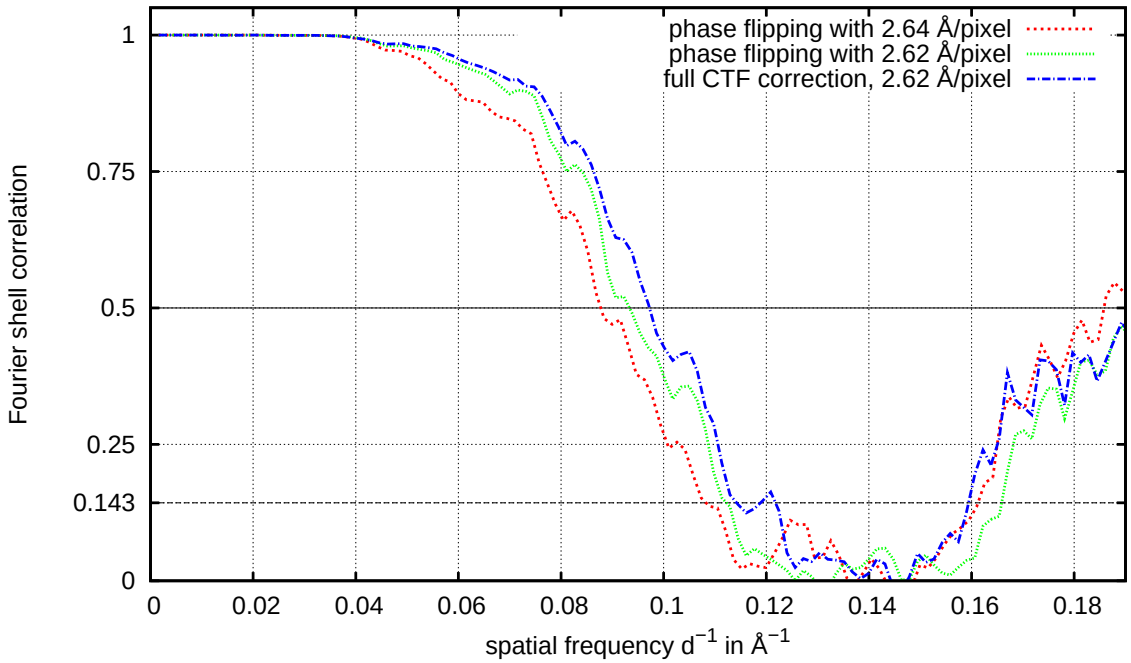


Figure 8.14: Fourier Shell Correlation curves of the odd and even data halves of the rigor data set. The different resolution criteria are indicated by the black lines at FSC(0.5) and FSC(0.143). The resolution could be improved by each subsequent approach to correct the data set for the effects of the contrast transfer function (CTF). The blue curve shows a final resolution of 10.2 Å / 8.2 Å (FSC(0.5) / FSC(0.143)). The artificial rising of the FSC curve at higher frequencies reflects effects of numerical embedding and masking.

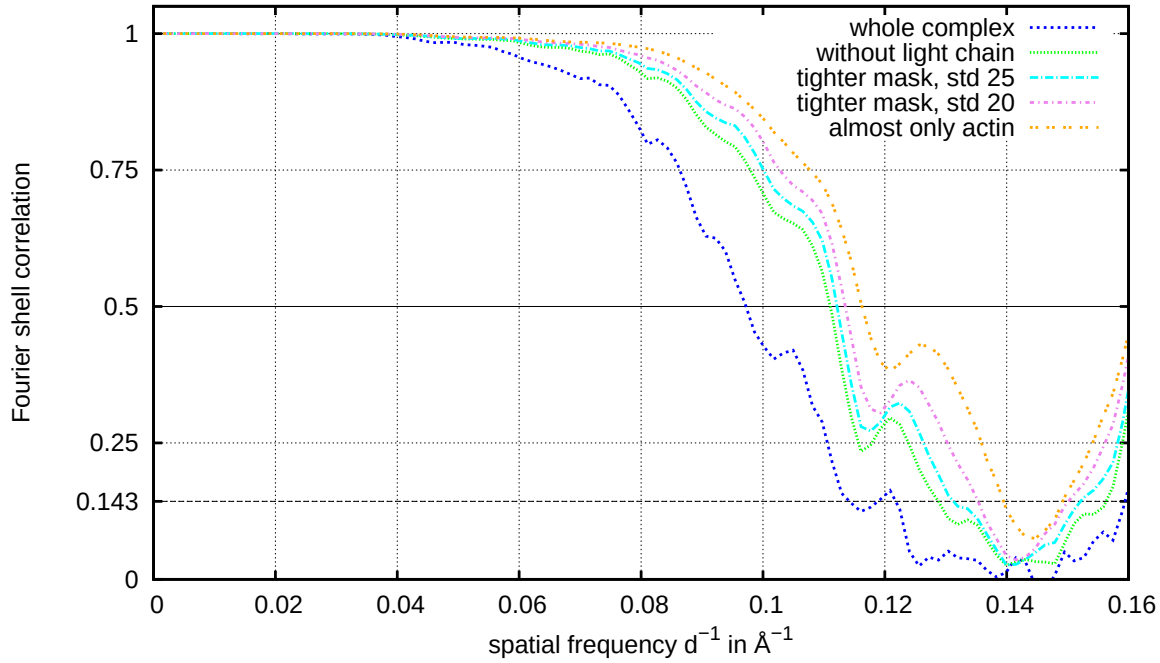


Figure 8.15: Fourier Shell Correlation curves for the fully CTF corrected data halves of the rigor data set. The density has been restricted in size by using different Gaussian masks. Resolutions have been calculated using the whole complex up to using the density radially restricted to almost only the actin density with the beginning of the myosin motor domain. The results reveal a clear resolution gradient as function of the radius with highest resolution for the actin and lowest resolution for the light chain density. The resolution ranges from 10.2 Å / 8.2 Å to 8.6 Å / 7.2 Å (FSC(0.5) / FSC(0.143)). The artificial rising of the FSC curve at higher frequencies again reflects effects of numerical embedding and masking.

not been divided into different halves. The reconstructions have been done independently of each other to avoid common reference artifacts. The FSC curves for the different approaches to correct the CTF are shown in Fig. 8.14. Despite an unwanted rising of the FSC curves at higher frequencies, it can clearly be seen that the reconstruction has been improved from one CTF correction and parameter set refinement to the next. The first approach that resulted in a slightly too large helical rise offers a resolution of 11.2 Å / 9.2 Å at FSC(0.5) / FSC(0.143) (red curve). The recalibration of the pixel size and a subsequent new estimation of the CTF parameters using CTFFIND3 allows to resolve the acto-myosin complex with 10.7 Å / 8.9 Å (FSC(0.5) / FSC(0.143)) (green curve). The full CTF correction allows an interpretation of the density map even further till 10.2 Å / 8.2 Å (FSC(0.5) / FSC(0.143)) (blue curve). The analysis shown in Fig. 8.15 reveals a resolution gradient as function of the radius. Third order Gaussian ellipsoids have been used with different standard deviations (std) in x,y and z in order to restrict the size of the density. Step by step more of the outer density has been masked. The light chain has the lowest resolution. Although the other curves are lying closer together there is still a gradient towards highest resolution for the central actin density. According to the values at FSC(0.5) / FSC(0.143) the resolution can be stated as: whole complex 10.2 Å / 8.2 Å, without light chain 9.0 Å / 7.8 Å, tighter mask (std 25) 8.9 Å / 7.6 Å, tighter mask (std 20) 8.8 Å / 7.4 Å and almost only actin 8.6 Å / 7.2 Å. The artificial rising of the correlation curves at higher frequencies reflects effects of the masking, which become even stronger with stronger masking.

8.2.2 Strong binding ADP state

Parallel to the processing of the rigor data set the same processing has been performed for the data set of the ADP state. Figure 8.16 shows the slices of the final reconstruction. The filament segments have been multiplied by their estimated CTFs and the reconstructed volume has been corrected in 3D for the amplitudes. The initial symmetry parameters have been the same as for the rigor processing: 27.5 Å helical rise and $-167.1°$ symmetry angle. They converged to 27.51 Å and $-167.143°$ using a $1°$ sampling both for the azimuthal reference projections and for the $\pm 18°$ out-of-plane tilts. The symmetry parameters exactly resemble the 28/13 crystallographic symmetry. In comparison to the slices of the rigor density it can be seen, that the two density shapes are quite similar except the light chain density. In the image slices for the ADP state this density is weaker. In both reconstructions a high negative density is present close to the protein density. Using the EM-BFACTOR program a B-factor of 326.7 Å^2 could be estimated that affects the density. Thus, for restoring of the amplitudes the negative B-factor of -326.7 Å^2 has been applied to the reconstructed volume. A surface representation of the final fully CTF- and B-factor corrected volume is shown in Fig. 8.17. It is difficult to visualize the differences of the two densities in 2D. Both reconstructions have a lot of similarities as it can

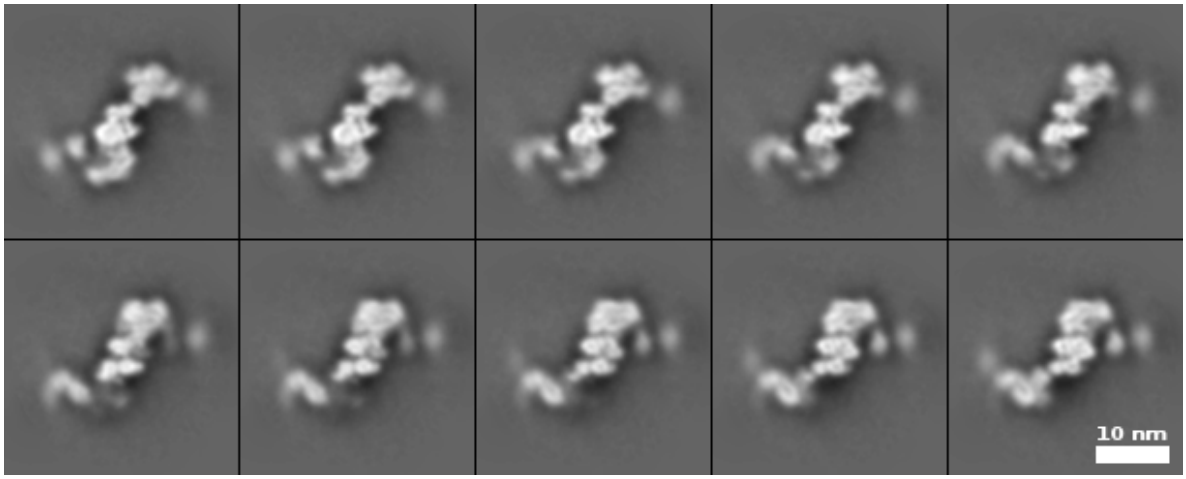
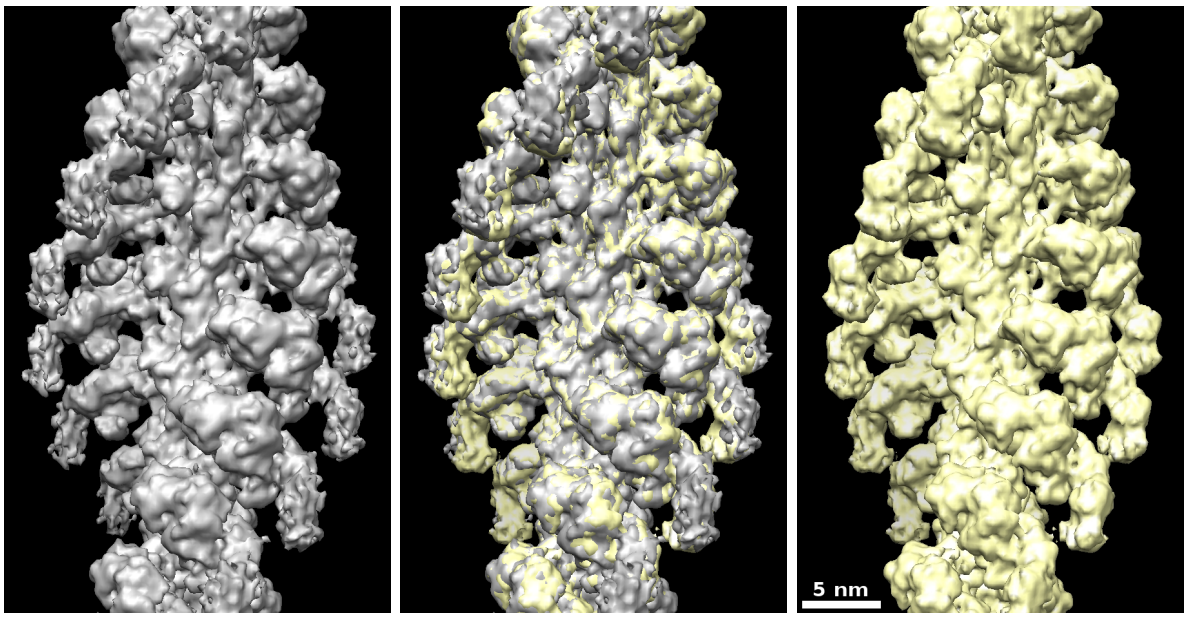


Figure 8.16: Slices of the phase- and amplitude corrected reconstruction of the strong binding ADP state. The reconstruction was done using the *sxihrsr* program from SPARX. A 3D amplitude correction of the volume was performed in SPIDER, but so far no B-factor correction or any other filtering has been used.



(a) ADP state (b) ADP vs. rigor state (c) rigor state

Figure 8.17: (a) Surface representation of the final, fully CTF corrected reconstruction of the ADP-state data set. Using the IHRSR method the symmetry parameters converged exactly to the 28/13 symmetry with a helical rise of 27.51 Å and a symmetry angle of $-167.143°$. The density was sharpened by a B-factor of -326.7 Å^2. In total the density is reconstructed by using approx. 243000 single particles / actin and myosin monomers. The surface rendering was done in Chimera. For comparison, (b) shows the overlay of both reconstructions and (c) the rigor density.

be seen in the overlay that shows alternately one or the other color. The difference that can even be seen in the two-dimensional representation is the position of the lever arm and of the light chain that is bound to it. In the ADP state the position is slightly higher and the light chain is more distant to the myosin motor domain, like an arm that is splayed out from the body in one case and in the other case it hangs down closer to the body. Additionally, the 3D representation shows that the light chain density is more noisy and less dense than the one in the rigor state. The number of particles used for the reconstruction is only slightly (about 7%) lower than of the rigor state (243000 vs. 259000 particles). The defocus- and out-of-plane tilt distribution are assessed in the following. Fig. 8.18 illustrates the defocus distribution as it has been estimated by the CTFFIND3 program using the pixel size of 2.62 Å. There are less images with defoci outside the defined range than before. Three sub-

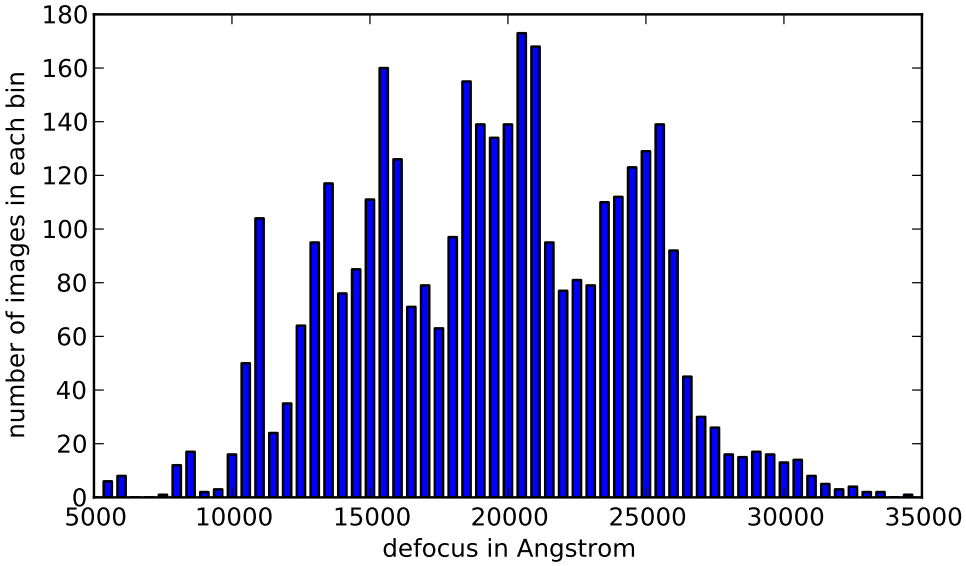


Figure 8.18: Distribution of estimated defocus values in the 3602 ADP state images that have been acquired with random defocus assignments by Leginon. The defined defocus range was the same as for the rigor state: 10000–25000 Å. All values were again applied in the so-called under-focus.

ranges are present that have been sampled more often than other defoci. The manual selection of 3602 images with good filaments out of 5618 images may again have changed the random distribution. An imprecise focussing may have led to the values outside the range. The out-of-plane tilt distribution of the last reconstruction iteration can be seen in Fig. 8.19. The distribution is much narrower than the rigor distribution and there are nearly twice as many segments that have been estimated without any out-of-plane tilt. Such different distributions were not expected. This one is clearly the preferred one. However, the highest considered out-of-plane tilts still have high scores. Fig. 8.20 illustrates the

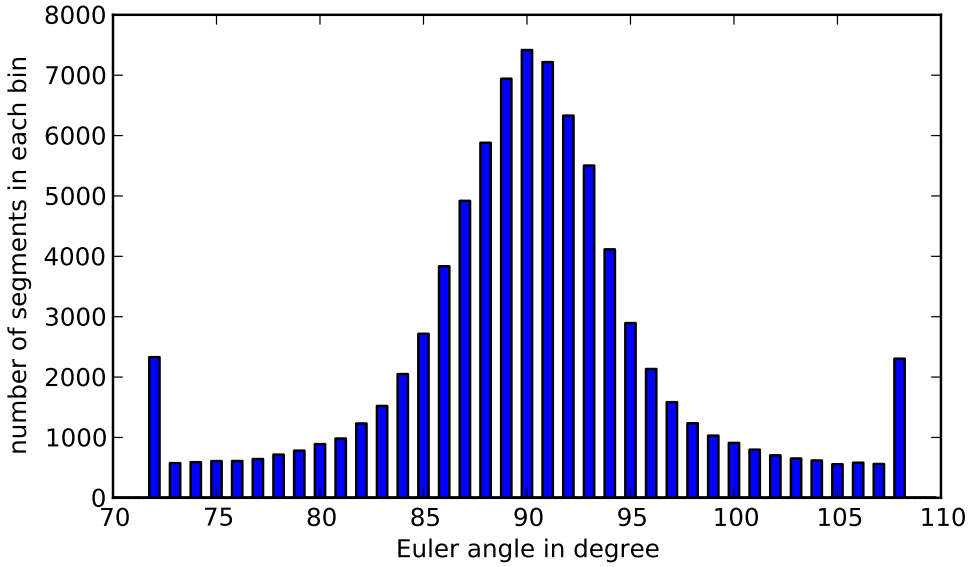
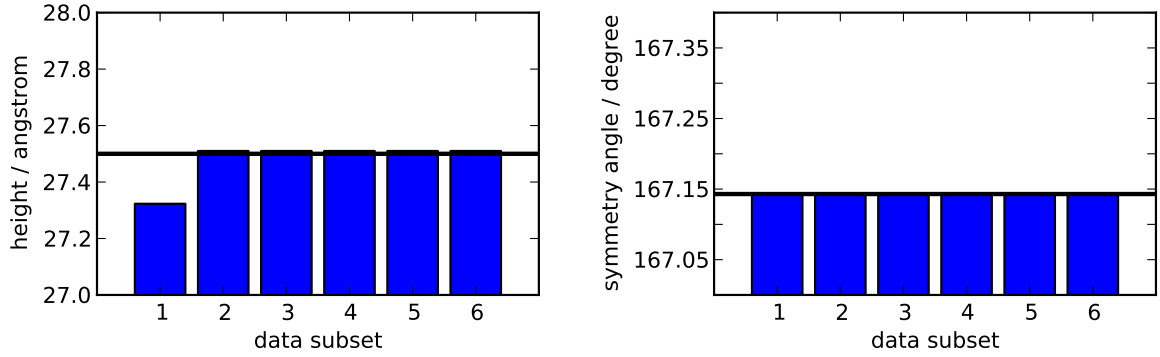


Figure 8.19: Distribution of estimated out-of-plane tilts considering all 84983 segment images, that have been used from the ADP data set. The Euler angle ϑ describes the tilt out of the projection plane. The number of segments with ϑ=90°, that corresponds to segments that lie within the projection plane, is nearly doubled compared to the rigor distribution.

analysis of the symmetry parameters of six data subsets, that again have been chosen in chronological order. As for the whole data set the parameters converged for all subsets to the parameters of an 28/13 helix except for the first subset. Its helical rise converged to 27.32 Å. Such size is the same as for filaments that are shortened due to an out-of-plane tilt of ±6.5°. Again the reconstruction algorithm cannot distinguish between them, if the pixel size is really slightly different in the first data subset. Suspicious is the fact that the data acquisition of this subset directly followed the second subset of the rigor data set, that has shown also a slightly lower helical rise. Indeed it might indicate a systematic error introduced by the microscopy and not by the specimen. The complex seems to be quite stable in its symmetry, since all other subsets converged exactly to the same parameters that has not been seen for the rigor state. The radial density profiles are shown in Fig. 8.21 together with a simulated profile for the ADP state. The profile modeling was performed as for the rigor state using rigid body docked crystal structures that have subsequently flexible fitted by molecular dynamics simulations (cf. next chapter for a detailed description of the method). The peak for the light chain density and the minimum before that peak is more pronounced in the simulated data than for the rigor state. Except that the profile values at high radius are less negative than for the rigor state, there is almost the same amount of negative density present in both reconstructions, that leads to minima at or even below



(a) helical rise

(b) symmetry angle

Figure 8.20: Symmetry parameters for six ADP data subsets, that have been reconstructed independently of each other. Like for the rigor analysis the data is chronologically divided into subsets to enable systematic errors to show up. The height of the symmetry unit converged to a mean value of (27.479 ± 0.070) Å and the symmetry angle to a mean value $(-167.143 \pm 0.000)°$. Lines indicate the symmetry parameters for a 28/13 crystallographic symmetry (27.5 Å and $-167.143°$). The symmetry angle is shown with positive sign.

zero. The positions of the maxima are also shifted slightly towards higher radius, except for the actin peak of the fully corrected reconstruction. The resolution assessment by means of Fourier shell correlations shows similar improvements for the different CTF correction approaches as has been seen for the rigor data set (see Fig. 8.22). The first approach that resulted in a slightly too large helical rise shows a resolution of 11.1 Å / 9.0 Å at FSC(0.5) / FSC(0.143) (red curve). The recalibration of the pixel size and a subsequent new estimation of the CTF parameters using CTFFIND3 allows to resolve the actomyosin complex with 10.2 Å / 8.8 Å (FSC(0.5) / FSC(0.143)) (green curve). The full CTF correction allows an interpretation of the density map to 10.1 Å / 8.6 Å (FSC(0.5) / FSC(0.143)) (blue curve), that is only slightly better. The resolutions are quite similar for the rigor and ADP data set. Interestingly for the ADP data set the phase flipped reconstruction with the correct pixel size yield already almost the same resolution as the fully corrected reconstruction. For the rigor data set the FSC curves for this two CTF corrections have been separated more strongly. The analysis shown in Fig. 8.23 reveals a similar resolution gradient as function of the radius that has been seen for the rigor reconstruction. Again there is a huge gap between the FSC curves that have been estimated without the light chain density and the one of the whole complex. The light chain has the lowest resolution and although the other curves are lying closer together there is still a gradient towards highest resolution for the central actin density. According to the values at FSC(0.5) / FSC(0.143) the resolution can be stated as: whole complex 10.1 Å / 8.6 Å, without light chain 8.9 Å / 8.0 Å, tighter mask (std 25) 8.7 Å / 7.9 Å, tighter mask (std 20) 8.6 Å / 7.8 Å and almost only actin 8.4 Å / 7.2 Å. The FSC curves

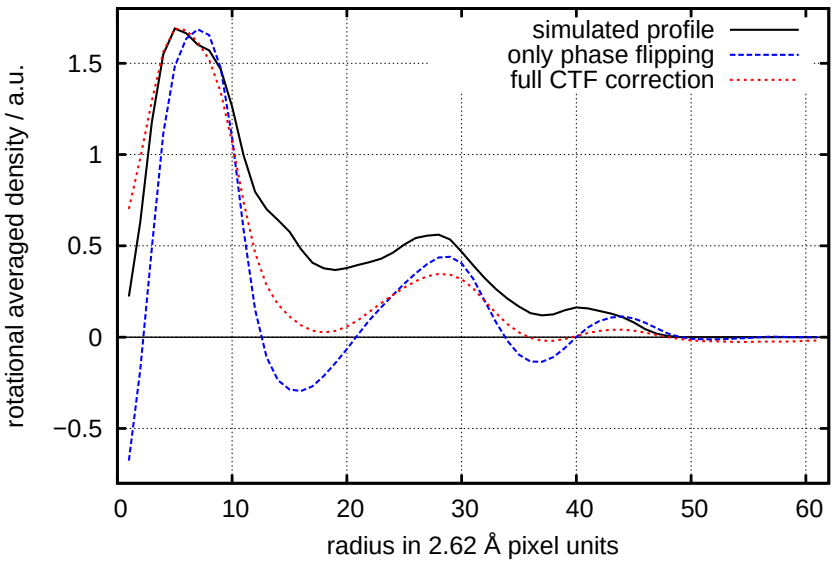


Figure 8.21: Radial density distribution of the reconstructed ADP-state density. Like in the rigor analysis the helices have been projected down their filament axis and the resulting image has been rotationally averaged to yield a one-dimensional profile. Profiles are shown both for the phase flipped reconstruction and for the fully CTF corrected density. For the assessment of the quality a simulated profile is shown as well.

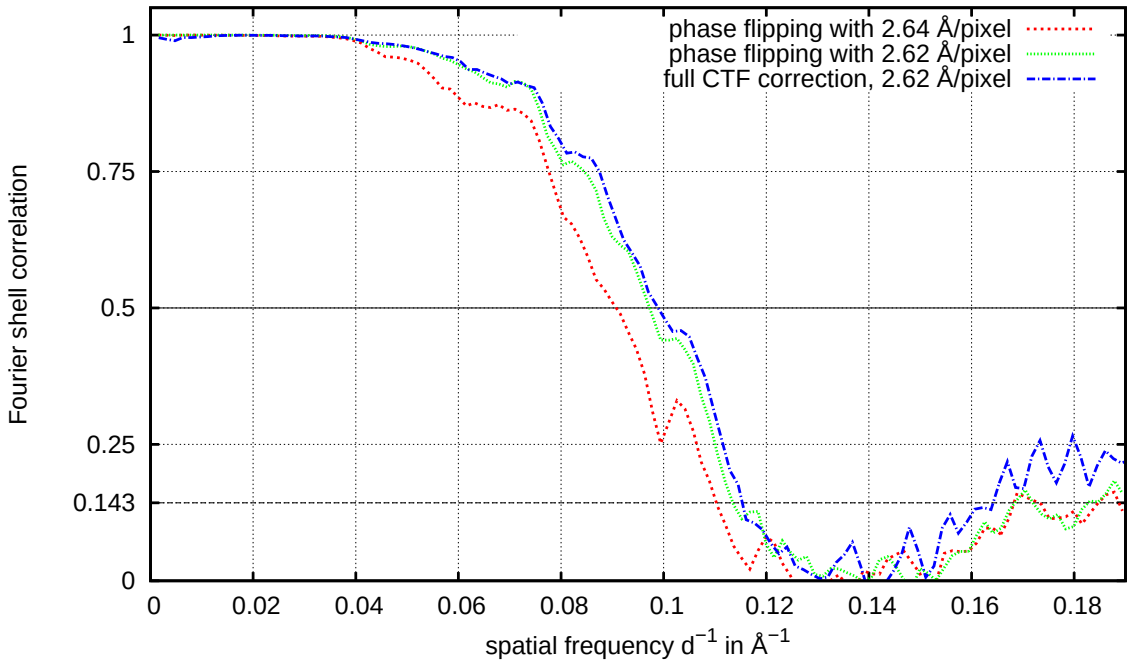


Figure 8.22: Fourier Shell Correlation curves of the odd and even data halves of the ADP data set. The different resolution criteria are indicated by the black lines at FSC(0.5) and FSC(0.143). The resolution could be improved by each subsequent approach to correct the data set for the effects of the contrast transfer function. The blue curve shows a final resolution of 10.1 Å / 8.6 Å (FSC(0.5) / FSC(0.143)). The artificial rising of the FSC curve at higher frequencies reflects effects of numerical embedding and masking.

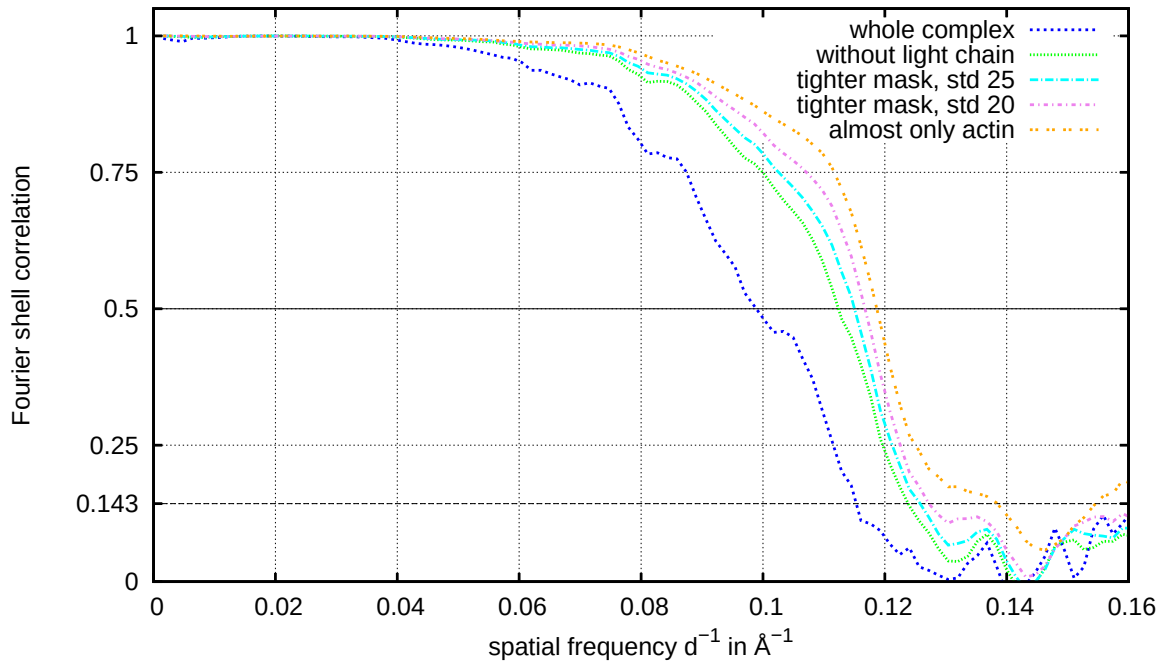


Figure 8.23: Fourier Shell Correlation curves for the fully CTF corrected data halves of the ADP data set. The density has been restricted in size by using different Gaussian masks. Resolutions have been calculated using the whole complex up to using the density radially restricted to almost only the actin density with the beginning of the myosin motor domain. The results reveal a similar resolution gradient as has been seen for the rigor reconstruction with highest resolution for the actin and lowest resolution for the light chain density. The blue curve for the whole complex is the same as in Fig. 8.22. The resolution ranges from 10.1 Å / 8.6 Å to 8.4 Å / 7.2 Å (FSC(0.5) / FSC(0.143)). The artificial rising of the FSC curve at higher frequencies reflects again effects of numerical embedding and masking.

of the ADP state show a steeper decline than the rigor curves. The resolutions at FSC(0.5) are slightly better for the ADP data set, whereas at FSC(0.143) the resolutions are slightly better for the rigor data set due to an additional shoulder of the correlation curves. The artificial rising of the curves towards higher spatial frequencies is less pronounced for the ADP data set.

8.2.3 Findings for the data set reconstructed by the filtered least squares algorithm

In order to assess the findings of the least squares method in a more objective way the old data set is analyzed and discussed also by means of the iterative helical reconstruction method. The reconstruction has been done using the *IHRSR++* program without parallel computing. In the first reconstruction approach the helical rise converged to 28.3 Å using the data, that has been interpolated for the least squares method to 2.75 Å/pixel. Obviously, the pixel size has been overestimated without taking into account the out-of-plane tilts. By considering $\pm 18°$ out-of-plane tilts the former used pixel size revealed to be rather 2.673 Å instead of 2.75 Å. For the IHRSR method a second refinement has been done to correct the CTF for the recalibrated pixel size, followed by a new reconstruction. The impact for the least squares method is assessed below in the discussion. The image preparation prior to the reconstruction has been done just as for the new data sets. In contrast to the preparation for the least squares method the phases have been flipped in the filament images. The CTF parameter estimation and the phase flipping has been done using SPIDER. The data set has been divided into six chronological subsets and been reconstructed independently. The final reconstructions have been aligned to each other and been averaged. Using the recalibrated pixel size the symmetry parameters converged to mean values of (27.46 ± 0.08) Å and $(-167.27 \pm 0.02)°$. The slices of the final, phase flipped reconstruction can be seen in Fig. 8.24.

Obviously, the quality of the reconstruction is much improved compared to the least squares reconstruction (cf. Fig. 6.14, page 61). The slices are comparable to those of the new data sets, but unfortunately different slices had to be chosen, they are not from the same position of the helix as before. No amplitude correction has been done and the negative density around the protein density is quite strong. For the B-factor estimation the spherically averaged structure factors have been noise-weighted considering the Fourier shell correlation curve of the data halves. The estimation in the resolution range from 18 Å to 9 Å using the EM-BFACTOR program results in a B-factor of 313.2 Å^2 that affects the data. Thus, the amplitudes have been restored using a negative B-factor of -313.2 Å^2. The surface rendering of the final, phase- and B-factor corrected reconstruction is shown in Fig. 8.25. The visual impression is comparable to the two new data sets. The position of the lever arm, as discussed later, seems to be more like the new ADP state. The light chain has less density than the new

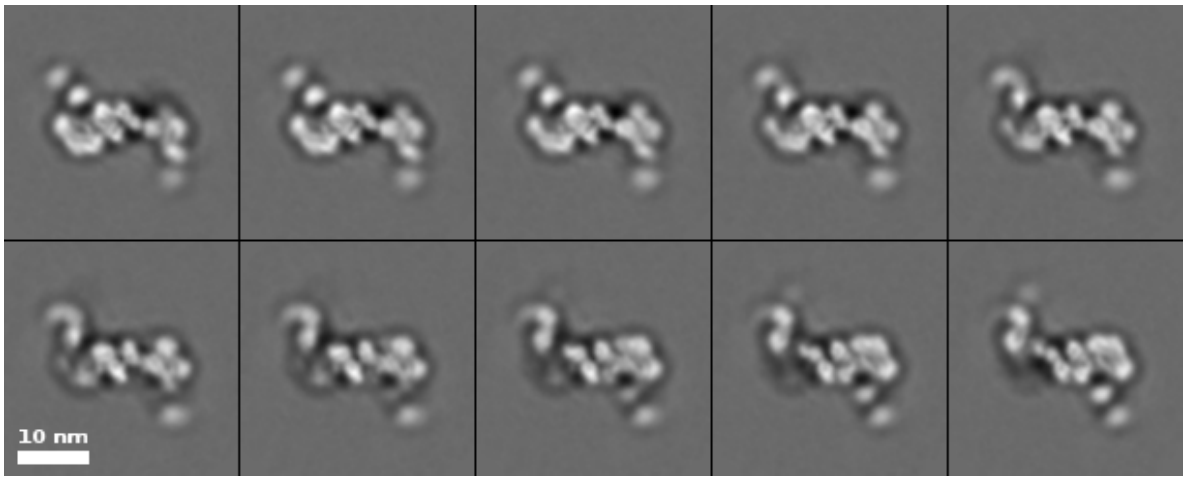


Figure 8.24: Slices of the phase flipped reconstruction of the old main data set, that has been discussed before when reconstructed by the filtered least squares reconstruction method. Now the reconstruction was done using the *IHRSR++* program. An amplitude- or B-factor correction of the volume has not been done. The CTF parameters have been estimated using SPIDER's operation TF ED, however, a large residual negative density (black) remains, which may point to a wrong CTF model and a required refinement.

rigor state but seems to be defined more clearly than the new ADP state. The good quality of this reconstruction reveals that the quality of the data set could by far not be exploited by the least squares method. Figure 8.26 depicts the symmetry parameters the different subsets converged to. Again there is one subset that exhibits a smaller helical rise and leads to a decreased mean value. Interestingly the mean value of the six subsets for the two new data sets have not been equal to the parameter found in the reconstruction of the whole data set. For this data a reconstruction of the whole data set has not been done. The symmetry angle converged to slightly varying values. The mean value lies in between the values found for the new rigor data set ($-167.405°$) and the new ADP data set ($-167.143°$). It resembles quite well the arithmetic mean of both.

A small second data set was acquired for the strong ADP binding state, as described in Chapter 6. Only 71 images could be obtained for the image processing. They comprise about 200 filaments, in total about 13300 symmetry units / single particles, and have been processed in the same way as the main data set, that originally was thought to be in the rigor state. Figure 8.27 shows density overlays of the different reconstructions. The fact that the densities of the two old data sets are not differing gave rise to the evidence that the former preparation might have led to a kinetically impure mixture of structural states. The fact that the old "rigor" data set resembles quite well the new ADP binding state suggests that the so-called "rigor" data set indeed is actually more or less in the ADP binding

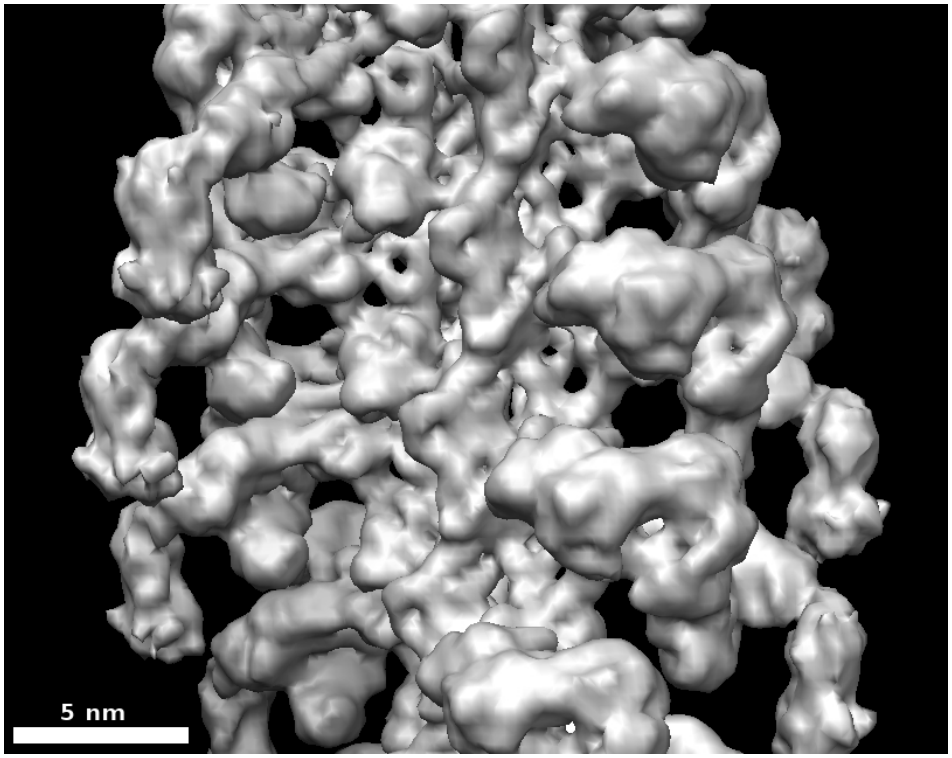
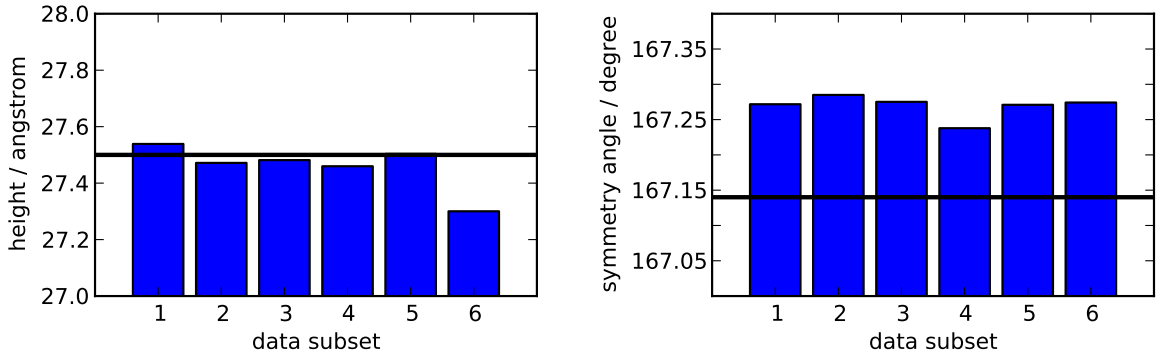


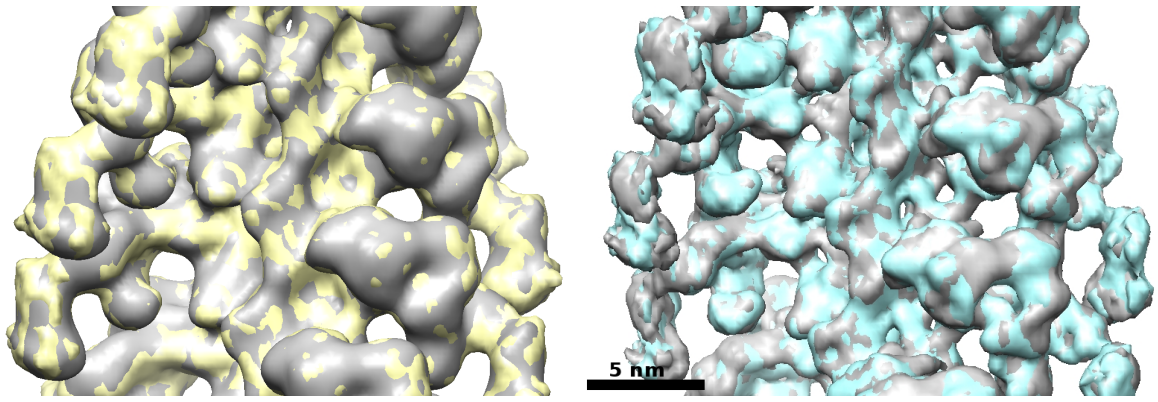
Figure 8.25: Surface representation of the final, phase flipped and B-factor corrected reconstruction of the data set that has been discussed by the filtered least squares method. Using the IHRSR method the symmetry parameters converged to a helical rise of 27.46 Å and a symmetry angle of $-167.27°$. The density was sharpened by a B-factor of -313.2 Å^2. In total the density is reconstructed by using approx. 212000 symmetry units / single particles / actin and myosin monomers. The surface rendering was done in Chimera.



(a) helical rise

(b) symmetry angle

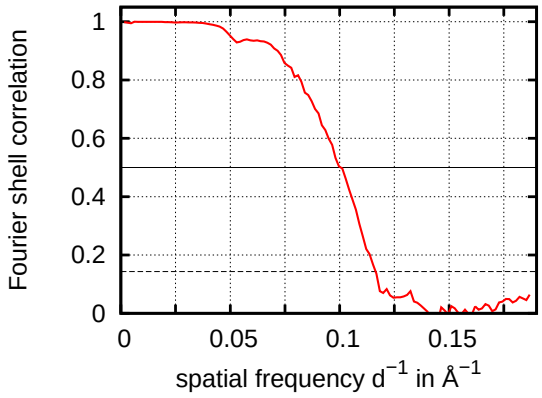
Figure 8.26: Symmetry parameters for six data subsets, that have been reconstructed independently of each other. The height of the symmetry unit converged to a mean value of (27.459 ± 0.076) Å and the symmetry angle to a mean value $(-167.269 \pm 0.015)°$. Lines indicate the symmetry parameters for a 28/13 crystallographic symmetry (27.5 Å and $-167.143°$). The symmetry angle is shown with positive sign.



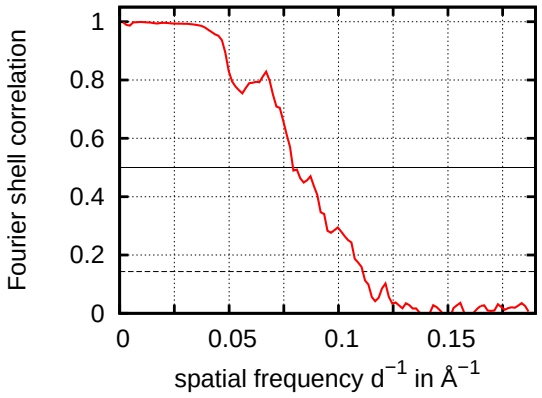
(a) Old "rigor" vs. small ADP data set

(b) Old "rigor" vs. new ADP data set

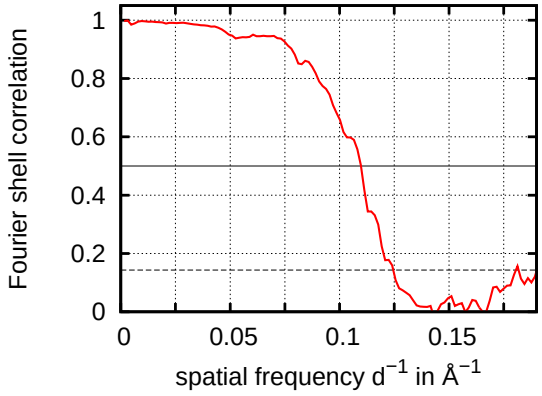
Figure 8.27: (a) Overlay of the surface representations of both old data set. The data set intended to be in the rigor state is shown in grey, the old small ADP data set in yellow. Both densities are without amplitude- and B-factor correction. (b) shows the old main data set ("rigor") in grey compared to the new ADP data set in cyan. Both densities have been filtered by a Gaussian high pass filter. All three data sets are quite similar, show no obvious differences and especially feature the same position for the lever arm and light chain density.



(a) Data halves old "rigor" data set



(b) Old "rigor" vs. small ADP data set



(c) Old "rigor" vs. new ADP data set

Figure 8.28: Fourier Shell Correlation curves. The different resolution criteria are indicated by the black lines at FSC(0.5) and FSC(0.143). (a) The phase corrected reconstruction of the old "rigor" data set offers a resolution of 10.0 Å / 8.6 Å (FSC(0.5) / FSC(0.143)). (b) a correlation between the two old states gives a resolution of 12.5 Å at FSC(0.5). It might be influenced by the low signal-to-noise ratio of the small ADP state. (c) the correlation of the whole old "rigor" data set with the new ADP state gives the best resolution obtained for the whole complex including the light chain density: 9.1 Å / 8.0 Å (FSC(0.5) / FSC(0.143)).

state. This result agrees with the kinetic evaluation of Prof. Dr. H. Lee Sweeney (personal communication), that the preparation favored an ADP bound state instead of the rigor state. The amount of mixture within the data set remains unclear. The *IHRSR++* reconstruction of the old small ADP data set converged to the helical rise of 27.52 Å and a symmetry angle of $-167.28°$. Again the angle does not match the one found for the new ADP data set. However, it is almost the same as for the old main data set. Finally the resolution is calculated by means of Fourier shell correlation. Fig. 8.28 shows different correlations between the data sets. The resolution of the phase corrected old "rigor" data sets is comparable to those of the new data sets that additionally have been amplitude corrected. The correlation of the two old states shows a lower resolution, but it is difficult to estimate if it is due to structural differences or probably only due to the low signal-to-noise ratio of the small ADP state. The correlation in (c) reveals the vast similarities of the old "rigor" state and the new ADP state and thereby confirms the evidence that the old state is at a very high percentage truly an ADP state. The correlation is better than those for the data halves of the new data sets for the whole complex including the light chain density. It might be influenced by the fact that whole data sets are correlated instead of data halves as before.

8.3 Discussion

Two large new data sets could be acquired that show different states of the actomyosin complex. Using the IHRSR method the reconstructions show a much refined quality. Additionally, the analysis reveals why the least squares method could not deal with the data sets in a productive manner. The obtainable resolution showed to be quite sensible to the quality of the correction for the contrast transfer function. The best possible CTF estimation is essential to have an optimal information recovery. The new data sets with a lower signal-to-noise ratio in the power spectra express the influence the instrumental equipment has. Most probably the lower quality of the camera and its smaller size let the CTF parameter estimation become quite difficult. The reconstructed densities feature a resolution between 10 Å and 7 Å depending on the resolution criterion used and depending on the radius of the density that is going to be included into the resolution estimation. In all cases a resolution gradient is present in the density, featuring highest resolution for the actin density close to the helical axis and lowest resolution for the light chain density at highest radius. The rigor data set reveals a small additional twist of the filaments compared to the 28/13 symmetry. The reconstruction converged to a symmetry angle of $-167.405°$. Compared to the strong binding ADP state of the actomyosin complex the rigor state shows the largest amount of reconstructed density for the light chain domain, although in terms of resolution there is no difference. The distributions of the out-of-plane tilts have surprisingly

been different for the two states. The ADP state shows a narrower distribution and at the same time the resolution of the phase flipped reconstruction has almost the same quality that the fully corrected reconstruction offers, where has been a larger difference for the rigor state. The symmetry parameters of the ADP state converged exactly to those of the 28/13 crystallographic symmetry (helical rise of 27.5 Å and symmetry angle of $-167.143°$). A difference in the position of the lever arm and the light chain density is present between the two states. It has already been seen long ago in a low resolution map reported by Whittaker et al. (1995). The ADP state shows a position slightly higher towards the pointed end of the actin filament, that corresponds to a power stroke intermediate, compared to the rigor state. The rest of the myosin motor domain seems to have already a lot of common features in both states. A closer look at that follows in the next chapter when molecular dynamics flexible fitting of crystal structures into the densities is discussed. The analysis of six data subsets revealed that there are deviations in the helical rise that might be introduced by systematic influences and might not originate from the structure itself. The problem is that the algorithm cannot distinguish them from projection shortenings due to an out-of-plane tilt. The high scores for the largest considered out-of-plane tilts also allow to raise the question if some kind of data classification or only excluding the images fallen into the highest tilt angles might improve the reconstruction. However, this is a known problem of the chosen reconstruction algorithm and needs to be considered in detail in future work.

The new ADP state shows to be quite similar to the former assumed “rigor” state of the old preparation. The Fourier shell correlation indicates a resolution of 9.1 Å / 8.0 Å. The old “rigor” state unfortunately has to be assumed to be a mixture of different states. Further high pass filtering of the new ADP state density shows less well resolved features inside the motor domain than the rigor state (data shown in next chapter). Therefore, one might raise the question if the preparation have yielded a pure state this time. The less well resolved light chain density might be due to an averaging of different positions or it might be populated less often in the filaments. Because of the non-covalent interaction of the light chain it might be not present in all myosins. On the other hand the estimated symmetry is quite stable for the new ADP state and three different preparations have yielded almost the same density distribution. Hence, it might also be the case that the kinetic tuning for the processivity of the myosin V favors a strong bound ADP state, that all three preparations led to the same state, although the symmetry angles are not exactly the same. This may point to a slight change of the actin binding site interactions which could affect helical symmetry. The old data sets feature a symmetry angle in between the two new states. The symmetry angle converged to $-167.27°$ and reveals that the symmetry angle estimated to have been best using the least squares method indeed is closest to the true angle present in the data set. Furthermore, the analysis showed that only a few percent of each data set are

filaments without any out-of-plane tilt. For the new rigor state it has only been about 4.5% and for the new ADP state about 9%. By neglecting the out-of-plane tilts the pixel size has not been estimated correctly. It showed to be rather 2.673 Å instead of 2.75 Å. Both effects are severe problems when applying the least squares method. The reading frame every 10 pixels has also been problematic for the small fraction of filaments without any out-of-plane tilt due to the deviation in the size of the pixel. The analysis of this chapter impressively shows how important it is to make no ad hoc assumptions and to estimate all values as free parameters in the best case.

Chapter 9

Molecular dynamics simulations allow further interpretation

Molecular dynamics (MD) simulations are used to combine the reconstructed densities of the actomyosin complex with structures that were obtained with higher resolution by protein crystallography. Flexible fitting (Trabuco et al., 2009) of the structures into the densities allows a further interpretation of the structural states beyond the resolution of the 3D reconstructions from the TEM images. The following results clearly show differences between the two, now newly reconstructed structural states. While the rigor complex has been investigated prior to this work for other myosins (for myosin II (14 Å) by Holmes et al. (2003) and for myosin Ie (8 Å) by Behrmann et al. (2012a)), it is the first time that the positioning of the myosin's internal structural elements is assessed for the strongly bound complex of actin and myosin with the nucleotide Adenosine diphosphate (ADP) present in the nucleotide binding pocket. This state can only be accessed due to the special myosin V kinetics, which is tuned to processive movement of single molecules along actin filaments. Thus, myosin V populates a structural state, that is also present in the cyclic movement of all other myosins, but only transiently.

9.1 Material and Methods

Normally protein crystal structures are incomplete. Some residues of the amino acid sequence are missing due to the flexibility of loops or other less structured regions. Generally also all hydrogen atoms and some other atoms are missing. Therefore, all structures have to be completed prior to MD simulations. The coordinates for the actin monomers were obtained already completed from Fujii et al. (2010). The initial rigor model was for this study here newly built out of three different structures. A sequence alignment between the crystal structures and the full myosin Va gallus gallus sequence (from

uniprot.org) identified the missing parts of all structures. Coordinates for the residues 595–627 could not be resolved for all structures. The residues belong to a large flexible loop, the so-called loop 2. To account for that the myosin model was split into two chains. Both pieces came from a mixture of following protein data files (pdb-files): 1W8J.pdb (chain A, Coureux et al., 2004), 2DFS.pdb (chain A, Liu et al., 2006) and 1OE9.pdb (Coureux et al., 2003). The heavy chain was built using residues 2–4, 382–385 and 628–634 from 1W8J, residues 5–381, 386–594 and 635–766 from 2DFS and residues 767-794 from 1OE9. The coordinates for the light chain were taken from 1OE9. The missing loops of about five residues have been modeled with Rosetta (https://www.rosettacommons.org/software). Large loop 2 is still missing. Its loose ends have been harmonically restrained during the simulations in order not to be teared apart from each other and to keep approximately their distance. The AutoPSF feature of the VMD software allows to complete the structures. Missing atoms of the amino acids, especially all hydrogens, are added. The modeled structure was checked for chirality errors and cis peptide bonds. One cis peptide bond was detected at a region that splits two helices. Since it came from the crystal structure, it was kept trusting that the crystal structure was correct in having a cis peptide bond there.

The myosin model with ADP present in the nucleotide binding pocket was built from the initial rigor model and two other structures that have kindly been provided by 1st class Research Director Dr. Anne Houdusse (Institut Curie, Paris, France). All residues were taken from the rigor model except for residues 702–795 and the light chain, they were taken from a structure similar to 1OE9 from chain A and the whole chain B for the light chain. Residues 1–239 and ADP coordinates were taken from a crystal structure of the myosin V motor domain in a complex with ADP-beryllium-trifluoride (ADP-BeF_3) (pdb-file similar to 1W7J). The chimeric myosin model was made by bonds between residue 239–240 and 701–702. The missing loops in the light chain were modeled by hand using coordinates from the completed rigor model. The ADP did not pass the structure completion by the AutoPSF feature. It had to be replaced. New coordinates were taken from 2CNQ.pdb, that had been aligned to the ADP within its binding site. The correction was done by hand. Again the structure was checked for chirality errors and cis petide bonds. Two cis peptide bonds were detected and kept, trusting again, that they are correct. While the protein data files (pdb-file) contain the following informations: atom ID, atom name, residue name, residue ID, x,y and z coordinates and some more parameters, a protein structure file (psf-file) contains all the molecule-specific information, that is needed to apply a particular force field to a molecular system. Contained are informations about atoms, bonds, angles, dihedral, impropers (dihedral force terms to maintain planarity of the peptide bond) and cross-terms. The psf-file is generated by the AutoPSF feature. Unfortunately, the angle parameters for the C2', C3'

and C4' atoms in the ribose of ADP were unknown. Thus, the psf-file had to be modified by hand in order to fix the problem. A Myosin Vc structure provided by 1st class Research Director Dr. Anne Houdusse contains a partial loop 2. It was used for a trial to built loop 2 of myosin Va. The loop might have a β-hairpin structure. Some simulations have been done including the guessed coordinates for loop 2, but the density seems not to give enough indication where it is positioned.

Rigid body docking of the structures into the densities was done using *colores* from the Situs software (Wriggers et al., 1999). It performs a comprehensive six-dimensional positional search using three translational and three rotational degrees of freedom and is suitable for one-at-a-time docking of single subunit structures, which are not necessarily expected to account for the full map. The densities were trimmed and both six actin monomers and six myosins were docked using 10 Å for the estimated resolution during docking. Not all monomers are fully engulfed by the density. For MD simulations it is preferred to have structures outside the density, that later on are not going to be interpreted, than to have large parts of density without structures. The second case will influence the neighboring monomers to drift into the vacant densities, thereby their structures will be deformed and cannot be interpreted as well. The docked monomers were combined into one protein data file using VMD scripting. The jointed structure has been completed by the AutoPSF feature to have a correct naming of all its segments and to create the necessary protein structure file. The simulations have been done using the NAMD software, version 2.9, together with the molecular dynamics flexible fitting (MDFF) extention. The executable program was built from source at the computer cluster and was compiled together with MPI (message passing interface) to enable parallel computation. As introduced in section 5.2 page 38 secondary structure restraints are needed to maintain the integrity of the structural elements during simulation. Also additional restraints have been used to prevent chirality errors and cis peptide bonds from arising. The harmonic restraints to account for missing loop 2 in myosin have been chosen with the same strength as for hydrogen bonds (force constant of value 20). Furthermore, symmetry restraints have been used. They are based on harmonic forces to maintain a symmetric structure during the simulations. A symmetric structure can be obtained by transforming and overlapping the atomic coordinates of all symmetric units. The average positions of the transformed atoms are calculated. Since symmetry restraints are so far working just with one symmetry group, only the myosins have been symmetrized using four myosins that are almost completely engulfed by the trimmed density. The first simulation step always has been an energy minimization for 2000 steps. Atom clashes with distances less than 1 Å, that might have been introduced by building the chimeric structures, are generally solved during the initial minimization. As second step a trajectory of all atoms has been simulated using 1 fs time steps for a simulation runtime of 3 ns. For most of the

simulations the scaling factor ξ, that adjusts the strength of the density (cf. Eq. 5.9 page 39), has been set to ξ=0.3. Because NAMD's *gridForces* feature requires an input 3D map defining the external potential in the DX file format, the densities have been converted into DX files.

9.2 Results

9.2.1 Simulations in vacuo

Due to the ease of calculation most simulations have been done in vacuo. The system of six monomers each for actin and myosin contains about 125000 atoms in total. The MDFF package offers the possibility to check the convergence of a simulation by means of calculating the root mean square deviation (rmsd) of the protein backbone with respect to the initial structure. A typical curve is shown in Fig. 9.1. Simulations have not only been performed once for both the rigor state and the strong binding

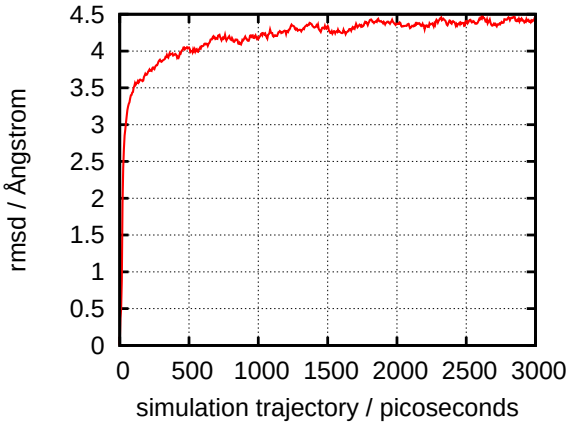


Figure 9.1: Backbone root mean square deviation (rmsd) per atom with respect to the initial structure for the MDFF simulation indicates convergence. When the deviation converges to a stable value the backbone movements to fit into the density are done and mostly only flexible loops and amino acid side chains are constantly populating slightly different positions.

ADP state, but also simulations have been calculated using differently filtered density representations of the structural states. Due to the non-physical force field obtained from the density map it was preferred to check the influence of different outlines of the maps. Four different simulation systems have been built: firstly using the unfiltered densities, only the phases were corrected, the amplitudes and the B-factor were not corrected. Secondly the densities have been filtered using a Gaussian high-pass filter. The filtering was done in the EMAN2 program with a cutoff frequency and sigma value of 0.2 in units where 0.5 is the Nyquist frequency. Thirdly the fully CTF corrected densities have been used. The filtering was applied by the B-factor correction. Fourthly the densities were filtered using a hyperbolic tangent (short: tanh) high-pass filter of EMAN2 with a cutoff frequency of 0.05 in units with 0.5 being the Nyquist frequency. The densities for this filtering have again only been corrected for the phases. Figures 9.2 to 9.9 present the simulation systems after 3 ns simulation time both for the rigor state and the strong binding ADP state. The threshold for the density representations has been

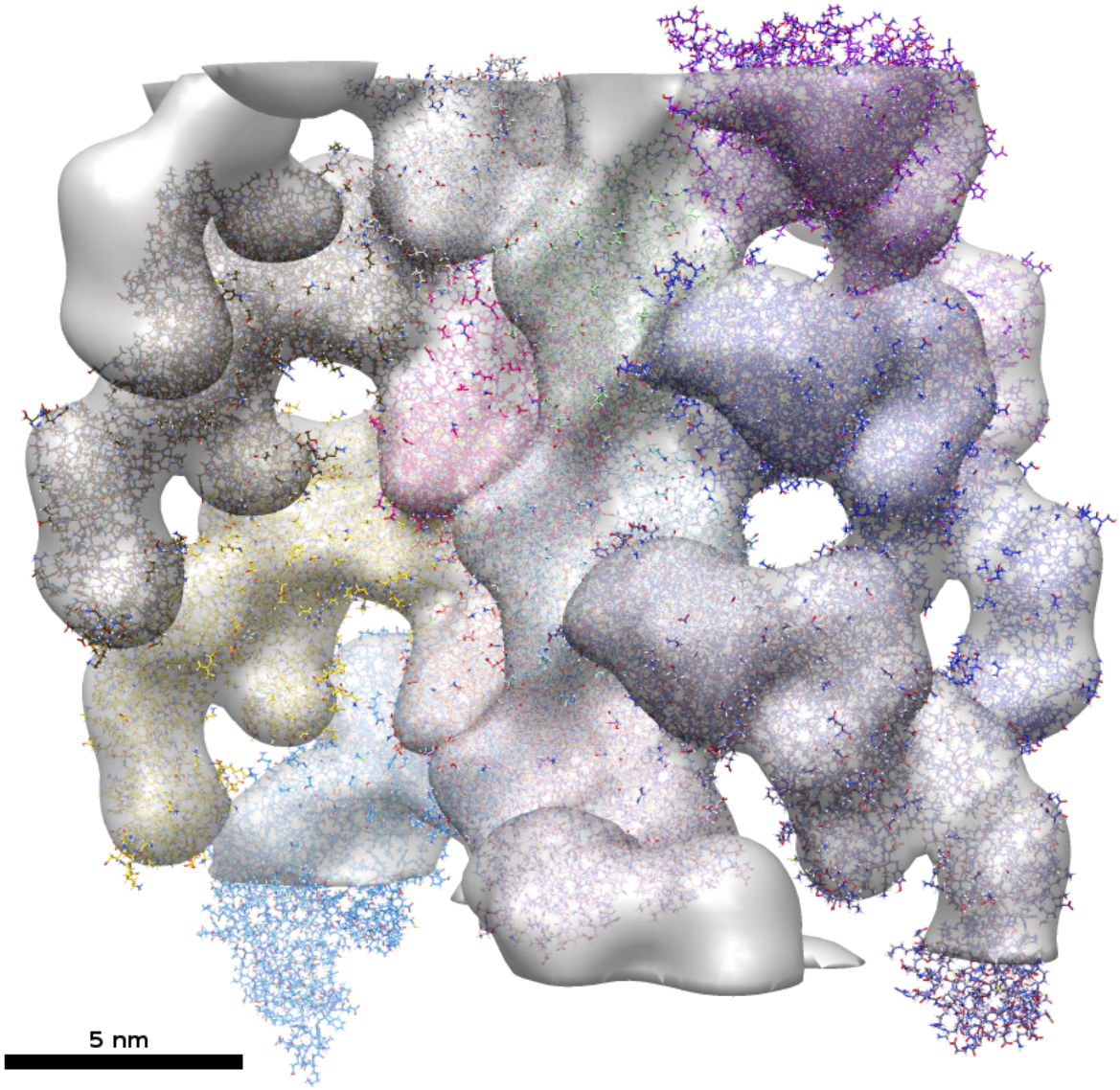


Figure 9.2: Simulation system with the unfiltered rigor density. Shown are the converged structures after 3 nanoseconds (ns) flexible fitting into the density. The protein structures are shown in atomic representation. Six actin monomers are situated in the central actin density, while three myosin monomers are located on the left and on the right. Two density regions can be seen that are left vacant during simulation, since they exhibit only a little connection to the main density. The structural parts that are sticking out of the trimmed density are deformed, thus the two myosins in turquoise and purple have not been used for the symmetry restraints and only the myosins in the middle that are fully engulfed by the density have been evaluated.

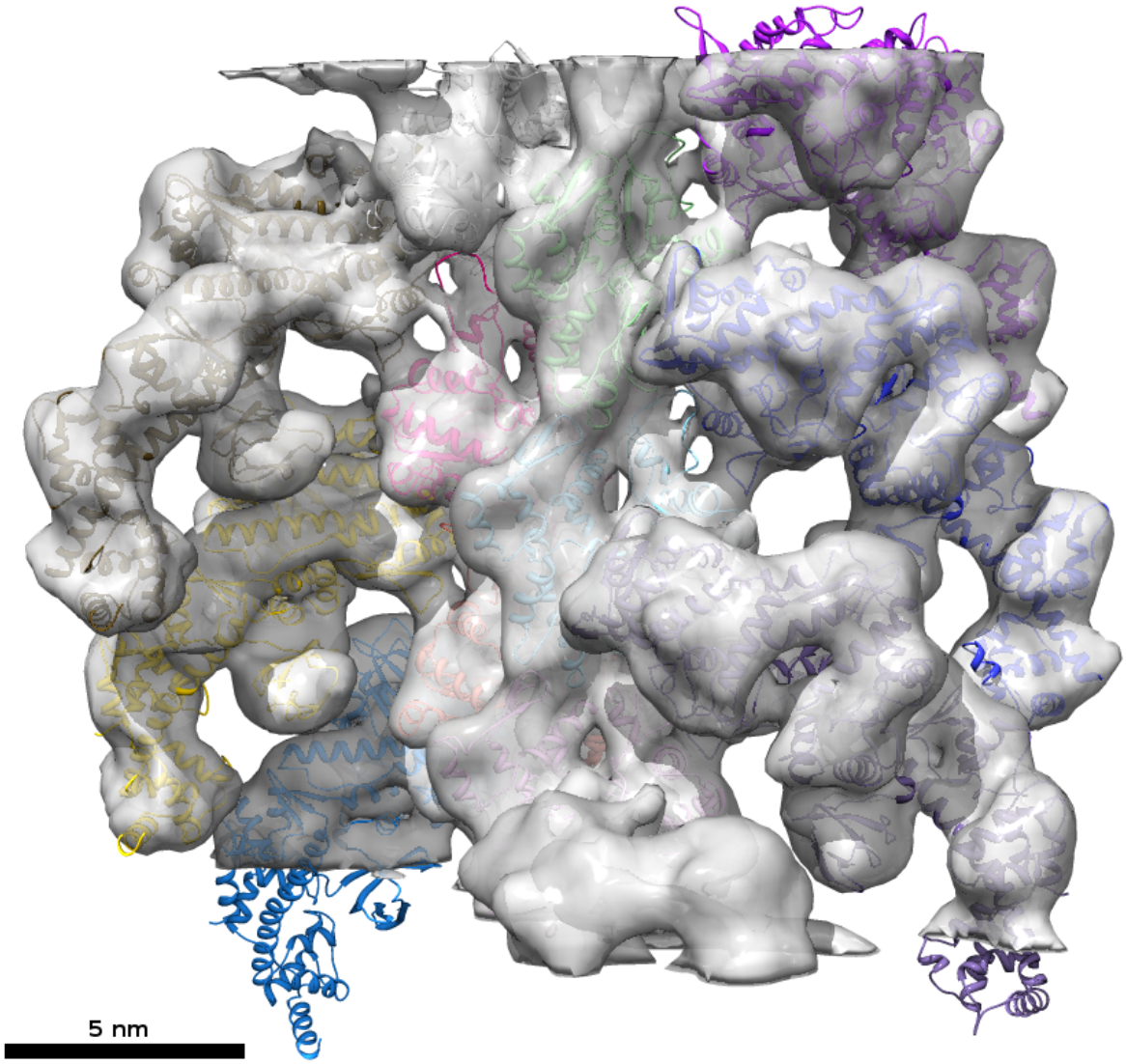


Figure 9.3: Simulation system with the Gaussian filtered rigor density after 3 ns flexible fitting using MDFF without solvent (simulation in vacuo). The protein structures are illustrated by a ribbon representation of their backbone, thus highlighting structural elements like helices. In this filtering the density exhibits a hole in the nucleotide binding site and reveals structural elements that are without density. The outline of the density resembles very well the shape of the actin and myosin monomers.

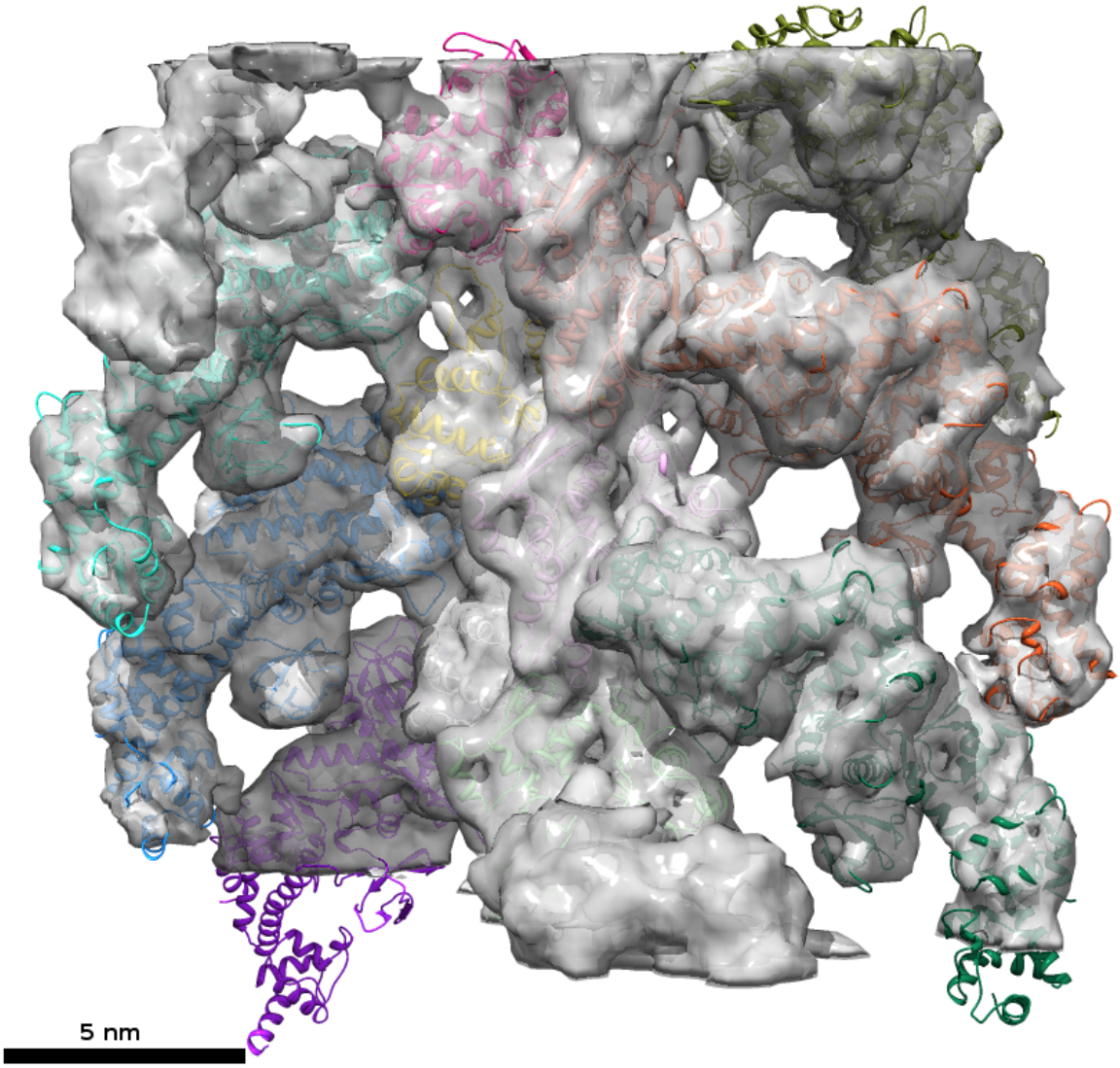


Figure 9.4: Simulation system with the B-factor corrected rigor density. Shown is the last frame of a simulated trajectory for 3 ns. The outer shape of the density shows more refined contours that represent very well the structural elements of the proteins. The loops at the interface of actin and myosin fit better into the holey density than using the Gaussian filtered density. The hole in the nucleotide binding site of the Gaussian filtered density is not present in this representation. With the used density threshold small loops or parts of helices are sticking outside, especially in the light chain region. The position of the lower part of the light chain seems not to be defined correctly. Analysis showed that it is remodeled in comparison to the known interaction with the lever arm.

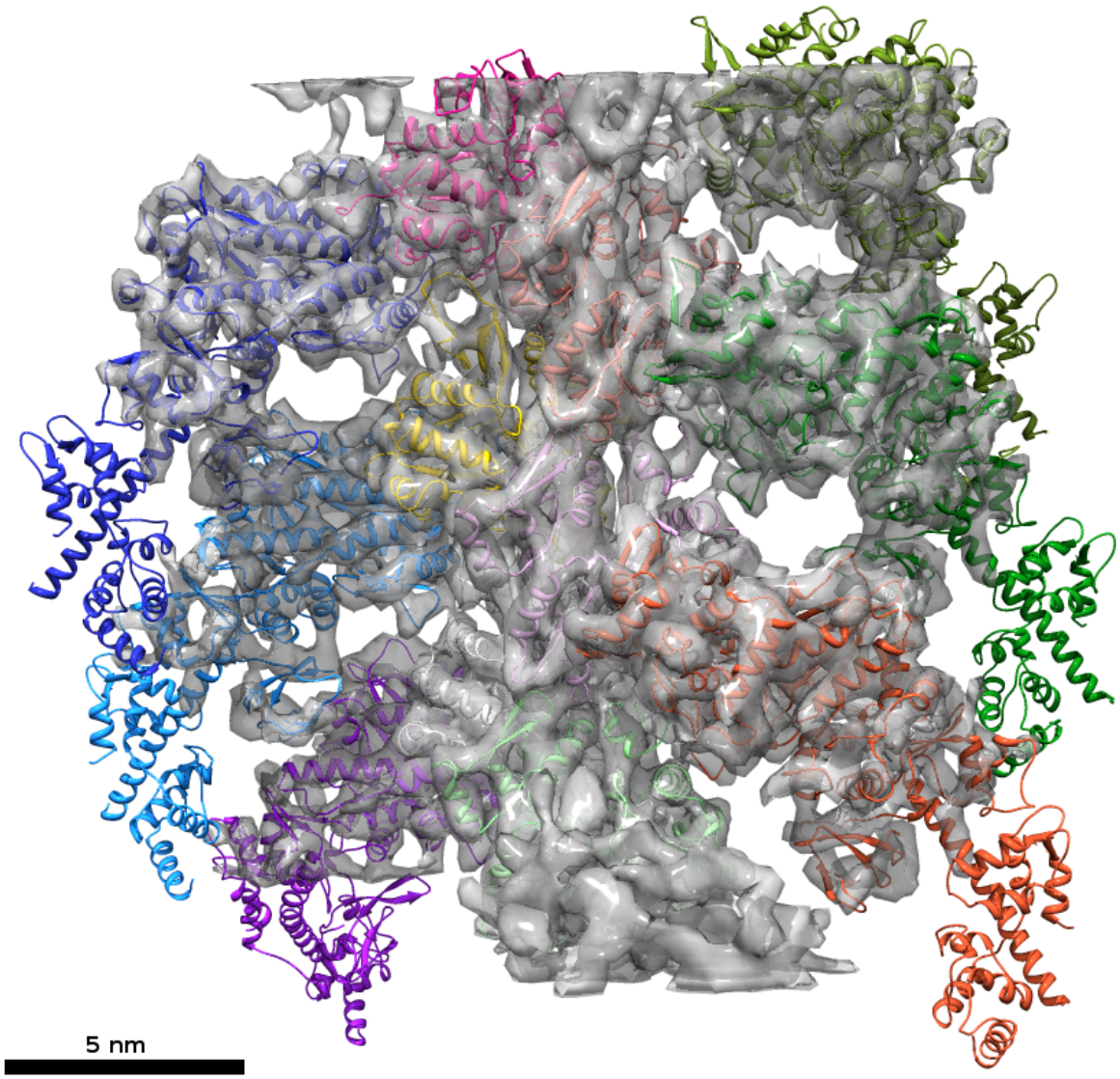


Figure 9.5: Simulation system with the hyperbolic tangent (tanh) filtered rigor density after 3 ns flexible fitting. The sharp filtering of the density allows further structural details of the density to show up. In this filtering the density for the light chain is not resolved as it is for the myosin motor domain. It is probably due to the lower resolution of the light chain, that becomes quite noisy upon sharp filtering and does not show up in this representation where small density blobs are hided. The helices of the actin are resolved very well, also most of the helices of the main part of the myosin motor domain are resolved quite well. The structural elements of the interface of actin and myosin are also nicely resolved. At higher radius from the filament axis some helices are sticking out of the density. Nevertheless, the density precisely defines the positions of the helices and thereby leaving less doubts concerning the validity of the obtained structure, except for the light chain position.

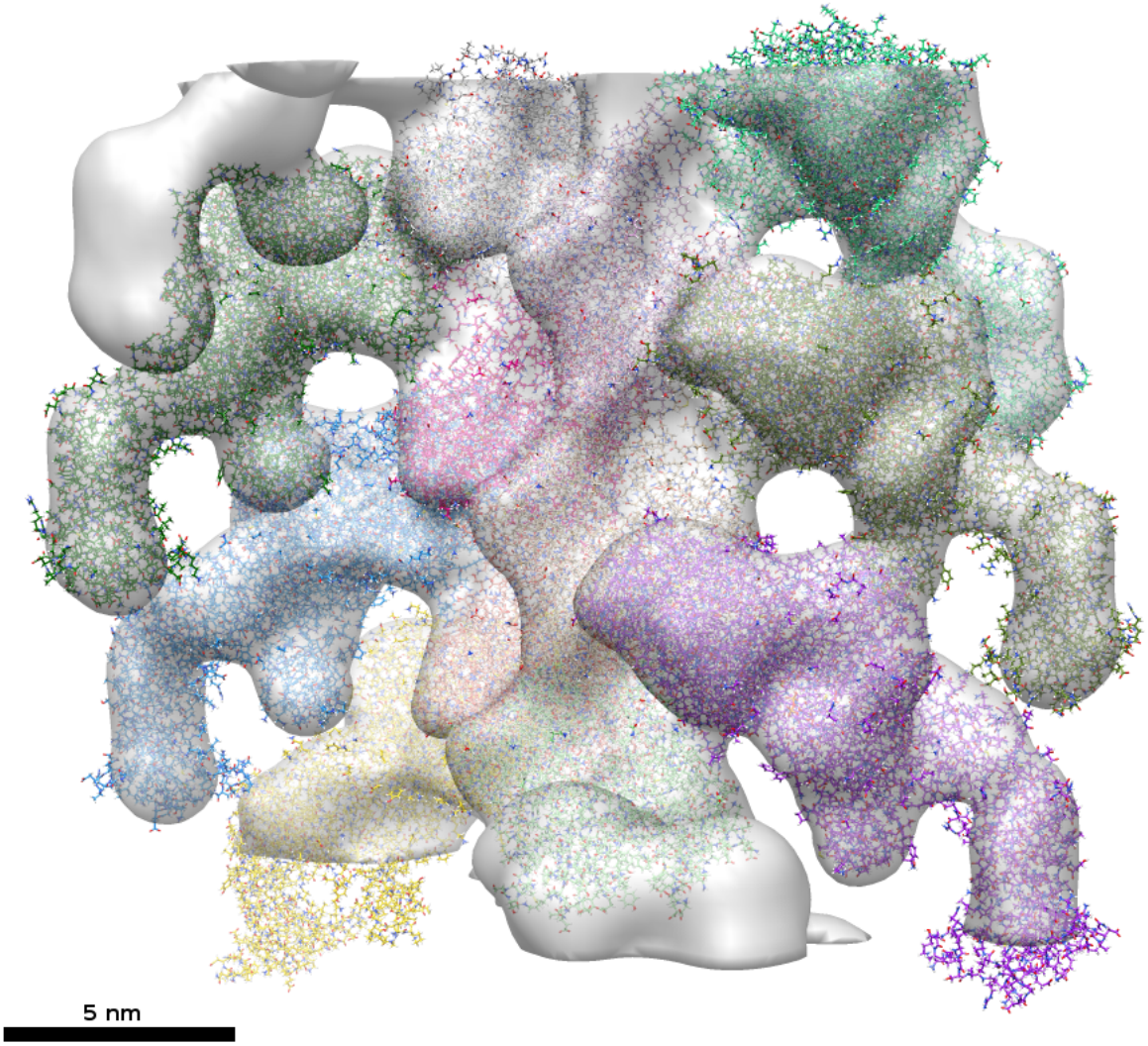


Figure 9.6: Simulation system for the unfiltered density of the strong binding ADP state. Shown are the converged structures after 3 ns flexible fitting into the density. The central six actin monomers and the two rows of three myosin monomers are shown in atomic representation. The structural parts that are sticking out of the trimmed density, the yellow and green myosins and the uppermost actin, are deformed. Again two density regions are left vacant during simulation, that exhibit only little connection to the main density. Only the four middle myosins have been used for the symmetry restraints. Compared to the rigor density the light chain has less density and more atoms are sticking out.

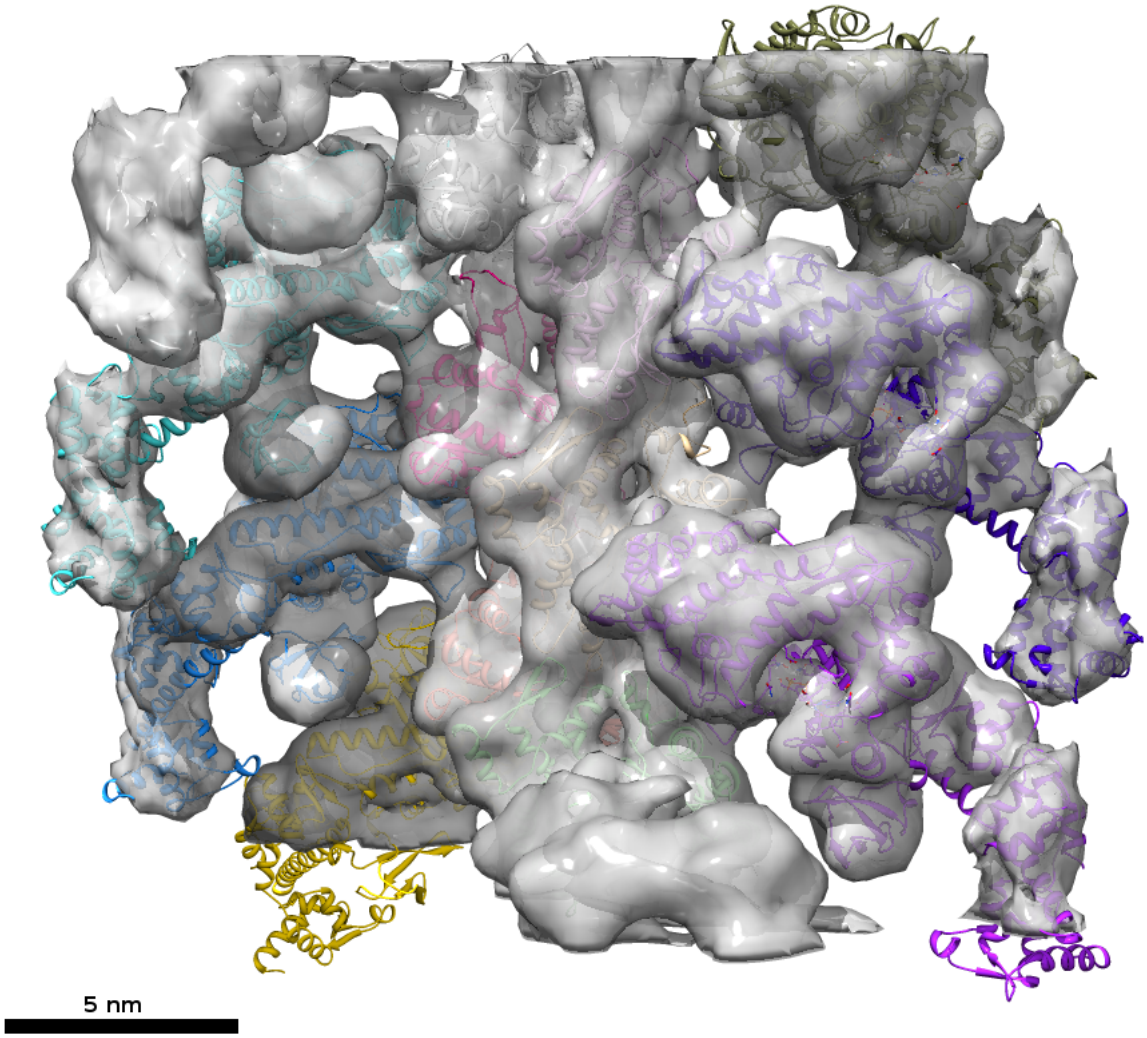


Figure 9.7: Simulation system for the Gaussian filtered density of the strong binding ADP state after 3 ns flexible fitting. The protein structures are illustrated by a ribbon representation of their backbone. Again the Gaussian filtering leads to an unexpected hole within the density in the nucleotide binding site presenting structural elements for which density is missing. In addition the lever arm at the junction between converter and light chain lacks density. Part of the helix is completely without density. Except for that the outline of the density resembles very well the shape of the actin and myosin monomers.

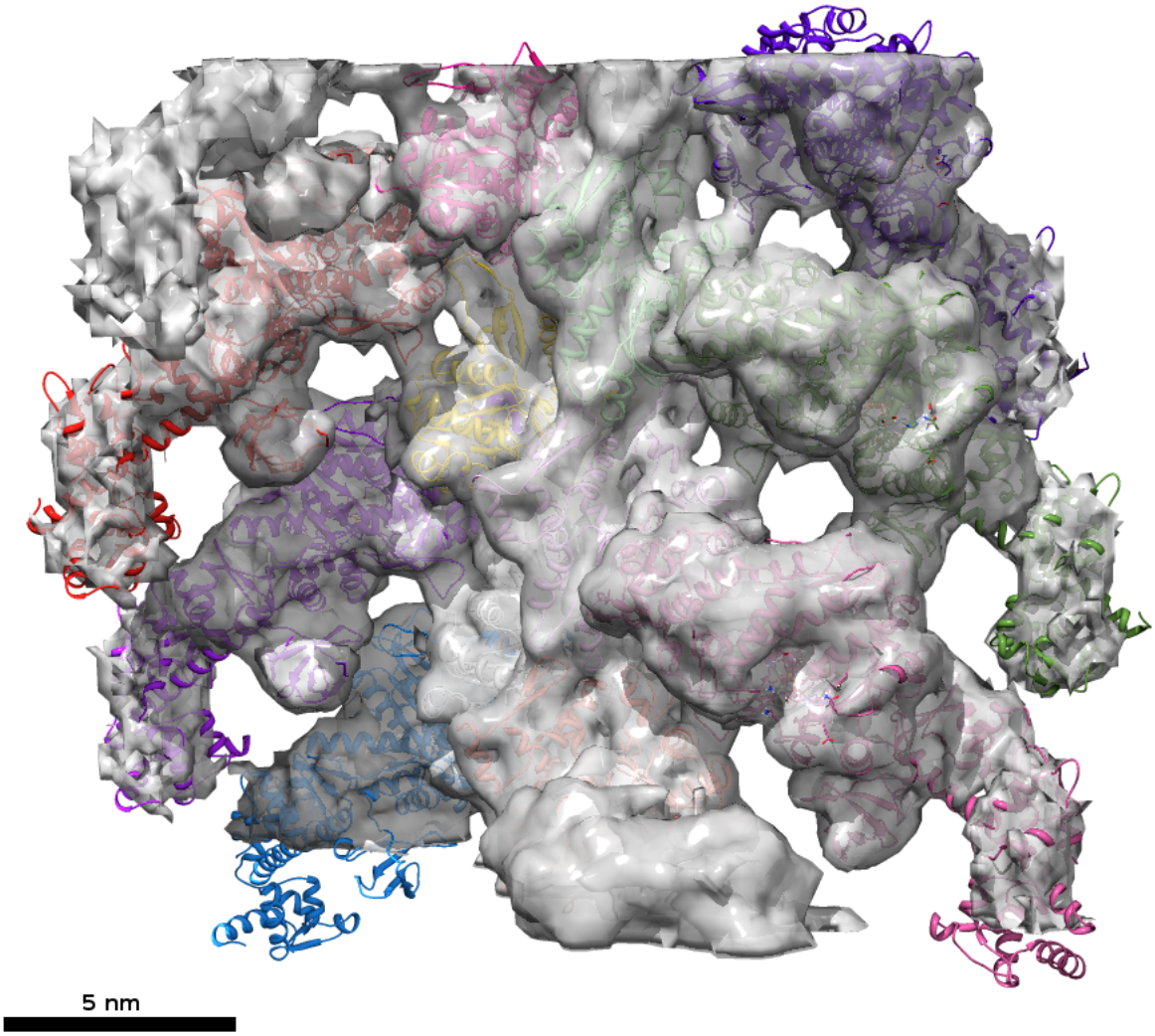


Figure 9.8: Simulation system for the B-factor corrected density of the strong binding ADP state of the acto-myosin complex. Shown is the last frame of a simulated trajectory for 3 ns. Again the outer shape of the density shows more refined contours just as for the B-factor corrected rigor density. The nucleotide binding side has density that covers its structural elements and also the lever arm is partially engulf by density. Most structural elements are covered and represented very well by the density. Despite the low density for the light chain its interaction has not been remodeled such as for the rigor state. Even though more parts are sticking out of the density, the position of the light chain seems to be accurate.

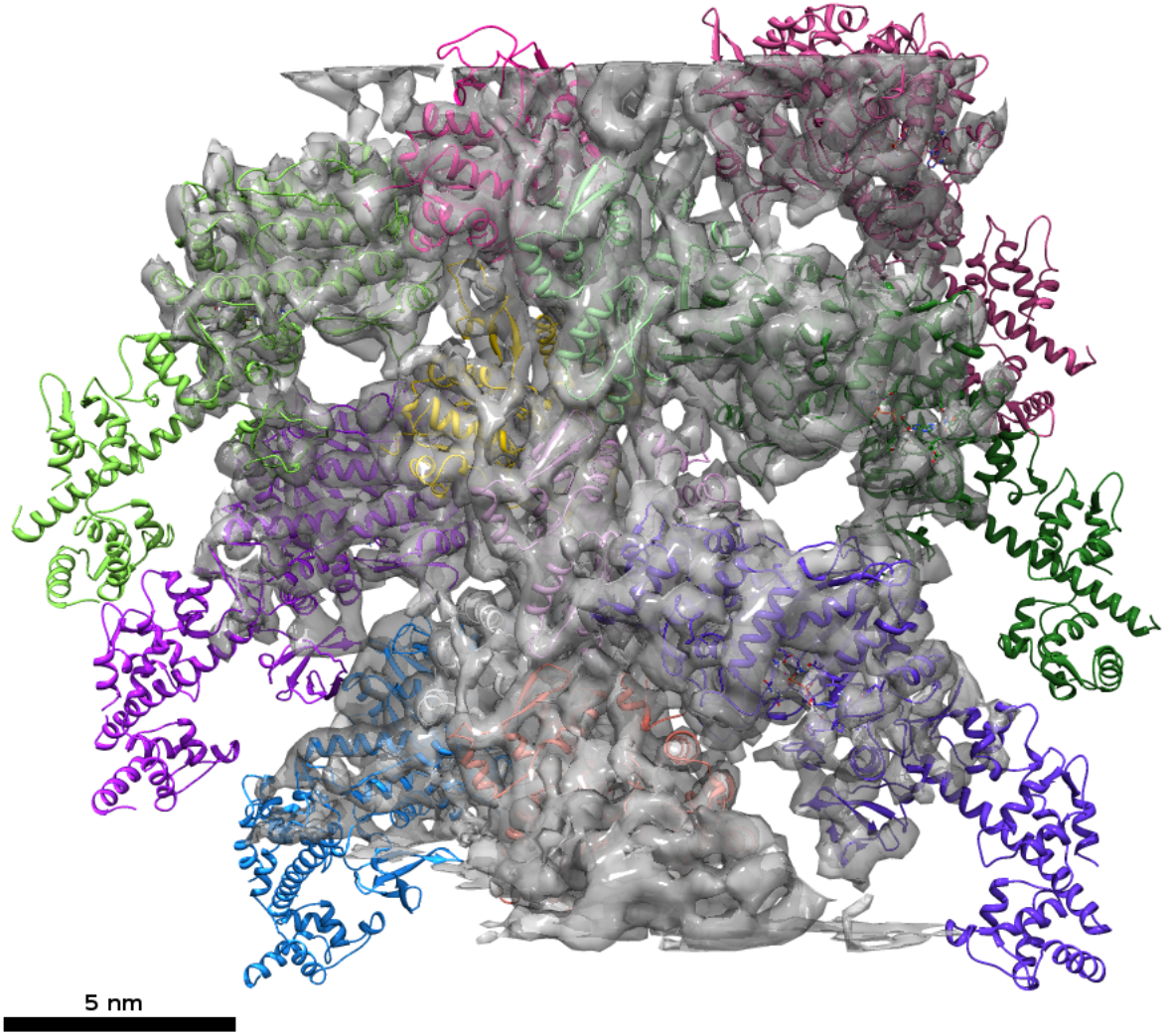


Figure 9.9: Simulation system with the hyperbolic tangent (tanh) filtered density of the strong binding ADP state after 3 ns flexible fitting. As for the rigor state the sharp filtering of the density allows further structural details to show up. The density of the light chain has the same problems as have been seen for the rigor state. It became quite noisy upon sharp filtering and does not show up in this representation where small density blobs are hided. Although both densities have comparable resolutions, the filtered representation has not the same quality as the rigor one. Some additional density protrusions can bee seen and at higher radius from the filament axis less density is present for the structural elements of the myosin. Nevertheless, the density defines the positions of the helices within actin and the main part of the myosin motor domain quite well and thereby again reinforces the validity of the obtained structure.

set arbitrarily. Each protein monomer is shown in a different color (arbitrary color code). All representations have been created using the program Chimera (Pettersen et al., 2004). For a description of the simulation systems see figure captions. The following analysis is focused on the structures the myosins converged to and investigates whether and how the two states differ. First they are compared to a crystal structure. Figure 9.10 shows the comparison between the final structures of both states and the nucleotide-free myosin V crystal structure (1OE9.pdb, Coureux et al. (2003)), that has been termed rigor-like by the authors, since it seems to be in the rigor conformation although it has been obtained without binding to actin. First of all the actin binding site in both states is comparable to

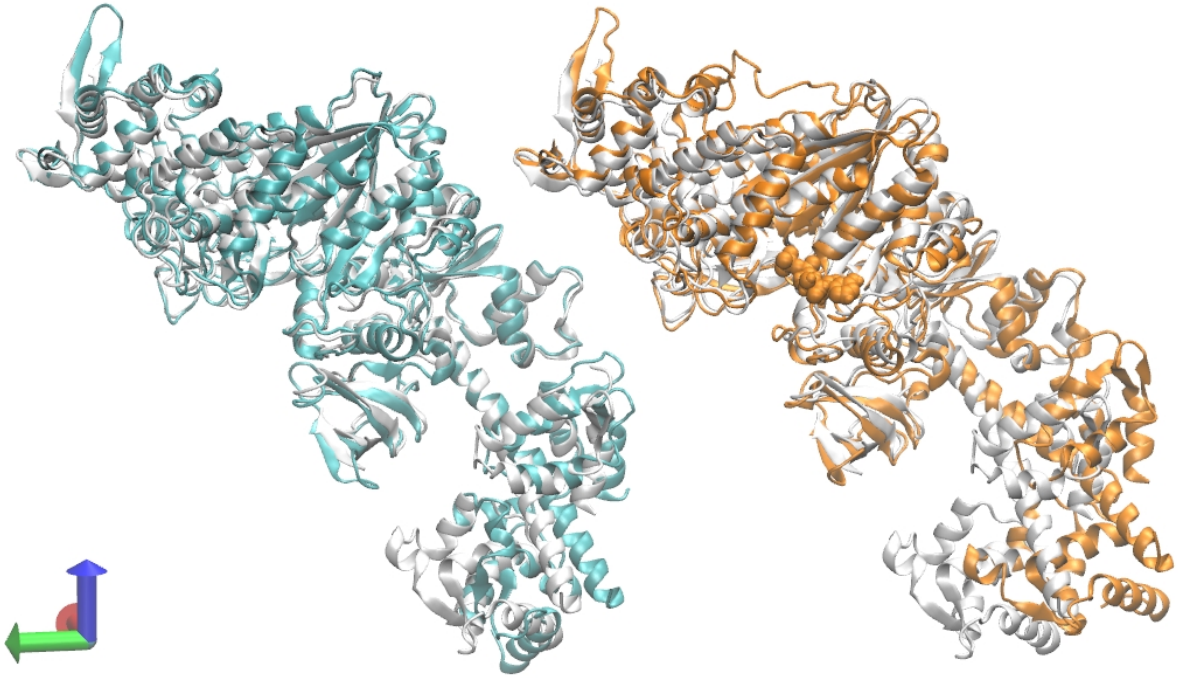


Figure 9.10: Comparison of the nucleotide-free myosin V crystal structure (pdb entry 1OE9, shown in white) with the flexible fitted structures: on the left compared to the rigor structure (cyan) that has been fitted into the hyperbolic tangent filtered density, and on the right compared to the structure of the ADP state (orange) also fitted into the hyperbolic tangent filtered density. The ADP molecule inside the nucleotide binding site is drawn in van-der-Waals representation. The orientation is indicated in the lower left corner: x,y,z-directions in red,green and blue. The z-direction coincides with the filament axis. Positive z-direction points to the pointed end of the actomyosin filament. The actin binding site is on the left and the light chain domain on the lower right. The images have been created using the program VMD.

that of the crystal structure. The actin binding cleft of myosin, that is known to close upon strong binding to actin, is already closed in the ADP state. The rigor structure shows throughout the whole molecule only slight deviations from the crystal structure. All important structural elements have

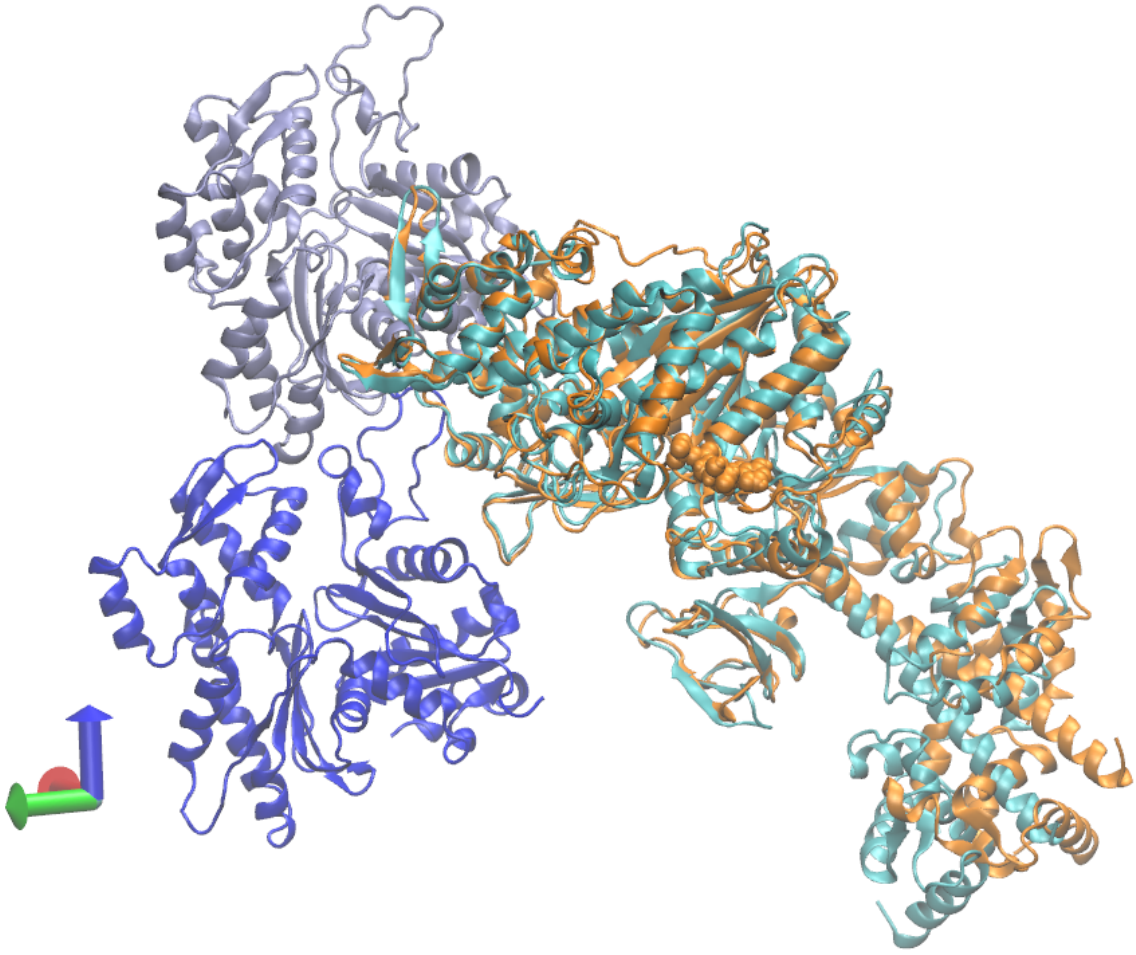
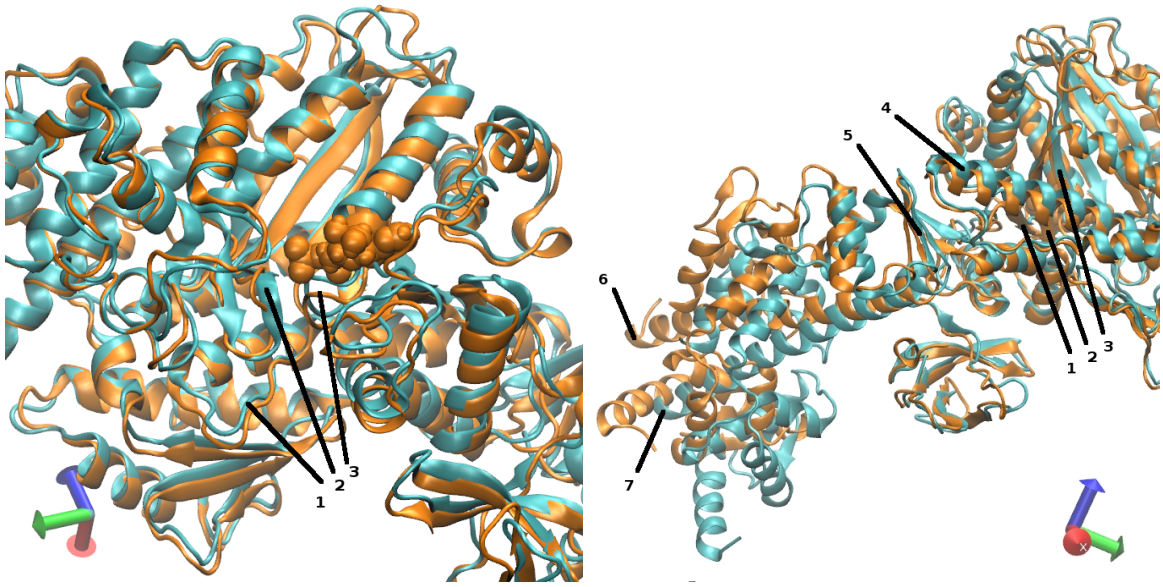


Figure 9.11: Ribbon representation of the rigor structure (cyan) and the strong binding ADP structure (orange) together with the two adjacent actin monomers (blue) they are interacting with. Both structures were flexible fitted into the hyperbolic tangent filtered densities. The blue coordinate again points towards the pointed end of the actomyosin filament. The image was created using the program VMD.

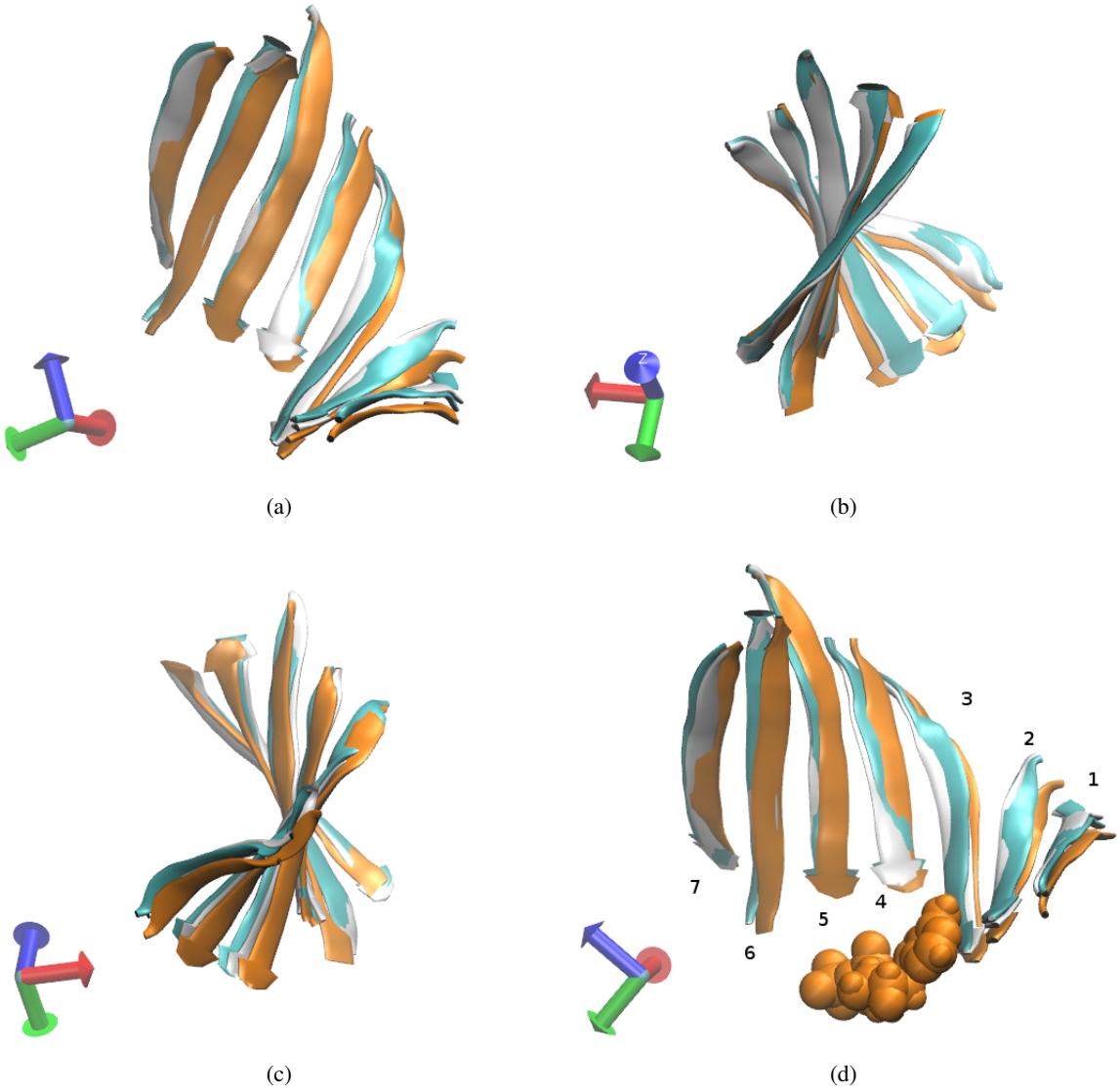


(a) Nucleotide binding site (b) Converter and lever arm

Figure 9.12: Detailed aspects of the myosin structures as shown in Fig. 9.11 (rigor state in cyan and ADP state in orange). The blue coordinate points towards the pointed end of the filament. (a) the structural elements of the nucleotide binding site differ only slightly. Switch 2 (indicated by number 1) and switch 1 (black number 2) are positioned quite similar. The p-loop (indicated by number 3), its adjacent helix and loops differ in their positions due to the nucleotide that is present only in the ADP state. Once the nucleotide is gone (rigor state) the p-loop moves closer to its adjacent helix. The central β-sheet of the myosin that is located in the back of this figure is going to be discussed in detail in Figure 9.13. (b) illustrates how small structural changes within the myosin motor domain are amplified. Number 1, 2, 3 label strands of the central β-sheet. Although the relay helix (4) is already straight, it has no longer a kink in its middle point, its position differs slightly in the ADP state compared to the rigor state. The β-strands behind it push the relay helix in the altered position, this in turn changes the three-stranded β-sheet of the converter (5) and consequently the lever arm exhibits larger deviations in its end points (6,7), since it is anchored in the β-sheet of the converter (5). The representations have been rendered in the program VMD.

almost the same positions, except for the distal C-terminal end of the lever arm and the light chain position. Since the beginning of the lever arm matches between rigor state and crystal structure very well, the altered position and interaction of the light chain might have led to the different position of the truncated end of the lever arm. The upper part of the light chain has the same interaction as in the crystal structure, only slightly shifted in space, but the lower part is remodeled. The helix-loop-helix motif on the lower left is folded up to come close to the lever helix, also the lowermost helix of the light chain right below the end of the lever arm was changed. Its direction and interaction with the end of the truncated lever arm are not the same as in the crystal structure. For the ADP state on the right it can be seen that going from the actin binding site towards the light chain the deviations from the crystal structure become larger. The lever arm is held in a position more upwards than in the rigor state, that corresponds to a powerstroke intermediate. Unlike the closing movements of the actin binding cleft the distal part of the myosin motor domain differs as long as the ADP nucleotide is kept bound. The details of the nucleotide binding site are discussed later. The interaction of the light chain is kept almost the same as it is seen within the crystal structure. It is a debatable point whether the proximity of the light chain to the next myosin monomer of the non-physiological decorated actin has resulted in the remodeled interaction in the rigor state. In order to avoid overloading the images the other structures fitted into the differently filtered densities are not shown. They differ slightly but overall show the same features. They are included into the analysis later on. Figure 9.11 shows the direct comparism of both experimental obtained states. Their similarities and differences can be seen, as just discussed. Their differences are highlighted in detail in Fig. 9.12.

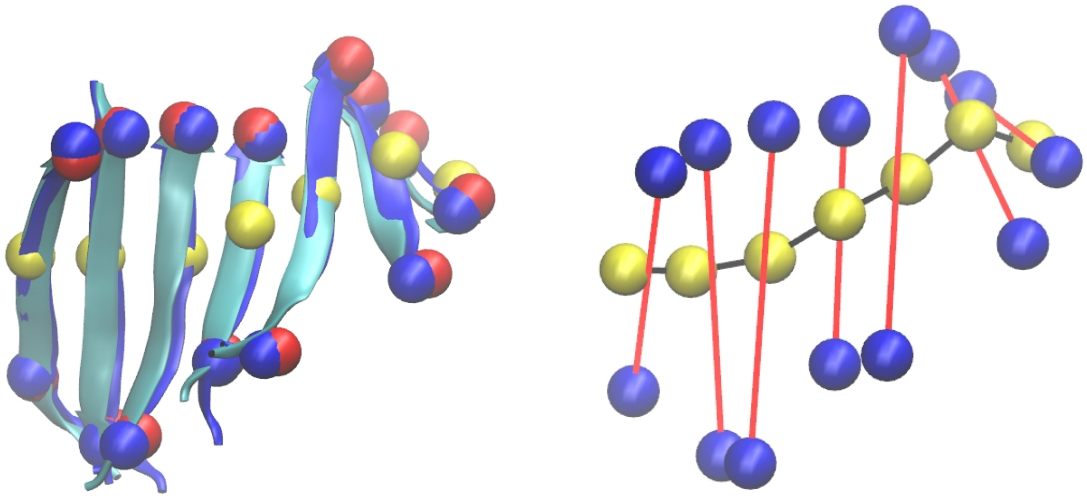
The central β-sheet of the myosin motor domain has been introduced in section 2.3.2 on page 13 as *transducer* as it integrates the structural changes within myosin. Therefore, it is going to be analyzed in more detail. Due to the differences of the relay helix and the lever arm, deviations of the strands of the central β-sheet may be expected. Figure 9.13 shows four different views of the so-called *transducer*. The β-strands of the rigor state and of the nucleotide-free crystal structure (rigor-like) are positioned quite similarly. In contrast, following the color change a different twist of the β-sheet can be identified for the strong binding ADP state. Not only the twist of the sheet is altered but also its bending. For the purpose of quantifying the deviations a simplified β-sheet model has been used as illustrated in Fig. 9.14. The vectors running along the β-strands through both end points are used to calculate the twist between the strands, while the polygonal lines are useful to give some indication of the bending of the β-sheet. The following C_α atoms of amino acid residues have been used for the end points of the strands: strand 1: residue 99 and 103, strand 2: residue 110 and 105, strand 3: residue 662 and 655, strand 4: residue 162 and 157, strand 5: residue 438 and 431, strand 6: residue 221 and 228, strand 7: residue 242 and 236. For the middle positions following C_α atoms have been chosen:



(a) (b)

(c) (d)

Figure 9.13: Different views of the central β-sheet, the so-called *transducer* of the myosin motor domain. Again the structures are shown that have been fitted into the hyperbolic tangent filtered densities, the rigor structure in cyan, the ADP state in orange and the nucleotide-free crystal structure (pdb entry 1OE9) in white. The coordinates are the same: x (red), y (green) and z (blue). Positive z-direction points towards the pointed end of the filament. Because the structural states differ in the presence or absence of the nucleotide, in figure (d) ADP is shown in its proximity to the transducer. Strand 4 is connected to the p-loop, strand 5 to switch II and strand 6 to switch I. All three structural elements are involved in coordinating the nucleotide. Additionally, the presence of ADP alters the position of the first three β-strands, that in turn push onto the relay helix that is located behind them. The ADP state shows an altered twist and a modified bending of the transducer. The scenes were rendered by the program VMD.

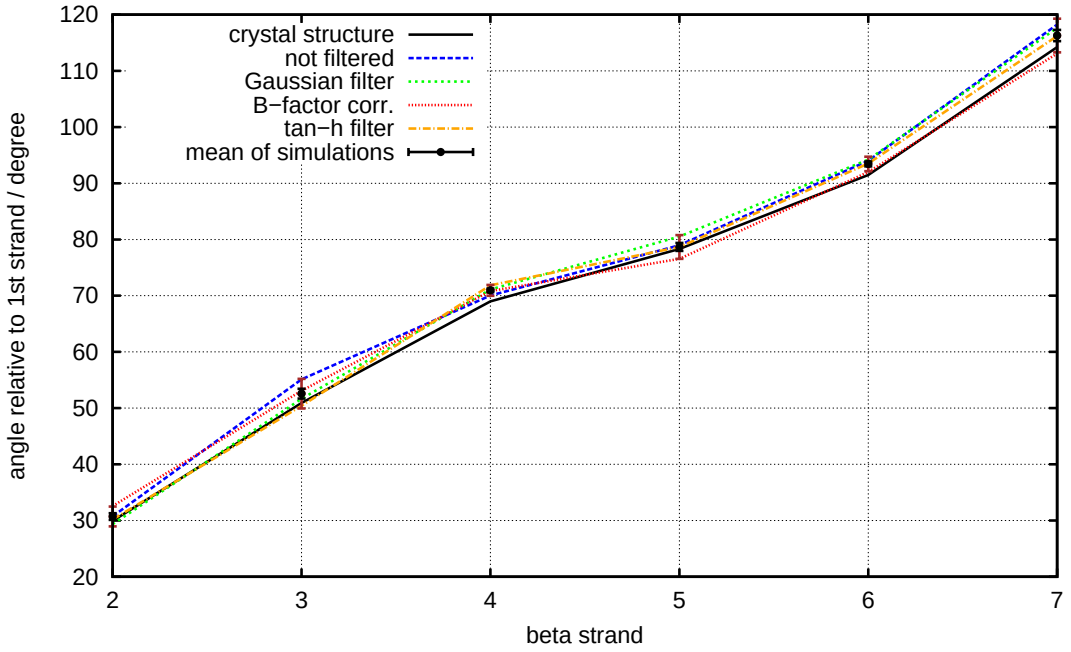


(a) coordinates of the β-strands

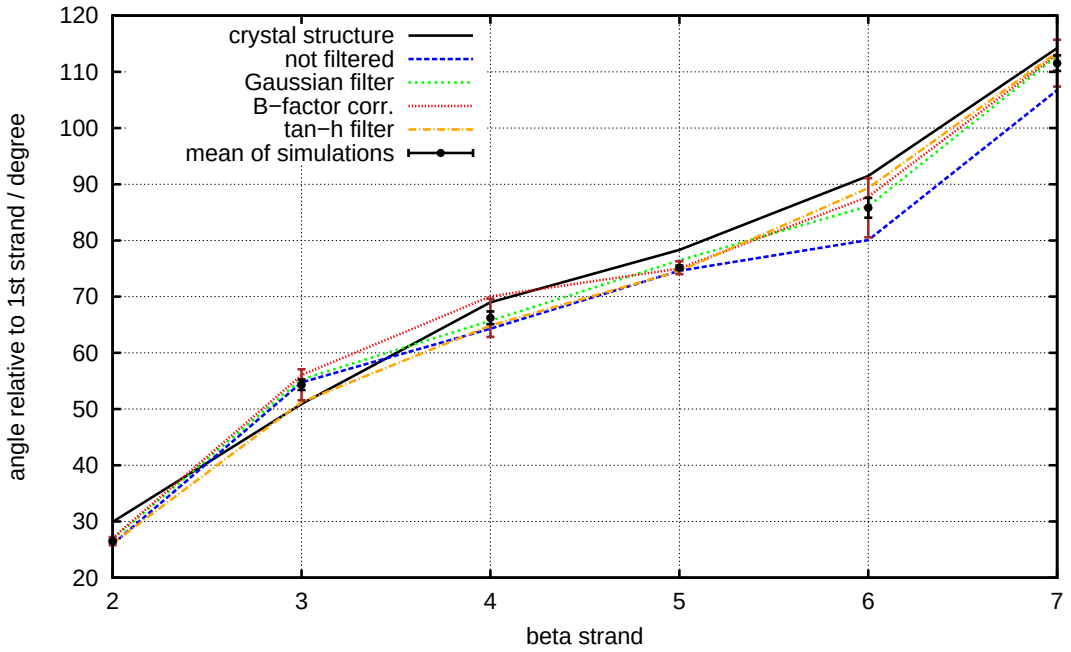
(b) vectors and polygonal lines

Figure 9.14: β-sheet model for the analysis of the deviations of the β-strands in the different structural states. The blue and red spheres indicate the coordinates of the C_α atoms of the end points of each strand. The yellow spheres indicate the middle points approximately, likewise they belong to C_α atoms of the myosin backbone. (b) shows the simplified model where each strand is described by a vector through both end points. Additionally, its middle points are connected by polygonal lines (scenes rendered by the program VMD).

strand 1: residue 101, strand 2: residue 108, strand 3: residue 659, strand 4: residue 160, strand 5: residue 435, strand 6: residue 224, strand 7: residue 239. Each angle has been calculated using the scalar product of a pair of vectors $(\vec{a}, \vec{b})$ and is given by $\theta = \arccos[(\vec{a} \cdot \vec{b})/(|\vec{a}||\vec{b}|)]$. Separately for the two states the twist has been analyzed both relative to the first β-strand and relative to the previous strand. The results are shown in Fig. 9.15 and 9.16. The bending of the β-sheet has been assessed by calculating the angles between the polygonal lines connecting the middle points of the β-strands. Figure 9.17 illustrates the resulting bending angles relative to the first polygonal line. For each structural state the four simulation systems with the differently filtered densities have been evaluated. Furthermore, the mean angles of the four simulation systems per state have been calculated. They are plotted as well together with their standard deviation and allow to recognize significant deviations of the twist and bending of the β-sheet for the ADP state compared to both the rigor state and the nucleotide-free crystal structure. Although the myosin analyzed by protein crystallography was not bound to actin, the β-sheet of the rigor state features no obvious deviations from the crystal structure. Indeed they seem to describe the same structural state. Despite some deviations between the fittings in the differently filtered densities the mean angles for the twist and bending are comparable to the angles of the crystal structure within the error margins. The twist of strand 4 relative to strand 1 deviates in all simulation

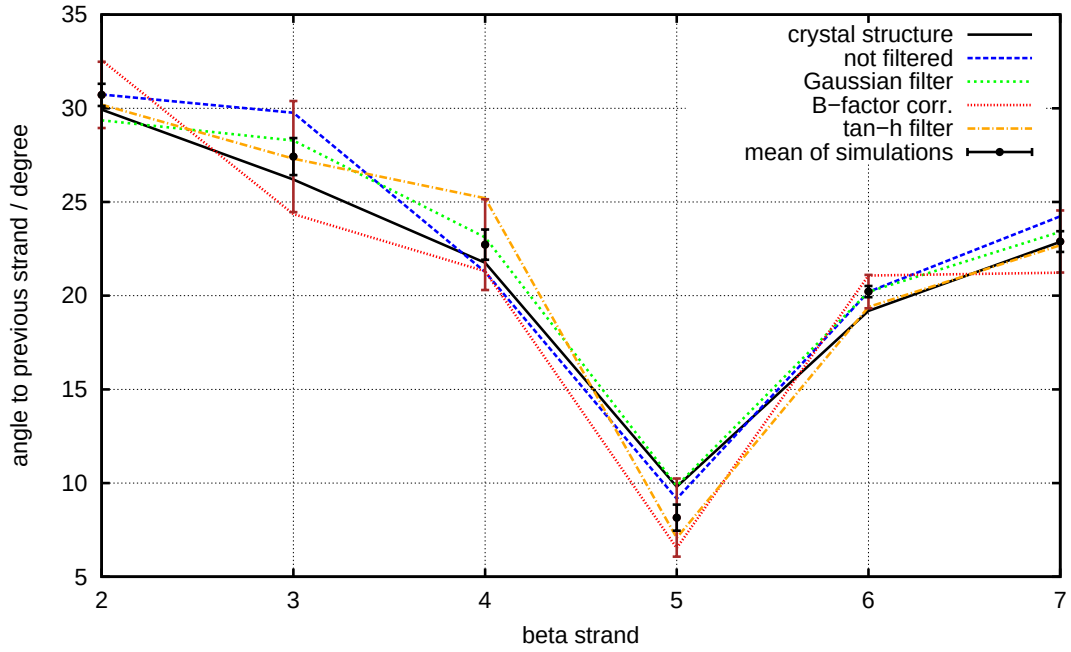


(a) rigor state – features no obvious deviations from the crystal structure

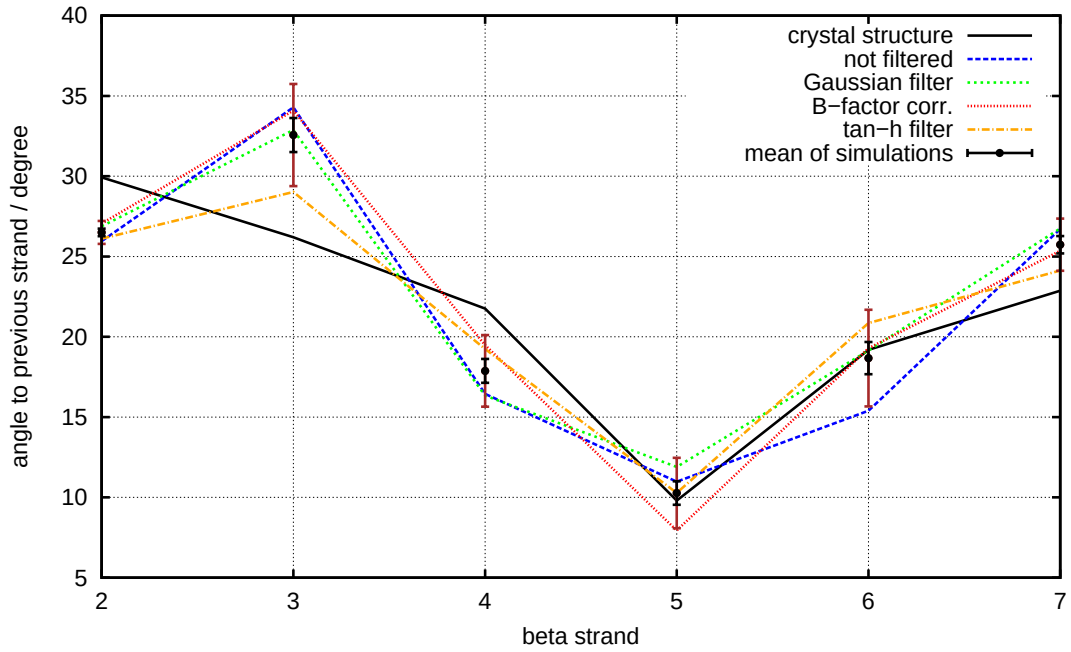


(b) ADP state – shows significant deviations from the crystal structure

Figure 9.15: Twist between β-strands of the central *transducer* of the myosin motor domain. The angles are calculated relative to the first strand. All structures fitted into the differently filtered densities have been evaluated. The mean angle of the simulations is shown with errorbars in black considering one standard deviation ($\pm\sigma$) and in brown considering three standard deviations ($\pm 3\sigma$).

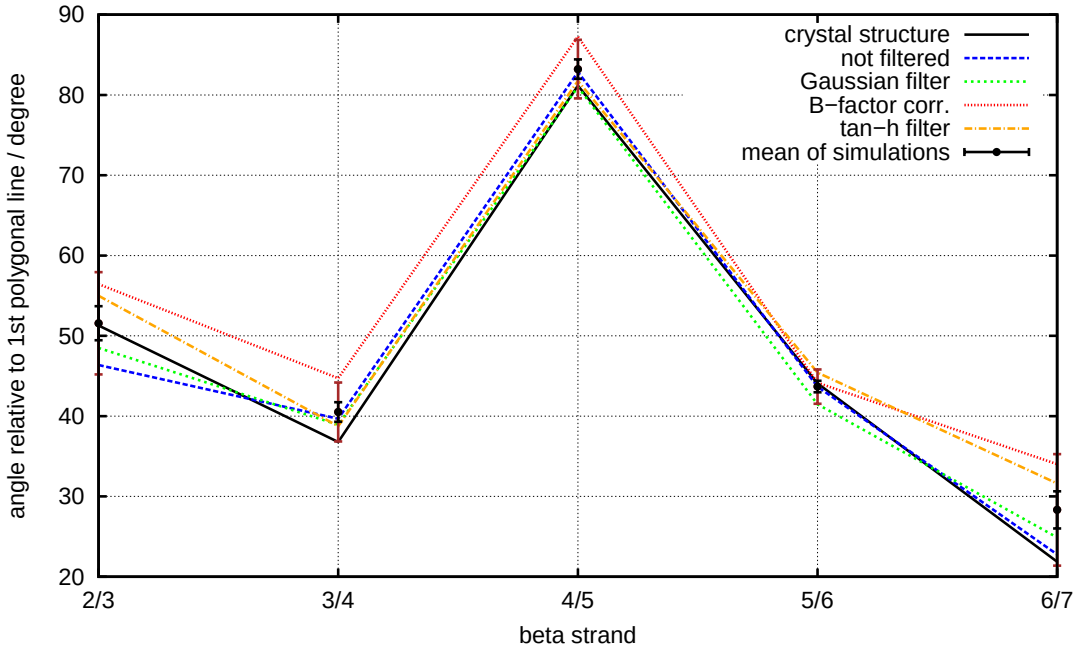


(a) rigor state – agrees with the crystal structure within error margins

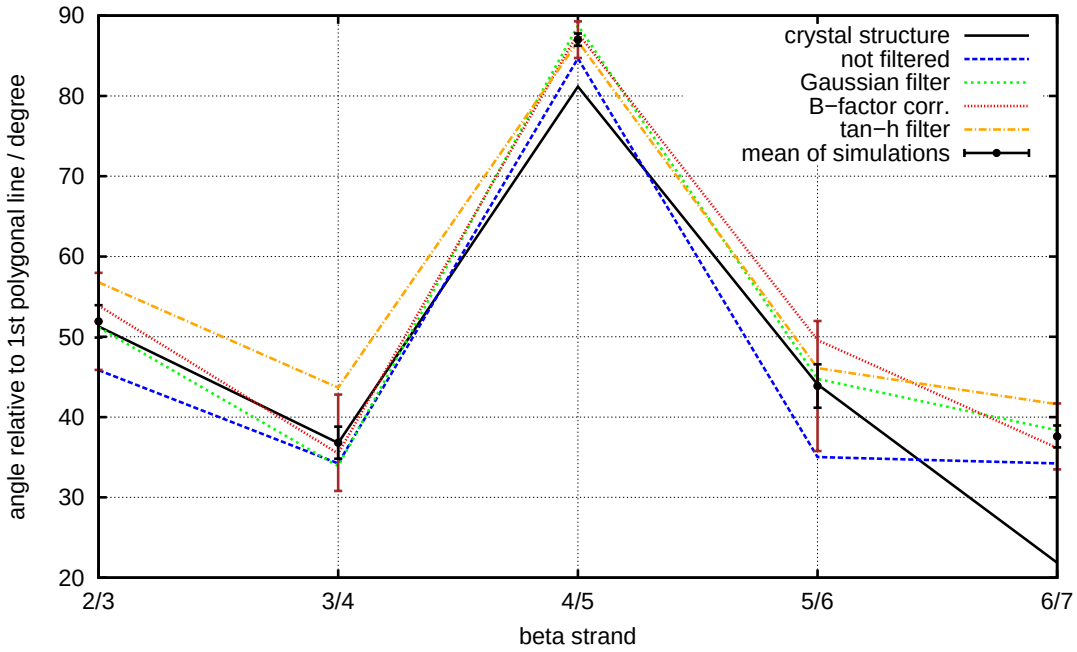


(b) ADP state – shows significant deviations from the crystal structure

Figure 9.16: Twist between β-strands of the central *transducer* calculated relative to the previous strand. The mean angle of the simulations is shown with errorbars in black considering one standard deviation ($\pm\sigma$) and in brown considering three standard deviations ($\pm 3\sigma$). The ADP state shows significant deviations from the crystal structure.



(a) rigor state – resembles within error margins the crystal structure



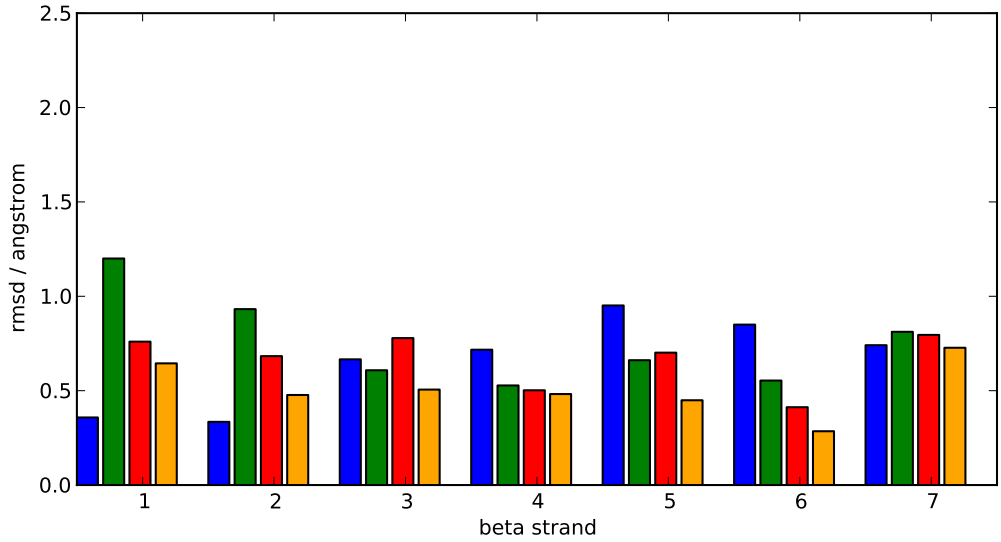
(b) ADP state – clearly differs from the crystal structure

Figure 9.17: Bending between β-strands of the central *transducer* calculated relative to the first polygonal line that connects strands 1/2. The mean angle of the simulations is shown with errorbars in black considering one standard deviation ($\pm\sigma$) and in brown considering three standard deviations ($\pm 3\sigma$). The bending of the *transducer* of the ADP state clearly differs from the crystal structure.

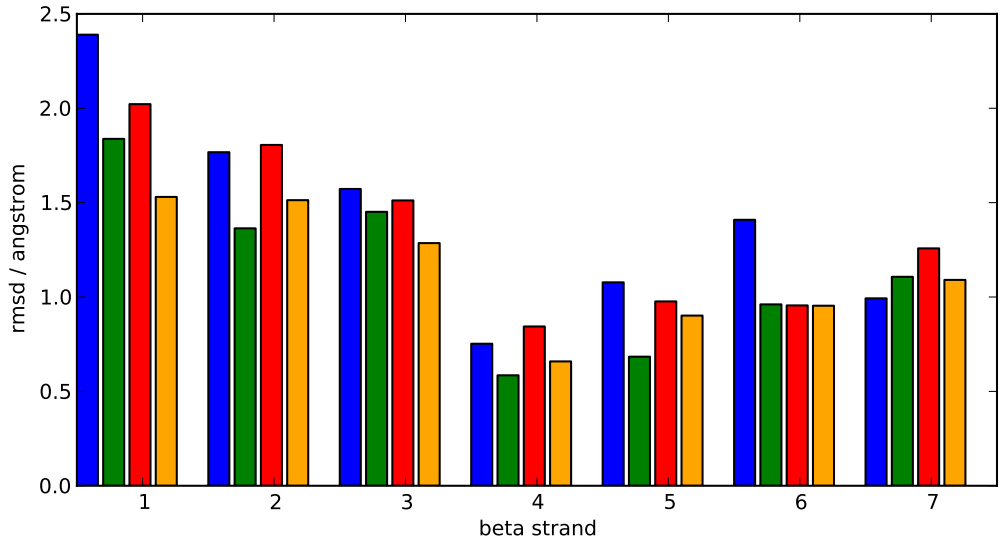
systems slightly from the angle found for the crystal structure. The ADP state features significant deviations both for the twist angles and the bending angles. The twist between strand 2 relative to strand 1 is significantly lower compared to the crystal structure (26.5°±0.2° vs. 29.9°). The same is true for the twist between strand 5 relative to strand 1 (75.1°±0.4° vs. 78.3°). The twists between strands 2/3 and strands 6/7 are significantly higher (32.6°±1.0° vs. 26.2° and 25.7°±0.5° vs. 22.9°), while the twist between the strands 3/4 is significantly lower (17.9°±0.7° vs. 21.8°). The deviation in the bending of the β-sheet is given by a significantly higher bending angle between the polygonal lines connecting strands 1/2 and 4/5 (87.0°±0.8° vs. 81.2°) as well as between the polygonal lines connecting strands 1/2 and 6/7 (37.6°±1.4° vs. 21.9°). The errors of the mean angles have been obtained from the standard deviation of the measured data and the number of measurements N: $\sigma_{mean} = \sigma/\sqrt{N}$, N=4. Additionally to the twist and bending analysis, the root mean square deviation of the β-strands relative to the nucleotide-free myosin V crystal structure has been evaluated, the results are shown by bar plots in Fig. 9.18. Only the C_α atoms of the protein backbone were included into the evaluation. The crystal structure has been aligned to each structure in order to calculate the minimal deviation. The structures have been used as they were fitted into the densities without further movements. Using only one aligned reference the deviations have been even larger. For the rigor state the deviations are rather small, whereas particularly the first three strands of the ADP state exhibit larger deviations.

The analysis of the differences of both structural states shall be concluded by the assessment of the amplification of the observed structural changes, also known as movement of the lever arm: the changes of the *transducer* amplified at the end of the truncated lever arm. Figure 9.19 shows the relative positions of the helix of the lever arm. The light chain is not displayed. In the upper left of (a) the set of curves can be seen the structures converged to in the differently filtered densities. The small spheres indicate the coordinates of the C_α atoms of residue 767 and 789, that have been used to evaluate the angle of the lever arm. (b) points out the different positions of the lever arm due the different filterings of the densities. The mean lever arm angle between the APD structures and the nucleotide-free myosin V (NF myosin V) crystal structure is given by 17.2°±2.0°. The mean angle for the rigor structures compared to the crystal structure is 11.9°±0.9°. Figure 9.19 however shows that the angular deviations are pointing in different directions. The mean lever arm angle between the ADP- and rigor states accounts for 14.8°±1.3°.

Below follows the examination of some of the simulation parameters to estimate how reliable the results are or whether they are strongly dependent on the set of chosen parameters. Figure 9.20 shows the comparison for one simulation system with two differently chosen parameters. The backbone of

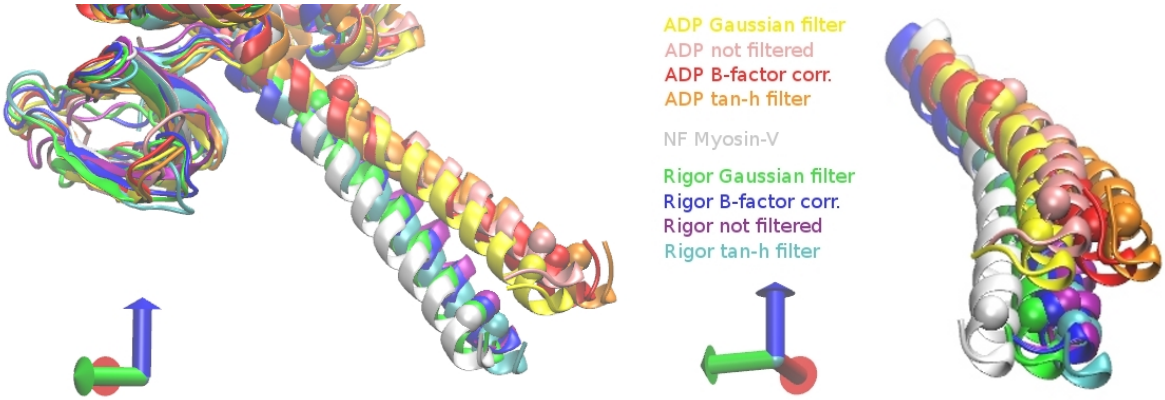


(a) rigor state



(b) ADP state

Figure 9.18: Root mean square deviation (rmsd) per atom of the β-strands relative to the nucleotide-free myosin V crystal structure. Only the C_α atoms of the strands have been included into the calculation. Shown are the results from the flexible fittings into the differently filtered densities: blue:not filtered, green: Gaussian filtered, red: B-factor corrected, orange: tanh filtered.



(a) (b)

Figure 9.19: Lever arm positions of the flexible fitted structures into the differently filtered densities. Again the blue coordinate points towards the pointed end of the filament. The light chain is not displayed to allow the visualization of the C_α atoms that were used to evaluate the angular deviation. (Representations created in VMD.)

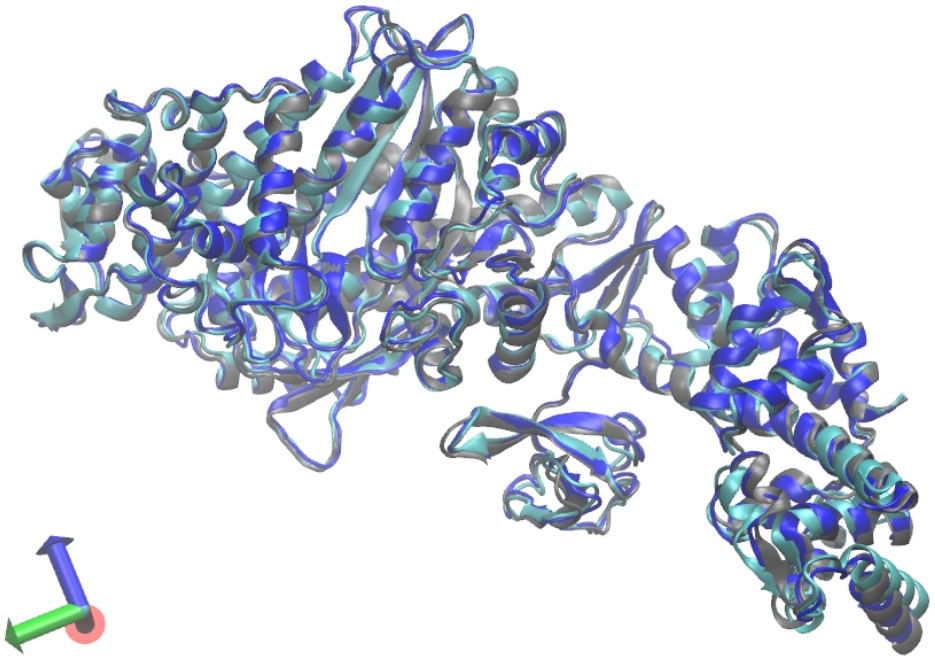
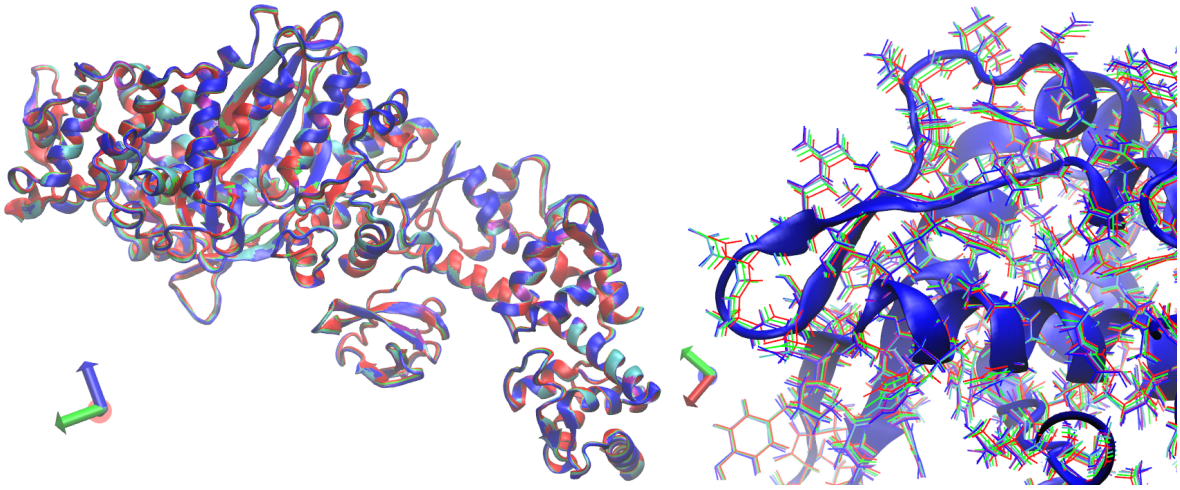


Figure 9.20: Influence of simulation parameters. Shown in blue is the rigor structure after 3 nanoseconds flexible fitting into the B-factor corrected density, in cyan the same but using a higher scaling factor $\xi = 0.5$ for the density during the simulation and in grey a structure with weaker secondary restraints (half the values than before) during the fitting. Representation created in VMD.

the myosin converged to almost the same positions except for some loops and the light chain. The influence of the simulation parameters is actually lower than the effect of the different filterings of the density. Therefore, this examination confirms the former results and indicates a high reliability. Furthermore, the effects of energy minimizations at the end of the simulations using higher scaling factors ξ has been examined. The results are presented in Fig. 9.21. The backbone is unaffected by



(a) backbone of the whole myosin (b) closer look at side chains

Figure 9.21: Effect of energy minimizations at the end of the simulation. Again shown in blue is the rigor structure after 3 nanoseconds flexible fitting into the B-factor corrected density without minimization. Cyan: with additional minimization using $\xi = 1$, purple: $\xi = 2$, green: $\xi = 5$, red: $\xi = 10$. Each minimization was done for 2000 steps. (Representations created in VMD.)

the energy minimization except for some loops. A closer look at the side chains show, that they are slightly repositioned as they are pushed by higher force into the density. As long as only the backbone is going to be interpreted and evaluated the energy minimization can simply be left out. Or the other way around: the energy is already such good without energy minimization that the backbone can be interpreted accurately.

Completing the ADP structure is going to be compared with the structure that was flexible fitted by MDFF into the old “rigor” state that turned out to be almost completely an ADP state as well. Figure 9.22 shows the overlay of both structures. The two structures are quite similar, showing once more the ADP state character of the old density. The variations found inside the nucleotide binding site have not been seen both for the new ADP state and the rigor state. If the old density is a mixture of both states or almost only the ADP state then it could be expected that this structure shows the same

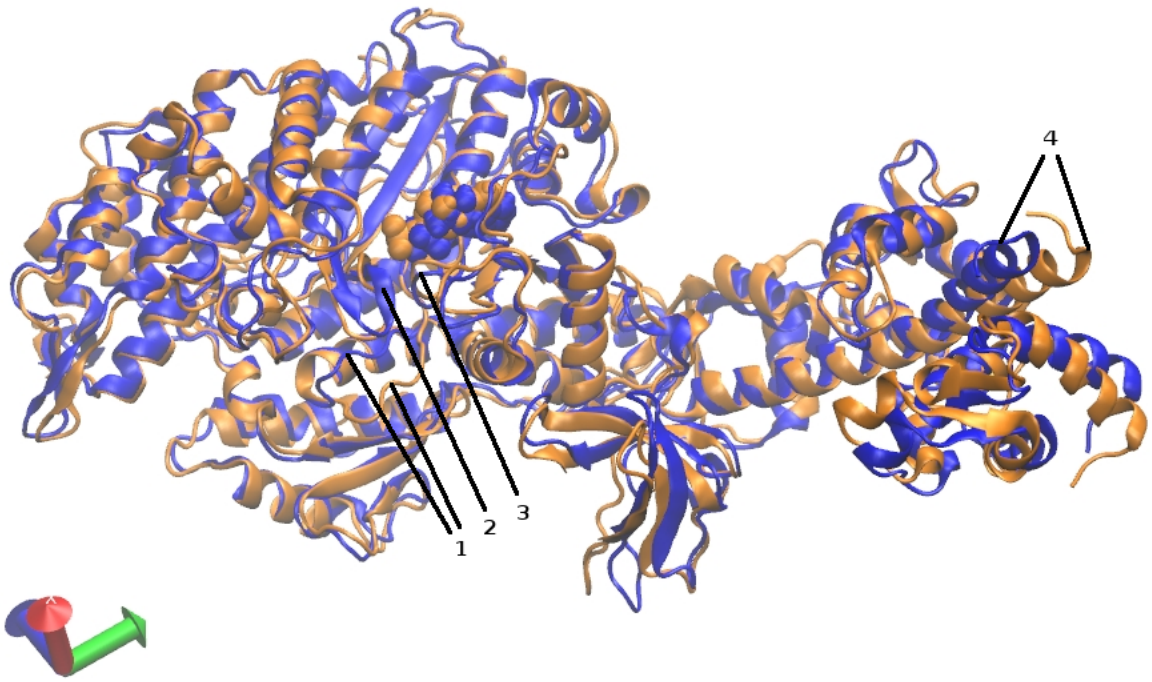


Figure 9.22: Overlay of the new ADP state (orange) and the structure from the old “rigor” density (blue), that essentially is more or less an ADP state. Again the new ADP state is the one, that has been fitted into the tanh filtered map. The old density has not been filtered and only the phases have been corrected. The orientation is rotated such that the structural elements of the nucleotide binding site can be seen. (1) points at the different positions of the switch 2 element. (2) Switch 1 shows only partially the same position, the adjacent helix and its connection with it are differing. (3) The p-loop and the nucleotide are positioned not similarly. (4) The end positions of the truncated lever arm are varying slightly, that has already been seen before for the different filterings of the density. The light chain is slightly repositioned as well. (Representation created in VMD.)

features as the two other states show. Thus, it is questionable whether the buffer conditions stabilized a different intermediate state or if the variations are without significance. Next the rigor state is compared to the recently published rigor state of myosin Ie with reported 8 Å resolution (Behrmann et al., 2012a). Their complex included in addition to actin and myosin also tropomyosin, that is known to regulate muscle activation. Figure 9.23 shows the myosin Ie structure (cyan) aligned to the myosin V rigor structure (blue) together with two actin monomers. The alignment of the structures was per-

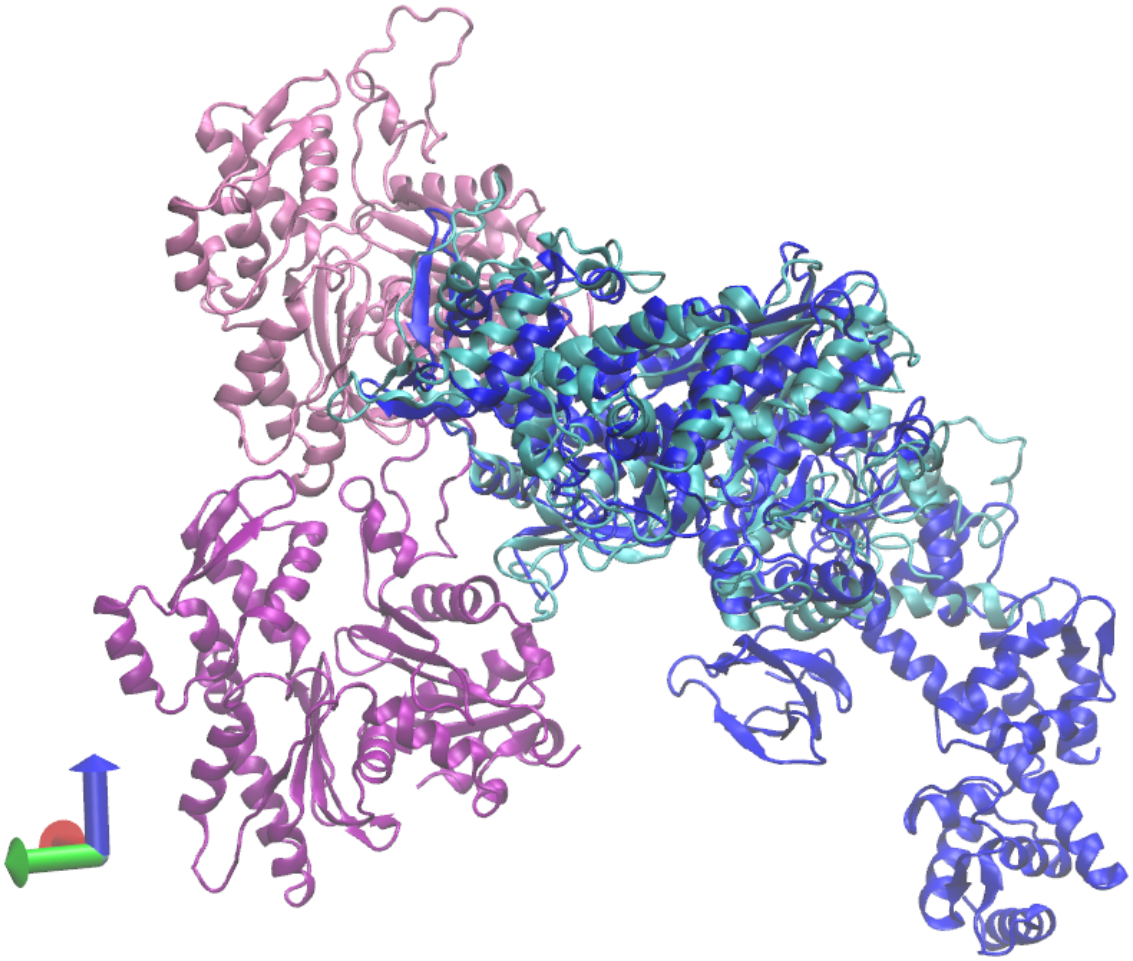


Figure 9.23: Comparison of the myosin V rigor structure (blue) with the myosin Ie rigor structure (cyan, pdb entry 4A7L). The two adjacent actin monomers are shown as well. The myosin V structure was flexible fitted into the B-factor corrected rigor density. Representation created in VMD.

formed in the program VMD using identical and similar parts of the amino acid sequence of both myosins. Despite being more compact without structural elements like the lever arm, the light chain

and the N-terminal β-barrel the myosin Ie shows a lot of differences in the positioning of its helices. Having such big variations it is actually hard to compare them. Also the interface between actin and myosin might not be the same. Apart from that, the amino acid sequences are quite different and it is unclear to what extent it effects the structure. Thus, the comparison is inappropriate to characterize the newly obtained myosin V rigor state based on similarities, but the differences indicate that most likely the actin-myosin interface with its side chain interactions differs, because not only the helices and a β-hairpin motif close to actin are positioned closer to actin for myosin Ie than for myosin V but also myosin Ie has extended loops near the actin-binding site. Owing to the presence of tropomyosin for myosin Ie it cannot be decided whether the repositioning of structural elements is only an effect of tropomyosin or the myosin interaction with actin might differ in general.

9.2.2 Simulations with solvent

In order to assess the quality of the simulations that have all been done in vacuo so far, it is essential to know about the influence of the missing solvent, or the other way around to examine the effect of the presence of solvent. Only if there aren't large variations the states obtained from simulations in vacuo are really reliable. For one simulation system an implicit solvent simulation has been performed. So far the first 1.4 ns have been calculated. Figure 9.24 shows the comparison between in vacuo- and implicit solvent simulation. Due to a longer computing time, the implicit solvent calculations require, only a short run structure is presented here. The result should be considered as preliminary result only, nevertheless it is obvious that the backbones of the structures are almost identical except for some loops and the light chain. As loops and light chain have not been interpreted so far, it is once more a confirmation of the validity of the backbone structures obtained and interpreted before. It is interesting whether the structures will offer still more identical elements as soon as the implicit solvent simulation converges or if some different positions will be found. So far the p-loop shows an odd conformation.

Simulations including explicit solvent resemble the natural protein's environment best. However they require immense computing power and time. Figure 9.25 presents a sketch of the simulation system and serves at the same time as outlook. It would be nice to obtain the result of an explicit solvent simulation for at least one simulation system. So far a water box has been added to the protein complex and the system has been neutralized by sodium chloride ions (Na^+Cl^-). Using a computer with 128 processors (Intel Xeon E5-4650L) equipped with 2 terabyte memory, parallel computation can be accomplished with threads instead of a Message Passing Interface, a trajectory of the first 167 picoseconds could be simulated. Using this computer non-stop for about one month the system might converge. A conservative estimate about how many processors might be useful for parallel computation using

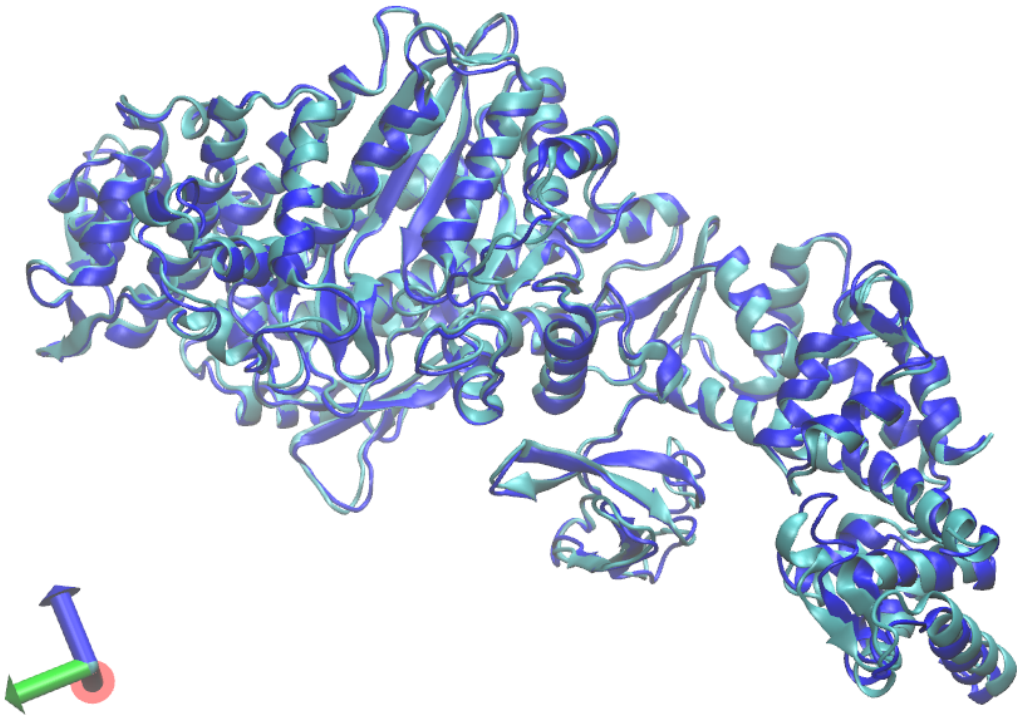


Figure 9.24: Implicit solvent simulation vs. in vacuo simulation, exemplarily shown for the rigor structure that has been fitted into the B-factor corrected density: blue: 3 nanoseconds flexible fitting in vacuo, cyan: 1.4 nanoseconds preliminary flexible fitting with implicit solvent. (Scene created in VMD.)

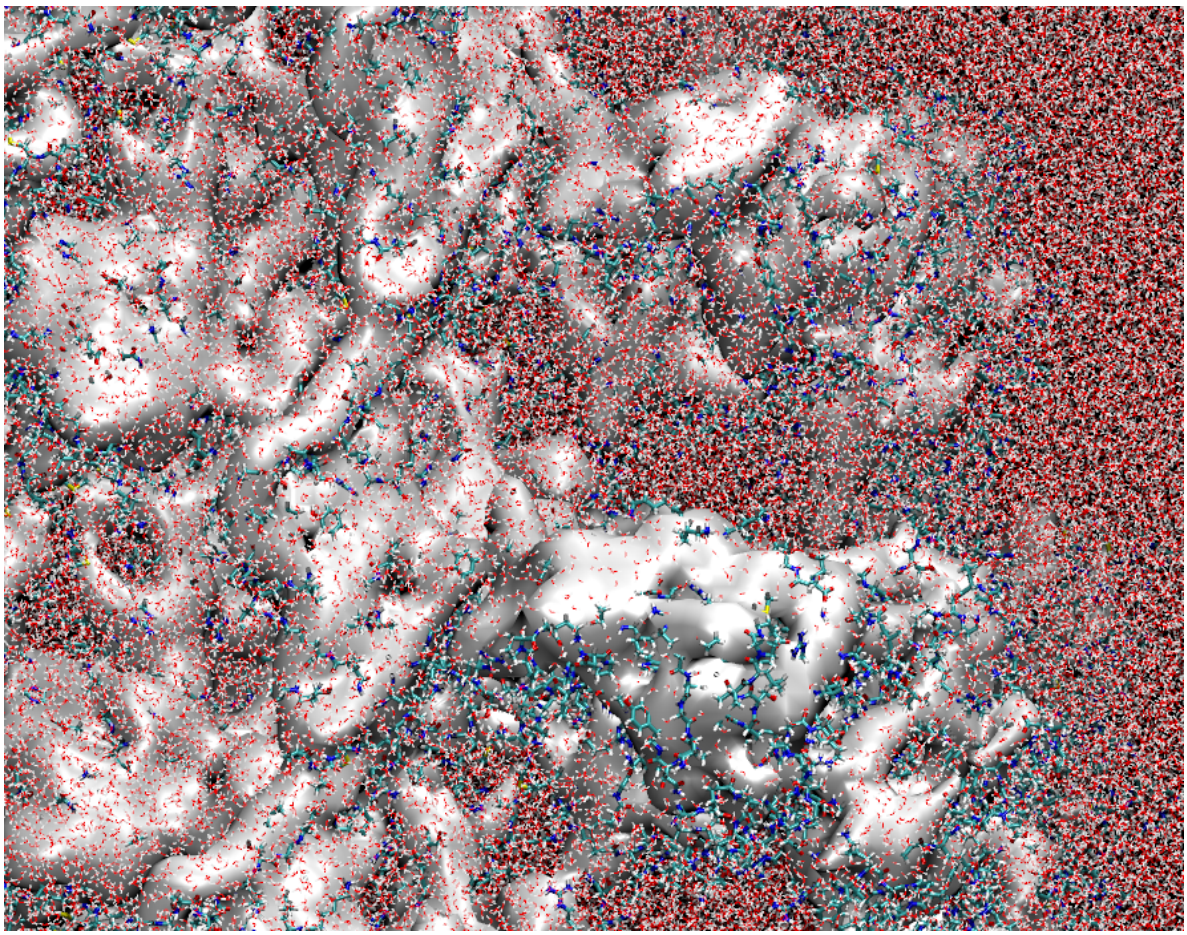


Figure 9.25: Simulation system including explicit solvent. The threshold of the density is chosen such to show some of the amino acids. The tiny red and white lines are the water molecules. The whole system consists of 1.8 million atoms. The proteins alone account for only 125000 atoms.

the software NAMD is given by 1000 atoms per processor. Newer estimates already state a value of 100 atoms per processor, since the software is built for a highly parallel usage. In most cases the infrastructure might be limiting. Using 128 processors for about 1.8 million atoms there have still been 14000 atoms per processor. In principle a computer cluster offers more processors, but until now the simulation does not run successfully on the cluster. The simulation using 128 threads resulted in an amount of 20 GB random access memory that is needed. It might even be necessary for each node using the cluster, depending on how the software was developed. Such modeling has unfortunately to be postponed to future studies.

9.3 Discussion

Molecular dynamics flexible fitting of crystal structures into the electron microscopy densities was successfully used to gain more detailed insight into the structural states of the actomyosin complex. Despite some calculations that still should be done, namely more variations of the simulation parameters, solvent simulations and simulations starting from other published crystal structures of myosin V, in order to obtain a complete assessment of the findings, the main results are without any questioning valid. The most important findings are: firstly the rigor structure confirms the nucleotide-free crystal structure (PDB entry 1OE9, Coureux et al. (2003)) to be in rigor state as almost the same structure was obtained in this work. Secondly the strong binding ADP state of actomyosin was characterized for the first time. The myosin offers almost the same actin binding site with a fully closed actin binding cleft as seen in the rigor state, but the twist and bending of the central β-sheet, the *transducer*, differs, leading subsequently to an intermediate position of the lever arm that indicates an incomplete power-stroke as long as the nucleotide is kept bound. The position of the light chain is problematic because it has been remodeled in the rigor state. The remodeling of the position is unlikely since both the rigor-like and post-rigor crystal structures show the same conformation for the converter / light chain interaction and light chain conformation. Considering that the light chain density for the rigor state is more complete than for the ADP state, it is questionable whether a non-physiological complex of decorated actin might have influenced the remodeling due to steric hindrance by the nearby myosin. Flexible fitting into the “rigor” density calculated from an initial data collection yielded almost the same structure as for the new ADP state and confirms the assumption that it is to a great extent or even completely an ADP state. It is interesting to note that the nucleotide binding site of that reconstruction differs in a way which has not been seen for the rigor and ADP state before. Therefore, further analysis is necessary to evaluate its significance.

Part III

Discussion and Outlook

Chapter 10

Towards Bayesian information retrieval

The algebraic least squares reconstruction method follows a strict mathematical deduction in the framework of linear imaging theory to solve the general inverse problem of 3D reconstructions. Its origins go back to the beginnings of 3D reconstructions from projection images obtained by transmission electron microscopy. The reconstruction approach describes a direct inversion of the set of projection equations without any iterations or other approximations. The solution of the ill-defined equation system is given by means of a least squares determination. During this dissertation a mathematical elegant matrix notation was obtained. Within the preceding diploma thesis the advantageous properties of the least squares method were demonstrated, i.e. a reconstruction of the sought object with better defined density localization than using a standard weighted backprojection – albeit assuming ideal oriented test data. Thus, the method was applied to the actomyosin complex with the intent to optimize information retrieval and thereby enhance the interpretable resolution of the reconstructed 3D object. Setting up the projection matrix and subsequently inverting it in the subspace of significant eigenfunctions is a computationally immense task. In order to limit the computational costs and thereby ease its applicability, the problem has been reduced from 3D to 2D. The current implementation was developed for an axial tomography geometry where 2D reconstructions as "slices" along the axis are calculated. The stacking of these "slices" then yields the 3D solution. A complete data processing workflow was established beginning with the image preparation prior to the reconstruction and ending with the 3D alignment of single filament segment volumes to each other to finally obtain the solution to the whole data set by averaging all aligned segment volumes. The application to a data set of the actomyosin complex displayed severe problems of the method due to the dimensional reduction of the reconstruction process. As a consequence the filaments are treated as ideally oriented axial tomograms. Absolute values have to be assigned to a set of parameters that will not be refined in later steps: the pixel size, the symmetry angle, the helical rise and hence likewise the relative orientation of parts of the helices to each other. This is necessary because the symmetry parameters are used during

read-in of the projection data for the least squares reconstruction.

Electron tomography investigations of filaments without the need to average over non-identical samples helped to clearly show the flexible properties of the actomyosin filaments. Adhesion to the support carbon film and filament-filament contacts lead to a flexible adjustment of the filaments that is accompanied by bendings out of the projection plane and wrapping around other filaments. Even the regular helical repeat pattern is disturbed. Such adaptability to obstacles certainly is of great value within living cells and organisms. On the other hand it clearly demonstrates how inappropriate the description of the filaments as axial tomograms can be. Later on it was estimated that less than 10% or with a higher amount of filaments even less than 5% of the filaments were lying within the projection plane. The large out-of-plane tilt distribution poses a massive drawback to the applicability of the current implementation of the least squares method. Without the possibility to treat out-of-plane tilt angles even the pixel size was not estimated correctly, consequently leading to an incorrect data read-in even for filaments that are lying within the projection plane. The problem of estimating the correct pixel size and symmetry angle seems to be solvable compared to considering the flexible properties of the filaments within the least squares reconstruction method.

As first step to reduce ad hoc assumptions the iterative helical reconstruction method IHRSR was used. In this method the symmetry angle and helical rise are treated as free parameters and are subject to an iterative refinement. The reconstruction is based on a normal weighted backprojection that allows to treat out-of-plane tilts also. The pixel size has been corrected using the known helical rise as calibration. A subsequent repeat of the whole reconstruction process has yielded a much improved 3D density, compared to the reconstruction attempts using the least squares method. Only this new density allows for a biological interpretation. A conceivable approach to make use of the benefits of the least squares reconstruction is to try a completely new implementation without the dimensional reduction to 2D. If the 3D projection matrix can be set up and subsequently solved despite its enormous size, then the iterative helical method can be used as basic reconstruction combined with a subsequent 3D least squares reconstruction. In this procedure the more accurate density localization could be used as final improvement.

The iterative helical reconstruction method needs less ad hoc assumptions compared to the current implementation of the least squares method, but its set of free parameters has to be chosen carefully. It is indeed unknown whether the chosen values are the most suitable. Critics mention an arbitrary tuning of free parameters might lack objectivity or might be biased. In the worst case a reconstruction might describe noise instead of the underlying signal (known as overfitting) if wrong parameters have been

chosen. Noisy features can even be enhanced during iterative refinement procedures (Scheres, 2012b). However, statistical reconstruction approaches reduce the necessary heuristic ad hoc assumptions and parameter tunings to a minimum. Only recently a statistical image processing and reconstruction approach based on Bayes' law emerged for 3D EM structure determination, while in other fields it is already widely used. First publications using cryo-EM data begin to show its potential. In the Bayesian approach most parameters of the statistical model are learned iteratively from the data. The reconstruction problem is solved by "finding the model that has the highest probability of being the correct one in the light of both the observed data and available prior information" (Scheres, 2012a). In order to find a unique solution, external or prior information is used to complete the experimental data, that in general do not completely determine the structure. The smoothness of a cryo-EM reconstruction is used as prior information in the approach of Scheres (2012b). It is incorporated into the refinement as regularization. The method is called maximum a posteriori (MAP) or regularized likelihood optimization and might be the next step for future enhancements of the interpretable resolution in 3D electron microscopy. A Bayesian information retrieval combined with helical symmetry parameter search would be promising. So far the statistical approach has not been combined with a helical parameter estimation and like other methods it depends on an accurate correction of the contrast transfer function (CTF).

However, the CTF correction addresses a huge problem of the data sets. Even the best algorithm cannot accomplish much if the CTF correction does not succeed. Zero-loss energy filtering during microscopy leads to an additional filter contrast that affects mainly the first CTF zero position (Angert et al., 2000). If the acquired images lack an interpretable CTF imprint in the power spectra, showing only the first maximum and not even the first zero position is clearly defined by the next maximum, then the CTF correction can only be a poor guess. This leads most likely to an improper correction of the higher spatial frequencies and thus restricts the achievable resolution. The strength of the CTF imprint depicts the quality of the instrumental equipment. In our case a no longer state-of-the-art detector had to be used (Gatan UltraScan 1000 CCD camera) which has a relatively high noise level and a small 2k x 2k detection area. Lower noise level in the data clearly would improve the 3D reconstruction from cryo-EM images as has recently been shown for the new generation direct electron detectors. For instance, Fernández et al. (2013) report a B-factor of 155 Å^2 using a FEI Polara microscope at 300 kV equipped with a FEI Falcon direct electron detector. Veesler et al. (2013) report even a B-factor of only 100 Å^2 using a FEI Krios microscope at 300 kV together with a FEI Falcon direct electron detector. Compared to B-factors about 350 Å^2 determined in this work. Thus, the availability of improved detectors – in particular for the otherwise excellent FEI Krios microscope here in Heidelberg – is quite important beside a good reconstruction algorithm to obtain the best possible result.

Chapter 11

New insights into the Actomyosin complex

Iterative helical reconstruction combined with molecular dynamics flexible fitting was successfully used to resolve functional differences of the general myosin macromolecular structure during the myosin crossbridge cycle. While the rigor state of the complex of actin and myosin has been investigated before using other myosins, the strong binding ADP state of the complex of actin, myosin and ADP was analyzed for the first time for the now achieved structural resolution of about 8 Å. Rigor densities have already been reconstructed to a resolution of 14 Å for muscle myosin II (Holmes et al., 2003) and with 8 Å resolution for myosin Ie in a complex combined with tropomyosin (Behrmann et al., 2012a). In this work, myosin V has now been investigated in two different binding states both strongly bound to actin. Its unique kinetics, tuned for "single" molecule processive movement along actin filaments, offers the possibility to study its predominantly populated state, the strong binding ADP state, that is only transiently present for other myosins. Despite a high noise level in the TEM images and a weak imprint of the microscope's contrast transfer function (CTF) in the data, due to a detector not being state of the art, the iterative helical reconstruction could achieve density representations of the actomyosin complex with resolutions ranging from 10 to 7 Ångström. While the actin density has a resolution of about 7 Å, both for the rigor state and for the strong binding ADP state reliable 8 Å resolution can be stated for the myosin motor domain without taking problems with the light chain into account – which most likely reflect partially ill-defined biochemical conditions during the sample preparation. A resolution gradient was detected resulting in the highest resolution for the central actin density and lowest resolution for the light chain density at highest radius from the filament axis. The number of individual actin and myosin molecules used for the reconstructions corresponds to approx. 259000 and 243000 for the rigor state and the strong binding ADP state, respectively.

Interestingly, the reconstructions of different nucleotide states converged to slightly different symmetry angles. The ADP state reflects exactly the angle that a crystallographic 28/13 helix possesses: $-167.143°$, while the rigor state converged to a slightly higher filament twist angle of $-167.405°$. This indicates also a difference of the actin-myosin interface, which needs to be investigated further.

Molecular dynamics flexible fitting allows the fitting of protein crystal structures into the TEM densities via molecular dynamics simulations while maintaining the integrity of secondary structure elements via appropriate restraints. Both actin and myosin have been solved via X-ray diffraction with resolutions of about 2 Å. Because the complex of both could not be crystallized so far, the hybrid approach of transmission electron microscopy combined with MD simulations provides the opportunity to study the functional complex close to the resolution of protein crystallography. The crystal structures of actin and myosin have been fitted flexibly into four differently filtered density representations for each structural state. For MD simulations the densities are interpreted as potential distribution thus the density gradient acts as force to drive the structures into the density. Due to the fact of its non-physical character the force may differ using different density outlines due to filtering.

Indeed, the resulting structures, obtained from MD simulations in vacuo into the differently filtered density representations, differ slightly but show the same overall features. Depending on the filtering, the density outline led to varying positions for some flexible loops and the light chain. However, tests of the influence of simulation parameters, energy minimizations and first results from implicit solvent simulations showed how reliable the protein backbone structure has now been determined (cf. figures 9.18, 9.20, 9.21 and 9.24). Detailed analysis of the influences of density filtering resulted in the finding that filtering actually doesn't affect the functional interpretation of the backbone structures of the myosin motor domain. In order to interpret side chain interactions as well, the kind of filtering should be re-evaluated and some more simulations still have to be done. Additional to larger variations of simulation parameters, solvent simulations and simulations starting from other published crystal structures are needed. Nevertheless, the key findings – based on the myosin backbone – are valid.

Comparison of the obtained rigor structure with the nucleotide-free myosin V crystal structure (PDB entry 1OE9, Coureux et al. (2003)) yields almost complete agreement of both. On the one hand the result reconfirms the rigor character of the crystal structure of myosin, that was obtained without being bound to actin, on the other hand it illustrates the quality of the used method of this work and its results. Deviations between both can only be seen for the light chain region that is going to be discussed below. The strong binding ADP state offers significant deviations from the nucleotide-free

myosin V crystal structure. While the actin binding site seems to be in almost the same configuration as in the rigor state, exhibiting a fully closed actin binding cleft, the twist and bending of the central β-sheet, the so-called *transducer* (Coureux et al., 2004), differs, leading to subsequently larger deviations towards the lever arm. Thus, the present investigation was able to visualize for the first time that the *transducer* occupies different states in the actomyosin complex. The deviations in the twist of the β-strands relative to the first or previous strand are ranging between $-4°\pm1°$ to $6.5°\pm1°$. The bending of the β-sheet was analyzed by polygonal lines connecting the middle points of the β-strands. The difference in bending compared to the rigor state is given by $6°\pm1°$ between the beginning and the middle of the β-sheet and by $16°\pm1°$ between the two ends of the β-sheet. The values were obtained by averaging the results of the simulations with the differently filtered densities. The observed conformation of the central β-sheet leads subsequently to an intermediate position of the lever arm that indicates an incomplete powerstroke as long as the nucleotide is kept bound. Although the structural elements close to the actin binding site are already close to rigor conformation, the presence of the nucleotide drives the central β-sheet into a conformation that pushes onto the already straight relay helix. Its slightly different position lets the lever arm take an intermediate position. The angular change of the lever arm was calculated to $15°\pm1°$ between rigor and ADP state. **Thus, myosin accommodates strong binding to both actin and ADP by a strained conformation of the central β-sheet**. Comparison with other crystal structures still need to be done just as MD simulations starting from different structural states. If they converge to the same final states it would be an additional validation of the results.

The quality of the central actin density is quite high and almost comparable to the result of Fujii et al. (2010) who have analyzed undecorated actin filaments and obtained the most recent f-actin model. A comparison of the newly obtained rigor state with the one Behrmann et al. (2012a) have reported using myosin Ie and tropomyosin proved to be difficult because even helices close to the actin-myosin interface differ in their position like many other helices of myosin. The amino acid sequences also differ in many residues that certainly affects the degree of structural analogy. A comparison of a myosin Ie crystal structure (PDB entry 1LKX, Kollmar et al. (2002), with an almost closed actin-binding cleft) with the newly obtained myosin V rigor structure showed as well a reduced structural analogy. However, helices close to actin are positioned in a similar distance to actin both for the 1LKX structure and the myosin V rigor state, which are positioned closer to actin in the myosin Ie rigor state obtained from hybrid microscopy. A comparison of the rigor actin-myosin interfaces of myosin Ie and myosin V is problematic owing to the presence of tropomyosin for myosin Ie. It can hardly be decided whether the repositioning of structural elements is only an effect of tropomyosin or the myosin interactions with

actin might differ in general. Likewise, the assessment of the actin-myosin interfaces for the myosin V rigor and ADP state is difficult, because, on the one hand, the slight deviations in helical symmetry indicate differences, otherwise the myosin backbones seem to be almost identical and different side chain interactions cannot be interpreted as long as the results from solvent simulations are still missing.

The position of the light chain is problematic because it has been remodeled during the MD simulation especially in the rigor state. While the beginning of the lever arm is still identical to the crystal structure, deviations occur only towards the C-terminal end and within the light chain interaction and conformation. Thus, neither does its origin seem to lie within the motor domain nor does it appear to be an artifact introduced only by simulation, as already the density for the light chain is less well resolved. Additonally, the ADP state offers less density for the light chain. Different reasons might account for both shortcomings including structural disorder of the sample, rotational misalignment during image processing or biochemical effects during the preparation. However, structural disorder in the sense of a mixture of different states and rotational misalignment would affect not only the light chain. Filtering of the densities shows depending on the strength of filtering the light chain density to appear or disappear; it is affected as a whole not gradually. Thus, structural disorder and rotational misalignment might be excluded. Classification of the data would be the method to test for heterogeneity. But it can only be productive, if one assumes that further resolution is hidden inside the data due to heterogeneity of the complex. However, without much doubt the poor signal-to-noise ratio, the weak CTF imprint and its correction attempts are restricting the resolution within the obtained data sets.

Less density for the light chain in the ADP state compared to the rigor state could be explained by a partial loss of the light chain during biochemical preparation as it has been discussed by Holmes et al. (2003) for the regulatory (second) light chain in myosin II. The preparations for the rigor- and ADP state differ in their ionic strengths. Apparently, kinetic data about the impact on the non-covalent binding affinity of the light chain to the lever arm is missing. It should be mentioned that an optical difference exists between strongly defocused filaments of both states. While rigor filaments have usually shown nice arrowhead patterns, “ADP filaments” offered a different impression. The remodeling of the rigor light chain might therefore have a different reason: Rigid body docking of the nucleotide-free crystal structure – that otherwise showed to be almost identical to the newly obtained rigor structure – resulted in a conformation wherein the light chain of one monomer comes quite close to the junction of the end of the relay helix and the converter of the next monomer. The non-physiological array of myosins forces charged and polar amino acids into distances in the range of 1–5 Å. The rigor density

of the light chain clearly has a different position than the docked light chain of the crystal structure. Thus, it is questionable, whether the proximity to the next monomer or even steric hindrance led already to a repositioning of the light chain in the sample solution. Possible external influences would represent a conceivable explanation for the fact of otherwise identical structures.

Two more data sets were initially recorded at the JEOL microscope in Osaka, which was used for the 6.6 Å f-actin reconstruction of Fujii et al. (2010). They consist of a larger "rigor" set and a quite small set presumably in the "ADP" state. They had to be put aside due to the evidence that the biochemical preparation might have led to a kinetically impure mixture of structural states. Flexible fitting into the "rigor" density yielded almost the same structure as for the "ADP" state. This confirms our assumption that all data is to a great extent or even completely in an ADP state. The iterative helical reconstruction converged to a symmetry angle of $-167.27°$ that corresponds to the mean value of the other two new densities. The signal-to-noise ratio of the acquired data is better due to a superior detector quality and furthermore the reconstructed density for the light chain is resolved better than in the new ADP state derived from the FEI Krios (Heidelberg) data. The small data set was obtained from an ADP state preparation and it is actually the third preparation variant that led seemingly to an ADP state density. It is interesting to note, that – from the comparison of all the reconstructed densities – the ADP state is favored strongly by the myosin V, i.e. it is going to populate this state as long as the buffer conditions don't explicitly prohibit it. The flexible fitting into the "rigor" density resulted in a structure like the new ADP state but features conformations in the nucleotide binding site that have not been seen for the new rigor and ADP state. It will therefore be highly significant to analyze well defined biochemical states at higher resolution in the future.

Chapter 12

Summary

The key achievements of this thesis are as follows: A new algebraic least squares reconstruction algorithm was applied to an actomyosin data set obtained by cryo transmission electron microscopy. The method was first investigated for model data in the preceding diploma thesis and turned out to have the potential to optimize information retrieval from projection images compared to a standard weighted backprojection method, because it allows for a better defined density localization in the reconstructed object density. The method was extended by a new kind of numerically necessary filtering, the implementation started already during the diploma thesis. In addition, during the dissertation, a new mathematical matrix notation could be derived which nicely illustrates how the reconstruction is built of significant eigenfunctions. Furthermore, a complete data processing workflow from the filament selection to the 3D alignment of the reconstructed volumes has been implemented (cf. chapter 6).

However, thorough analysis combined with electron tomography studies (cf. chapters 6 and 7) showed that flexible specimen properties limit the applicability of the current implementation of the least squares method. In future, a new implementation without a dimensional reduction from 3D to 2D will be necessary to exploit its otherwise advantageous properties, without being limited by ad hoc assumptions of the orientations of the filaments or their symmetry parameters.

Therefore in this work, iterative helical reconstruction using backprojection was applied as a first step to reduce the ad hoc assumptions (cf. chapter 8). The correction of the contrast transfer function showed to be quite problematic owing to a high noise level in the data, because a no longer state-of-the-art detector had to be used. Most probably this fact restricted the obtainable resolution.

Nevertheless, iterative helical reconstruction yielded densities with 7 Å resolution for actin and 8 Å resolution for the myosin motor domain. For the myosin light chain a lower resolution was obtained, possibly to some extent caused by biochemically ill-defined binding conformations of it. About

259000 actin and myosin molecules were used for the reconstruction for the rigor state and about 243000 molecules for the strong binding ADP state. The symmetry angles converged to $-167.405°$ for the rigor state and to $-167.143°$ for the strong binding ADP state. While the ADP state reflects exactly the angle of a crystallographic 28/13 helix, the rigor state exhibits a slightly higher filament twist angle. Sharp filtering of the reconstructed densities allowed to visualize almost all α-helices as secondary structure elements.

Subsequent molecular dynamics simulations (cf. chapter 9) allowed a further interpretation and were able to resolve functional differences of myosin in two subsequent binding states of the crossbridge cycle in the complex with actin. All simulations were performed in vacuo so far, but tests of the influence of the simulation parameters, energy minimizations and first results of implicit solvent simulations showed how reliable the protein backbone structure was already determined. Detailed analyses of the MD results depending on the filtering of the density maps showed that the findings, based on the myosin backbone, are already valid without doubt.

The myosin V rigor state offers the same conformation as the nucleotide-free crystal structure (PDB entry 1OE9, Coureux et al. (2003)) which was supposed to be in rigor conformation even without being bound to actin. Only the position of the light chain of the newly obtained rigor structure is problematic, which might possibly be based on an effect similar to a crystallographic contact. The strong binding ADP state, that precedes the rigor state, was analyzed for the first time with the now achieved structural resolution. While the actin binding cleft is already closed as in the rigor state, the twist and bending of the central β-sheet differs significantly, leading subsequently to an intermediate position of the lever arm. Thus, the important finding is, that myosin accommodates strong binding to both actin and ADP by a strained conformation of the central β-sheet. The present investigation was able to visualize for the first time that the central β-sheet, myosin's *transducer* (Coureux et al., 2004), occupies different states in the actomyosin complex. In addition, differences in the actin-myosin interface are indicated by a slight deviation in helical symmetry between the two states. However, in order to be able to interpret amino acid side chain interactions, further investigations are needed, especially simulations including solvent and possibly reconstructions offering higher resolutions.

Bibliography

Adcock, S. A. and McCammon, J. A. Molecular Dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.*, 106:1589–1615, 2006.

Adrian, M., Dubochet, J., Lepault, J., and McDowall, A. W. Cryo-electron microscopy of viruses. *Nature*, 308:32–36, March 1984.

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., and Sorensen, D. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999.

Angert, I., Majorovits, E., and Schröder, R. R. Zero-loss image formation and modified contrast transfer theory in EFTEM. *Ultramicroscopy*, 81:203–222, 2000.

Baker, L. A. and Rubinstein, J. L. Radiation damage in electron cryomicroscopy. *Methods Enzymol.*, 481:371–388, 2010.

Behrmann, E., Müller, M., Penczek, P., Mannherz, H., Manstein, D., and Raunser, S. Structure of the rigor actin-tropomyosin-myosin complex. *Cell*, 150:327–338, 2012a.

Behrmann, E., Tao, G., Stokes, D. L., Egelman, E. H., Raunser, S., and Penczek, P. A. Real-space processing of helical filaments in SPARX. *J. Struct. Biol.*, 177:302–313, 2012b.

Bierbaum, S. *Analysis of actin-myosin filaments with electron cryo tomography*. Ruprecht-Karls-Universiät Heidelberg, Bachelor thesis, 2012.

Branden, C. and Tooze, J. *Introduction to Protein Structure*. Garland Publishing, 2nd edition, 1998.

Brawley, C. M. and Rock, R. S. Unconventional myosin traffic in cells reveals a selective actin cytoskeleton. *PNAS*, 106:9685–9690, 2009.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R. A., Potter, C. S., Pulokas, J., and Reilein, A. Leginon: an automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.*, 132:33–45, 2000.

Chiu, W. What does electron cryomicroscopy provide that x-ray crystallography and NMR spectroscopy cannot? *Annu. Rev. Biophys. Biomol. Struct.*, 22:233–255, 1993.

Cooke, R. Milestone in Physiology, The Sliding Filament Model: 1972-2004. *J. Gen. Physiol.*, 123: 643–656, 2004.

Cooper, J. A. Effects of cytochalasin and phalloidin on actin. *J. Cell. Biol.*, 105:1473–1478, Oct 1987.

Coureux, P. D., Wells, A. L., Ménétrey, J., Yengo, C. M., Morris, C. A., Sweeney, H. L., and Houdusse, A. A structural state of the myosin V motor without bound nucleotide. *Nature*, 425:419–423, September 2003.

Coureux, P. D., Sweeney, H. L., and Houdusse, A. Three myosin V structures delineate essential features of chemo-mechanical transduction. *EMBO J.*, 23:4527–4537, 2004.

Crowther, R. A., DeRosier, D. J., and Klug, A. The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proc. R. Soc. A, Mathematical and Physical Sciences*, 317:319–340, Juni 1970.

Darden, T., York, D., and Pedersen, L. Particle mesh Ewald: An n·log(n) method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

De La Cruz, E. M., Wells, A., Rosenfeld, S. S., Ostap, E. M., and Sweeney, H. L. The kinetic mechanism of myosin v. *Proc. Natl. Acad. Sci. USA*, 96:13726–31, 1999.

DeRosier, D. J. and Klug, A. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*, 217:130–134, Januar 1968.

Diamond, R. Filtering in the Method of Least-Squares. *Acta. Cryst.*, B25:805–806, 1969.

Dominguez, R. and Holmes, K. C. Actin Structure and Function. *Ann. Rev. Biophys.*, 40:169–186, 2011.

Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A., and Homo, J.-C. Electron microscopy of frozen water and aqueous solutions. *J. Microsc.*, 128:219–237, 1982.

Egelman, E. H. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy*, 85:225–234, 2000.

Egelman, E. H. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J. Struct. Biol.*, 157:83–94, 2007.

Egelman, E. H., Francis, N., and De Rosier, D. J. F-actin is a helix with a random variable twist. *Nature*, 298:131–135, 1982.

Engelhardt, W. A. and Ljubimova, M. N. Myosine and Adenosinetriphosphatase. *Nature*, 144:668–669, 1939.

Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 369:253–289, 1921.

Fernández, I. S., Bai, X. C., Hussain, T., Kelley, A. C., Lorsch, J. R., Ramakrishnan, V., and Scheres, S. H. Molecular architecture of a eukaryotic translational initiation complex. *Science*, 342:1240585, 2013.

Fernández, J. J., Luque, D., Castón, J. R., and Carrascosa, J. L. Sharpening high resolution information in single particle electron cryomicroscopy. *J. Struct. Biol.*, 164:170–175, October 2008.

Frangakis, A. S. and Hegerl, R. Noise reduction in electron tomographic reconstructions using non-linear anisotropic diffusion. *J. Struct. Biol.*, 135:239–250, 2001.

Frank, J. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, 2nd edition, 2006a.

Frank, J. Introduction: Principles of Electron Tomography. In Frank, J., editor, *Electron Tomography, Methods for Three-Dimensional Visualization of Structures in the Cell*. Springer, 2nd edition, 2006b.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.*, 116:190–199, 1996.

Fujii, T., Iwane, A. H., Yanagida, T., and Namba, K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature*, 467:724–728, 2010.

Geeves, M. A. and Holmes, K. C. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.*, 68:687–728, 1999.

Geeves, M. A. and Holmes, K. C. The molecular mechanism of muscle contraction. *Adv. Prot. Chem.*, 71:161–193, 2005.

Gilbert, P. PhD thesis, University of Cambridge, 1972.

Gillespie, P. G. Myosin I and adaptation of mechanical transduction by the inner ear. *Phil. Trans. R. Soc. B*, 359:1945–1951, 2004.

Glaeser, R. M. Electron microscopy. In Ehrenstein, G. and Lecar, H., editors, *Methods in Experimental Physics*, volume 20, pages 391–444. New York: Academic Press, 1982.

Hahn, E. L. and Maxwell, D. E. Spin Echo Measurements of Nuclear Spin Coupling in Molecules. *Phys. Rev.*, 88:1070–1084, 1952.

Hammer III, J. A. and Sellers, J. R. Walking to work: role for class V myosins as cargo transporters. *Nat. Rev. Mol. Cell Bio.*, 13:13–26, 2012.

Hawkes, P. W. The Electron Microscope as a Structure Projector. In Frank, J., editor, *Electron Tomography*. Springer, 2006.

Henderson, R. Cryo-Protection of Protein Crystals against Radiation Damage in Electron and X-Ray Diffraction. *Proc. R. Soc. Lond. B*, 241:6–8, 1990.

Henderson, R. The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.*, 28:171–193, May 1995.

Henderson, R. Realizing the potential of electron cryo-microscopy. *Q. Rev. Biophys.*, 37:3–13, February 2004.

Henderson, R., Baldwin, J., Ceska, T., Zemlin, F., Beckmann, E., and Downing, K. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.*, 213:899–929, June 1990.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P., Huang, Z., Penczek, P., Yang, C., Glaeser, R., Adams, P., and Ludtke, S. J. SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.*, 157:47–55, 2007.

Holmes, K. C. Muscle proteins - their actions and interactions. *Curr. Opin. Struct. Biol.*, 6:781–789, 1996.

Holmes, K. C. Fiber diffraction. http://www.mpimf-heidelberg.mpg.de/ holmes/fibre/branden.html, 1998.

Holmes, K. C. Introduction - Discussion Meeting “myosin, muscle and motility”. *Phil. Trans. R. Soc. B*, 359:1813–1818, 2004.

Holmes, K. C. and Schröder, R. R. Axial tomography by filtered least squares.

Holmes, K. C., Popp, D., Gebhard, W., and Kabsch, W. Atomic model of the actin filament. *Nature*, 347:44–49, 1990.

Holmes, K. C., Angert, I., Kull, F. J., Jahn, W., and Schröder, R. R. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature*, 425:423–427, September 2003.

Holmes, K. C., Schröder, R. R., Sweeney, H. L., and Houdusse, A. The structure of the rigor complex and its implications for the power stroke. *Phil. Trans. R. Soc. B*, 359:1819–1828, 2004.

Humphrey, W., Dalke, A., and Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

Huxley, A. F. and Niedergerke, R. Structural changes in muscle during contraction: Interference microscopy of living muscle fibers. *Nature*, 173:971–973, 1954.

Huxley, H. E. and Hanson, J. Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature*, 173:973–976, 1954.

Jenkins, G. M. and Watts, D. G. *Spectral Analysis and Its Applications*. Holden-Day, Oakland, CA, 1968.

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

Keeler, J. *Understanding NMR Spectroscopy*. Wiley, 2nd edition, 2010.

Klug, A. and Crowther, R. A. Three-dimensional Image Reconstruction from the Viewpoint of Information Theory. *Nature*, 238:435–440, August 1972.

Kodera, N., Yamamoto, D., Ishikawa, R., and Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature*, 468:72–76, November 2010.

Kollmar, M., Dürrwang, U., Kliche, W., Manstein, D. J., and Kull, F. J. Crystal structure of the motor domain of a class-I myosin. *EMBO J.*, 21:2517–2525, 2002.

Kremer, J. R., Mastronarde, D. N., and McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.*, 116:71–76, 1996.

Kron, S. J. and Spudich, J. A. Fluorescent actin filaments move on myosin fixed to a glass surface. *Proc. Natl Acad. Sci. USA*, 83:6272–6276, 1986.

Kühne, W. *Untersuchungen über das Protoplasma und die Contractilität*. Leipzig : W. Engelmann, 1864.

Lander, G., Stagg, S. M., Voss, N. R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P. W., Lyumkis, D., Potter, C., and Carragher, B. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.*, 166:95–102, 2009.

Langmore, J. P. and Smith, M. F. Quantitative energy-filtered electron microscopy of biological molecules in ice. *Ultramicroscopy*, 46:349–373, 1992.

Laplace, P. S. *Théorie analytique des probabilités*. 1820.

Lee, J. W. and Prenter, P. An Analysis of the Numerical Solution of Fredholm Integral Equations of the First Kind. *Numer. Math.*, 30:1–23, 1978.

Liu, J., Taylor, D. W., Krementsova, E. B., Trybus, K. M., and Taylor, K. A. Three-dimensional structure of the myosin V inhibited state by cryoelectron tomography. *Nature*, 442:208–211, 2006.

Ludtke, S. J. FSC Curve Interpretation. http://blake.bcm.edu/emanwiki/EMAN2/FAQ/FSC, 2011.

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

Luecke, G. R. and Hickey, K. R. A Note on the Filtered Least Squares Minimal Norm Solution of First Kind Equations. *J. Math. Anal. Appl.*, 100:635–641, 1984.

Lymn, R. W. and Taylor, E. W. Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry*, 10:4617–4624, 1971.

Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.*, 25:1584–1604, 2004.

McCammon, J. A., Gerlin, B. R., and Karplus, M. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

Mehta, A. D., Rock, R. S., Rief, M., Spudich, J. A., Mooseker, M. S., and Cheney, R. E. Myosin-v is a processive actin-based motor. *Nature*, 400:590–593, 1999.

Mindell, J. A. and Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.*, 142:334–347, 2003.

Oda, T., Iwasa, M., Aihara, T., Maéda, Y., and Narita, A. The nature of the globular- to fibrous-actin transition. *Nature*, 457:441–445, Januray 2009.

Penczek, P. A. Resolution measures in molecular electron microscopy. *Methods Enzymol.*, 482:73–100, 2010.

Penczek, P. A. and Frank, J. Resolution in Electron Tomography. In Frank, J., editor, *Electron Tomography*. Springer, 2006.

Penrose, R. On best approximate solutions of linear matrix equations. *Proc. Camb. Phil. Soc.*, 52: 17–19, 1956.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem.*, 25:1605–1612, 2004.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26: 1781–1802, 2005.

Pollard, T. D. and Korn, E. D. The "Contractile" Proteins of Acanthamoeba castellanii. Cold Spring Harbor Symposia on Quantitative Biology. *The Cold Spring Laboratory, Cold Spring Harbor, NY*, XXXVII:573–583, 1972.

Rademacher, M. Weighted Back-projection Methods. In Frank, J., editor, *Electron Tomography*. Springer, 2006.

Rayment, I., Holden, H. M., Whittaker, M., Yohn, C. B., Lorenz, M., Holmes, K. C., and Milligan, R. Structure of the actin-myosin complex and its implications for muscle contraction. *Science*, 261: 58–65, 1993a.

Rayment, I., Rypniewski, W. R., Schmidt-Base, K., Smith, R., Tomchick, D. R., Benning, M. M., Winkelmann, D. A., Wesenberg, G., and Holden, H. M. Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science*, 261:50–58, 1993b.

RCSB Protein Data Bank. PDB statistics. http://www.rcsb.org/pdb/statistics/holdings.do, 2013.

Reimer, L. and Kohl, H. *Transmission Electron Microscopy: Physics of Image Formation*. Springer, 5th edition, 2008.

Rhodes, G. *Crystallography made crystal clear*. Elsevier Inc., 3rd edition, 2006.

Rock, R. S. Molecular motors: A finicky myosin v chooses its own path. *Current Biology*, 22(15): R606 – R608, 2012.

Rose, H. H. Optics of high-performance electron microscopes. *Science and Technology of Advanced Materials*, 9, 2008.

Rosenthal, P. and Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.*, 333:721–745, 2003.

Royal Swedish Academy of Sciences. Development of multiscale models for complex chemical systems. Press release, 2013.

Saad, A., Ludtke, S. J., Jakana, J., Rixon, F. J., Tsuruta, H., and Chiu, W. Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. *J. Struct. Biol.*, 133:32–42, 2001.

Sachse, C., Chen, J. Z., Coureux, P.-D., Stroupe, M. E., Fändrich, M., and Grigorieff, N. High-Resolution Electron Microscopy of Helical Specimens: A Fresh Look at Tobacco Mosaic Virus. *J. Mol. Biol.*, 371:812–835, 2007.

Saxton, W. O., Baumeister, W., and Hahn, M. Three-dimensional reconstruction of imperfect two-dimensional crystals. *Ultramicroscopy*, 13:57–70, 1984.

Scheraga, H. A., Khalili, M., and Liwo, A. Protein-Folding Dynamics: Overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.*, 58:57–83, 2007.

Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol.*, 180:519–530., 2012a.

Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.*, 415:406–418, 2012b.

Schirra, H. J. Determination of Protein Structure with NMR Spectroscopy. http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/2dnmr.htm, 1996.

Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer, 2010.

Schröder, R. R., Hofmann, W., and Ménétret, J.-F. Zero-loss energy filtering as improved imaging mode in cryoelectronmicroscopy of frozen-hydrated specimens. *J. Struct. Biol.*, 105:28–34, 1990.

Schröder, R. R., Manstein, D. J., Jahn, W., Holden, H., Rayment, I., Holmes, K. C., and Spudich, J. A. Three-dimensional atomic model of F-actin decorated with Dictyostelium myosin S1. *Nature*, 364: 171–174, 1993.

Squire, J. M. and Knupp, C. X-ray diffraction studies of muscle and the crossbridge cycle. *Adv. Protein Chem.*, 71:195–255, 2005.

Squire, J. M., Al-Khayat, H. A., Knupp, C., and Luther, P. K. Molecular architecture in muscle contractile assemblies. *Adv. Protein Chem.*, 71:17–87, 2005.

Stewart, M. Electron microscopy of biological macromolecules. In Duke, P. J. and Michette, A. G., editors, *Modern Microscopies*. Plenum Press, New York, 1990.

Sweeney, H. L. and Houdusse, A. The motor mechanism of myosin V: insights for muscle contraction. *Phil. Trans. R. Soc. B*, 359:1829–1841, 2004.

Sweeney, H. L. and Houdusse, A. Structural and Functional Insights into the Myosin Motor Mechanism. *Annu. Rev. Biophys.*, 39:539–557, 2010.

Szent-Györgyi, A. The contraction of myosin threads. *Stud. Inst. Med.Chem. Univ. Szeged*, I:17–26, 1942.

Tang, G., Peng, L., Baldwin, P. R., Mann, D., Jiang, W., Rees, I., and Ludtke, S. J. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.*, 157:38–46, 2007.

Thon, F. Phase Contrast Electron Microscopy. In Valdrè, U., editor, *Electron Microscopy in Material Sciences*, pages 570–625. Academic Press: New York and London, 1971.

Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B., and Schulten, K. Molecular Dynamics Flexible Fitting: A practical guide to combine cryo-electron microscopy and x-ray crystallography. *Methods*, 49:174–180, 2009.

Unser, M., Trus, B., and Steven, A. C. A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy*, 23:39–51, 1987.

Unwin, P. N. T. and Henderson, R. Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.*, 94:425–440, May 1975.

Vale, R. D. and Milligan, R. A. The Way Things Move: Looking under the Hood of Molecular Motor Proteins. *Science*, 288:88–95, 2000.

Veesler, D., Ng, T., Sendamarai, A. K., Eilers, B. J., Lawrence, C., Lok, S., Young, M., Johnson, J., and Fu, C. Y. Atomic structure of the 75 MDa extremophile Sulfolobus turreted icosahedral virus determined by CryoEM and X-ray crystallography. *Proc. Natl. Acad. Sci. USA*, 110:5504–5509, 2013.

Warren, B. E. *X-Ray Diffraction*. Dover Publishing, reprint edition, 1990.

Watson, J. D. and Crick, F. H. C. Molecular structure of nucleic acids. *Nature*, 171:737–738, April 1953.

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

Whittaker, M., Wilson-Kubalek, E., Smith, J. E., Faust, L., Milligan, R. A., and Sweeney, H. L. A 35-A movement of smooth muscle myosin on ADP release. *Nature*, 378:748–51, 1995.

Williams, D. B. and Carter, C. B. *Transmission Electron Microscopy: A Textbook for Materials Science*. Springer, 2nd edition, 2009.

Wriggers, W., Milligan, R. A., and McCammon, J. A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.*, 125:185–95, 1999.

Wulf, S. E. F. *Vergleich zwischen konventionellen und least-squares 3D-Rekonstruktionsalgorithmen für axiale Elektronentomograpie*. Ruprecht-Karls-Universiät Heidelberg, Diploma thesis, 2009.

Xia, B., Tsui, V., Case, D. A., Dyson, H. J., and Wright, P. E. Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized born model, and with explicit water. *J. Biomol. NMR*, 22(4):317–331, 2002.

Yang, C., Jiang, W., Chen, D.-H., Adiga, U., Ng, E. G., and Chiu, W. Estimating contrast transfer function and associated parameters by constrained non-linear optimization. *J. Microsc.*, 233:391–403, 2009.

Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E., and Selvin, P. R. Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. *Science*, 300:2061–2065, 2003.

Zeeman, P. The Effect of Magnetisation on the Nature of Light Emitted by a Substance. *Nature*, 55: 347–347, 1897.

Zhang, X., Settembre, E., Xu, C., Dormitzer, P. R., Bellamy, R., Harrison, S. C., and Grigorieff, N. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. U S A*, 105:1867–1872, 2008.

Zhang, X., Jin, L., Fang, Q., Hui, W. H., and Zhou, Z. H. 3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell*, 141:472–482, April 2010.

Zhou, Z. H. and Chiu, W. Determination of icosahedral virus structures by electron cryomicroscopy at subnanometer resolution. *Adv. Prot. Chem.*, 64:93–124, 2003.

Acknowledgement

At this point, I like to take the opportunity to thank all those people who supported me during the time of my dissertation.

Firstly, I thank Prof. Dr. Rasmus R. Schröder, doctoral adviser and referee of my thesis, for offering me this interesting and challenging research project and for his instructive mentoring. I also thank Prof. Dr. Ulrich Schwarz for acting as second referee. Additionally, I thank Prof. Dr. Kenneth C. Holmes for the discussions about the least squares algorithm. I cordially thank Prof. Dr. Keiichi Namba and Prof. Dr. Rasmus R. Schröder for giving me the opportunity for a three-week reseach stay in Osaka, Japan. For the kind advice and assistance during that time I thank Dr. Takashi Fujii.

Special thanks go to Dr. Setsuko Fujita-Becker for providing purified proteins and for the biochemical preparations and protocols I could follow. Additionally, I appreciate her help in protein quality screenings and in filament selection. I am much obliged to Prof. Dr. H. Lee Sweeney (University of Pennsylvania, Philadelphia, USA). I thank him for providing different aliquots of myosin V and in addition for his valuable kinetic evaluation and biochemical preparation advice. I cordially thank 1st class research director Dr. Anne Houdusse (Institut Curie, Paris, France) for her kind support and advice. I highly appreciate her interest in my work, the very valuable discussions and her optimism that greatly helped me during the last year.

Furthermore, I thank Dr. Leonardo G. Trabuco for the introduction to molecular dynamics flexible fitting and for building the initial rigor model using Rosetta. I also want to thank all people that have been working in the group during my dissertation, especially Marco Oster for his great support using Leginon and for his trouble-shooting concerning computers and Dr. Götz Hofhaus for his support operating the FEI Krios Microscope. Extra thanks to Sebastian Bierbaum for sharing the rough and the smooth of preparing grids containing decorated actin filaments and for his electron tomography work. Many thanks to Melinda Feucht and Dr. Marc Hemberger from the IT support.

I cordially thank my father Prof. Dr. Günther Gercken for his support and his proofreading of my thesis. For her great support throughout my life I thank my mother Dipl.-Psych. Ruth Wulf. I also thank my boyfriend Patrick Ejilugwu. She and Patrick are always there to give advise and help whenever they can. I thank both for the strength they gave me during the last years.