Three-dimensional electron microscopy

of myriapod hemocyanin, planorbid snail acetylcholine-binding protein and cyanobacterial Vipp1

**Dissertation zur Erlangung des Grades
Doktor der Naturwissenschaften**

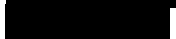MICHAEL SAUR
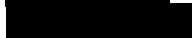geboren am ▮▮▮▮▮▮ in ▮▮▮▮▮▮

Mainz, den 14.05.2013

Elektronenmikroskopische Aufnahme negativ-kontrastierter Liposomen.

Contents

CHAPTER D: DISCUSSION 113

ABBREVIATIONS

1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
Å	Ångstrom
AcAChBP	*Aplysia californica* acetylcholine-binding protein
ACC	Amorphous calcium carbonate
ACCBP	Amorphous calcium carbonate-binding protein
AChBP	Acetylcholine-binding protein
ACF	Auto-correlation function
BgAChBP	*Biomphalaria glabrata* acetylcholine-binding protein
BgAChBP1	*Biomphalaria glabrata* acetylcholine-binding protein type 1
BgAChBP2	*Biomphalaria glabrata* acetylcholine-binding protein type 2
C3	Three-fold cylindrical
CTF	Contrast transfer function
D3	Three-fold dihedral
ΔCT-Vipp1	Vesicle-inducing protein in plastids 1 with truncated C-terminus
DOPG	Dioleoylglycerophosphoglycerol
EM	Electron microscope
EPF	Extra-pallial fluid
Hc	Hemocyanin
HdhAChBP	*Haliotis discus hannai* acetylcholine-binding protein
HdhACCBP	*Haliotis discus hannai* amorphous calcium carbonate-binding protein
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
LB medium	Lysogeny broth medium

LGIC	Ligand-gated ion channel
MDGD	Monogalactosyldiacylglycerol
ml	Milliliter
mm	Millimeter
mM	Millimol per liter
µl	Microliter
ML	Maximum likelihood
MRA	Multiple-reference alignment
MSA	Multivariate statistical analysis
nm	Nanometer
OD	Optical density
PAGE	Polyacryamide gel electrophoresis
pLGIC	Pentameric ligand-gated ion channel
PspA	Phage shock protein A
RMSD	Root mean square deviation
SCF	Sinogram correlation function
ScoHc	*Scutigera coleoptrata* hemocyanin
ScoHcA-D	*Scutigera coleoptrata* hemocyanin subunits A through D
SDS	Sodium dodecyl sulfate
SNR	Signal-to-noise ratio
SQDG	Sulfoquinovosyldiacylglycerol
TEM	Transmission electron microscope
TEV	Tobacco etch virus
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Vipp1	Vesicle-inducing protein in plastids 1
wtVipp1	Wild type of vesicle-inducing protein in plastids 1

List of Figures

XII

List of Tables

Chapter A: Introduction

1 Prelude

They are the giants among soluble proteins and yet, large quaternary protein structures in the Megadalton range are approximately one million times smaller than the diameter of a one cent Euro coin. For this thesis, five protein complexes of three protein families showing intriguing quaternary assemblies were investigated structurally, with the objective to biologically interpret the results and gain an increased understanding of their function inside their respective organisms. The method of choice for the structural investigation was single particle analysis, a bioinformatical method that uses images of proteins taken with a transmission electron microscope to calculate three-dimensional reconstructions. The proteins and the method will be introduced in this chapter, followed by the elaboration and a discussion of the obtained results in the subsequent chapters. Large efforts have been made during the analysis of the data – by using different published algorithms – in order to prevent the method's inherent vulnerability to produce noise bias, which will be addressed in the following chapters as well.

2 Electron microscopy of proteins

Three-dimensional analysis of biological macromolecules has thrived since the investigation of sperm whale myoglobin (Kendrew et al. 1958) and the publication of the first atomic resolution structure of a protein (hen egg-white lysozyme; Blake et al. 1965), both of which used X-ray crystallography. In 1964, Finch et al. presented the possibility of applying the

approach followed by X-ray diffraction to transmission electron microscopy (TEM). The first biological structure to be analyzed using computational 3-D reconstruction of transmission electron microscopic projections was the tail of the T4 bacteriophage, which yielded a resolution of approximately 30 Å (DeRosier and Klug 1968). In the early stages, this method was mainly used in combination with highly symmetric proteins or protein crystals to achieve satisfactory resolutions (e.g. Crowther et al. 1970; Henderson and Unwin 1975). Rapid improvements of computational systems facilitated the development of a method called single particle analysis, which uses TEM projections of randomly distributed particles (Frank et al. 1978; van Heel and Frank 1981). The images of thousands of single particles are combined to calculate a three-dimensional reconstruction of a macromolecular complex, taking advantage of the various orientations of the particles encountered in the sample. However, transmission electron microscopy of biological samples imposes great challenges, as image contrast is weak and sensitivity to electron radiation is high. The latter forces the application of low electron doses (typically in the range of 15-25 $e^-/Å^2$), which in turn leads to very noisy images.

In order to improve the signal-to-noise ratio (SNR) of the images, samples may be stained with heavy metal salts (for a review see Crowther 2004). This technique is called negative staining, since the structures of interest appear light against a dark background (Hall 1955; Huxley 1956; Brenner and Horne 1959). The heavy metal salts form a structure around the protein that is comparable to a plaster cast for a broken arm (Caspar 1966), improving contrast and reducing beam sensitivity. The fact that there will be internal spaces into which the salts are not able to penetrate implies that images of a stained sample mostly describe the outer hull of the structures of interest (Crowther 2004). For image formation, the heavy metal salts are the agents producing contrast, as the beam looses intensity during penetration of these electron-dense regions while easily penetrating the unstained inner sections of the biological sample. The advantages of this technique are its ease of use and rapidly obtainable results, whereas limited achievable resolution, preferential views and artifact formation account for its disadvantages (e.g. Bremer et al. 1992; Ohi et al. 2004).

For high-resolution three-dimensional reconstructions, biological samples must remain in their respective physiological conditions (Taylor and Glaeser 1974). This can be achieved by fast-freezing a thin layer of sample-containing solution on a holey carbon EM grid in liquid ethane (e.g. Adrian et al. 1984; Dubochet et al. 1988). Nowadays, semi-automatic vitrification machines allow higher reproducibility and vitrification in different atmospher-

ic conditions. Still, if specific atmospheric conditions are not required, scientists often rely on manual preparation for higher quality samples. Subsequent storage and imaging of the samples are dependent on liquid nitrogen temperature for the preservation of the samples' integrity (Crowther et al. 2004). Since the work of pioneers like Jacques Dubochet, Robert Glaeser and Richard Henderson in the 1970s and 1980s, this so-called cryo-EM method has been rapidly improved and is nowadays able to yield 3-D reconstructions of near-atomic resolution (e.g. Gonen et al. 2004; Yu et al. 2011; Bai et al. 2013). This said, the inherent disadvantages of cryo-EM are the cumbersome sample preparation and imaging procedures, which require a lot of training, the low SNR in the final images and the overall immensely increased amount of time necessary for a cryo-EM project.

3 Myriapod hemocyanin

3.1 Hemocyanin of arthropods

Hemocyanins are found in the hemolymph of many arthropods and molluscs, where they function as extracellular oxygen carriers (e.g. Gaykema et al. 1984; van Holde and Miller 1995; Decker et al. 2007). While molluscan hemocyanins are arranged in cylindrical quaternary structures, hemocyanins of arthopods are composed of a variable amount of hexamers (e.g. Markl and Decker 1992; for a recent review on molluscan hemocyanins see Markl 2013). As the structural differences suggest, hemocyanins in both phyla are not related and should have evolved independently (e.g. Burmester 2001).

Arthropod hemocyanins belong to a protein superfamily that also contains arthropod phenoloxidases, hexamerins (storage proteins of insects), pseudo-hemocyanins (cryptocyanins), which are the storage proteins of decapod crustaceans, and hexamerin receptors (for a review see Burmester 2002). Synthesis takes place in so-called cyanocytes which accumulate hemocyanin molecules until cell rupture releases the protein (Fahrenbach 1970). As mentioned above, arthropod hemocyanins consist of varying amounts of hexamers. For instance, 1x6-mers can be found in the Californian spiny lobster *Panulirus interruptus* (Gaykema et al 1984), whereas other crustaceans, like the manits shrimp *Squilla mantis* or the freshwater crayfish *Astacus leptodactylus*, show 2x6-meric assemblies (e.g. Markl et al. 1979; van Bruggen 1983). In chelicerates, a 4x6-mer can be found in the tarantula *Eurypelma californicum*, while the horse-shoe crab *Limulus polyphemus* bears the largest arthropod hemocyanin found

to date, being comprised of eight hexamers (e.g. MARKL ET AL. 1979). Hemocyanin subunits have a molecular mass of approximately 75 kDa (~660 amino acids) and can be divided into three structural domains (Figure 1; GAYKEMA ET AL. 1984). Domain #1 is comprised of α-helices and has been suggested to be important for the transmission of allosteric movement (HAZES ET AL. 1993; MAGNUS ET AL. 1994), whereas domain #3 is mainly comprised of β-barrels and is thought to provide structural stability (e.g. VOLBEDA AND HOL 1989). Oxygen binding is achieved by an active site in domain #2, which bears a type-III copper-binding center containing two Cu^+ ions (CuA and CuB) that are coordinated by three histidines each (e.g. BEINTEMA ET AL 1994; Figure 2). Upon oxygen binding as O_2^{2-}, the Cu^+ ions undergo a formal shift to Cu^{2+} (FREEDMAN ET AL. 1976). As such, the 8x6-mer found in *Limulus polyphemus* houses 48 active sites which bind oxygen in a highly cooperative manner (e.g. MARKL 1986; MARTIN ET AL. 2007).

For a long time it was believed that Myriapoda did not carry hemocyanin, due to their systematic classification into the Tracheata (e.g. BRUSCA AND BRUSCA 1990 *cited by* KUSCHE AND BURMESTER 2001). However, first evidence of the occurence of hemocyanin in centi-

Figure 1   Molecular model of a *Scutigera coleoptrata* hemocyanin monomer depicting the three domains #1 (blue), #2 (red), #3 (yellow)
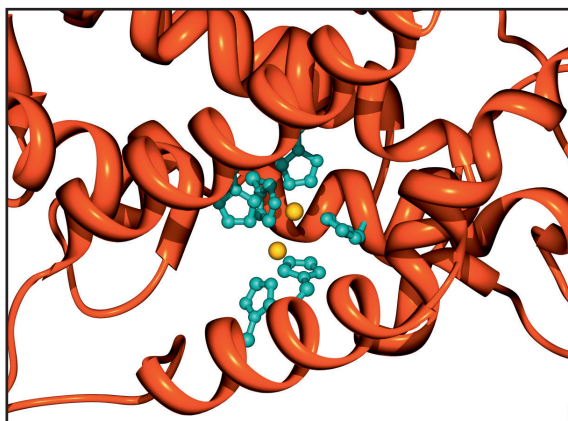
Figure 2 Active site of a *Scutigera coleoptrata* hemocyanin subunit. The two copper ions are coordinated by three histidines each.

pedes was already provided by RAJULU (1969) for *Scutigera longicornis* and by MANGUM ET AL. (1985) for *Scutigera coleoptrata*. The discovery of hemocyanin in those Chilopoda was attributed to their very short tracheae (e.g. MARKL ET AL. 2007) and their life as ambush predators, which requires explosive movements and a fast oxygen uptake. Yet, these arguments were not applicable when hemocyanin was also found in Diplopoda, which bear a well developed tracheal system (BRUSCA AND BRUSCA 1990 *cited by* JAENICKE ET AL. 1999). It was suggested that hemocyanins could contribute to the ongoing controversy about the phylogenetic relationships inside the phylum of Arthopoda (KUSCHE AND BURMESTER 2001). More specifically, it is still highly controversial were Myriapoda are to be placed. Evidence suggests that they are a sister-group of the Chelicerata (HWANG ET AL. 2001), whereas other authors place them into the Mandibulata, which also comprises Hexapoda and Crustacea (GIRIBET AND EDGECOMBE 2012). Sequence comparison of hemocyanins indicates that Myriapoda are indeed a sister-group of chelicerates, given the fact that both subphyla lack the second helix of domain 1 (KUSCHE AND BURMESTER 2001). Nevertheless, an independent deletion of this helix in the two groups is also possible (*ibid.*).

3.2 *SCUTIGERA COLEOPTRATA* HEMOCYANIN

Initially found in the Mediterranean region, *Scutigera coleoptrata* reached the United States and, recently, southern Germany as well. Typically, it bears 15 pairs of legs and can grow between 2.5-5 cm in length (Figure 3). As in all Chilopoda, the first pair of walking legs is modified to venomous claws which serve to immobilize prey (e.g. UNDHEIM AND KING 2011). This trait and the enormous speed *Scutigera coleoptrata* can achieve (compared to its body length) help the animal in its life as an ambush predator. Additionally, a 6x6-meric hemocy-

Figure 3 The centipede *Scutigera coleoptrata*. The right-hand side shows a close-up picture of the cranial region. The first pair of legs, which is modified to venomous claws, is clearly visible (picture by Kevin Collins).

anin (ScoHc) with Hill coefficients up to 10.7 (at pH 8.0) supplies the tissues with sufficient oxygen (Mangum et al. 1985). Further biochemical analyses provided evidence that ScoHc is comprised of four subunit types, termed ScoHcA, ScoHcB, ScoHcC and ScoHcD (Gebauer and Markl 1999; Kusche et al. 2003). Initially, an octrahedral quaternary structure was proposed for the 6x6-mer (Mangum et al. 1985). However, analyses using single particle electron microscopy of oxygenated ScoHc (oxy-ScoHc) could show that the molecule is, in fact, a D3-symmetrical particle (Figure 4 A+B; Markl et al. 2009). On the basis of phylogenetic relationships to other myriapods, Markl et al. (2009) derived a pseudo-atomic model for oxy-ScoHc which divided the particle into core and mantle subunits (Figure 4 C). Furthermore, five different inter-hexamer interfaces could be described (Figure 4 C). In my diploma thesis, I three-dimensionally reconstructed the deoxygenated form of ScoHc (deoxy-ScoHc) and described the differences between the conformationally different ScoHc 6x6-mers, as well as between the individual inter-hexamer interfaces. However, in retrospective, the reconstructions of oxy-ScoHc and deoxy-ScoHc show signs of over-fitted noise (see section D1.1 for details).

3.3 *Polydesmus angustus* hemocyanin

In contrast to centipedes, diplopods are detritivores which are distributed in temperate and tropical forests and do not accelerate to high velocities (e.g. Sierwald and Bond 2007). Jaenicke et al. (1999) provided evidence that *Spirostreptus* sp., a member of the Diplopoda, bears a 6x6-meric hemocyanin and that the concentration of the protein in the hemolymph is very high (~50 mg/ml). Striking differences can be observed when comparing the oxygen binding properties of *Scutigera coleoptrata* and *Spirostreptus* sp. hemocyanin, since the former
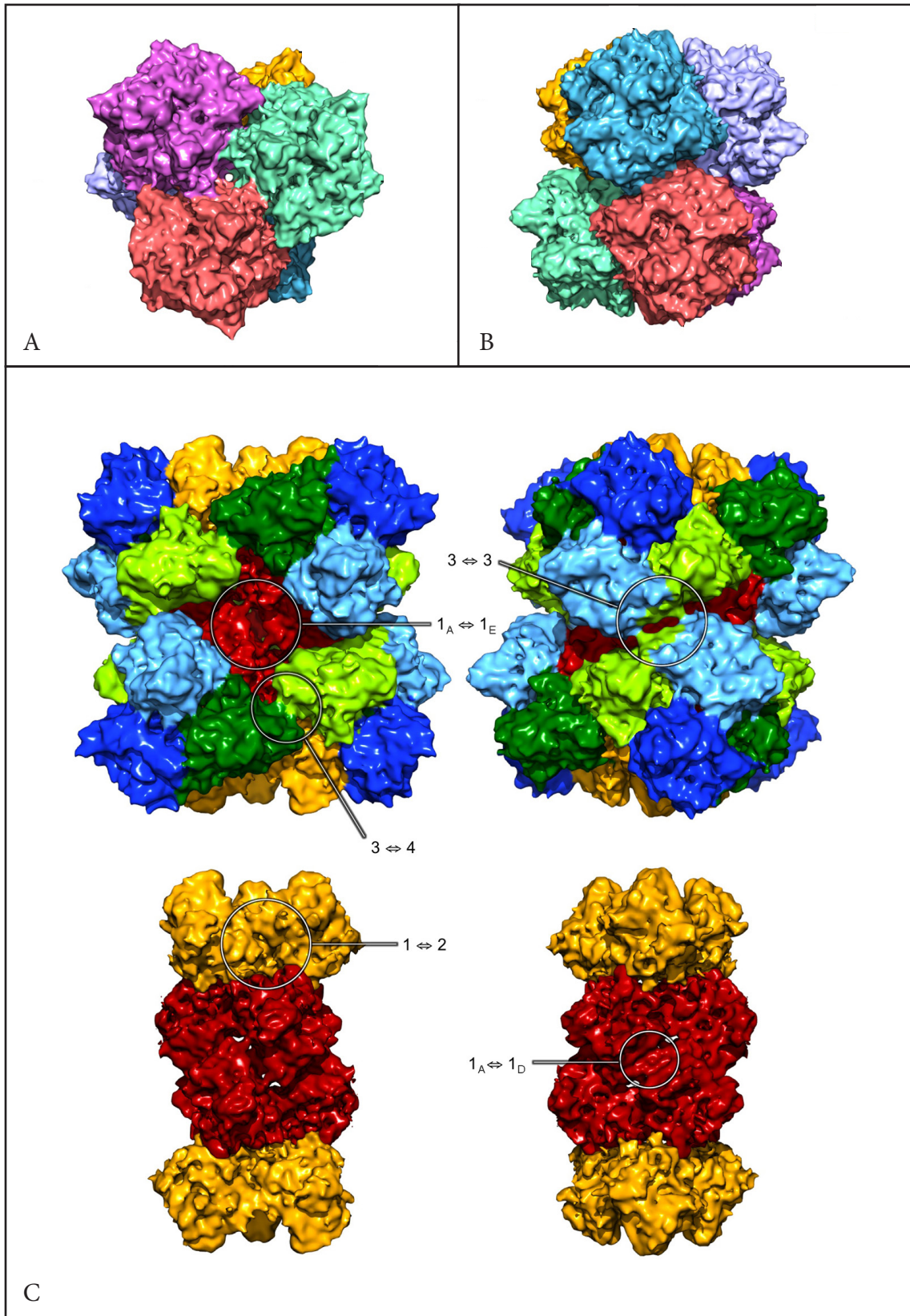
Figure 4 *Scutigera coleoptrata* 6x6-mer hemocyanin. A) Top view of ScoHc. Each hexamer is colored differently. B) Side view of ScoHc. C) Side views of ScoHc. The four ScoHc subunit types divide the molecule into a core (orange and red) and a mantle (blue and green). The bottom of C) shows only the core of ScoHc. The five inter-hexamer interfaces are indicated. For detailed information see MARKL ET AL. (2009).
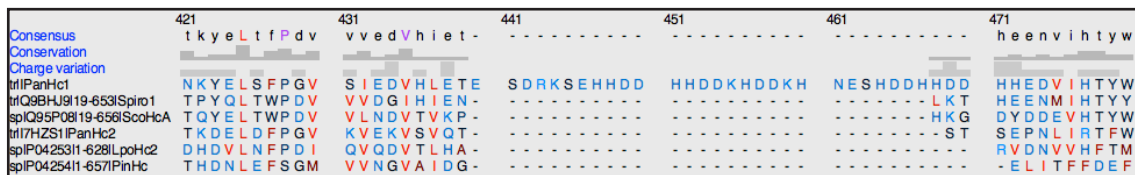
Figure 5   Multiple-sequence alignment between *Polydesmus angustus* subunit 1 (PanHc1), *Spirostreptus* sp. subunit 1 (SpiHc1), *Polydesmus angustus* subunit 2 (PanHc2), *Limulus polyphemus* subunit 2 (LpoHc2) and *Panulirus interruptus* (PinHc). Note the long loop that is only found in PanHc1. For the accession numbers of the respective sequences see Appendix B.

shows an O_2 affinity of p_{50}=~55 Torr and a Hill coefficient of 8.9 at pH 7.5, whereas the latter has an affinity of p_{50}=4.7 Torr and a Hill coefficient of 1.3 at pH 7.5 (MANGUM ET AL. 1985; JAENICKE ET AL. 1999).

Like in *Spirostreptus* sp., the hemocyanin of the diplopod *Polydesmus angustus* is comprised of at least two different subunit types (PanHc1 and PanHc2; see Appendix B for Uniprot accession numbers). However, PanHc1 bears a peculiar insertion that is mainly comprised of histidines and aspartates (histidine loop; Figure 5). The implications of this insertion are not known, due to the fact that no structural and functional studies have been performed on PanHc.

3.4 RESEARCH OBJECTIVES

For ScoHc, the objectives of the present study consisted of an entirely new CTF-estimation and re-calculation of the oxy-ScoHc and deoxy-ScoHc datasets, in order to substantiate the differences previously found between them. To further reinforce the results, the calculations were carried out in several software packages that are extensively used in the field of single particle analysis. In contrast to the interface-specific analyses provided in previous work, the emerging differences between oxy- and deoxy-ScoHc were to be analyzed on a more global scale, in order to locate the regions of the protein that drive allosteric information transfer.

For PanHc, a first three-dimensional reconstruction was calculated *via* single particle analysis, in order to gain information about the quaternary structure of this molecule and possible structural implications of the histidine loop. Given the large differences of the oxygen binding properties between *Scutigera coleoptrata* and diplopods, a comparison of the structural results obtained for ScoHc and PanHc might also provide further understanding about the transfer of allosteric information within these large oxygen carriers.

4 Vesicle-inducing protein in plastids 1

Membrane organization in photosynthetic organisms is a highly interesting mechanism, since lipids, proteins and pigments have to be transported across different cell organelles and inserted into their destination membranes (e.g. Vothknecht et al. 2012). For instance, thylakoid membranes of higher plants consist of lipids that might be produced in the biogenic membranes of the endoplasmic reticulum, the outer plastid envelope or the inner plastid envelope (Benning 2009). Due to the fact that no connection between the inner envelope and thylakoid membranes has been found for differentiated plastids thus far, a lipid trafficking system must exist (Westphal et al 2001a). Vesicles distributed in the stroma of *Pisum sativum* chloroplasts have indeed been found and attributed to the lipid exchange between thylakoid membranes and the inner envelope (Westphal et al. 2001b). Interestingly, the occurrence of lipids seemed to be abolished in a mutant form of *Arabidopsis thaliana* which also showed a very light-green and small phenotype in comparison to the wild-type plant and the disability to grow photoautotrophically (Kroll et al. 2001). Furthermore, structural analyses of the chloroplasts revealed the inability to form an ordered thylakoid membrane system (*ibid.*). This mutant was characterized by a 5-fold decrease of a 33 kDa protein that was subsequently termed vesicle-inducing protein in plastids 1 (Vipp1) and grew normally when the plant was complemented for the protein by an inserted Vipp1 sequence (Westphal et al. 2001a).

Vipp1 has been found in most photosynthetic organisms and all higher plants that were analyzed for the presence of the protein thus far (for a recent review see Vothknecht et al. 2012). It had been previously described as membrane-associated protein of 30 kDa (M30) which was shown to bind to the inner envelope and thylakoid membranes *in vitro* (Li et al. 1994). The secondary structure of the protein is predicted to be mostly α-helical with a high propensity to form coiled coils (Hankamer et al. 2004; Fuhrmann et al. 2009a; Bultema et al. 2010). Furthermore, Vipp1 has been shown to assemble into large oligomeric rings which are most probably formed by the interaction between multiple tetramers (Aseeva et al. 2004). More accurately, evidence suggests that Vipp1 tetramers are, in fact, dimers of dimers (Fuhrmann et al. 2009a). Vipp1 rings show a constant height of 220 Å, whereas the variable amount of tetramers incorporated into Vipp1 rings entails a variable width between 250 Å and 330 Å (*ibid.*). Multiple rings have been described to assemble "top-to-bottom" into rod-like structures (e.g. Aseeva et al. 2007; Aseeva et al. 2009; Fuhrmann et al. 2009a).

4.1 Phage-shock protein A

The phage-shock-protein system found in many Archaea, Cyanobacteria, as well as Gram-negative and Gram-positive bacteria, includes the phage shock protein A (PspA) which may share sequence similarities of 50 % with Vipp1 (Brissette et al. 1990; Bultema et al. 2010; Vothknecht et al. 2012). The simultaneous occurrence of PspA and Vipp1 in cyanobacteria and the spatial proximity of the gene loci suggest that Vipp1 emerged in a duplication event of PspA (Westphal et al. 2001b). The latter is a stress protein that is up-regulated upon invasion by filamentous phages and the occurrence of stress factors, such as heat, hyper-osmotic conditions or blockage of the export machinery (Brissette et al. 1990; DeLisa et al. 1994). The protein has been described to be associated with the cytoplasmic membrane of bacterial cells (Brissette et al. 1990), where it can prevent proton leakage across the membrane (Kobayashi et al. 2007). Furthermore, an increased PspA concentration can be found in bacteria carrying defects in the Sec or Tat pathways, both of which are important for protein translocation (Kleerebezem and Tommassen 1993; DeLisa et al. 1994). Due to the increased efficiency of Tat export experienced with elevated levels of PspA, the protein might be important for protein translocation (*ibid.*). Interestingly, Vipp1 can replace PspA in this function (DeLisa et al. 1994). Structurally, PspA can be very variable, given the fact that it has been found as a 36-meric ring – comparable to Vipp1 rings – constituted by tetramers (Hankamer et al. 2004), as a nearly dodecameric complex in combination with PspF (Joly et al. 2009) and as a clathrin-coat-type assembly (Standar et al. 2008).

4.2 Biological functions of Vipp1

Apart from its importance for the ordered assembly of thylakoid membranes, the exact biological function of Vipp1 remains unknown. Its similarity to PspA, however, might indicate certain functions. Indeed, Vipp1 is up-regulated upon occurrence of stress factors like strong irradiation (Im and Grossman 2001) and ranges among the mostly increased proteins in hyper-osmotic conditions by showing a five-fold increment in concentration (Huang et al. 2006). In contrast, Vipp1 concentration is decreased in cold conditions (Suzuki et al. 2001). Furthermore, it has been shown that Vipp1 binds to chaperones (Liu et al. 2005; Liu et al. 2007), which could indicate a functional importance for protein translocation. However, a functional interaction with chaperones remains questionable due to the fact that the latter catalyze the assembly and disassembly of Vipp1 oligomers (Liu et al. 2007; Ramundo et al. 2013). Hence, it was proposed that chaperones might facilitate the translocation of

Vipp1 into the plastid (Vothknecht et al. 2012). Generally, the deduction of a specific Vipp1 function on the basis of PspA is not possible, given that PspA cannot substitute Vipp1, while Vipp1 is able to functionally replace PspA (Westphal et al. 2001b). This functional discrepancy might be the result of a C-terminal extension encountered in all Vipp1 proteins – not present in PspA – and which is defined by approximately 20-40 amino acids comprising a linker of variable length which is followed by a short α-helix (Westphal et al. 2001; Aseeva et al. 2004; Aseeva et al. 2007; Bultema et al. 2010). Multiple analyses concerning Vipp1 function led to diverging results. Generally it was found that, although being a soluble protein, Vipp1 is mostly associated with membranes (Kroll et al. 2001; Fuhrmann et al. 2009a). For *Arabidopsis thaliana*, an association with the inner envelope and thylakoid membranes could be shown (Kroll et al. 2001; Aseeva et al. 2007). For *Synechocystis* sp. strain PCC6803, an association of Vipp1 could only be demonstrated for the inner envelope by Westphal et al. (2001b), whereas other studies showed a membrane association comparable to *A. thaliana* (Fuhrmann et al. 2009b). Given its direct membrane interaction, Vipp1 has been proposed to be actively involved in thylakoid bilayer formation (e.g. Aseeva et al. 2007; Vothknecht et al. 2012). This hypothesis is substantiated by the fact that a 10-fold increase in Vipp1 concentration can be observed in maize mesophyll stroma in comparison to bundle sheath stroma, supposedly due to the former's larger thylakoid system (Majeran et al. 2005). Zhang and Sakamoto (2013), however, do not support the model that Vipp1 is directly involved in thylakoid formation. Other putative functions are the stabilization of highly bent membranes during vesicle budding (e.g. Vothknecht et al. 2012) or an interaction with photosynthetic proteins, as studies suggest an influence of Vipp1 on photosystem I - photosystem II ratio (Fuhrmann et al. 2009b), photosystem oligomerization (*ibid.*) and the coupling of the light harvesting complexes with the photosystems (e.g. Aseeva et al. 2007; Zhang and Sakamoto 2013).

Structurally, Vipp1 rings have been attributed to the function of the protein, as membrane-associated fractions only contained the oligomeric form (e.g. Aseeva et al. 2007; Otters et al. 2012). Although the C-terminal extension is supposedly essential for the protein function, it is not important for oligomerization (Aseeva et al. 2004; Otters et al. 2012). Rather, deletion of the first α-helix has a destructive effect on Vipp1 rings (Otters et al. 2012), as has the deletion of a central helical domain (Aseeva et al. 2004). The latter is supported by a study of C-terminally truncated PspA which showed a shift from oligomers to homo-dimers (Joly et al. 2009), whereas in the study performed by Otters et al. (2012)

oligomers shifted to trimeric or tetrameric assemblies. Furthermore, the rod-like structures formed by Vipp1 rings have been subject to functional hypotheses as well, since rod-like structures – or microtubule-like structures (MTLs) – have been found in chloroplasts and cyanobacteria (e.g. Lawrence et al. 1984; van de Meene et al. 2006). Hence, these assemblies were speculated to function as cytoskeleton-like elements that are used as tracks for directed lipid or protein transport (e.g. Liu et al. 2007). However, it has yet to be proven that MTLs are indeed constituted by Vipp1 (Vothknecht et al. 2012). Finally, Vipp1 might exist in a "dynamic equilibrium between monomers, dimers, and higher-ordered oligomers" (Fuhrmann et al. 2009a), indicating a structure-function relationship that is more complex than presumed thus far.

4.3 Research objectives

The objectives of the present study were to structurally and functionally analyze wild type Vipp1 (wtVipp1) of the cyanobacterium *Synechocystis* sp. strain PCC6803, using transmission electron microscopy. The study was conducted in collaboration with Raoul Hennig, working in the group of Prof. Dirk Schneider at the Institute of Biochemistry of the JGU Mainz, who performed the biochemical analyses (for details see section B5). For structural analysis, a 3-D reconstruction of Vipp1 was to be calculated using single particle analysis. As will be described below, the heterogeneity of the current Vipp1 samples did not allow cryo-EM, thus limiting the current study to negatively stained specimen. However, being the first three-dimensional analysis of Vipp1, lower-resolution density maps should provide a deeper understanding of the molecule in its oligomeric state and the spatial distribution of individual Vipp1 monomers than previous 2-D analyses were capable of. Moreover, 2-D analyses should provide an understanding of the stability of oligomeric rings against point-mutations targeted at the N-terminal region of Vipp1. Considering that the importance of the first α-helix for the assembly of Vipp1 rings has been previously shown (Otters et al. 2012), this study aimed at identifying specific amino acids involved in the quaternary assembly. Two-dimensional analysis of C-terminally truncated Vipp1 should substantiate the results obtained by Aseeva et al. (2004) that indicated a normal assembly of Vipp1 rings. Lastly, lipid-binding interactions of Vipp1 with different lipids were surveyed by transmission electron microscopy. Different combinations of glycoglycerolipids and phosphatidylglycerol were used in this study, as these are the main lipids found in thylakoid membranes (Benning et al. 2009).

5 *Biomphalaria glabrata* acetylcholine-binding protein

Acetylcholine-binding protein (AChBP) has been first described by Smit et al. (2001) and Brejc et al. (2001) as a protein found in the freshwater snail *Lymnaea stagnalis* (LsAChBP). It showed structural similarity to nicotinic acetylcholine receptors (nAChR), which are important drug targets due to their implication in diseases like schizophrenia, Alzheimer's, epilepsy, and Parkinson's, as well as disorders like cognitive and memory defects, and nicotine addiction (for a review see Paterson and Nordberg 2000). The auto-immune disease myasthenia gravis targets AChRs by binding to a certain loop (main immunogenic region loop; MIR loop), leading to accelerated degradation of the receptor (Tzartos et al. 1991).

5.1 Nicotinic acetycholine receptors

Nicotinic AChRs are targets for drugs such as tranquilizers (e.g. propophol), anti-convulsants and anti-emetics, amongst others (Jayakar et al. 2013; Zhang et al. 2013). Additionally, some of the most potent toxins currently known – such as α-conotoxin, d-tubocurarine and α-bungarotoxin – bind to nAChRs with high affinities (e.g. Paterson and Nordberg 2000; Sixma and Smit 2003). Nicotinic acetycholine receptors belong to the Cys-loop receptor super-family, which, for example, also includes the γ-aminobutyric acid (GABA) receptors, 5-HT_3 serotonin receptors and glycine receptors, all of which share a conserved sequence region hosting a disulfide bridge termed the Cys-loop (e.g. Sixma and Smit 2003). This super-family is also known as pentameric ligand-gated ion channels (pLGIC) which is a more accurate and recent term (Corringer et al. 2012). The structure of closed channel nAChRs has been solved and refined to 4 Å by 2-D electron crystallography, showing a pentameric assembly divided into an extracellular, a transmembrane and an intracellular domain (Figure 6; Miyazawa et al. 2003; Unwin 2005). Generally, nAChRs can be divided into muscle and neuronal nAChR, the former being situated at the neuromuscular junction where it depolarizes the post-synaptic membrane upon acetylcholine binding and the latter being predominantly situated in pre-synaptic membranes of nerve terminals regulating the release of neurotransmitters (Rucktooa et al. 2009). Nicotinic AChRs can be composed of several subunit types. In vertebrates, for example, muscle nAChRs typically show a combination of $(\alpha_1)_2\beta_1\gamma\delta$ subunits, whereas neuronal nAChRs may consist of homo-pentameric (e.g. α7 or α9) or various hetero-pentameric assemblies (e.g. Celie et al. 2005). However, one alpha subunit is always essential for ligand binding (Rucktooa et al. 2009). Different confor-
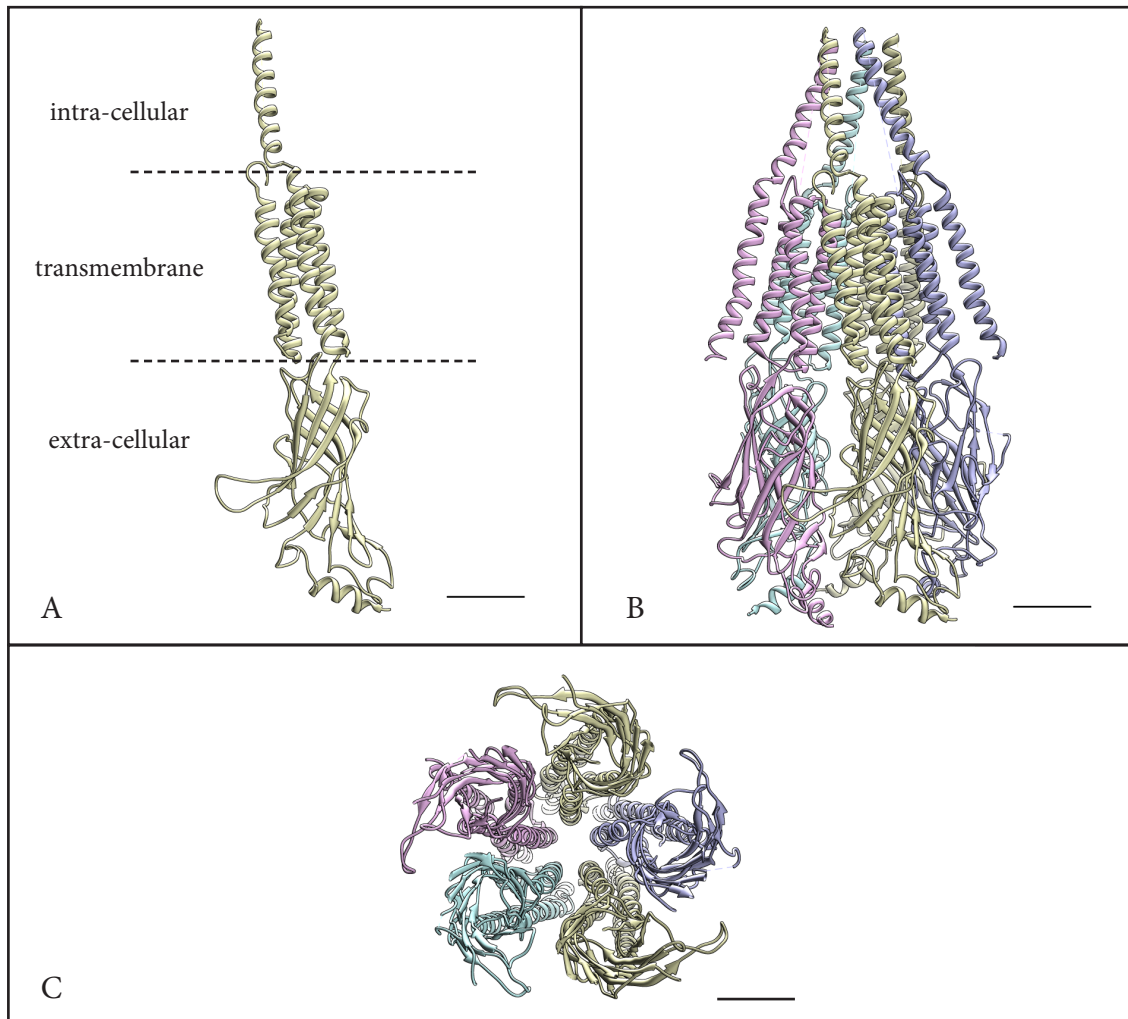
Figure 6   Structural representation of *Torpedo marmorata* AChR (PDB code: 2BG9). A) Monomer with its extra-cellular, transmembrane and intra-cellular domains. B) Fully assembled pentamer in its side-view. C) same as B) but showing the top-view (N-face). Scale bars: 20 Å.

mational states result in substantial movement of the α subunits and the transmembrane helices, ultimately opening or closing the channel for cations (Unwin 2005). Typically, three different conformational states can be described for nAChRs, namely the resting (closed), activated (open) and desensitized (closed) states (e.g. Hansen et al. 2005). The latter is triggered by a prolonged exposure to agonists (e.g. Langley 1905 *cited in* Corringer et al. 2012; Zhang et al. 2013). In closed conformation, ion flow is inhibited by a hydrophobic ring that is constituted by the transmembrane helices, whereas ion flow in these cation-selective channels is controlled by the anionic intracellular vestibule in the open conformation (Unwin 2005). This contrasts anion-selective receptors (e.g. homomeric glycine receptors), where these structures are mainly cationic (*ibid.*). Despite the success of nAChR structure determination, the closed channel structure of the nAChR remains the only high-resolution

structure of a fully assembled nAChR to date, due to the difficulties associated with structural studies of membrane proteins.

5.2 STRUCTURAL SIMILARITIES BETWEEN nAChRs AND AChBP

The structural similarity of AChBPs to nAChRs is confined to the extra-cellular domain of the latter which, however, is the ligand-binding domain (LBD) associated with agonist and antagonist binding (BREJC ET AL. 2001). Ligand binding analyses using methods such as iso-thermal titration calorimetry (e.g. CELIE ET AL. 2004) or tryptophan fluorescence (e.g. HANSEN ET AL. 2004) have shown similar binding characteristics of AChBP and nAChRs, many ligands binding with nanomolar and sub-nanomolar affinities (SMIT ET AL. 2001; HANSEN ET AL. 2004). Furthermore, recently established chimeras between AChBP and the transmembrane domain of a 5-HT$_3$ receptor yielded a fully functional pLGIC (BOUZAT ET AL. 2004). Hence, the two proteins can be described as structural and functional homologues concerning ligand interaction (BREJC ET AL. 2001). The importance of AChBP for nAChR research is given by its solubility in water, which simplifies structural investigation. It was first structurally described as a homo-pentameric assembly with a molecular mass of ~24 kDa (excluding the ~888 Da glycosyl group) which resembles the LBD of α7 nAChRs (BREJC ET AL. 2001; SMIT ET AL. 2001). Therefore, it comprises structural elements that are conserved among the pLGIC super-family, such as immunglobulin-like β-sandwich motifs which are stabilized by inner hydrophobic residues, the "aromatic box" of the ligand-binding site (RUCKTOOA ET AL. 2009) and the Cys-loop (Figure 7). The latter, however, differs in length and hydrophilicity from the ones found in Cys-loop receptors (SIXMA AND SMIT 2003). Thus far, AChBP has only been described in mollusks and the polychaete *Capitella teleta* (MCCORMACK ET AL. 2010), exhibiting pentameric or dipentameric assemblies (CELIE ET AL. 2005). Despite the generally low sequence identity of approximately 20-30 % between different AChBPs and between AChBPs and AChRs, structural details have been well conserved (e.g. SIXMA AND SMIT 2003; RUCKTOOA ET AL. 2009). The ligand-binding site is situated between two subunits, which contribute principle face amino acids from the plus side and complementary face amino acids from the minus side (BREJC ET AL. 2001). Its structural details will be more deeply described in sections C3.1.5 and D4.4. Despite its pentameric assembly and five putative binding sites, AChBP does not exhibit cooperativity (SIXMA AND SMIT 2003), some studies even showing indications of negative cooperative effects (SMIT ET AL. 2001). Moreover, the conformational changes upon ligand binding described by UNWIN
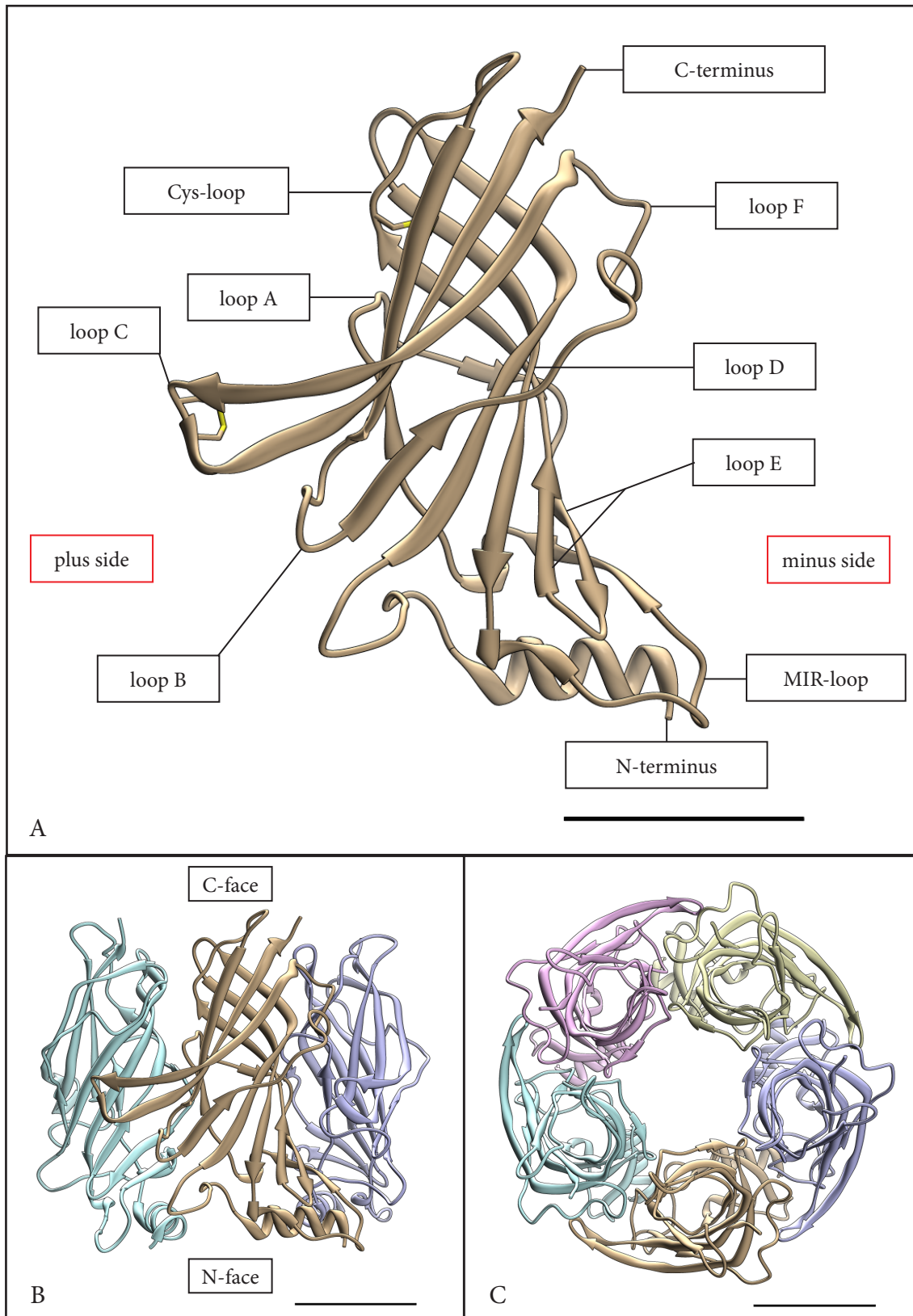
Figure 7   Structural representation of LsAChBP (PDB code: 1I9B). A) Monomer, comprised of 210 amino acids, showing the most important structural details. Loops A-C are the principal face of the LBD which is situated on the plus side of the monomer, loops D-F are the complementary face of the LBD, being situated on the minus side of the monomer. Note the two disulfide bridges on loop C and the Cys-loop, respectively. B) Trimer showing the intra-pentamer interface constituted by the plus side of one monomer and the minus side of the other. The LBD is situated between two monomers. Note that the top plane is termed C-face and the bottom plane N-face. C) Top-view of the full pentameric LsAChBP assembly showing the C-face. Scale bars: 20 Å.

ET AL. (2005) cannot be reproduced with AChBP (CELIE ET AL. 2004). The only significant structural change of the latter is confined to loop C, which moves as much as 11 Å between agonist and antagonist binding, and loop F (HANSEN ET AL. 2005). Comparisons of the conformational states between AChRs and AChBPs led to the conclusion that the apo state of *Aplysia californica* AChBP (AcAChBP) resembles the resting state of AChRs, showing low affinity for nicotinic ligands (HANSEN ET AL. 2005). Altogether, AChBP is likely to show three distinct conformations of loop C, namely (i) a peptide agonist-bound open state, (ii) an apo or organic antagonist-bound intermediate state and (iii) an agonist-bound closed state (*ibid.*).

5.3 BIOLOGICAL FUNCTIONS OF AChBP

In twelve years of AChBP research, a variety of putative biological functions have been proposed for the protein. To date, several postulated functions are published which are based on more or less convincing indications. Originally, AChBP was thought to be a modulator of molluskan synaptic transmission. SMIT ET AL. (2001) showed by *in vitro* studies that glial cells situated at synapses contained AChRs which, upon binding of ACh during synaptic signaling, would result in the release of AChBP into the synaptic cleft, hence modulating the actual post-synaptic signal. Similar modulatory functions of peri-synaptic glial cells have been reported for numerous receptor types, including the cholinergic neuromuscular junction (ROBITAILLE 1998; SIXMA AND SMIT 2003). In 2009, the modulatory function of AChBP was challenged by BANKS ET AL. who conducted *in vivo* studies and showed that AChBP was mostly expressed in glial cells which were located at neuronal cell bodies, as opposed to the synapse-rich neuropile. Furthermore, they found AChBP to be mainly distributed in glial projections that invaginated into the cell bodies. The conclusion of the authors was that AChBP most likely modulates non-synaptic interactions, although they do not exclude the possibility that synaptic modulation by AChBP between axons and neuronal cell bodies is possible. A third function was presented by MA ET AL. (2007), who found an AChBP-like protein in the pearl oyster *Pinctada fucata* that is able to stabilize super-saturated solutions of amorphous calcium carbonate (ACC) and inhibit calcite formation. Given the fact that the hemolymph and the extra-pallial fluid (a liquid-filled space between the outer epithelium of the mantle and the shell) are super-saturated with different ions – and especially amorphous calcium carbonate – the protein might be able to prevent uncontrolled calcification and control shell growth. Additionally, it was shown that the protein could prevent the growth of abnormal nacre lamellae. Due to minor differences to AChBP (an N-terminal unpaired

cysteine leading to dipentameric assemblies and multiple putative N-glycosylation sites) and its peculiar function, the protein was termed amorphous calcium carbonate-binding protein (ACCBP). The findings were substantiated by the discovery of two AChBP-like proteins in the abalone *Haliotis discus hannai* which were identified as AChBP (not being able to bind to ACC) and ACCBP, respectively, and which showed diverging expression patterns (HUANG ET AL. 2009). Lastly, AChBP was shown to bind certain dinoflagellate toxins (Spirolides and Gymnodimines) with picomolar affinity (BOURNE ET AL. 2010). However, the authors did not present a hypothesis concerning the putative functional relevance of these results for the organism.

5.4 *BIOMPHALARIA GLABRATA* ACETYLCHOLINE-BINDING PROTEIN

The rosette-like protein found in the hemolymph of the planorbid freshwater snail *Biomphalaria glabrata* has been observed in several planorbid snails and was initially identified as an erythrocruorin (MOSBY AND WOOD 1975; TERWILLIGER ET AL. 1976). This was refuted by LIEB ET AL. (2006) due to the biochemical properties of the former that were not attributable to these respiratory proteins. Recently, it has been shown that the rosette-like protein is, in fact, an acetylcholine-binding protein (SAUR ET AL. 2012). Genomic and biochemical analyses unveiled that *Biomphalaria glabrata* comprises two distinct 25 kDa AChBP types, which share a sequence identity of 52 % (termed BgAChBP1 and BgAChBP2). Both types are heavily glycosylated and able to bind amorphous calcium carbonate (*ibid.*). Furthermore, non-reducing SDS-PAGE showed that a minor component of the protein migrated at 50 kDa, which indicates the stable presence of a disulfide-bridged subunit dimer (*ibid.*). Furthermore, a dodecahedral structure was suggested for the rosette-like protein found in the planorbid snail *Helisoma trivolvis* (ILAN ET AL. 1986).

5.5 RESEARCH OBJECTIVES

In this study, I conducted detailed structural analyses of BgAChBP using cryo-EM, single particle analysis and template homology modeling. The results of these analyses have been elaborated in SAUR ET AL. (2012) and will be presented in detail in the subsequent chapters. Furthermore, a structural comparison between recombinant BgAChBP1 and BgAChBP2 samples will be presented and discussed based on recently obtained data.

Chapter B:
Materials and Methods

1 General Remarks

In the following sections, the basic concepts of image processing and 3-D reconstruction will be elaborated in detail, largely using the software suite IMAGIC-5 as an example. Algorithms differing from the ones presented in this context will be analyzed in the discussion. Finally, project-specific materials and methods are depicted in sections B4 - B6.

2 Software

Appion is a free software which combines different image processing packages in one database system (Lander et al. 2010).

Chimera (v1.7) is a free 3-D data visualization, analysis and modeling program, which was introduced to scientific society in 2004 by a group of the University of California (Pettersen et al. 2004). In this work, it was used for visualization, segmentation, and fitting of molecular models.

ClustalW2 is a server-based alignment tool for protein sequences (Larkin et al. 2007).

Bfactor (v1.04) is a free program to apply amplitude correction to 3-D density maps by multiplying higher frequencies by certain values derived from a temperature factor (download link: http://emlab.rose2.brandeis.edu/grigorieff/download_b.html).

EMBfactor is a free program comparable to the program bfactor. It incorporates algorithms for automatic amplitude correction and noise suppression (Rosenthal and Henderson

2003; Fernández et al. 2008).

Ctffind3 (v3.2) is a free program for the estimation of the CTF of experimental images (Mindell and Grigorieff 2003).

EMAN1 (v1.9) is a free software package for three-dimensional reconstruction using non-weighted, correlation-based alignment algorithms. (Ludtke, Baldwin and Chiu 1999).

EMAN2 (2.0.6) is a free software package for three-dimensional reconstruction and the successor of EMAN1. In this work, it was mainly used for particle selection and 2-D classification purposes (Tang et al. 2007).

Findctf2d (v1.0) is a free program comparable to ctffind3. It is operated by using a graphical user interface (Grant 2007).

FREALIGN (v8.10) is a free software package for the refinement of already existing 3-D reconstructions using a weighted correlation-based algorithm. Alignments and 3-D reconstructions are entirely calculated in fourier space. The algorithm will be described more closely in section D1.2.2 (Grigorieff 1998; Grigorieff 2007).

IMAGIC-5 (v110325; ImageScience, Berlin, GER) is a commercially available software package for the analysis of electron microscopic images, including 2-D classification, 3-D reconstruction and 3-D refinement (van Heel et al. 1996). It incorporates angular reconstitution algorithms, as well as projection matching functionality using traditional cross-correlation-based alignment algorithms without weighting functions.

LinoColor 6.0.12 – Apple Macintosh version – is a scanning program used to digitize the micrographs obtained from cryo-TEM.

Modeller (v. 9.9 and 9.10) is a free software used for homology modeling and loop refinement (Sali and Blundell 1993; Fiser, Do and Sali 2000).

Molprobity (v. 3.18) is a free-to-use web server for the structural validation of protein models (Davis et al. 2007).

RELION (v1.1) is a free software package for the analysis of electron microscopic images, including 2-D classification, 3-D reconstruction of heterogeneous datasets and automatic refinement of 3-D reconstructions. Contrary to the traditional one-image-to-one-reference alignment approach, it follows a Bayesian-type algorithm, calculating maximum-a-posteriori estimates by using an expectation-maximization algorithm (Scheres 2012a and 2012b). The algorithm will be described more closely in section D1.2.3.

Segger is a plug-in for UCSF Chimera, which uses the water-shed method to segment density maps (Pintilie et al. 2009; Pintilie et al. 2010)

Silverfast (v6.6.1r2; LaserSoft Imaging AG, Kiel, GER) is a commercial software suite used as a scanner control program for the digitization of micrographs.

3 BIOINFORMATICAL METHODS

3.1 DIGITIZATION

Images taken on the FEI Tecnai 12 transmission electron microscope were either recorded with a TVIPS TemCam-F416 post-column 4K CCD camera (TVIPS, Gauting, Germany) or with an in-column Megaview II 1K CCD camera (Olympus SIS, Münster, Germany). The latter was used in early EM work, prior to the arrival of the TVIPS camera. The project-specific sections will indicate which camera was used for the respective project.

Images of cryo-EM work on the FEI Tecnai F20 and F30 microscopes were recorded on photographic film, which was digitized with the Primescan D7100 drum scanner, built by Heidelberger Druckmaschinen (Heidelberg, Germany). Depending on the scan program used, the images were scanned either with an 8 bit (Linocolor) or a 16 bit (Silverfast) gray scale.

In order to prevent the formation of Newton rings, the micrographs were mounted onto the drum using a thin layer of oil (Mounting Fluid, SDS AG, Germany) between the drum and the micrographs (Figure 8).

3.2 PARTICLE SELECTION

After digitization, the single particles of each micrograph have to be tagged. This task was either completed with the boxer sub-routine of EMAN1 or the e2boxer sub-routine of EMAN2. For selection, a tight box was used for good particle centering. Tagged particles were digitally cut out of the micrograph and stored in an array of individually addressable pictures in an IMAGIC-5 image file. The box size used for the extrapolation was chosen to be approximately 150 % of the particle size, in order to include information displaced beyond the particle boundary during image acquisition (see section B3.3.3).

3.3 THE PHASE CONTRAST TRANSFER FUNCTION

The phase contrast transfer function (PhCTF, or short: CTF) is a microscope-specific function which is one of the most important factors when considering image formation in the
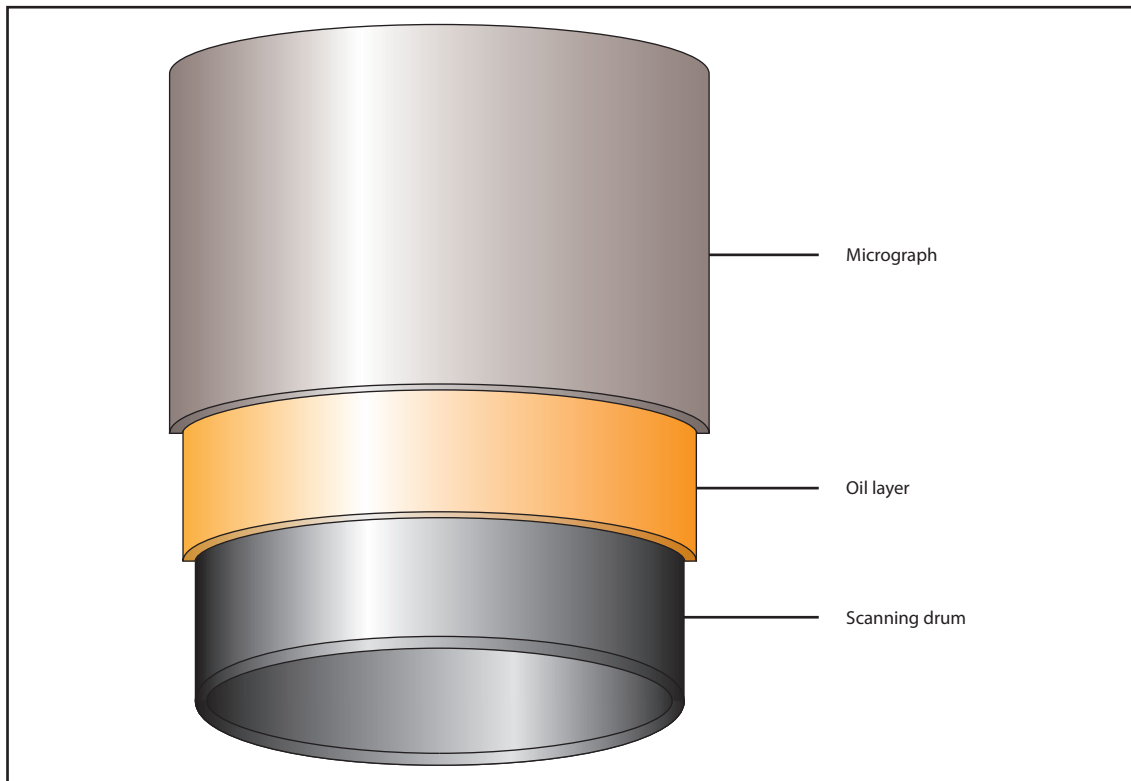
Figure 8   Illustration of the oil layer used between the scanning drum and the micrographs to prevent the formation of Newton rings.

transmission electron microscope. The following sections will (i) describe phase contrast and (ii) the nature of the CTF, and (iii) highlight some of the crucial factors influencing this function. For an in-depth analysis of this topic, I highly recommend a book by Rolf Erni, termed "Aberration-corrected Imaging in Transmission Electron Microscopy" (ERNI 2010).

3.3.1 PHASE CONTRAST

As mentioned in the introduction, biological samples behave as weak phase objects in optical systems, requiring deliberate defocussing of the image in order to see the sample (ZERNICKE 1955). Considering light microscopy, an amplitude object changes the intensity of the light passing through the sample, whereas a phase object only changes the phase of the light. Since beams of light can be described as harmonic vibrations, the interaction between the vibrations of diffracted and direct (unchanged) beams can be depicted in vector diagrams (*ibid.*; Figure 9). For an amplitude object, the incident vibration can be described by vector MA. The scattered vibration, MD, shows less intensity due to the absorbing properties of the sample, whereas the phase remains unchanged (Figure 9 A). The relation between MA and MD can be described by MD'. A phase object, however, changes the phase of the incident vibra-
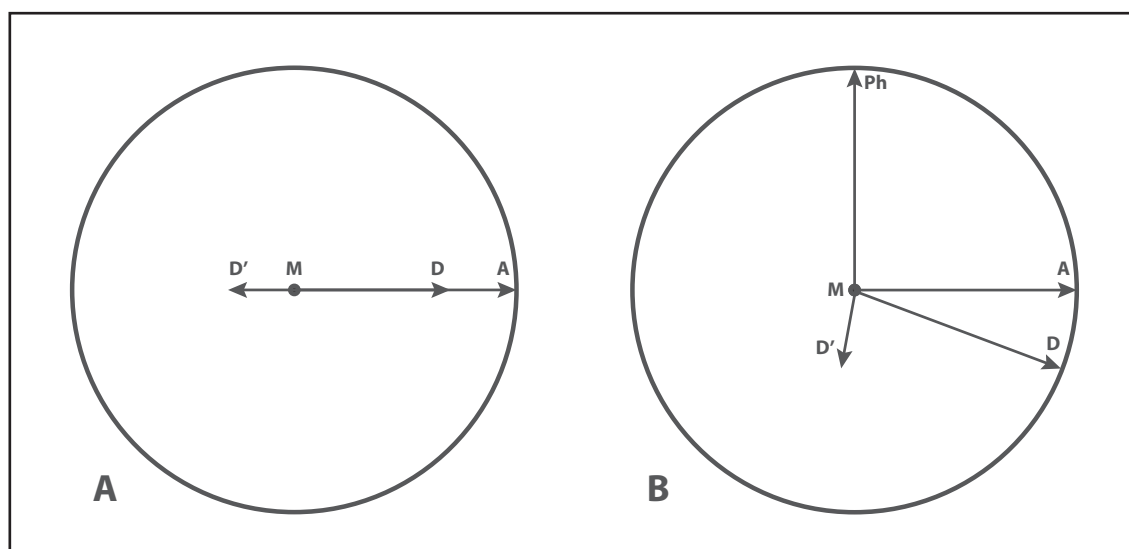
Figure 9   Vector diagrams showing the interactions between the vibrations of diffracted and direct beams of amplitude objects (A) and phase objects (B; with and without phase plate). According to ZERNICKE 1955.

tion MA, leaving the intensity almost unchanged (Figure 9 B). The resulting vector for the scattered vibration can be described by MD and the sample-induced change of MA by MD', which lies almost perpendicular to MA. In this case, the background and the sample do not differ in intensity and hence the sample seems invisible. Only if the phase modulations are transformed into amplitude modulations, phase contrast is obtained and the sample becomes visible. When defocussing, the relative phases between direct and diffracted vibrations change differently and the sample becomes visible due to interference effects, though blurred due to the deviation from the focal point. If a phase plate is used, the phase of the direct vibration is changed by 90° to MPh in the back focal plane of the objective lens. The background – composed of the phase-shifted direct vibration MPh – will thus be lighter than the sample – composed of the vector sum of the direct vibration (MPh) and the change of the incident vibration MA by the sample (MD').

The problems of weak phase objects, which are described above according to ZERNICKE (1955) and which where originally characterized for light microscopy, generally also hold true for electron microscopy of ice-embedded biological samples. Here, amplitude contrast is almost nonexistent (in the order of 7-10%) and the sample remains invisible. Phase plates, which would achieve similar results in the electron microscope as the ones described above, have been under investigation for many years (e.g. UNWIN 1970; DANEV AND NAGAYAMA 2001; MAJOROVITS ET AL. 2007; see Figure 10). The need of meticulous microscope alignments and their ephemerality due to charging effects, however, has confined electron microscopic phase plates to a small amount of laboratories until now. Consequently, acquiring de-
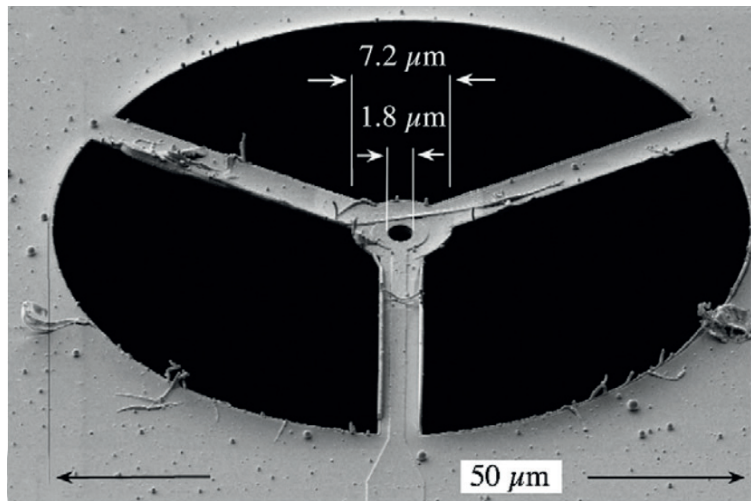
Figure 10 SEM image of a Boersch phase plate (reprinted from Major-
ovits et al. 2007 with permission from Elsevier)

focused images for contrast enhancement is the method of choice for most of the community, implicating certain side-effects for the image quality.

3.3.2 Description of the CTF

The image of a sample obtained in a transmission electron microscope is not an exact copy of the sample's appearance. In general, the spatial frequency spectrum of the image intensity can be described by the one of the sample itself multiplied by the CTF (Wade 1992; Downing and Glaeser 2008). The latter modulates the frequency spectrum of the sample due to lens aberrations and the amount of defocus chosen. It can be described by:

$$CTF(s) = \sin \gamma(s)$$
$$\gamma(s) = 2\pi \left[C_S \lambda^3 s^4 / 4 - \Delta z s^2 / 2 \right],$$

where s is the spatial frequency, $\gamma(s)$ the wave aberration due to the spherical aberration C_S and the defocus Δz, and λ the wave length (Downing and Glaeser 2008). Ideally, the CTF would need to have a constant finite value for all spatial frequencies, up to the aperture-defined maximum (Erni 2010; Figure 11 A). Given the equation above, it becomes apparent that the CTF would be 0 throughout the spatial frequency spectrum, if imaging took place without spherical aberration and under in-focus conditions. This, in turn, would entail the absence of any phase contrast and the sample would be invisible (*ibid.*).

Every TEM, however, has a spherical aberration defined by the design of its respective ob-
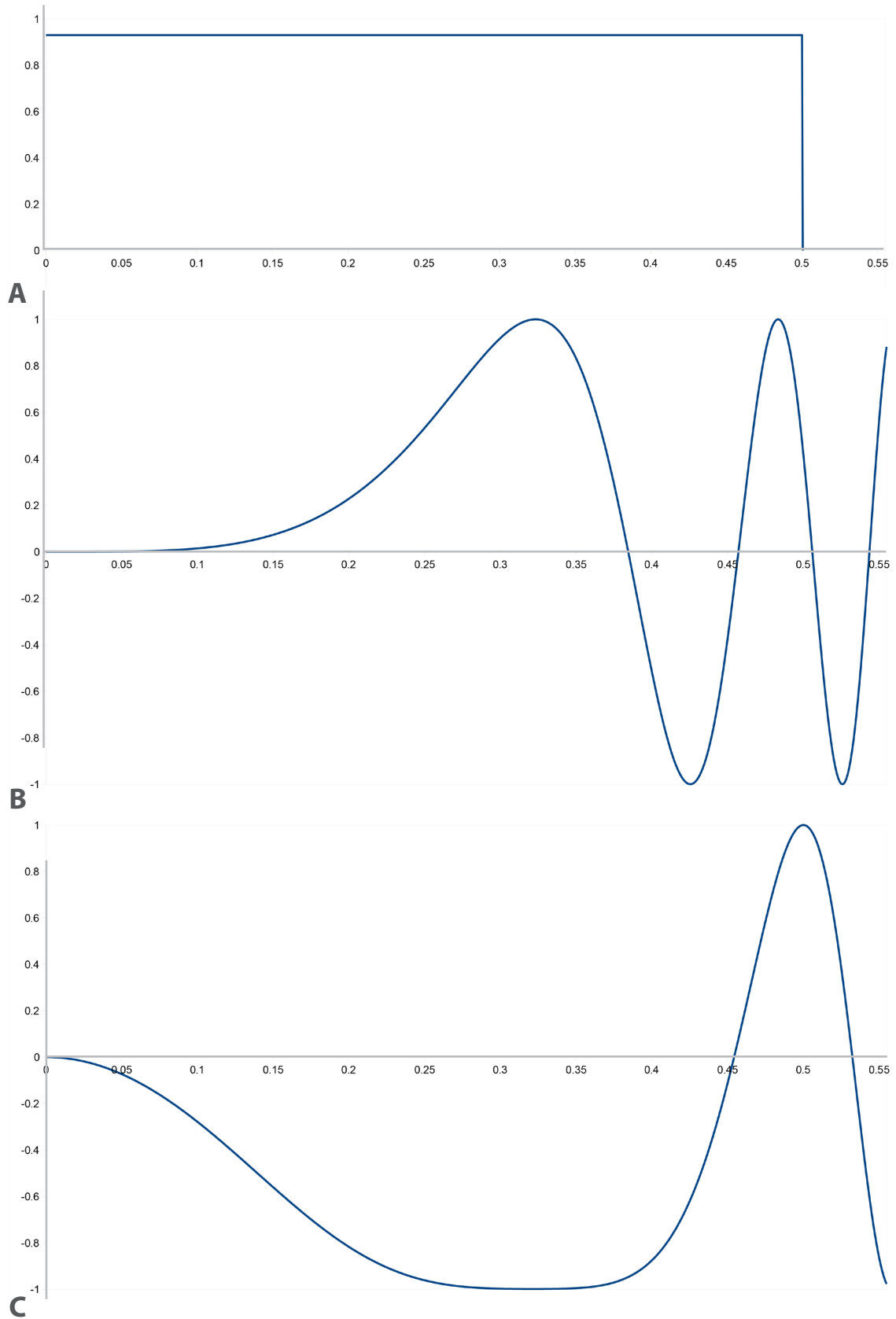
Figure 11 Examples of the contrast transfer function for different conditions: A) The perfect contrast transfer just being limited by the objective aperture's diameter. B) 300 kV microscope with a $C_S = 1.2$ mm and a defocus of 0 nm; C) 300 kV microscope with a $C_S = 1.2$ mm and a defocus of -48 nm; x-axis: spatial frequency (in 1/Å); y-axis: contrast intensity (in fractions of 100 % and -100 %).

jective lens. The implication of a spherical aberration of 1.2 mm in a 300 kV microscope on the CTF can be seen in Figure 11 B. The function only shows contrast for spatial frequencies higher than 7.5 Å and an oscillation along the x-axis, rendering it invaluable for biological samples. Although introducing modulations into the image, spherical aberrations combined with certain defoci can result in a phase-plate-like effect for the image. The CTF for a 300 kV microscope with a spherical aberration of 1.2 mm at an underfocus of -48 nm shows high contrast transfer for a broad scale of spatial frequencies, which approaches the ideal case shown in Figure 11 A up to approximately 2.17 Å (Figure 11 C). This ideal defocus is slightly different for every given microscope type. It was described by Otto Scherzer in 1949 and is hence called the Scherzer focus (Scherzer 1949 *cited in* Erni 2010). Yet again, significant contrast is only seen for spatial frequencies higher than 20-10 Å. Due to the low dose conditions necessary for the imaging of biological samples in ice and their unavoidable low SNR pictures, molecules imaged in Scherzer focus would remain invisible. For these low SNR imaging conditions, contrast must be brought to lower spatial frequencies. As in light microscopy, this can be achieved by defocusing the image. The CTF for the microscope described above, but with an underfocus set to -2000 nm, shows substantially more contrast at the low spatial frequencies than the previous functions (Figure 12 A). This enhanced contrast, however, comes with the price of an increased frequency of oscillations around the x axis.

Figure 12 A shows the contrast transfer function for coherent imaging conditions, meaning that the phases and intensities of all electrons are equal. This approximation can, however, never be achieved in a real microscope due to spatial and temporal incoherences, which generate envelope functions that attenuate the signal of the CTF towards higher spatial frequencies (Figure 12 B). Partial spatial coherence arises due to the finite size of the electron source and results in a non-parallel illumination of points in the sample (Wade and Frank 1977 *cited in* Erni 2010). The envelope function created by partial spatial coherence is partly defined by the defocus, evoking an increased dampening of the CTF at higher defoci. Partial temporal coherences mostly emerge from the chromatic aberration, the finite energy spread of the electrons and the interactions between different electrons of a beam, called the Boersch effect (Erni 2010; Wade and Frank 1977 *cited in* Erni 2010). A number of other circumstances – e.g. horizontal drift or non-ideal recording devices – can cause similar envelope-like image deterioration, although generally not being of comparable importance as the ones described above.
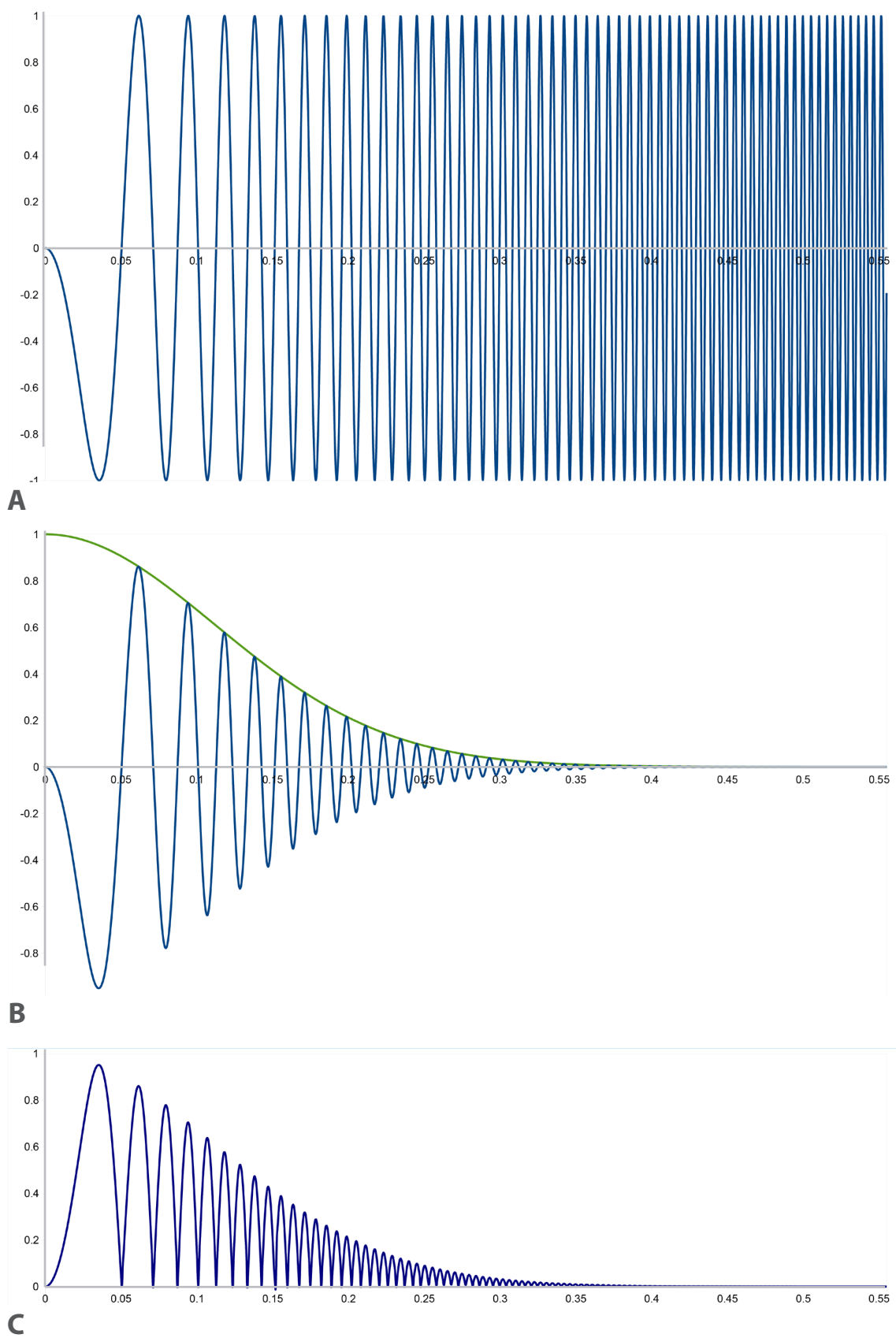
Figure 12 Examples of the contrast transfer function for different conditions and after phase flipping: A) 300 kV microscope with a C_S = 1.2 mm and a defocus of -2000 nm. B) Same as A, but with consideration of the envelope function (green) imposed by partial spatial and temporal coherence. C) CTF shown in B after phase flipping; x-axis: spatial frequency (in 1/Å); y-axis: contrast intensity (in fractions of 100 % and -100 %).

3.3.3 IMAGE CORRUPTION BY THE CTF

The oscillations of the contrast transfer function around the x-axis lead to contrast inversion of the image, as is easily observable in Figure 13. When studying small single particles, e.g. multi-protein complexes, images are further corrupted by a displacement of high frequency information beyond the particle size, which increases with increasing spatial frequency (DOWNING AND GLAESER 2008). The origin of this latter corruption lies in the interactions of both the diffracted beam and its Friedel mate with the undiffracted beam, which create fringes that are shifted in opposite directions (*ibid.*). Provided that the shift is less than half of the particle diameter, these fringes overlap and contrast transfer is still described by the CTF. High frequency information shifted more than half of the particle diameter, however, it can not be described as modulated by sin $\lambda(s)$, but as shifted by $\lambda(s)$ (*ibid.*; see equation above).

3.3.4 ESTIMATION AND CORRECTION OF THE CTF

For high-resolution cryo-electron microscopy, the correction of the described corruptions is essential. The problem arises that the CTF has to be exactly known in order to correct for its implications on the image, which requires exact knowledge of the defocus values of an image. Any deviation from the real CTF could result in destructive interferences at higher spatial frequencies due to the increasing frequency of oscillations around the x-axis. The estimation of the defocus values is performed *a posteriori* from the micrographs or particles, depending on the preferred method. In this work, per micrograph estimation was chosen due to the low SNR in most of the images. By calculating power spectra of the micrographs and comparing them with simulated contrast transfer functions, the defocus and astigmatism values can be accurately described (MINDELL AND GRIGORIEFF 2003; Figure 13 C).

The predominantly used way of subsequent correction for the CTF during this work is known as phase flipping. The contrast inversions imposed on the image by the CTF are corrected for by taking the abolute value of the function (Figure 12 C). This method also restores approximately 65 % of the delocalized information, although the latter is now twice as defocussed (DOWNING AND GLAESER 2008). The remaining 35 % is displaced by twice the amount as it was before (*ibid.*). The information missing between the bell-shaped parts of the function depicted in Figure 12 C can be restored by using different defoci during image acquisition (CONWAY AND STEVEN 1999).
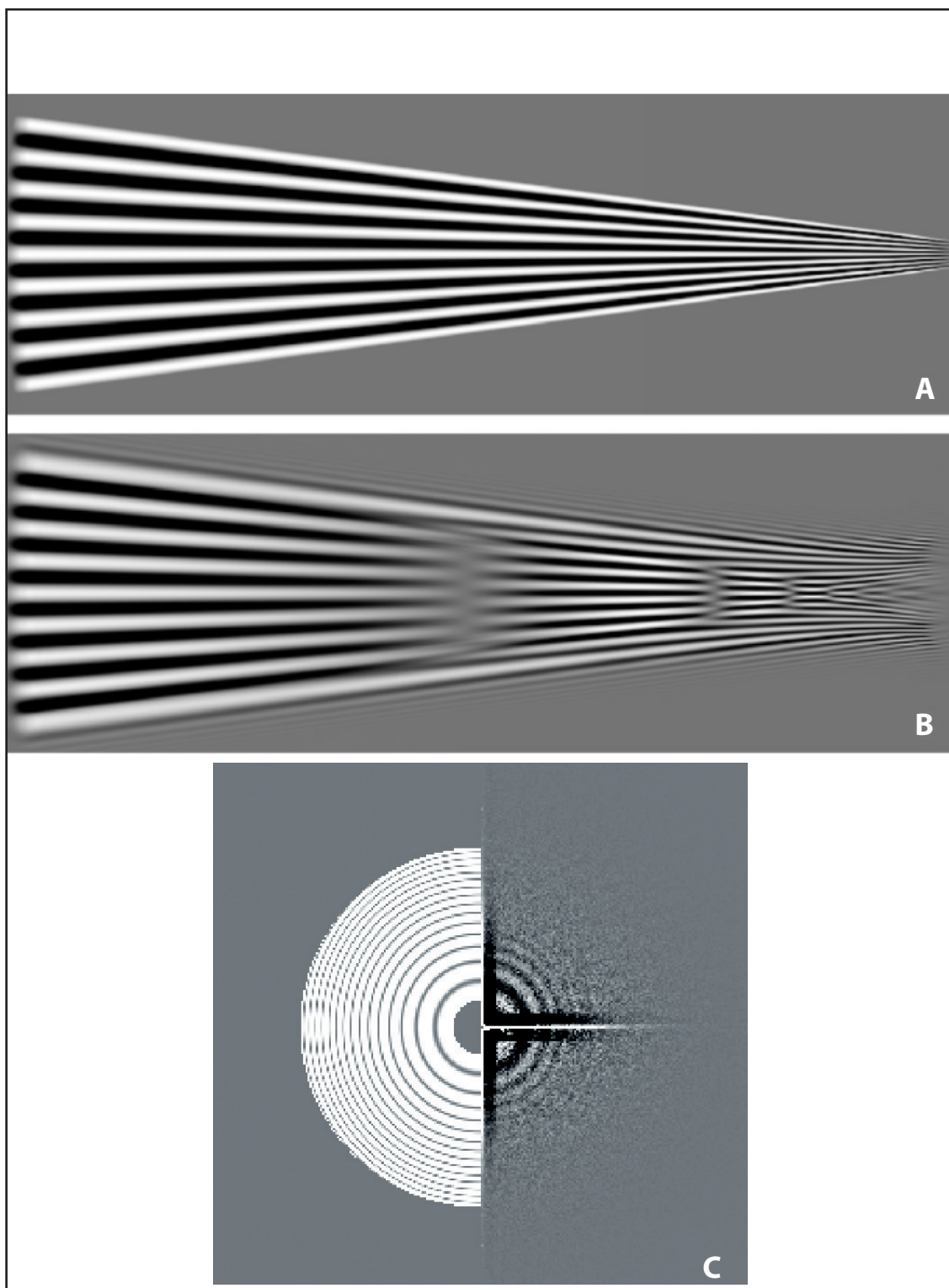
Figure 13 Image corruption by the CTF and *in silico* estimation of the CTF: A and B: Rays of alternating black and white densities simulating different frequencies of a picture (increasing from left to right). A) Original image. B) Image after corruption by the CTF of a 300 kV microscope at 2000 nm defocus unknown C_S (adapted from DOWNING AND GLAESER 2008 with permission from Elsevier). C) Estimation of the CTF by comparing the empirical power spectrum of a cryo-sample (right) to a simulated one (left).

3.4 Particle filtering and normalization

After selecting and correcting the particles, there are still two major issues to be dealt with before being able to process the dataset. Firstly, the images contain a lot of noise, due to the low electron dosage used for taking the pictures. Noise will inevitably hinder the reconstruction process and has to be removed at least partially. Furthermore, every piece of information that is bigger than the protein in question is useless for later steps. Hence, a band-pass filter is used, consisting of a low-pass and a high-pass filter (e.g. Schatz 1992). The low-pass filter cuts off high frequencies in which the noise is situated, whereas the high-pass filter cuts off the low frequency, low-resolution containing sections (*ibid.*). Throughout the iterative process of reconstruction, at least the high cut-off value is modified frequently for obtaining the best possible high-resolution sections without letting through a high degree of noise.

Secondly, the particle images' gray values have to be matched. While every particle image will nearly show a Gaussian distribution of gray values, this distribution tends to be closer to the white or the black side of the gray spectrum. Matching the distribution location is done by a normalization process in which the peak of the Gaussian distribution of the gray values is set to a standard output value without altering the distribution per se, thus being able to compare each image.

3.5 Reference-free alignment

For the subsequent image processing steps, the filtered and normalized particle images have to be well centered. Although great care was taken for the correct centering of the particles during manual particle selection, deviations occur which have to be removed *via* mass-centering. In IMAGIC-5, this is achieved by first averaging all particle images of the dataset, rotating the average in 1° steps and averaging the 360 resulting images again, yielding a statistical center of mass of all images (Figure 14). Further, each particle is translationally aligned to this center of mass of the whole dataset. After the particles are aligned, a new average is created and the particles are aligned again in a second iteration. After three iterations, the particles should be well centered.

3.6 Multivariate statistical analysis

Although band-pass filtering the dataset removed some noise, the particle images' SNR remains quite low. However, it can be improved by averaging images of exactly the same particle orientation due to the summation of the discrete structural information and the suppres-
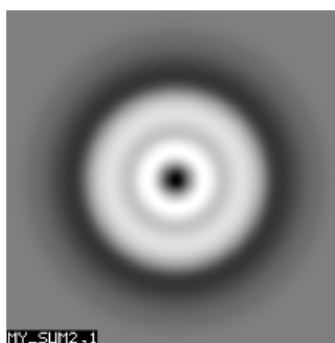
Figure 14 Rotationally averaged total sum of a dataset.

sion of the almost statistically distributed noise. It is critical not to average differing images, as this would result in blurring of the structural information. Consequently, the images have to be classified according to their similarity prior to the averaging. The similarity measurement and classification of images in single particle analysis are generally calculated by performing a multivariate statistical analysis (MSA).

Typically, the multivariate statistics algorithms implemented in the above mentioned software suites for SPA are based on principle component analysis. It is assumed that the gray values of similar images will show less differences than the ones distinctly different from each other. In IMAGIC-5, the procedure consists of three steps, namely (i) the representation of the images in hyperspace, (ii) the compression of hyperspace and the classification of the images, and (iii) the averaging of classes. A mathematically detailed description of the underlying algorithm can be found in van Heel, Portugal and Schatz (2009).

For illustration, let's assume 3-pixel images with differing gray values (Figure 15). Each pixel can be described as an axis in a Cartesian coordinate system, whose value on the axis is the pixel's gray value. Each image can thus be represented by one point in hyperspace whose location is defined by a 3-element vector composed of the image's gray values. Typically, datasets are comprised of more than three images and more than three pixels per image. Assuming n images with m pixels, an m-dimensional hyperspace will be spanned and a 'cloud' of n image points formed, in which similar images will be situated closer to each other than will be differing ones.

Just picturing the numbers of a dataset comprising 200,000 images with a 256 pixel edge length, which would result in 65,536 axes and 200,000-element vectors, illustrates that computational analysis of the results would be overly demanding. As a consequence, the results are compressed by redefining the m-dimensional coordinate system. Using the so-called
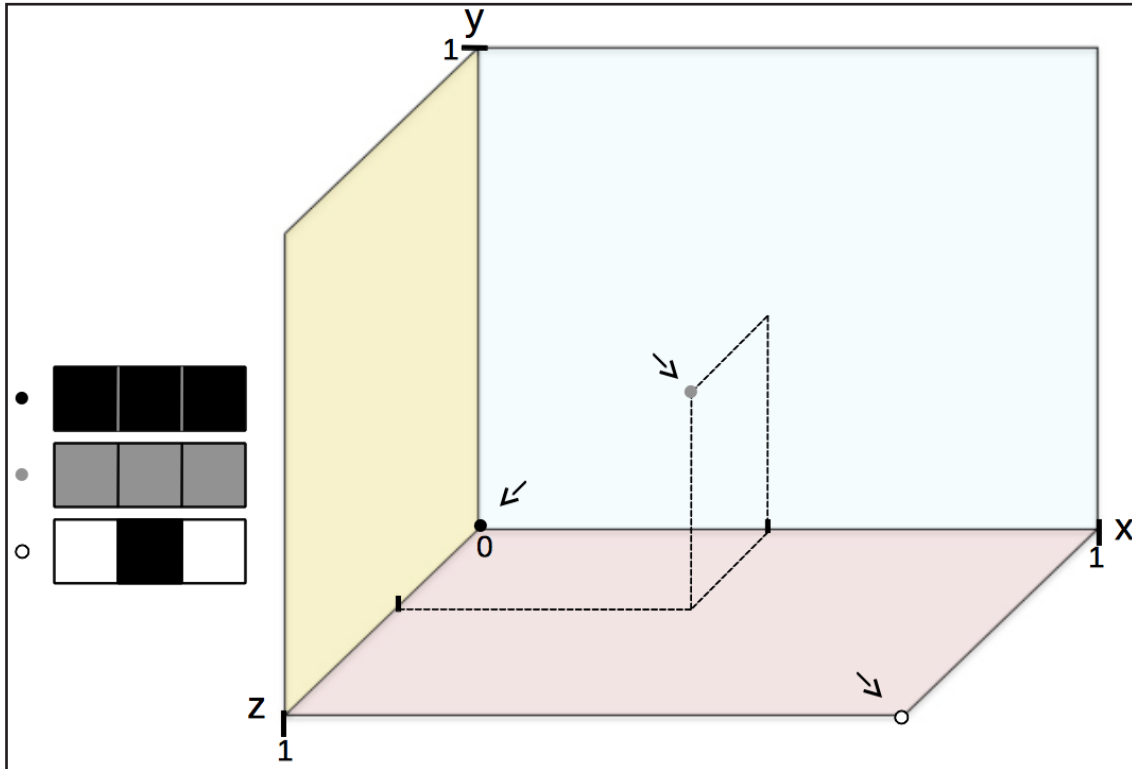
Figure 15 Image representation in MSA hyperspace visualized by three images containing three pixels each. Axes x, y and z depict the brightness values of pixels 1, 2, and 3, respectively, where 1 is white and 0 is black. Note the different distribution of the three images in hyperspace (arrows). The dashed lines clarify the position of image 2 in the three-dimensional space.

modulation metric (BORLAND AND VAN HEEL 1990), the dataset is normalized by its standard deviation and the inner product between a defined number of randomly chosen orthonormalized eigenimages and the images of the dataset is calculated (VAN HEEL, PORTUGAL AND SCHATZ 2009). After calculating weighted sums of the dataset images, a new set of vectors (eigenimage approximations) is calculated (*ibid.*). Finally, the algorithm will converge towards the true eigenvectors/-images. These operations lead to a rotation of the original coordinate system in a way that the first axis describes the center of mass of the dataset, the second axis the greatest dimension of the data cloud, i.e. the greatest variance within the dataset, the third axis the second greatest variance and so forth (SCHATZ 1992). The compression of the dataset is achieved by choosing a small amount of initial random unit vectors. Typically, up to 50 eigenvectors will suffice for the accurate description of the variance in a cryo-EM dataset (VAN HEEL, PORTUGAL AND SCHATZ 2009), leading to a reduction of axes of 100 orders of magnitude!

Following the compression of the hyperspace, the dataset has to be classified. In single particle analysis, automatic classification is normally either calculated using (i) K-means clustering or the (i) hierarchical ascendant classification (HAC) algorithm in combination with

a moving elements post-processor. For a review about K-means clustering see, for example, Bock (2008). Briefly, for a user-specified number k of desired classes, k seeds are randomly placed in hyperspace and each image of the dataset is classified to its nearest seed (van Heel, Portugal and Schatz 2009). For each of the k classes, the seed is replaced by the center of mass of each class, which then serves as new reference for classification and so forth, until image movement between classes stabilizes (*ibid.*). Statistical analyses in this work, however, were calculated with IMAGIC-5 using the second classification method. The HAC algorithm is a "bottom-up" approach that defines each image as one class at first (*ibid.*). The classes are subsequently merged according to the Ward criterion (Ward 1982), which aims at minimizing the intra-class variance, i.e. the difference between images of one class. This automatically results in the maximization of the inter-class variance (van Heel, Portugal and Schatz 2009). The merging process can be visualized in a classification tree (Figure 16).
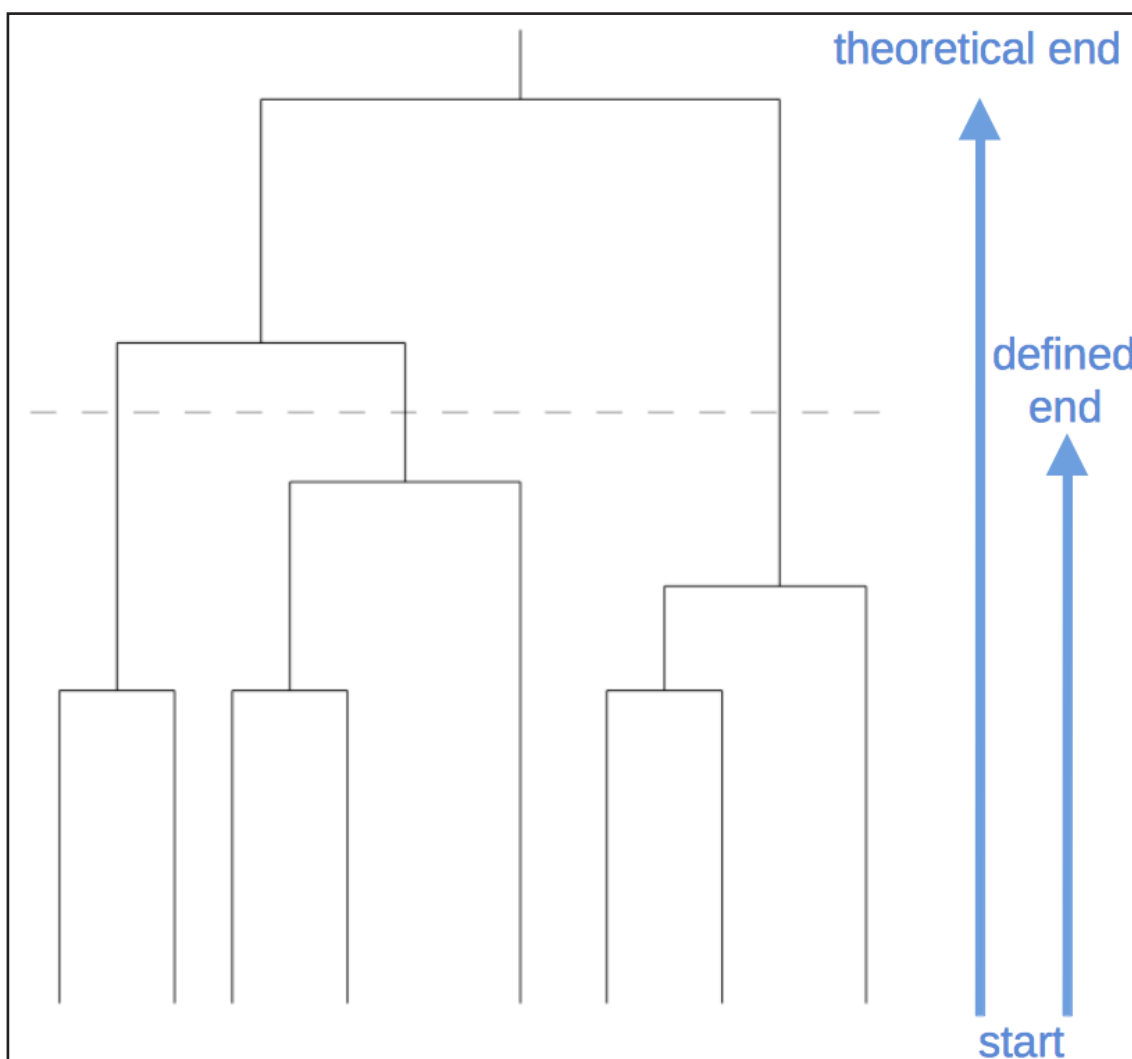
Figure 16 Hierarchical ascendant classification, starting with one image per class and theoretically ending with every image in one class. The dashed line marks the 'user abort' at the defined maximum number of classes (adapted from Schatz 1992).

A user-defined number of final classes has to be specified, as the algorithm would ultimately result in one single class containing all of the images. Typically, this number is chosen to arithmetically yield an average of roughly 10 images per class. Nonetheless, in large datasets, this number can be much higher (*ibid.*). In order to prevent classification of local minima, the moving elements post-processor relaxes the bond between images and their respective classes and iteratively compares the distances between each image and each class (VAN HEEL AND STÖFFLER-MEILICKE 1985; VAN HEEL 1989). Should an image show a longer distance to its current class compared to another one, it is regrouped to the latter (VAN HEEL, PORTUGAL AND SCHATZ 2009). The post-processing typically results in a migration of 20-30% of the images and approximately 10% less intra-class variance (SCHATZ 1992).

Finally, the classes are averaged into class averages (or class sums). Under idealized conditions, in which the structural information in different images is identical and the noise does not show any correlation, the latter is reduced by a factor of 1/(# of images averaged) (*ibid.*).

3.7 MULTIPLE-REFERENCE ALIGNMENT

Images depicting the same view of the particle are not invariably eligible for averaging, if they are shifted or rotated in the image plane. Reference-free alignment (see section B3.5) is able to correct for translational differences, enabling the calculation of class averages out of rotationally equal images of the same particle view. However, a dataset will contain many images of the same particle view which are rotated against each other. These images cannot contribute to one high-SNR class average, if their rotational imparity is not corrected for. Multiple-reference alignment (MRA) correlates an image of the dataset with references and rotationally and translationally matches it to the reference with the highest correlation. In early steps of image processing, these references will be class averages of characteristic views obtained by MSA. In later steps, re-projections (see section B3.11) of 3-D reconstructions will serve as references. Correlations in MRAs are generally calculated *via* a cross-correlation function (CCF; SAXTON AND FRANK 1977) and the basic fundamental algorithm of cross-correlation based translational and rotational alignments will be described subsequently.

In translational alignment, both reference and image are transformed into Fourier space by Fast Fourier Transformation (FFT; VAN HEEL ET AL. 1996). Thereafter, the reference is multiplied with the complex conjugate of the image and the result is transformed back into real space, yielding a cross-correlation function (SCHATZ 1992; Figure 17 A). If a particle is shifted against the centered reference, this function shows a bright maximum whose position
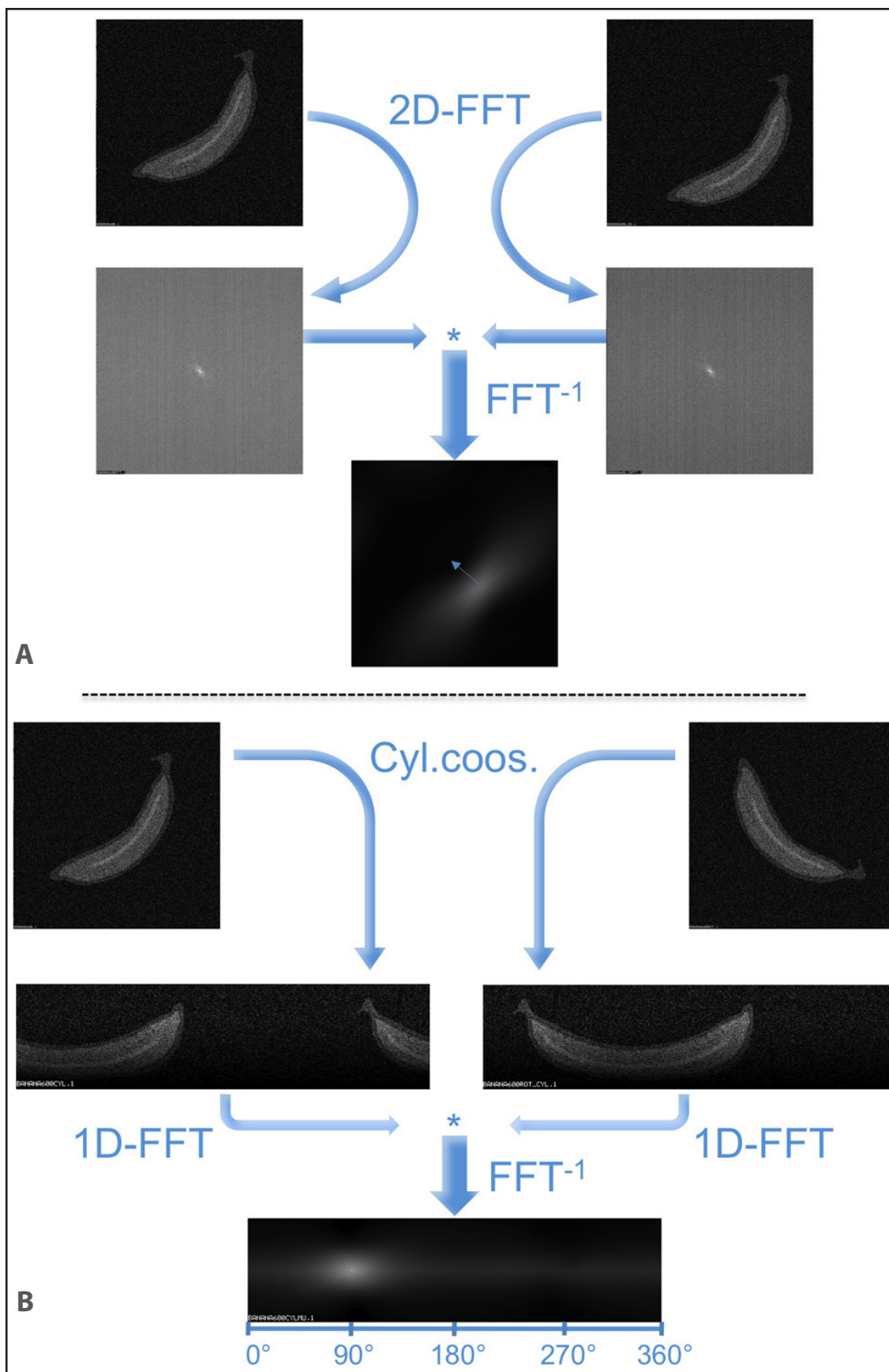
Figure 17 Basic algorithm of an MRA: A) Translational alignment. B) Rotational alignment; see Text for more information; FFT: Fast Fourier Transform; *: complex conjugate multiplication of Fourier transformed images; FFT^{-1}: reverse FFT (1-D or 2-D) into real space; Cyl.coos.: transformation into cylindrical co-ordinate system.

is shifted from the center of the image by the same amount as the particle is shifted from the reference (*ibid.*). The vector of this shift can be read automatically and the particle then shifted towards the reference. The quality of the alignment is assessed by comparing correlation coefficients, which are given by the brightness of the cross-correlation function maximum (*ibid.*).

For rotational alignment, both reference and image are transformed into one-dimensional representations with a cylindrical coordinate system (*ibid.*). Like in translational alignment, the one-dimensional images are subjected to FFT, the reference is multiplied by the complex conjugate of the image and the result is transformed back into real space, yielding an angular CCF (*ibid.*; Figure 17 B). The position of the peak indicates the clock-wise rotation of the image against the reference (*ibid.*).

Both alignment procedures have to be iteratively optimized by a repetitive calculation of alternating rotational and translational alignments. The algorithm compares the correlation coefficients that resulted from the alignment of an image against every reference and uses the highest correlation coefficient as the alignment output (van Heel et al. 2000). Algorithms addressing problems that result from this strict reference-to-image binding will be elaborated in the discussion (section D1).

3.8 Euler angles and symmetry

The ultimate goal of reconstructing a three-dimensional structure from two-dimensional images can only be achieved, if the 2-D images can be accurately described in 3-D space. This is achieved by the introduction of the so-called Euler angles, a set of three angles which rotate the 2-D images around a tri-axial Cartesian coordinate system into the position that corresponds to the 3-D structure of the particle, e.g. top or side view. Inversely, a three-dimensional structure can be rotated to any view by using these angles. The difficulty persisting in the field of single particle analysis when working with Euler angles arises from the implementation of different conventions into the variety of existing software packages. While, for example, IMAGIC-5 and FREALIGN use the zyz-convention, the EMAN packages use the zxz-convention (Baldwin and Penczek 2007; Figure 18). The conventions describe the first angle rotating along the z axis, the second along the y-/x-axis and the third around the z-axis again. Although having the same axial convention, IMAGIC-5 will rotate the object counter-clockwise, whereas FREALIGN will rotate clockwise (*ibid.*). A rotation direction, e.g. around z, is defined by the movement of the 3-D structure seen, if the positive z-axis of a right-handed tri-axial Cartesian coordinate system points towards the viewer (*ibid.*;

Counter-clockwise rotations of the coordinate system are positive rotations per definition, for which the movement of the object appears clockwise (HEYMANN ET AL. 2005).

Caution has also to be taken when symmetrizing a particle with different software suites. For a D3-symmetrical particle, for example, three dihedral point-group symmetry axes exist, around which a structure can be turned by 180° resulting in the same view. Software suites using the zyz-convention, place one of these dihedral axes on the y-axis, whereas suites using the zxz-convention use the x-axis. This results in an imparity of the dihedral symmetry axes towards the symmetrization axis of 30° (for D3 symmetry) when switching from zyz to zxz.
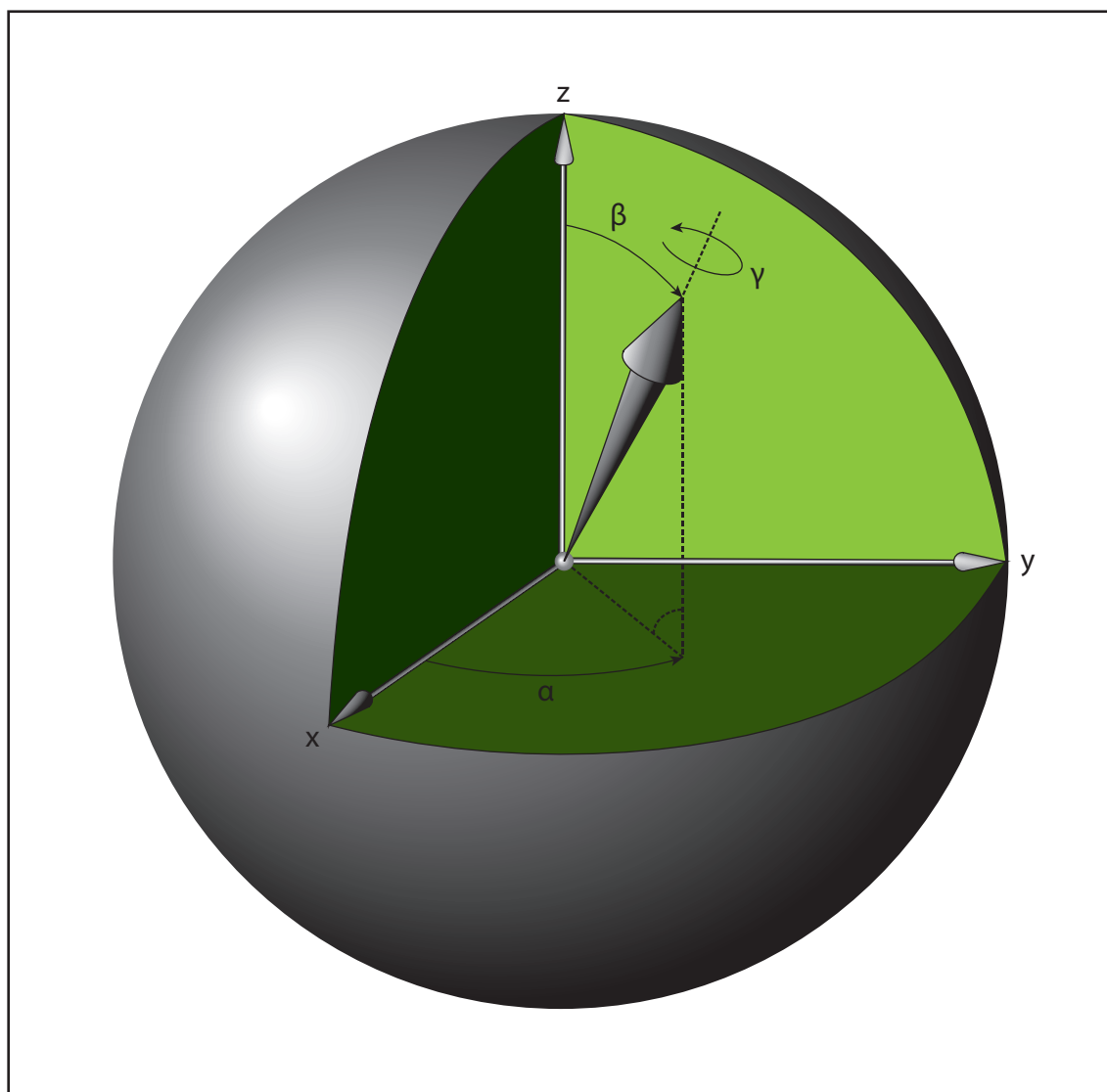
Figure 18 Rotation of an object around the three Euler angles in IMAGIC-5 as described in HEYMANN ET AL. (2005). Rotation directions are given for the object, whereas the coordinate system rotates in the opposite direction. The ZYZ-convention makes the coordinate system rotate around the z-axis (α), then around the newly formed y-axis (β) and finally around the new z-axis (γ). Note that the newly formed z-axes remain being the viewing direction.

3.9 ANGULAR RECONSTITUTION

The calculation of a three-dimensional reconstruction from two-dimensional projections requires the knowledge of the relative orientations between the projection images. In contrast to methods in which the relative orientations are known *a priori*, e.g. single-axis tomography (e.g. HARAUZ ET AL. 1987), single particle analysis requires an *a posteriori* determination due to the random position of the macromolecular complexes on the EM grid (VAN HEEL 1987). Angular reconstitution is a method to determine the relative orientation between two projection images based on a 1-D real-space adaptation of the central section and common line theorems (VAN HEEL 1987; DEROSIER AND KLUG 1968; CROWTHER ET AL 1970; CROWTHER 1971). These theorems postulate that the Fourier transform of any given 2-D projection of a three-dimensional structure is a section passing through the center of the 3-D Fourier transform of its structure. If two 2-D projections are produced from the same 3-D structure, they will necessarily share a line in the center of the 3-D Fourier space that will be identical for those two projections (Figure 19 A). Although the common line describes some of the relative orientation between the two projections, they are still able to freely rotate around it (Figure 19 A). At this point, the common line can thus be described as common tilt axis as well (VAN HEEL 1987). Only if a third projection is added, which has to be genuinely different from the other two, the relative orientations between the three projections can be fixed by the two additional common lines (Figure 19 B). Similarly, a real space 2-D projection will share one 1-D line projection with a second 2-D projection originated from the same 3-D
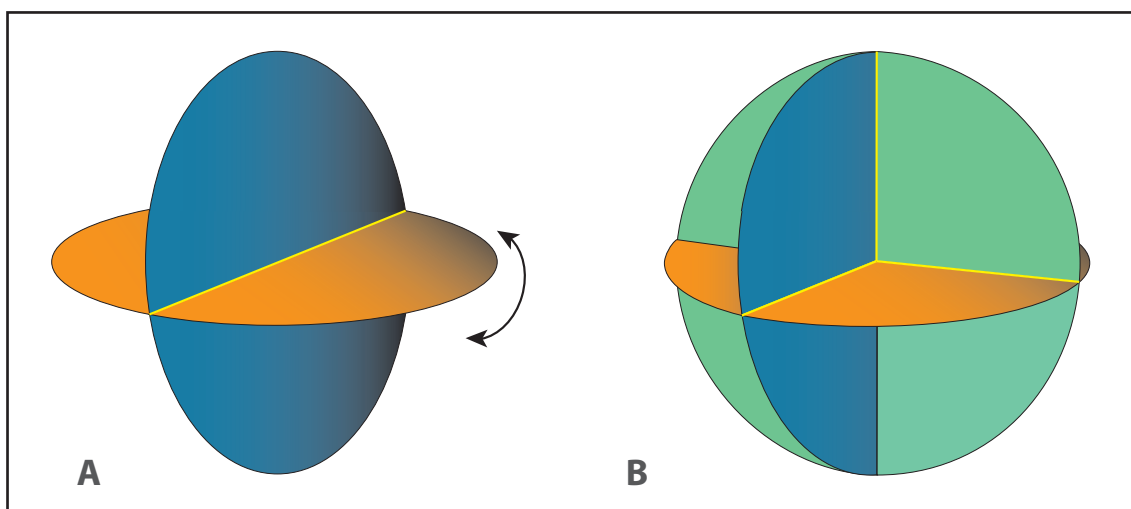
Figure 19 Illustration of 2-D projections of the same object in 3-D Fourier space according to VAN HEEL (1987); A) two slices have one line bearing the same information in common (yellow line) and their relative orientations cannot be fixed because of a possible rotation around this line; B) if a third projection is added, two more common lines fix the relative orientation between the three projections.

structure (VAN HEEL 1987; Figure 20). In order to find this 1-D common line, sinograms are generated for both 2-D projections that consist of 360 line projections each (Figure 20). These sinograms serve for the calculation of a sinogram correlation function (SCF), which depicts the correlation-coefficient between each line of one sinogram with each line of the other one (*ibid.*). The highest correlation-coefficient will determine the 1-D common line and the β-γ-combination of the preliminary relative orientation, which is then fixed through the SCFs between projections one and three and projections two and three. This allows more projections to be included in the search, for which the first three projections serve as a so-called anchor-set (VAN HEEL ET AL. 1996).

In consideration of the vulnerability of the angular reconstitution algorithm to noise, it is
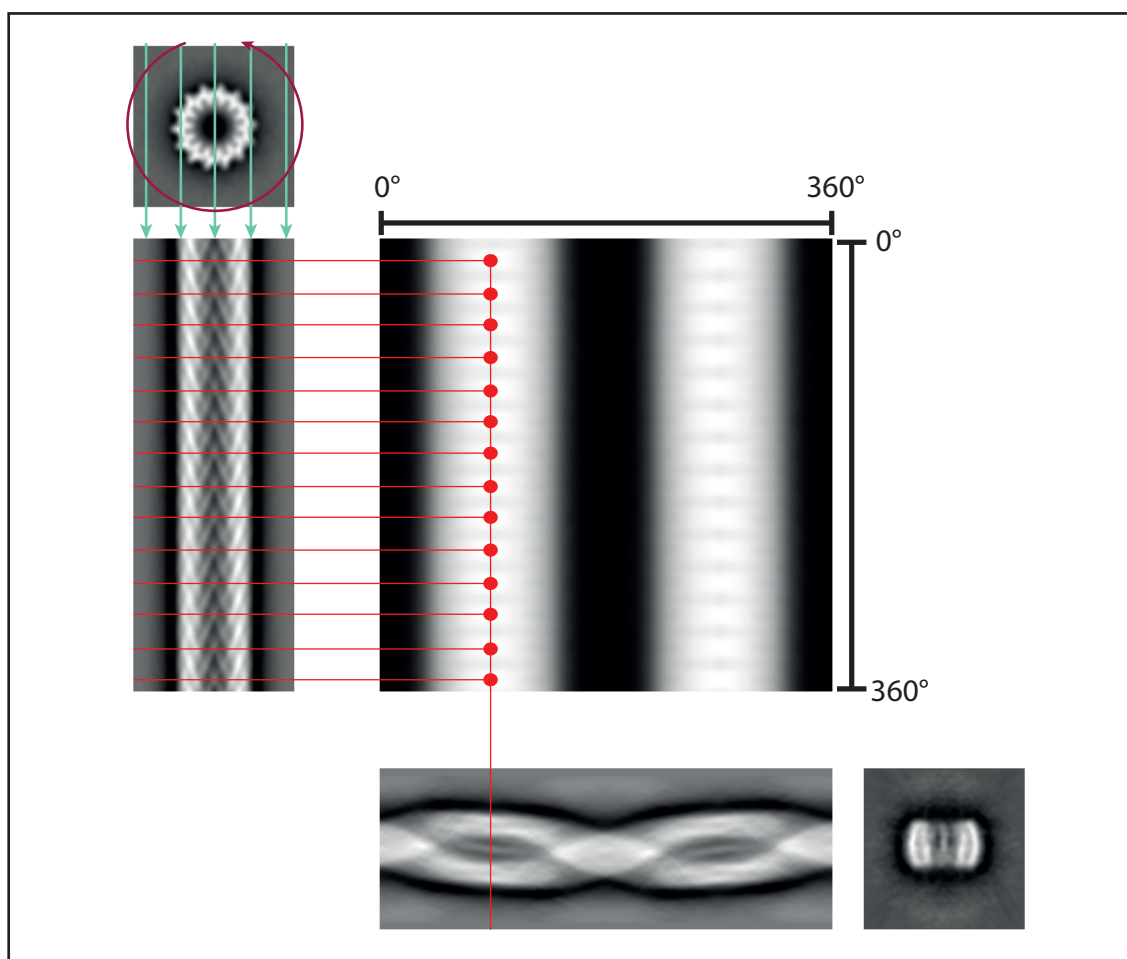
Figure 20 The principle of finding the 1-D common line in angular reconstitution; 360 line projections form a sinogram, of which every line is cross-correlated with every line of another sinogram. These cross-correlations are depicted as a sinogram correlation function (square image in the center) which in this case shows 14 maxima due to the C14 symmetry of the particle (red dots). The maxima correspond to the common 1-D lines (red lines) and give information about the relative β- (top angular ruler) and γ-angles (left angular ruler) between the two projections. The SCF shows a second band of high correlations due to the almost D-symmetrical nature of the particle, but with maxima lower than for the actual common lines.

imperative to use class averages when calculating relative orientations to prevent miscalculations. In this context, it further has to be noted that inherent particle symmetry will lead to multiple maxima in the SCF, rendering the algorithm less susceptible against this determination of false maxima.

3.10 THREE-DIMENSIONAL RECONSTRUCTION

The final step to obtaining a three-dimensional structure from the processed, noise-reduced projection images with determined relative orientations, is the 3-D reconstruction itself. As mentioned above, each 2-D projection can be represented in 3-D Fourier space as a plane that goes through the center of the three-dimensional space (central section theorem; e.g. DEROSIER AND KLUG 1968; CROWTHER ET AL. 1970). When adding many 2-D projections in 3-D Fourier space, they will, to a certain extent, describe the underlying 3-D structure, which can then be transformed into real space according to the back projection theorem (SMITH 1981, *cited in* SCHATZ 1992).

A consideration necessary to be taken into account, is that the structures considered in single particle analysis have a finite size in real space. The central sections are, consequently, not planes being infinitesimally thin, but 'slabs' with a defined thickness of 1/D, where D is the largest linear dimension of the object (HARAUZ AND VAN HEEL 1986; VAN HEEL 2000). The finite dimension of the slab leads to an overlap of different sections in the low-frequency region of the 3-D Fourier space, which amplifies the low frequencies and defines the resolution limit of the reconstruction (*ibid.*; Figure 21). In order to minimize the amplification of the low-frequency components, high-pass filters were used for the reconstructions at first, and were substituted by projection-specific filters (exact-filter algorithm; *ibid.*). In the most convenient case of an equal angular distribution of noiseless projections, the highest obtainable isotropic resolution can be defined by the region where the central sections just cease to overlap (DEROSIER AND KLUG 1968). This is

$$1/f_d \leq \pi D \, / \, n * N_{SYM},$$

where f_d is the maximum achievable resolution, n the number of projections and N_{SYM} the point-group symmetry of the structure (VAN HEEL 2000).

3.11 PROJECTION MATCHING

Once a 3-D reconstruction is calculated, the MRA references will be created by re-projecting different views of it *in silico* (simulated TEM projections). Aligning the dataset to references

which were re-projected from a 3-D structure has two advantages, namely that (i) the SNR is much better than for references from class averages and that (ii) the references already bear the Euler angles that the algorithm re-projected them from. When calculating an MRA, each image aligned to a certain reference can take over the Euler angles from this reference and be reconstructed without further processing (HARAUZ AND OTTENSMEYER 1984). Multivariate statistical analysis or angular reconstitution are not necessary in this case. Although being a typical method for the later stages of an iterative structural analysis, projection matching can theoretically be used right from the start by using for example a sphere with added noise as reference. Principally, projection matching yields better results than angular reconstitution, but is also dependent on a good alignment and accurate references, thus being prone to artifacts. The draw-backs of this method and the different approaches found among software suites to minimize them, will be presented and critically evaluated in the Discussion.

3.12 MASKING THE 3-D RECONSTRUCTION

A three-dimensional reconstruction will contain a certain amount of noise surrounding the actual structure. When used as a reference, the accuracy of the alignment would decrease, if the noise would remain in the objects periphery. Different approaches exist to eliminate this
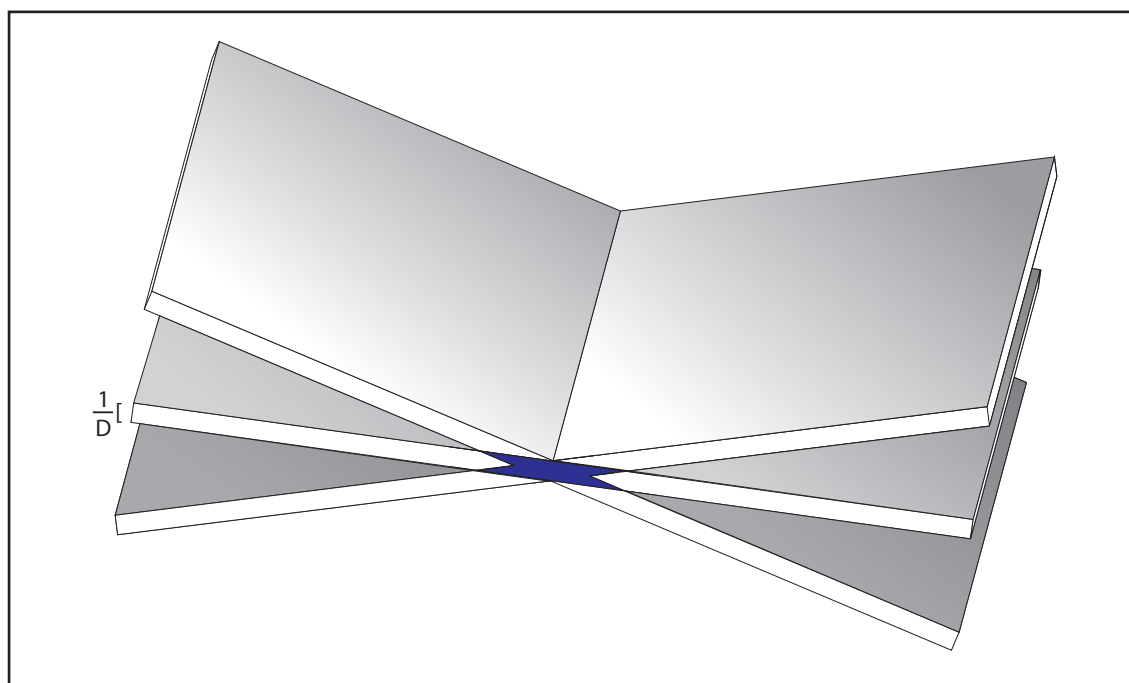
Figure 21 Illustration of overlapping 2-D projections in 3-D Fourier space; the region of overlap is colored in blue and denotes the maximum achievable resolution and the area of spatial frequency amplification. Note the wedges not filled by the central sections, showing the necessity of uniform and dense projection distribution. Adapted from VAN HEEL 2000.

noise, ranging from spherical masking and manual masking to the application of automatic contoured masking. The approach used for this work was the one followed by EMAN and EMAN2, using a flood filling algorithm on the basis of a given radius, threshold, dilation (generally 3 times) and Gaussian drop-off (generally 16 pixels; see the program documentation for details).

3.13 Resolution determination

Determining resolution has been a source for discussions ever since the beginning of three-dimensional reconstruction of electron-micrographic single particles (see for example van Heel 2005). At the beginning of a project, the Fourier Shell Correlation was used primarily to assess improvement in resolution, whereas reaching more elevated resolution levels required the combination of FSC with the assessment of the actual 3-D reconstruction (general appearance, secondary structures, filtered PDB models, etc.; e.g. Zhou 2008), or even using the so-called gold standard FSC, which will be covered more deeply in the discussion (e.g. Scheres 2012a and 2012b). The non-gold-standard FSCs were always calculated using masked density maps.

The process of obtaining a three-dimensional reconstruction is displayed in Figure 22. The approach shown there is the general IMAGIC-5-type work flow when no prior three-dimensional information is known. Other approaches follow brute-force-like work flows by starting refinement from randomly oriented projections or using a sphere with added noise as reference. Heterogeneity cannot be considered with the depicted work flow, but it will be addressed at a later stage.
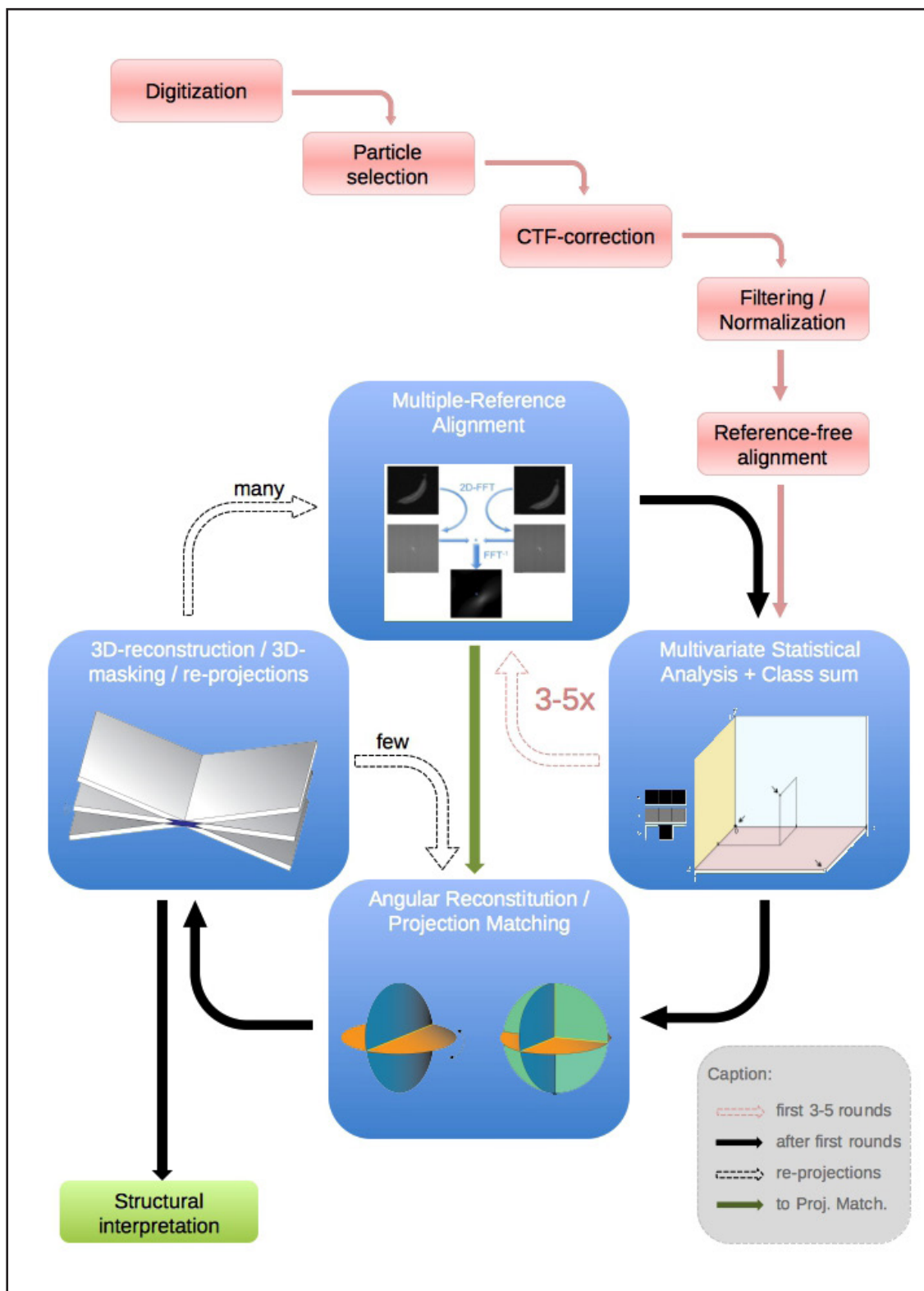
Figure 22 Flow-chart of an IMAGIC-5-style reconstruction process: After preprocessing the images, an iterative cycle of MSA and MRA is calculated to yield high SNR class averages, which are then subjected to the angular reconstitution algorithm. After determining the relative orientations, a 3-D reconstruction is calculated and references are re-projected (few references for angular reconstitution, as many as possible for MRA). In a later stage of refinement, MSA and angular reconstitution can be substituted by projection matching (green arrow). Finally, the refined 3-D reconstruction is structurally interpreted.

4 Myriapod hemocyanin

4.1 *Scutigera coleoptrata* hemocyanin

Three different datasets of *Scutigera coleoptrata* hemocyanin were compared for this study, two of which show an oxygenated form (oxy-ScoHc) and one a deoxygenated form (deoxy-ScoHc). The control dataset consisting of oxy-ScoHc was scanned at the JGU Mainz but imaged and processed independently at Brandeis University, Boston (MA, USA). Due to the fact that the results of three different software packages for three-dimensional reconstructions should be compared, a detailed list of the variables used for each program and iteration will be provided in Appendix C.

4.1.1 Specimen preparation

The first dataset of oxy-ScoHc was obtained as described by Markl et al. (2009). Briefly, animals collected on the island of Giglio, Italy, were dorsally punctured in a non-lethal procedure with a syringe after being anesthetized by cooling them down for 30 minutes at 4°C. The harvested hemolymph was suspended in 1 mM Pefabloc®, a protease inhibitor, and centrifuged in an Eppendorf 5424 centrifuge (5 min; 10,000g; room temperature) to exclude cell remnants. The supernatant was centrifuged over-night in an air-driven Beckman ultracentrifuge (air-fuge; 130,000 g; 21°C) to pellet the hemocyanin, which was subsequently dissolved in Tris buffer A (see Appendix A) and vitrified under oxygen-rich conditions (3.5 µl per grid at 3 g/l; 95-97 % relative humidity; 3-5 s double-sided blotting; atmospheric conditions: 25% O_2 and 75% N_2). Electron microscopic images were taken with an FEI Tecnai F30 transmission electron microscope (acceleration voltage: 200 kV; C_S=1.2; nominal magnification: 59,000x; nominal electron dose: 28 e^-/Å²; nominal underfocus: 1.4-2.5 µm) and Kodak SO-163 film (Agfa-Gevaert, Mortsel, Belgium). The micrographs were developed for 12 min and fixed for 10 min at 21°C (see Appendix A).

The protein for the deoxy-ScoHc dataset was purified by ▮▮▮▮▮▮▮▮▮▮ from animals collected in the proximity of Boston (MA, USA). The hemolymph was purified as described above, but using Tris buffer B as diluting agent (see Appendix A). Protein concentration and structural integrity were measured *via* UV-Vis spectroscopy with the assumption of 1 g/l at an OD_{280} of 1 and the presence of an absorption peak at OD_{340}, respectively. After over-night incubation of the sample in a pure nitrogenic atmosphere, the sample was vitrified by ▮▮▮▮▮▮ in a modified Gatan Cryoplunge3™ which allows maintenance of a con-

trolled atmospheric condition until freezing (sample volume, concentration, humidity and blotting as above; atmospheric conditions: 100% N_2). Transmission electron microscopy was performed on a Tecnai F30 microscope (acceleration voltage: 300 kV, spherical aberration: 1.2, magnification: 75,000x; nominal electron dose: 28 $e^-/Å^2$; nominal underfocus: 1-4 μm) on Kodak SO-163 film by ███████████. The development of micrographs was carried out as described above.

For the control dataset of oxy-ScoHc, kindly provided by ████████████████, hemolymph of animals collected in the proximity of Boston (MA, USA) was obtained as described above, although diluted in Tris buffer C (see Appendix A). Using an Eppendorf 5417C centrifuge, the hemolymph-buffer mix was centrifuged for 15 min at 6000 RCF and 4°C. Ultracentrifugation of the supernatant (3 hours; ~130,000 g; 4°C) was performed using a Beckmann Centrifuge TLA 120.1 rotor and 8 x 35 mm polyallomere tubes (centrifuge type not known). The pellet was resuspended in Tris buffer C. The sample was vitrified under standard atmospheric conditions (sample volume, concentration and blotting as above) and images were taken with an FEI Tecnai F30 microscope (acceleration voltage: 300 kV; C_S=2.0; nominal magnification: 59,000x; nominal electron dose: not known; nominal underfocus: 1.5-4.5 μm) and SO-163 film.

4.1.2 DIGITIZATION AND PREPROCESSING

The three datasets were scanned at the JGU Mainz with the Heidelberger Primescan D7100 drum scanner (see above). The densitometer step size was chosen to yield a resolution of 1 Å/pixel (oxy-ScoHc datasets) or 0.5 Å/pixel (deoxy-ScoHc) in the magnified images, which required scanning with 4305 dpi for oxy-ScoHc and the control dataset of oxy-ScoHc, and 6718 dpi for deoxy-ScoHc.

Particle selection was carried out in the boxer subroutine of EMAN1.9 for oxy-ScoHc and deoxy-ScoHc, whereas the selection tool for the control dataset is not known. The contrast transfer functions for the two datasets processed at the JGU Mainz were estimated in CTFFIND3. The datasets were subsequently averaged to a resolution of 2 Å/pixel (edge length 256x256) and normalized using the relion_preprocess routine of RELION1.1 with a bg_diameter variable set to 100.

4.1.3 EMAN1.9 REFINEMENTS

Prior to normalizing the datasets, they were corrected for the image corruption introduced

by the CTF *via* phase flipping in IMAGIC-5. For both datasets, a 3-D reconstruction of deoxy-ScoHc was used as a reference in the refine subroutine of EMAN1, which was set up to take advantage of the 'usefilt' option. The latter should minimize the over-refinement of noise components in the images by calculating the alignment steps with a dataset low-pass filtered to 12 Å and the reconstruction with an unfiltered dataset. After each iteration, the resulting 3-D reconstruction was low-pass filtered to the resolution measured and masked with a contoured shell using EMAN1.9. The variables given for each iterative cycle can be seen in Appendix C.

4.1.4 FREALIGN REFINEMENTS

Similar to the EMAN refinements, a 3-D reconstruction of deoxy-ScoHc was chosen as a reference. The parameter file was created *via* IMAGIC-5 after a multiple-reference alignment using the deoxy-ScoHc reference low-pass filtered to 20 Å. The refinement procedure used in FREALIGN also included CTF correction by the program's own algorithm which bears resemblance to a Wiener filter (Grigorieff 2007). Analogous to the EMAN refinements, the reconstructions resulting after each iteration were low-pass filtered according to the resolution measured and masked by a contoured shell created by the automask2 subroutine of EMAN1.9. The adjustment of different variables during the refinement process can be seen in Appendix C.

4.1.5 RELION REFINEMENTS

For both datasets, STAR files comprising the particle parameters were created by using the parameter file of the FREALIGN refinements. RELION's 3-D auto refine algorithm was used with the same reference as above with the parameters depicted in Appendix C.

4.1.6 STRUCTURAL INTERPRETATION OF THE DATA

The results were visualized in UCSF Chimera and structural interpretation followed after aligning the reconstructions to each other using the "fit in map" subroutine. The homology models of the open state of ScoHc described in Markl et al. (2009) were used for the assembly of pseudo-atomic models for each final reconstruction. Individual subunits were arranged into the hexameric form by template-matching them with the structure of *Panulirus interruptus* hemocyanin (Volbeda and Hol 1989; PDB code: 1hcy), and subsequently placed into the respective reconstructions as whole hexamers by rigid body fitting. Simulated maps were

created from the pseudo-atomic models by using the molmap subroutine in UCSF Chimera, in order to enhance the visibility of structural variations between the different physiological states.

4.2 *Polydesmus angustus* hemocyanin

Purified hemocyanin was kindly provided by the ▮▮▮▮▮▮▮ at the University of Hamburg, Germany and prepared for electron microscopy by ▮▮▮▮▮▮ at the JGU Mainz, Germany. Digitization, preprocessing and refinement was conducted under my supervision by ▮▮▮▮▮▮▮ as part of his Bachelor thesis (Neumann 2013).

4.2.1 *Specimen preparation*

Animals were collected near Hamburg, Germany, and hemolymph was purified as described in Jaenicke et al. (1999). Negatively stained samples were prepared by using a solution of 2% uranyl acetate at pH 2 and a method resembling a combination of the single droplet and the OpNS methods described by Harris and Agutter (1970) and Zhang et al. (2012), respectively. Briefly, 5 μl of sample was pipetted onto a negatively glow discharged continuous carbon grid, incubated for one minute, washed with 3x20 μl of destilled H_2O, blotted on the edge of the grid with a filter paper, contrasted for 45 s with 5 μl of staining solution and then blotted dry on the edge of the grid with filter paper. Images were taken on an FEI Tecnai 12 transmission electron microscope (acceleration voltage: 120 kV; C_S=6.3; nominal magnification: 131,000x; nominal underfocus: 0.5-1.5 μm) with the Megaview II 1K CCD camera.

Samples for cryo-electron microscopic images were prepared by Frank Depoix with the Gatan Cryoplunge3™ mentioned above, using the settings described for the oxy-ScoHc sample, but in normal atmospheric conditions. The samples were imaged with an FEI Tecnai F30 microscope (acceleration voltage: 300 kV; C_S=1.2; nominal magnification: 49,000x; nominal electron dose: 20-26 $e^-/Å^2$; nominal underfocus: 4.5-6 μm) and SO-163 film. The development of micrographs was carried out as described above for ScoHc.

4.2.2 *Digitization and preprocessing*

For the negatively stained images, digitization had evidently taken place inside the microscope with a final resolution of 4.98 Å/pixel. Particle selection was carried out in the e2boxer subroutine of EMAN2. The picked particles were written into IMAGIC-5 format, normalized and band-pass filtered between 300 Å and 20 Å.

The micrographs of the cryo dataset were digitized using a Heidelberger Primescan D7100 drum scanner (see B3.1) and a densitometer resolution of 3907 dpi to yield a final resolution of 1.33 Å/pixel in the magnified images. E2boxer was used for particle selection and the particles were written into IMAGIC-5 format. Subsequently, the CTF was estimated by CTFFIND3 and the images were corrected for it *via* phase flipping in IMAGIC-5. The data was normalized and band-pass filtered in IMAGIC-5, where the high-pass filter values were first chosen to be 20 Å and later 15 Å.

4.2.3 CRYO-ELECTRON MICROSCOPIC DATA

The cryo data was subjected to MSA-MRA cylces after being averaged to 5.32 Å/pixel (binning factor of 4) and yielded class averages which described a highly heterogeneous dataset. Consequently, a multiple 3-D refinement was calculated in 35 iterations with EMAN1.9 and C1 point group symmetry, using four equal Gaussian blobs with different amounts of added Gaussian noise as references. For reasons outlined in section C1.2.2, only particles belonging to 3x6-mer assemblies were further processed in a 20 iteration EMAN1.9 refinement. In a second refinement in 50 iterations, this restricted dataset was further confined to micrographs showing three or more concentric Thon rings. For the latter, the final reconstruction of the first refinement was used as a reference.

4.2.4 NEGATIVE STAINING DATA

The preprocessed image stack was subjected to multi-variate statistical analysis in IMAGIC-5 as described in B3. Due to the apparent heterogeneity of the dataset, only class members of classes obviously showing a 3x6-meric assembly were extracted and used for further processing. Angular reconstitution of three class averages showing different orientations of a 3x6-mer yielded an initial model, which was improved by adding the remaining class averages to the angular reconstitution algorithm. The result was used in an EMAN1.9 refinement which was calculated in 40 iterative cycles.

4.2.5 STRUCTURAL INTERPRETATION

A multiple-sequence alignment between the hemocyanins of *Polydesmus angustus* subunits 1 and 2 (UniProt entries: I7HPY6 and I7HZS1), *Limulus polyphemus* subunit 2 (UniProt entry: P04253), *Panulirus interruptus* chain A (Uniprot entry: P04254) and *Spirostreptus* sp. subunit 1 (UniProt entry: Q9BHJ9) was calculated using ClustalW2. A homology model of

PanHc subunit 1 was created with Modeller 9.10 using the previously generated model of ScoHcA as a template (MARKL ET AL. 2009). PanHc2 was modeled by using to PanHc1 as a template. A hypothetical pseudo-atomic model was created according to the phylogenetic and structural argumentation covered by MARKL ET AL. (2009).

5 VESICLE-INDUCING PROTEIN IN PLASTIDS 1

All biochemical procedures (i.e. recombinant expressions, purifications, mutations, liposome preparations) were performed by ▮▮▮▮▮▮▮ of the Institute of Biochemistry at the JGU Mainz, whereas I prepared the samples for electron microscopy, imaged and analyzed them *in silico*.

5.1 WILD TYPE VIPP1

5.1.1 SPECIMEN PREPARATION

The procedure of heterologous expression of Vipp1 in *Escherichia coli* BL21 (DE3) cells is described in detail by FUHRMANN ET AL. (2009A and B). Briefly, a His_{10}-tag was N-terminally added to the protein sequence and the construct was inserted into a pET19b plasmid. After insertion into *E. coli*, the bacteria were grown in LB medium and the protein was then purified by sonication, centrifugation, and Ni-NTA purification. For electron microscopy, the samples were negatively stained according to the method described in B4.2.1 with a solution of 2% uranyl formate at pH 2 and imaged with an FEI Tecnai 12 electron microscope (acceleration voltage: 120 kV, C_S=6.3 mm; nominal magnification: 71,540x; nominal underfocus: 0.5-1.5 μm) on a TVIPS TemCam-F416 4K CCD camera.

5.1.2 IMAGE PROCESSING

For a first dataset, images were digitized with a resolution of 4.36 Å/pixel and particles selected in EMAN1.9, followed by band-pass filtering and normalization in IMAGIC-5 without prior CTF correction. Due to the extreme heterogeneity of the dataset observed by calculating and refining class averages *via* MSA and MRA, the multirefine algorithm of EMAN1.9 was used to separate different ring sizes. For this purpose, class averages of top-views showing particular ring sizes (10-ring - 19-ring) were two-dimensionally symmetrized according to their ring size, in order to enhance the SNR and structural information. Each symmetrized

class average was then stacked to the appropriate height in 3-D space and used as a reference (stacked 3-D) for the multirefine algorithm in 50 iterations. One of the resulting reconstructions of the multirefine calculation, an 11-ring, was symmetrized and segmented using Segger in UCSF Chimera. One ring "module" was extracted, manually aligned against the stacked 3-Ds and symmetrized according to the ring size of the references, yielding simulated 3-D structures of different ring sizes (simulated 3-Ds; 10-ring - 19-ring).

For a second dataset, images were digitized with a resolution of 2.18 Å/pixel. Particle selection, CTF estimation *via* CTFFIND3 and CTF correction was performed in Appion. The dataset was averaged to 4.36 Å/pixel, band-pass filtered (600 Å - 15 Å) and normalized in IMAGIC-5. The simulated 3-D structures were used as references for a second multirefine in 30 iterations exploiting the appropriate cylindrical symmetry, with subsequent refinement of the results in IMAGIC-5 in 40 iterations and masking the results of each iteration by a contoured shell in EMAN1.9.

5.2 MUTATIONAL ANALYSES

5.2.1 SPECIMEN PREPARATION

Mutations were inserted into the vector plasmid by primer variation and amplification of the vector plasmid with the Stratagene Quikchange® Site-Directed Mutagenesis Kit (Agilent Technologies, Waldbronn, Germany) and expression of the protein followed as described for wild type Vipp1 (B5.1.1). Sample preparation for electron microscopy and imaging was conducted as described for wtVipp1 (see B5.1.1), but with an image resolution of 4.36 Å/pixel.

5.2.2 IMAGE PROCESSING

Images of Vipp1 mutants with a deleted C-terminus or the N12 mutation were further processed by iterative MSA and MRA cycles. The CTF of the images was estimated in CTFFIND3 and corrected in IMAGIC-5, followed by band-pass filtering and normalization in IMAGIC-5. For the former sample, images belonging to class averages showing top-views, side views of a double-ring assembly, or peculiar squared structures were extracted and separately refined in MSA-MRA cycles. At the end of this class average refinement, the class averages showing top views were 2-D symmetrized according to their ring size, in order to enhance the SNR and structural features.

5.3 VIPP1-LIPOSOME INTERACTIONS

5.3.1 SPECIMEN PREPARATION

For liposome preparation, 0.2 μg of the desired lipids were stored in Chloroform and dried under a nitrogen atmosphere. Residual organic solvent was removed by using a mechanical vacuum pump over night. For a final concentration of 4 μg/ml, the sample was resuspended in 50 μl HEPES buffer (see Appendix A) and subsequently vortexed for 5 min at room temperature to emulsify the sample. For the production of unilamellar vesicles, five cycles of freezing and thawing were performed, in which the samples were first repeatedly submerged in liquid nitrogen over a period of 1 min and then thawed in a water-bath of 37°C. The liposomes were sized to an average diameter of 50 nm by extruding them through a 0.05μM Nucleopore™ Track-etched membrane (Watman, Maidstone, UK). Liposome samples had a final concentration of 1mM, where mixtures of different lipids contained 50% of each lipid. The exact concentrations of Vipp1 samples could not be determined, but were kept stable throughout the experiments. For electron microscopy, dioleoylglycerophosphoglycerol (DOPG; 1,2-di-(9Z-octadecenoyl)-sn-glycero-3-phospho-(1'-rac-glycerol); Avanti polar lipids Inc., Alabaster, Ala, USA), sulfoquinovosyldiacylglycerol (SQDG; Lipid Products, South Nutfield, UK) and monogalactosyldiacylglycerol (MGDG; 1,2-diacyl-3-O-β-D-galactosyl-sn-glycerol; Avanti polar lipids Inc., Alabaster, Ala, USA) were studied in different combinations. Samples of DOPG, SQDG, DOPG+SQDG, DOPG+MDGD, DOPG+Vipp1, SQDG+Vipp1, DOPG+SQDG+Vipp1, and DOPG+MGDG+Vipp1 were prepared as described for wtVipp1. However, 5% ammonium molybdate + 0.1 % trehalose at pH 7.1 was used as staining reagent because of its advantages for liposome studies (see for example HARRIS 1986).

5.3.2 ELECTRON MICROSCOPY

An FEI Tecnai 12 (acceleration voltage: 120 kV; C_S=6.3 mm) was used for imaging the samples at varying magnifications and defoci. The interaction of Vipp1 with the liposomes and possible structural changes induced by the protein were assessed by visual inspection of the sample. It was consistently attempted to image regions showing different thicknesses of staining, due to the probability of their influence on liposome morphology.

6 *Biomphalaria glabrata* acetylcholine-binding protein

6.1 Hemolymph BgAChBP

Protein purification was conducted by ██████████ according to the protocol followed by Lieb et al. 2006. The first batch of cryo-EM samples was prepared and imaged by ████ ██████, and digitized by ███████████████. Particle selection and CTF-correction of the first batch was conducted under my supervision by ████████████████████ as part of her diploma thesis.

6.1.1 Specimen preparation

Cryo-EM samples were prepared and imaged in two batches, of which the first one was processed by Frank Depoix. The sample (1.2mg/ml) was pipetted onto C-flat holey carbon grids (Protochips Inc., Raleigh, NC, USA) which were situated in a modified Gatan Crypo-plunge™ (see oxy-ScoHc; relative humidity: 95-97%; atmospheric conditions: 100% N_2; double-sided blotting time: 3s). Images were taken with an FEI Tecnai F20 TEM (acceleration voltage: 200 kV; C_S=2.0 mm; nominal magnification: 50,000x; electron dose: 28-36 e^-/$Å^2$; nominal underfocus: 2-5 μm) on Kodak SO-163 film, developed for 12 minutes at 21°C in Kodak D19 high contrast developer and fixed in Kodak AGEFIX fixer (Agfa-Gevaert, Mortsel, Belgium).

6.1.2 Digitization and preprocessing

The micrographs were digitized using a Heidelberger Primescan D7100 drum scanner (see above) with a working resolution of 5080 dpi, resulting in a resolution of 1 Å/pixel in the magnified images. Particle selection was carried out with the boxer subroutine of EMAN1.9. The CTF was estimated and corrected for using the software findCTF2D and the dataset was normalized during band-pass filtering in IMAGIC-5.

6.1.3 Further processing

The first batch of the dataset was averaged to a resolution of 4 Å/pixel, low-pass filtered to 20 Å and subjected to MRA-MSA cycles and the multirefine subroutine of EMAN1.9. For the latter, four Gaussian blobs with different amounts of Gaussian noise added to each of them were used as multirefine references and, after 30 iterations, yielded four reconstructions of which two belonged to a dodecahedron-shaped particle, one was a cuboid and one an amor-

phous mass (blob). A subsequent multirefinement in ten iterations was calculated using the cuboid, the blob and the dodecahedron as references, followed by a refinement in ten iterations of each of the three resulting structures with their assigned particle images. In order to verify the result of the multirefinement, the particle images corresponding to the cuboid and the blob reconstructions were each refined against the dodecahedron-shaped reconstruction as reference in 30 iterations and yielded a cuboid and a blob, respectively. The particle images not corresponding to the dodecahedron-shaped structure were discarded. Particles belonging to the dodecahedral structure were extracted and refined against the former. After calculating 13 iterations, gradually increasing the point-group symmetry from C1 to C5 and D5, an icosahedral symmetry was imposed which was in accordance with the last D5-symmetrical reconstruction calculated. After adding the second batch of images to the first one, the full dataset (9770 single particles selected from 151 micrographs) was used to refine the structure in two different ways.

Firstly, a refinement strategy was followed that should prevent – or minimize – the over-refinement of noise components in the images. To achieve this goal, the 'usefilt' option of EMAN1.9 was used with a dataset for the alignments (low-pass filter at 8 Å) and a dataset for the reconstructions (low-pass filter at 3 Å). Furthermore, the reference for each subsequent iterative cycle was low-pass filtered to 10 Å. Fifteen iterations were calculated using the full resolution at 1 Å/pixel and 1° references. Amorphous protrusions, which from the beginning appeared at the exposed face of the pentamers, were carefully masked out after each iterative cycle in order to prevent noise bias in this region. The final reconstruction (8058 particles) was low-pass filtered to 6Å and submitted to EMdatabank.org (accession number: EMD-2055).

Secondly, a traditional refinement was calculated, in which the high-frequency cut-off of the filtered dataset is adjusted according to the resolution measured. In 85 iterations, of which 25 were calculated using the full resolution and 1° references, the protrusions mentioned above were not masked out. The final reconstruction (8347 particles) was low-pass filtered to 6Å.

Resolution estimates were obtained by calculating the FSC of the reconstructions of two halfsets of the data and using the 0.5 threshold. A correction of the envelope function was partly achieved by imposing a negative temperature factor with the programs bfactor (refinement one) or EMBfactor (refinement two) to the final reconstructions.

6.1.4 Structural interpretation

Homology models of both BgAChBP1 and BgAChBP2 were generated by first calculating a multiple sequence alignment with ClustalW2, using the sequences of *Lymnaea stagnalis* AChBP (PDB code: 1i9b), *Bulinus truncatus* AChBP (PDB code: 2bj0), *Aplysia californica* AChBP (PDB code: 2br7) and *Biomphalaria glabrata* AChBP (type 1 and 2). Modeling was performed *via* template homology modeling with Modeller v9.9, using *Aplysia californica* AChBP in complex with HEPES (PDB code: 2br7) as a template. The best model was chosen according to its DOPE score and the result was controlled *via* rigid body fitting in UCSF Chimera. Small adjustments were calculated with Modeller v9.9 *via* loop refinement and rotamers were optimized with UCSF Chimera, before validating the result in MolProbity. Two pseudo-atomic models were created by fitting individual subunits into the density map and assuming homomeric BgAChBP1 and BgAChBP2 assemblies. The models were submitted to the protein databank (PDB) as entries 4aod and 4aoe, respectively.

6.2 Recombinantely expressed BgAChBP1 and BgAChBP2

Recombinant expression and purification of the protein was realized by ▮▮▮▮▮▮▮▮. Digitization, preprocessing and refinement of BgAChBP1 was conducted under my supervision by ▮▮▮▮▮▮▮▮ as part of her Diploma thesis (Braukmann 2012).

6.2.1 Specimen preparation

The results for BgAChBP1 and BgAChBP2 obtained by cDNA sequencing were C-terminally tagged with a His$_6$-tag, cloned into pDEST™-14 expression vectors and expressed in BL21™-AI cells (see Saur et al. 2012 for details). The inclusion bodies containing the expressed protein were purified, dispersed and sonified. The protein was subsequently purified *via* Ni-NTA affinity chromatography (GE Healthcare) and refolded. After concentration of the protein, imidazole was added with a final concentration of 50 mM to prevent protein aggregation *via* the His-tag. As a second approach, proteins were N-terminally His-tagged with a TEV restriction site, allowing for the removal of the His-tag, except for one N-terminally remaining Glycine residue. For a detailed description of the biochemical methodology see Saur et al. (2012) and Möller (2012).

Negatively stained samples of BgAChBP1 and BgAChBP2 were prepared using the method described in B4.2.1 and 2% uranyl formate at pH 2 as staining agent. Images were taken with an FEI Tecnai 12 electron microscope (acceleration voltage: 120 kV, C_s=6.3

mm; nominal magnification: 71,540x; nominal underfocus: 0.5-1.5 μm) and a TVIPS Tem-Cam-F416 4K CCD camera.

Cryo-electron microscopic images of BgAChBP1 were obtained after vitrifying and imaging the samples similar to the ones of native BgAChBP.

6.2.2 DIGITIZATION AND PREPROCESSING

BgAChBP1 images of negatively stained samples were digitized with a resolution of 2.18 Å/pixel. The CTF was estimated and corrected for with the software findCTF2D and the resulting images were band-pass filtered and normalized in IMAGIC-5. Particle selection was carried out in boxer (EMAN1.9); only those particles were selected that were clearly separated from any other particles. Cryo-electron microscopic images were digitized as described for native BgAChBP with 5080 dpi, resulting in a resolution of 1 Å/pixel in the magnified images. Again, boxer was used for particle selection, although the high occurrence of particle aggregates in the dataset did not allow for the selection of many separately lying particles. The CTF was estimated with CTFFIND3 and corrected for with the ctf2d-operations subroutine of IMAGIC-5.

BgAChBP2 images were digitized with a resolution of 4.36 Å/pixel due to a two-fold averaging during image acquisition. The images were not CTF-corrected, which is why a low-pass filter of 20 Å was used during band-pass filtering and normalization in IMAGIC-5. Particle selection was carried out in EMAN2 and only particles were selected that resembled a side-view of a dipentameric assembly.

6.2.3 NEGATIVE STAINING: INITIAL MODELS AND REFINEMENTS

A first impression of the BgAChBP1 dataset was gained by creating class averages in IMAGIC-5, which were iteratively refined as described in B3. To account for any incompletely assembled dodecahedrons during 3-D reconstruction, a multiple structure refinement was calculated with the dataset averaged to 4.36 Å/pixel, using the multirefine subroutine of EMAN1.9. Three equal Gaussian spheres with different amounts of Gaussian noise added to them served as initial references for the procedure, which was calculated in 10 iterations. A subsequent refinement of the particles belonging to the dodecahedral result of the multirefine procedure was calculated in EMAN (50 iterations, C1 symmetry + 113 iterations, icosahedral symmetry) and IMAGIC-5 (29 iterations, icosahedral symmetry). For both refinements, the resulting 3-D reconstructions after each iteration were masked by a contoured mask in EMAN1.9 as described in B3.12.

For BgAChBP2, the selected particles were assessed by their class averages, after calculating iterative MRA and MSA cycles (see above). A first refinement in EMAN1.9 was calculated in 80 iterations using a cylinder as reference structure. A point-group symmetry of D5 was assumed for a dipentameric assembly from the beginning. Since particle selection aimed at side views of dipentameric assemblies, references were confined to tilt angles (alt in EMAN) of 85° to 90°. A further refinement in 40 iterations was calculated using the mask.auto3d-processor for contoured masking of EMAN2.

6.2.4 CRYO-ELECTRON MICROSCOPIC DATA

Similar to the negatively stained datasets, a first impression of the cryo dataset was gained by creating class averages *via* MSA and MRA. Following, the final reconstruction of the negatively stained BgAChBP1 dataset was used as a reference for refinements in EMAN (75 iterations at 2 Å/pixel + 17 iterations with 1 Å/pixel; traditional refinement (see native BgAChBP)) and IMAGIC-5 (34 iterations with 1 Å/pixel) imposing icosahedral symmetry from the beginning. Again, contoured masking was used after every iteration. For the IMAGIC-5 refinement, particles were two-dimensionally masked to improve the results of the MRA. Resolution estimates were obtained by calculating the FSC of the reconstructions of two halfsets of the data and using the 0.5 threshold. A correction of the envelope function was partly achieved by imposing a negative temperature factor with the program EMBfactor.

6.2.5 STRUCTURAL INTERPRETATION

The homomeric pseudo-atomic models created for the structural interpretation of native BgAChBP were fitted into the final reconstructions for recombinantely expressed BgAChBP1 and BgAChBP2. The fitting of single subunits was not possible, which enforced the fitting of whole pentamers. For BgAChBP2, the pixel size had to be adapted to 5 Å/pixel for a fitting of two pentamers without overlap. A simulated map was created for BgAChBP2 using its pseudo-atomic model and the molmap subroutine of UCSF Chimera.

Chapter C: Results

1 Myriapod hemocyanin

1.1 *Scutigera coleoptrata* hemocyanin

Due to prevailing concerns within our group that previous reconstructions of oxy- and deoxy-ScoHc might have been over-refined, the datasets of the two physiological states were re-processed by re-estimating the CTF of all images and calculating 3-D refinements in different software packages. The molecular models for oxy-ScoHc and deoxy-ScoHc have previously been described in detail (Markl et al. 2009; Saur 2009) and were used for structural analysis. This section will not repeat the analysis of the two molecular models and their respective interfaces, but will focus on the validation and further specification of the conformational movements previously observed between density maps of the two physiological states. The oxy- and deoxy-ScoHc datasets used for this analysis comprised 6,689 and 12,842 particles, respectively. It has to be noted that all 3-D refinements were calculated using a deoxy-ScoHc reconstruction as a reference, which was previously determined in FREALIGN.

1.1.1 *EMAN1 refinement*

The datasets used for refinement in EMAN1 were CTF-corrected *via* phase-flipping prior to the refinement procedure. The final 3-D reconstructions were calculated in 30 iterations, comprised 4,996 (oxy-ScoHc) and 9,520 (deoxy-ScoHc) particles and achieved resolutions of 10.7 Å (oxy-ScoHc) and 11.1 Å (deoxy-ScoHc) according to the FSC 0.5 threshold, respectively. The dimensions of ~25 nm in height and ~23 nm in width, which were described by Markl et al. (2009), are maintained in the newly calculated density maps. It is apparent that the two reconstructions do not fit into each other exactly, as deoxy-ScoHc protrudes to

the right in the upper 3x6-mer, whereas it protrudes to the left in the lower one (Figure 23).

Altogether, the structural differences appear to be a concerted torsion between the two 3x6-mers. However, placing ScoHc hexamers into both density maps *via* rigid-body fitting allowed a more detailed analysis of the movements, which revealed that, in fact, each hexamer rotates individually around an axis that lies in the vicinity of the amino acids constituting interface 1↔2 (Figure 24). Applying the convention of negative clockwise rotations (e.g. Heymann et al. 2005), oxy-ScoHc hexamers rotate by approximately -2° against deoxy-ScoHc hexamers and vice versa. Furthermore, a slight shift along the rotation axis can be measured (Table 1). Consequently, conformational changes between oxy-ScoHc and deoxy-ScoHc amount for a rotation of ~4° between two hexamers of distinct 3x6-mers, due to the fact that the latter rotate by ~2° in opposite directions. This 4° rotation translates into an approximately 8-10 Å shift of amino acids that are located at the periphery of the molecule.
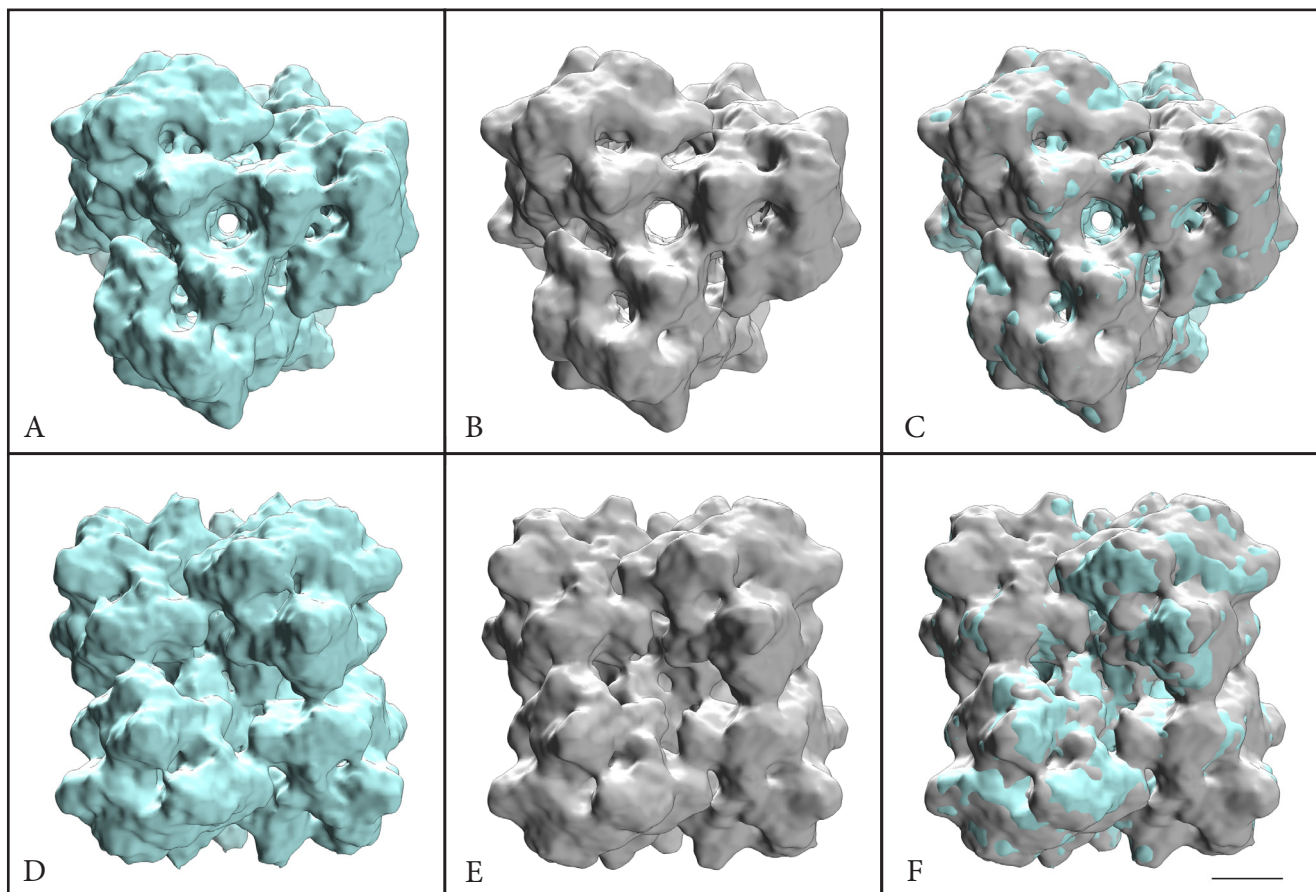
Figure 23 EMAN1.9 refinement of oxy-ScoHc (cyan) and deoxy-ScoHc (gray). A-C) top views of ScoHc. D-F): side views of ScoHc. C+F): superposition of oxy- and deoxy-ScoHc showing an apparent anti-clockwise and clockwise rotation of the upper and lower deoxy-ScoHc hexamers against oxy-ScoHc, respectively (viewed from the molecule's main symmetry axis). Scale bar: 50 Å.
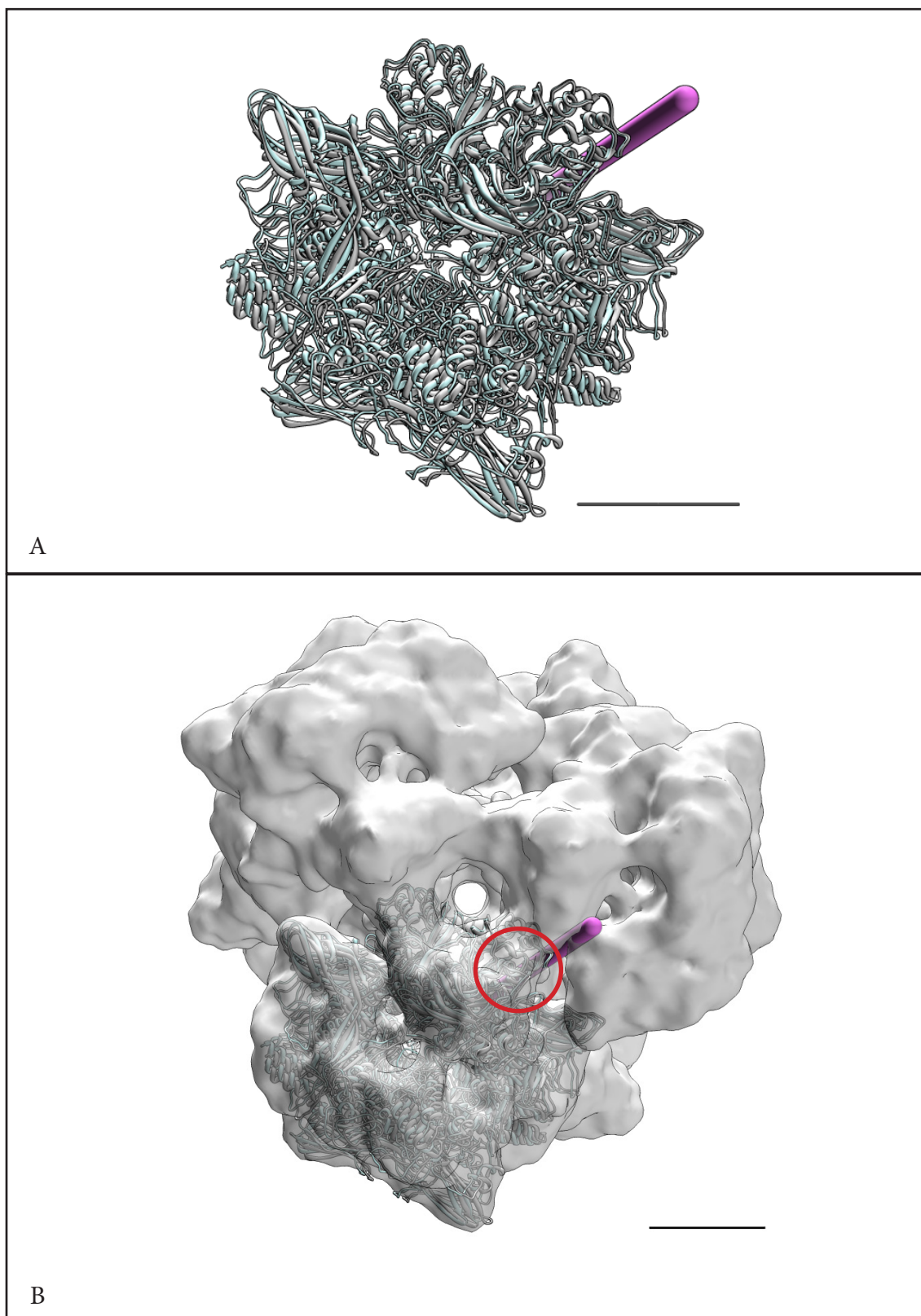
Figure 24 EMAN1.9 refinement of oxy-ScoHc (cyan) and deoxy-ScoHc (gray). A) Rotational difference between an oxy-ScoHc hexamer (cyan) and a deoxy-ScoHc hexamer (gray). The rotation axis is marked in purple. B) The two hexamers from A) fitted into a ScoHc 6x6-mer. The red circle marks the location of interface 1↔2. Scale bars: 50 Å.

Table 1   Comparison of the orientational changes between the molecular model of an oxy-ScoHc and a deoxy-ScoHc hexamer which were docked into the respective EMAN1.9 density map via rigid-body fitting. RMSD: root mean square deviation.

	RMSD (Å)	Rotation (°)	Axial shift (Å)
oxy-ScoHc ↔ deoxy-ScoHc	2.947	2.16	1.48

1.1.2 **FREALIGN** *REFINEMENT*

For three-dimensional analysis of the ScoHc datasets in FREALIGN, the images were corrected for the CTF using the program's internal Wiener-filter-like algorithm. The final reconstructions were calculated using 3,957 (oxy-ScoHc) and 7,559 (deoxy-ScoHc) particles, respectively. Before utilizing mode 1 for the refinement of the datasets, seven iterations of global orientational searches (mode 4) were calculated for each dataset with the parameters outlined in Appendix C. The final resolutions of the maps were measured to be 10.5 Å (oxy-ScoHc) and 11.9 Å (deoxy-ScoHc), respectively. They were determined using the FSC 0.143 threshold, which is comparable to the FSC 0.5 threshold applied for the results of other software packages (Nikolaus Grigorieff, personal communication).

When comparing the density maps calculated in FREALIGN with those obtained in EMAN1 and IMAGIC-5, it becomes apparent that the former lack density in the outer regions of the molecule (Figure 25). Furthermore, when compared to simulated density maps at different resolutions by eye, the 3-D reconstructions seem to show a lower resolution than would be assumed according to the resolution measurements. This effect is due to the impact that phase-flipping and Wiener-filter-like CTF corrections have on the image amplitudes and should be annihilated by imposing a B-factor on the FREALIGN reconstructions (Nikolaus Grigorieff, personal communication). This, however, did not lead to reconstructions that were visually comparable to the ones obtained in EMAN1 and IMAGIC-5. Yet, reconstructing the refined datasets with their final orientations and an amplitude contrast of 70% for CTF-correction, yielded FREALIGN density maps that were very similar to the others. In order to minimize errors, the analysis of the conformational movements was processed using both the standard output of FREALIGN (with an additional B-factor of -1000 Å^2) and the reconstructions obtained imposing an amplitude contrast of 70% (Figure 26). For clarity, the reconstructions will henceforth be termed "bfac" and "amp70" FREALIGN reconstructions, respectively.

The results of this analysis are displayed in Table 2. It is apparent that the rotational differences, as well as the axial shifts, are almost identical between the bfac and amp70 recon-

Table 2 Comparison of the orientational changes between the molecular model of an oxy-ScoHc and a deoxy-ScoHc hexamer which were docked into the respective FREALIGN density map via rigid-body fitting. RMSD: root mean square deviation; bfac/amp70/Brandeis: bfac/amp70/Brandeis FREALIGN reconstruction.

	RMSD (Å)	*Rotation (°)*	*Axial shift (Å)*
oxy-ScoHc (bfac) ↔ deoxy-ScoHc (bfac)	2.681	2.08	1.38
oxy-ScoHc (amp70) ↔ deoxy-ScoHc (amp70)	2.856	2.17	1.55
oxy-ScoHc (Brandeis) ↔ deoxy-ScoHc (bfac)	4.207	2.67	2.6
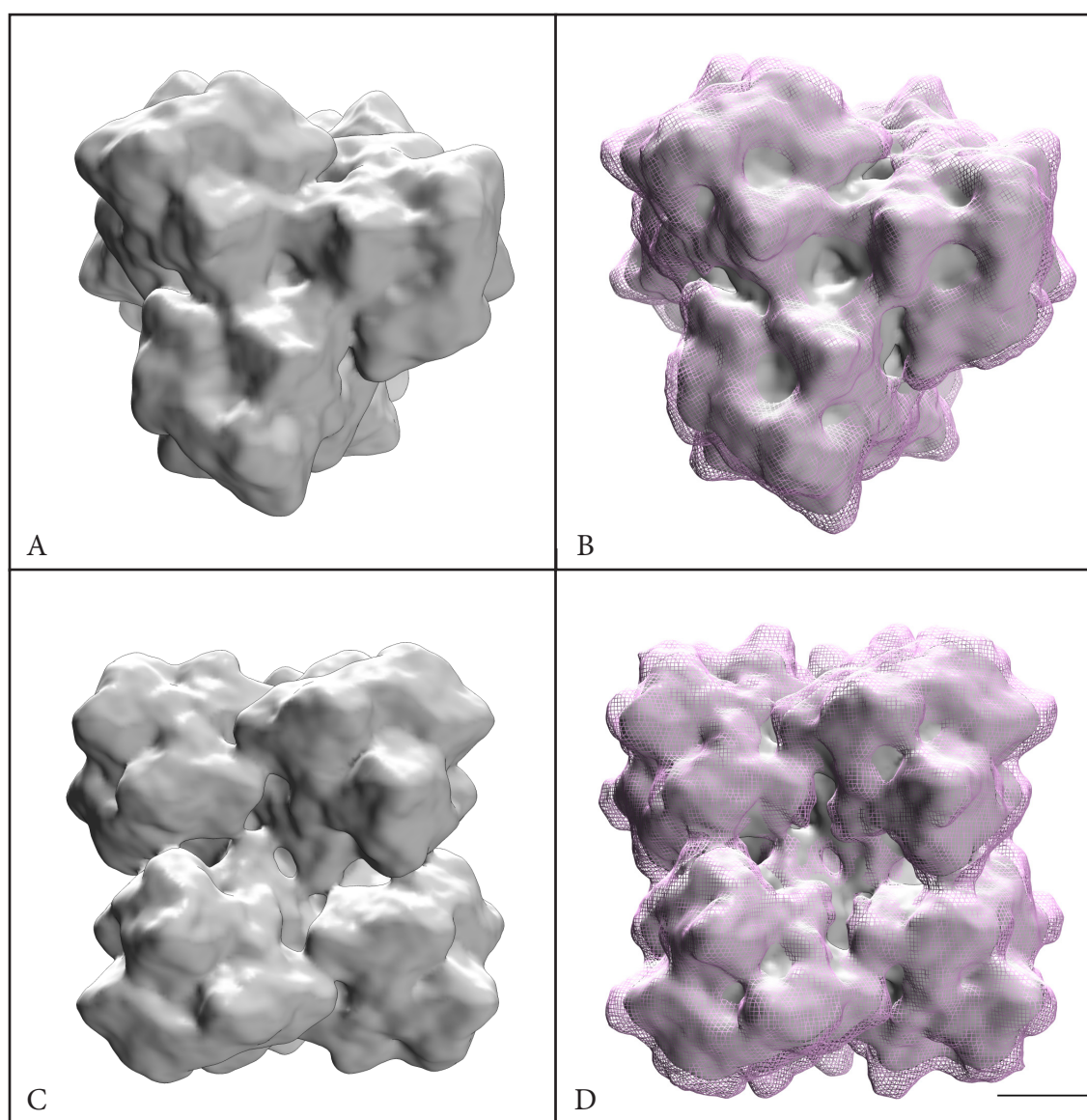
Figure 25 Comparison between the "standard" FREALIGN reconstruction (gray) of deoxy-ScoHc (without B-Factor) and the EMAN1.9 deoxy-ScoHc reconstruction (purple mesh). Scale bar: 50 Å.
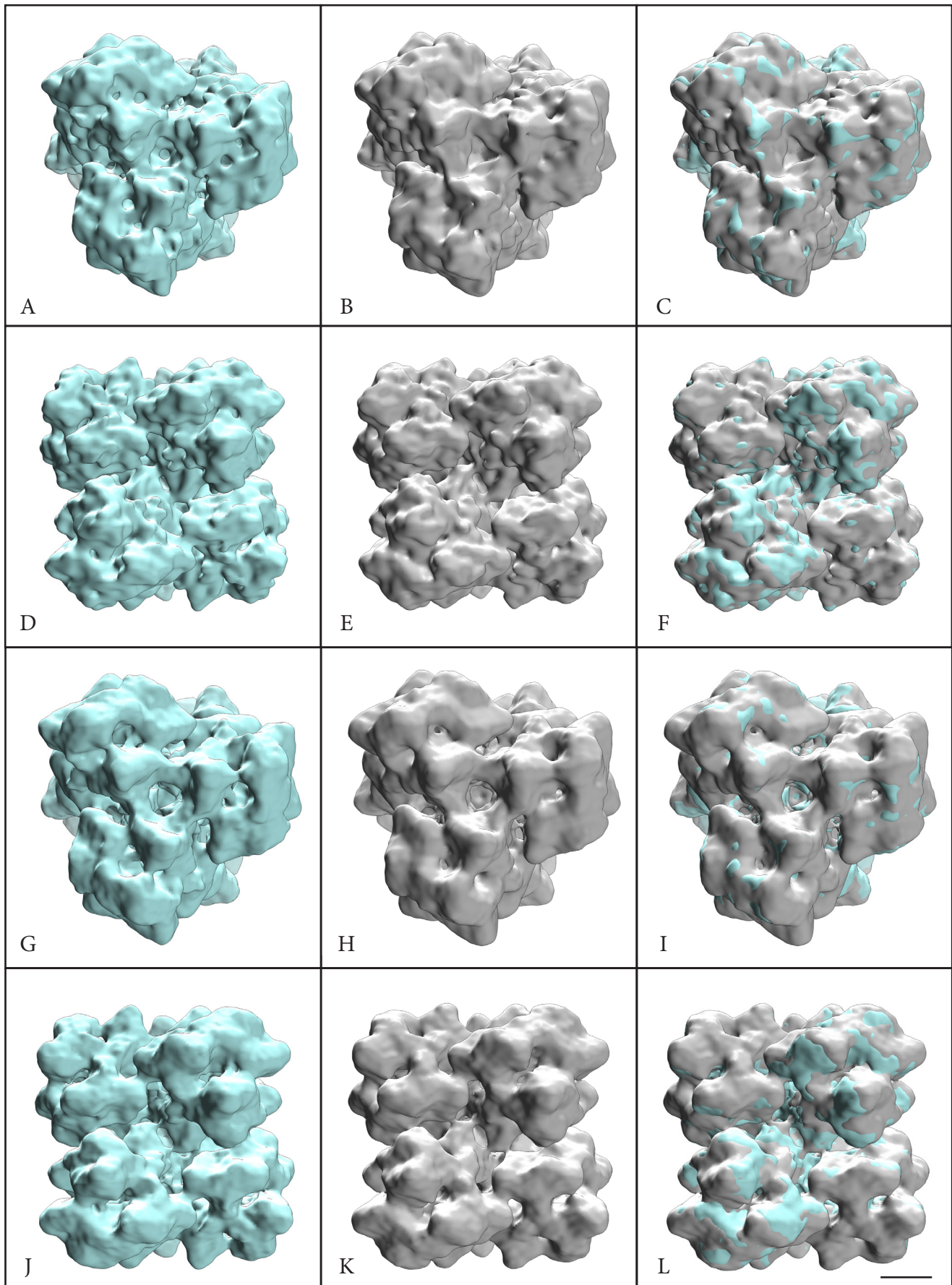
Figure 26 FREALIGN refinement of oxy-ScoHc (cyan) and deoxy-ScoHc (gray). A-F) "Bfac" FREALIGN reconstruction (with additional B-factor of -1000 Å^2). G-L) "Amp70" FREALIGN reconstruction. See text and Figure 23 for details. Scale bar: 50 Å.

structions. The rotational axes of both lie close to interface 1↔2 and show a small difference in location (Figure 27). In order to substantiate these conformational movements, a third ScoHc reconstruction was examined. The latter was independently imaged and calculated using FREALIGN at the Grigorieff lab (Brandeis University, Boston, MA, USA). Due to the preparation in atmospheric conditions, the sample should contain oxygenated ScoHc. Moreover, this "Brandeis" reconstruction is most likely to compare with the bfac FREALIGN reconstructions presented above. Indeed, the density map resembles oxy-ScoHc, showing an even larger clock-wise rotation (~2.6°) of its hexamers against deoxy-ScoHc (Figure 28; Table 2). This conformational change would amount to a rotation of ~5° between two hexamers of distinct 3x6-mer planes.

1.1.3 RELION REFINEMENT

RELION offers a fully automated refinement algorithm which was also used in the context of this thesis. Since the program does not discard particles during refinement, the final 3-D reconstructions were calculated using the full datasets and reached a resolution of 14.2 Å (oxy-ScoHc) and 11.6 (deoxy-ScoHc), respectively. The rotations between the two conformational states can, just like in the other programs, readily be seen by superimposing the density maps (Figure 29). However, due to the fact that the program also applies a Wiener-filter-like CTF-correction, similar effects as already described for the FREALIGN reconstructions can be seen (results not shown). For the comparison of rotations between oxy- and deoxy-ScoHc, I consequently used the standard output of the program with an additional B-factor of -1000 Å2 (bfac reconstruction) and 3-D reconstructions that were calculated using the orientations of the last iterative cycle, but an amplitude contrast of 70 % (amp70 reconstruction). The latter guaranteed visual comparability with the phase-flipped reconstructions of EMAN1 and IMAGIC-5, whereas the former was used for controlling the implications of using high amplitude contrasts. One could argue that RELION also incorporates the option of only phase-flipping the data. However, in order to maintain comparability to the reconstructions calculated in FREALIGN, the higher amplitude contrast was chosen. As will be elaborated in the Discussion, phase-flipped reconstructions and reconstructions calculated using 70 % amplitude contrast did not result in substantially different orientations of the molecular model. The conformational movements measurable between oxy- and deoxy-ScoHc amount for a rotation of ~2° along axes that pass through the amino acids constituting interface 1↔2 (Table 3; Figure 30) and consequently to a rotation of ~4° between two hexamers of distinct 3x6-mer planes. The rotation axes of both bfac and amp70 RELION reconstructions display a very close proximity.
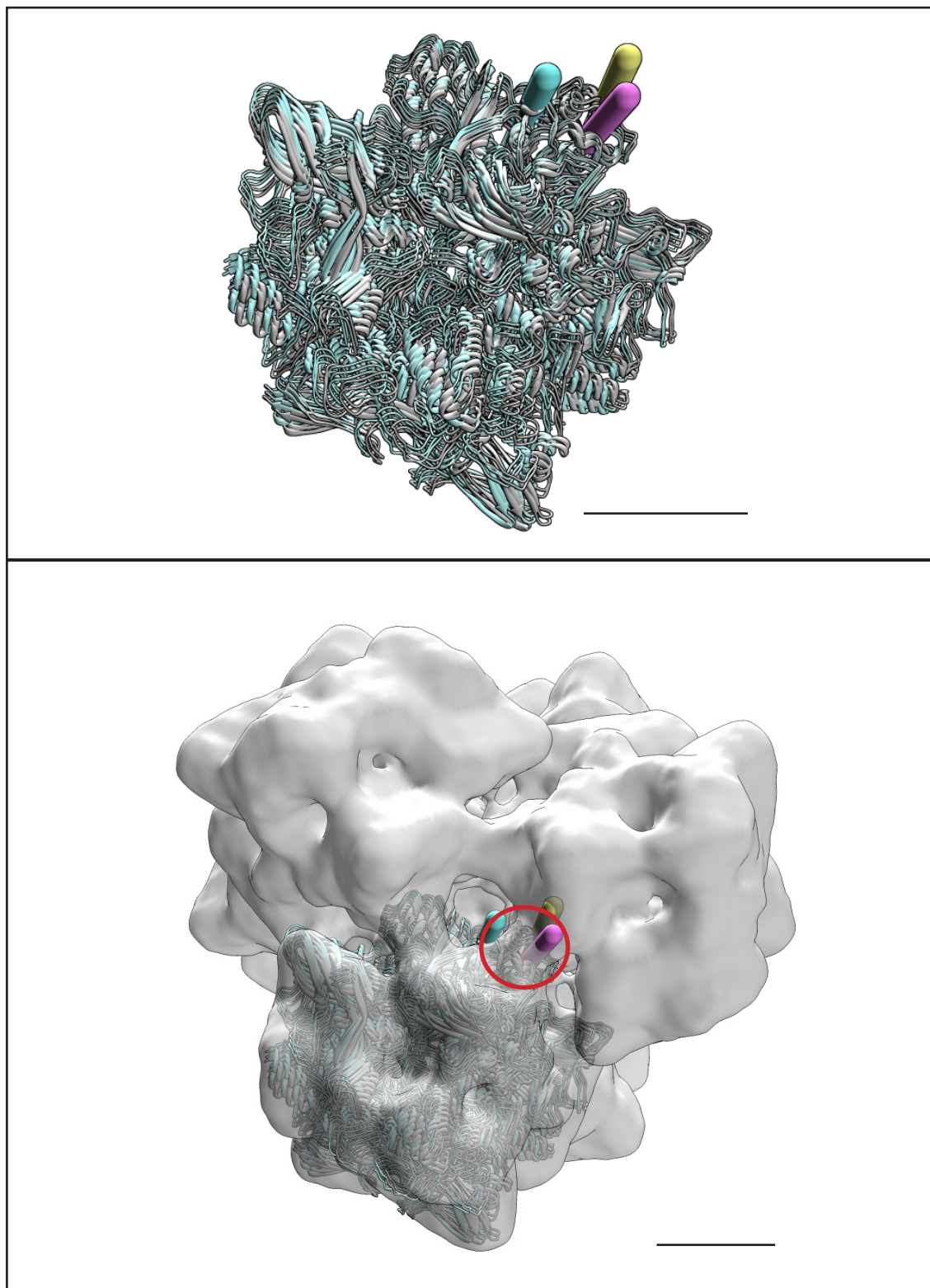
Figure 27 FREALIGN refinement of oxy-ScoHc (cyan) and deoxy-ScoHc (gray). A) Orientational differences between hexamers of oxy-ScoHc and deoxy-ScoHc fitted into "bfac", "amp70" or "Brandeis" FREALIGN reconstructions. The rotation axis is marked in cyan for the bfac reconstructions, purple for the amp70 ones and in yellow for the comparison of Brandeis oxy-ScoHc and bfac deoxy-ScoHc. B) The hexamers from A) fitted into a ScoHc 6x6-mer. The red circle marks the location of interface 1↔2. Scale bars: 50 Å.
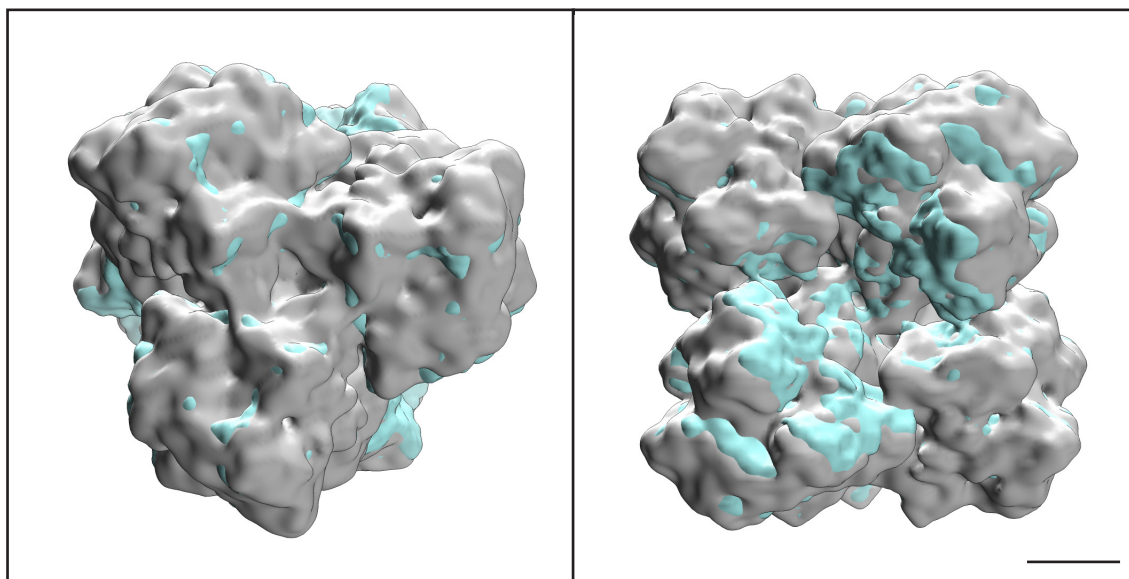
Figure 28 Comparison between the "bfac" FREALIGN reconstruction of deoxy-ScoHc (with an additional B-Factor of -1000 Å²) and the oxy-ScoHc reconstruction that was independently processed in the Grigorieff lab at Brandeis University. See text for details. Scale bar: 50 Å.

Table 3 Comparison of the orientational changes between the molecular model of an oxy-ScoHc and a deoxy-ScoHc hexamer which were docked into the respective RELION density map via rigid-body fitting. RMSD: root mean square deviation; bfac/amp70: bfac/amp70 RELION reconstruction.

	RMSD (Å)	Rotation (°)	Axial shift (Å)
oxy-ScoHc (bfac) ↔ deoxy-ScoHc (bfac)	2.844	2.29	0.82
oxy-ScoHc (amp70) ↔ deoxy-ScoHc (amp70)	2.433	2.11	0.84
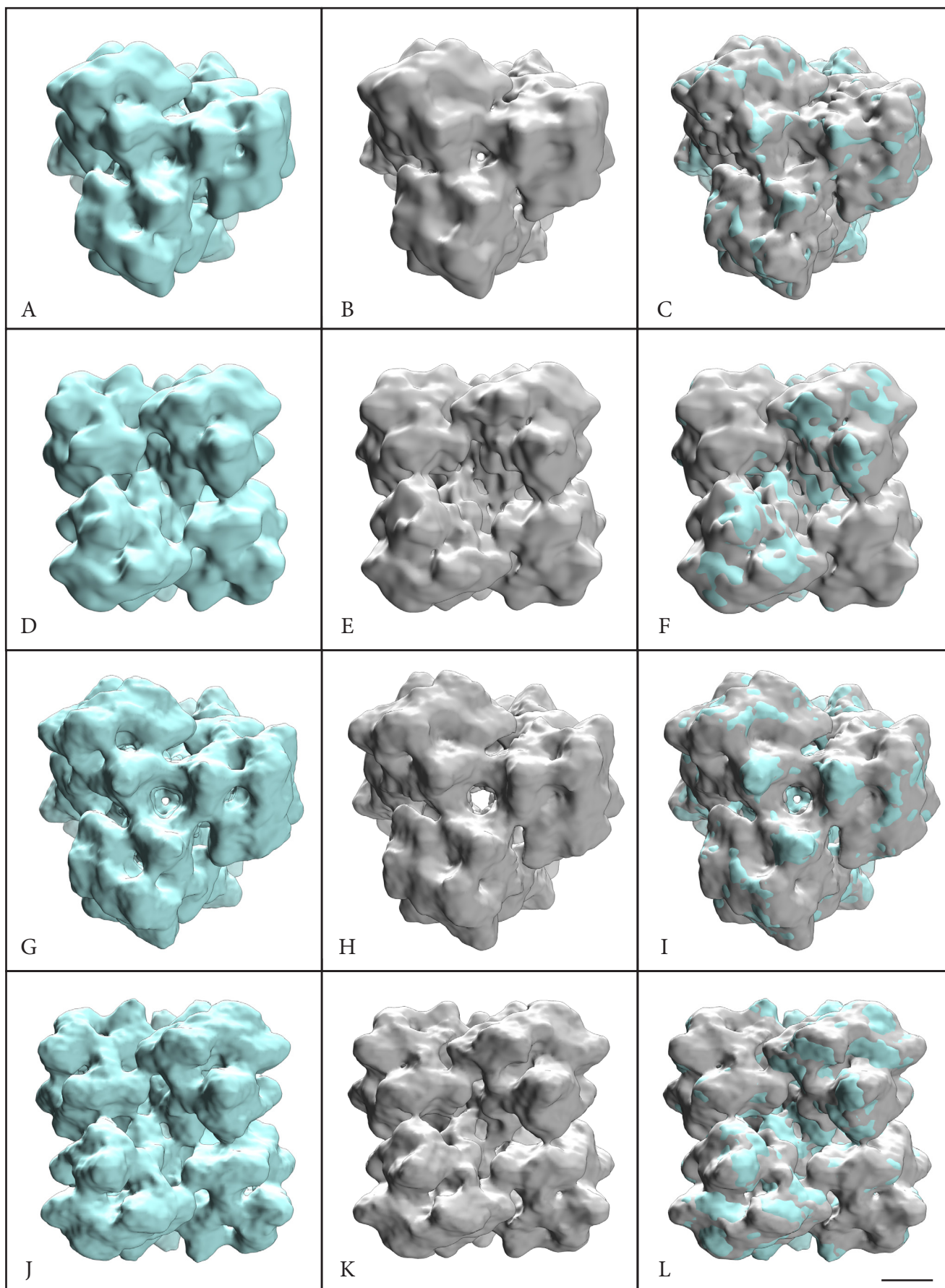
Figure 29 RELION refinement of oxy-ScoHc (cyan) and deoxy-ScoHc (gray). A-F) "Bfac" RELION reconstruction (with additional B-factor of -1000 Å²). G-L) "Amp70" RELION reconstruction. See text and Figure 23 for details. Scale bar: 50 Å.
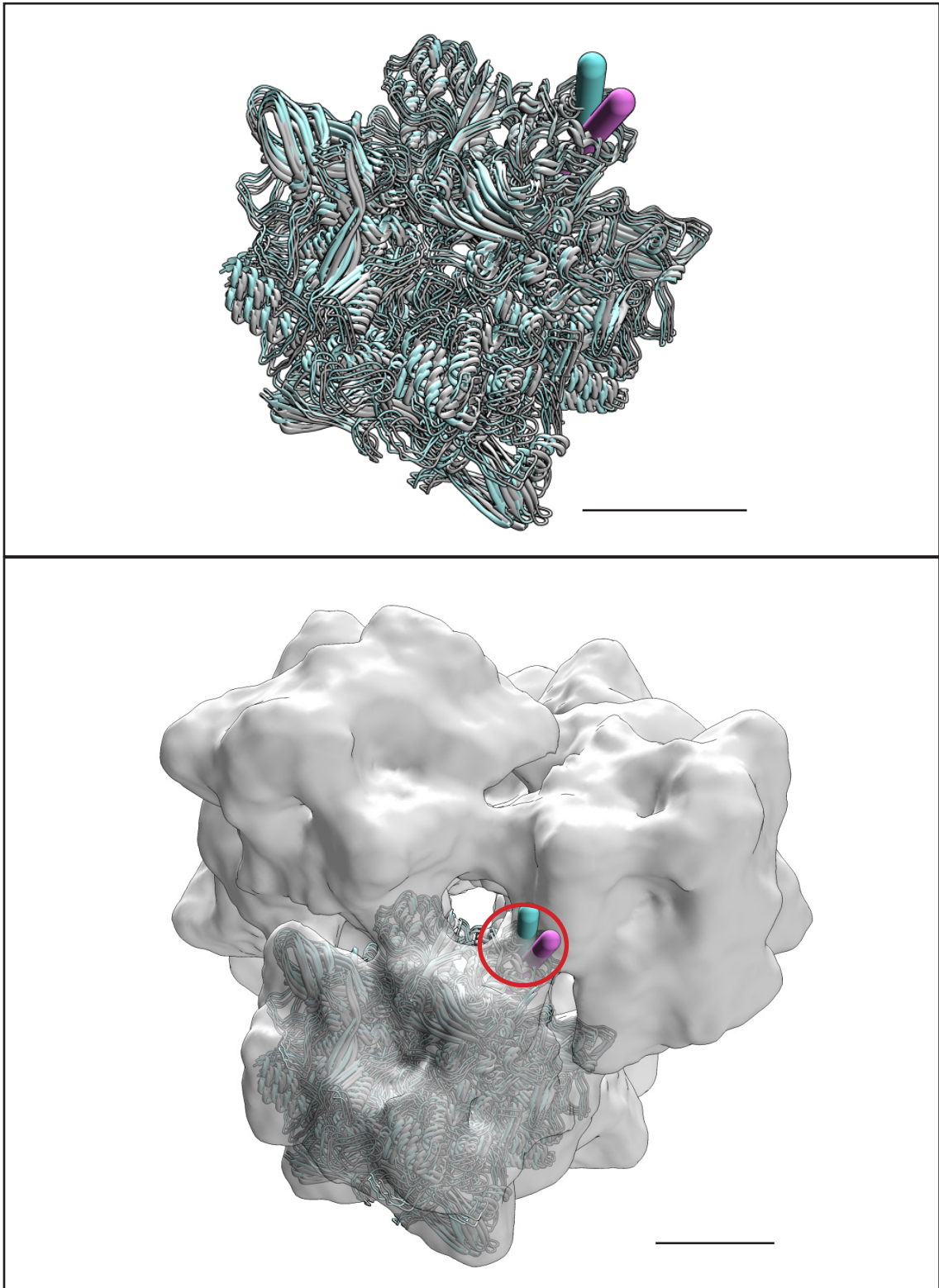
Figure 30 RELION refinement of oxy-ScoHc (cyan) and deoxy-ScoHc (gray). A) Orientational differences between hexamers of oxy-ScoHc and deoxy-ScoHc fitted into "bfac" or "amp70" RELION reconstructions. The rotation axis is marked in cyan for the bfac reconstructions and in purple for the amp70 ones. B) The hexamers from A) fitted into a ScoHc 6x6-mer. The red circle marks the location of interface 1↔2. Scale bars: 50 Å.

1.2 *Polydesmus angustus* hemocyanin

Being relatively small animals, the amount of hemolymph extractable from *Polydesmus angustus* is little. Earlier attempts to reconstruct PanHc from a negatively stained dataset were not successful and when new specimens were obtained, it was decided to use the small amount of sample that was available for cryo-EM. I chose relatively large defocus values (between 4.5-6 μm), due to the fact that no initial model was available for PanHc and the images should have the greatest possible contrast. This said, higher defocus values entail a lower achievable final resolution, which was accepted given the priority of a first general understanding of the PanHc quaternary structure over achieving a high resolution. As mentioned in section B4.2, digitization, preprocessing and refinement was conducted under my supervision by ████████ ████████ as part of his Bachelor thesis (Neumann 2013).

1.2.1 2-D Classification of the cryo-EM dataset

Cryo-EM images of the sample showed a considerable amount of heterogeneity, being comprised of hexameric structures, small rod-shaped structures, assemblies with a trimeric shape, Y-shaped structures and, occasionally, rectangular structures (see Figure 31 for examples). Only particles that were larger than the hexameric structures were included into the PanHc dataset, as diplopod hemocyanins had been shown to be comprised of 6x6-mers and a stable 3x6-meric dissociation product (Jaenicke et al. 1999). This yielded a dataset of 7,996 particles which was subsequently used to calculate MSA-MRA cycles to allow a better under-
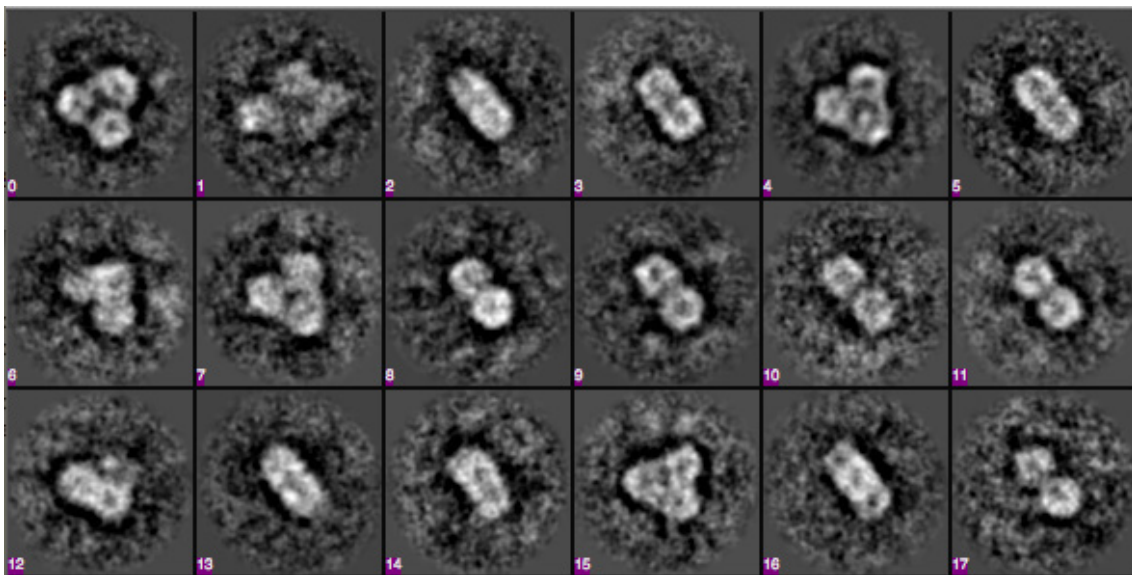
Figure 31 Typical class sums of the PanHc cryo-EM dataset. Triangular-shaped, rod-shaped and Y-shaped structures were the most common views of the dataset. Rectangular views (image number 1) were sparsely visible.

standing of the structural properties (Figure 31). The triangular assemblies show pronounced ring-shaped connections between the main densities, as well as a central hole. The rod-shaped structures have the size of a 2x6-mer side view and display divergent density distributions. Some show two main densities of equal brightness, whereas in others one of the main densities is brighter than the other. A third type can be described by the absence of two clear main densities, bearing a brighter section in the center of the rod. Rectangular structures can sparsely be seen as class averages. If so, they lack clearly discernible structural features.

1.2.2 Multiple 3-D Refinement of the cryo-EM dataset

Due to the possibly heterogeneous dataset, a multiple 3-D refinement was calculated in 35 iterations with EMAN1.9 and C1 point group symmetry, using four equal Gaussian blobs with different amounts of added Gaussian noise as references. The multiple 3-D refinement resulted in density maps of three seemingly trimeric assemblies (Figure 32 A-F) and one structure that mostly shows a 2x6-meric shape (Figure 32 G/H). The reconstructions shown in Figure 32 comprise – from left to right – 1,925, 1,501, 1,386 and 1,690 particles, respectively. The only structures which did not display significant structural heterogeneity – due to shapeless or faint densities and densities detached from the main structure (e.g. Figure 32 G/H) – are the triangular-shaped 3-D reconstructions shown in Figure 32 A-D. Due to their similar shape, the particles used for these 3-D reconstructions were combined in a new
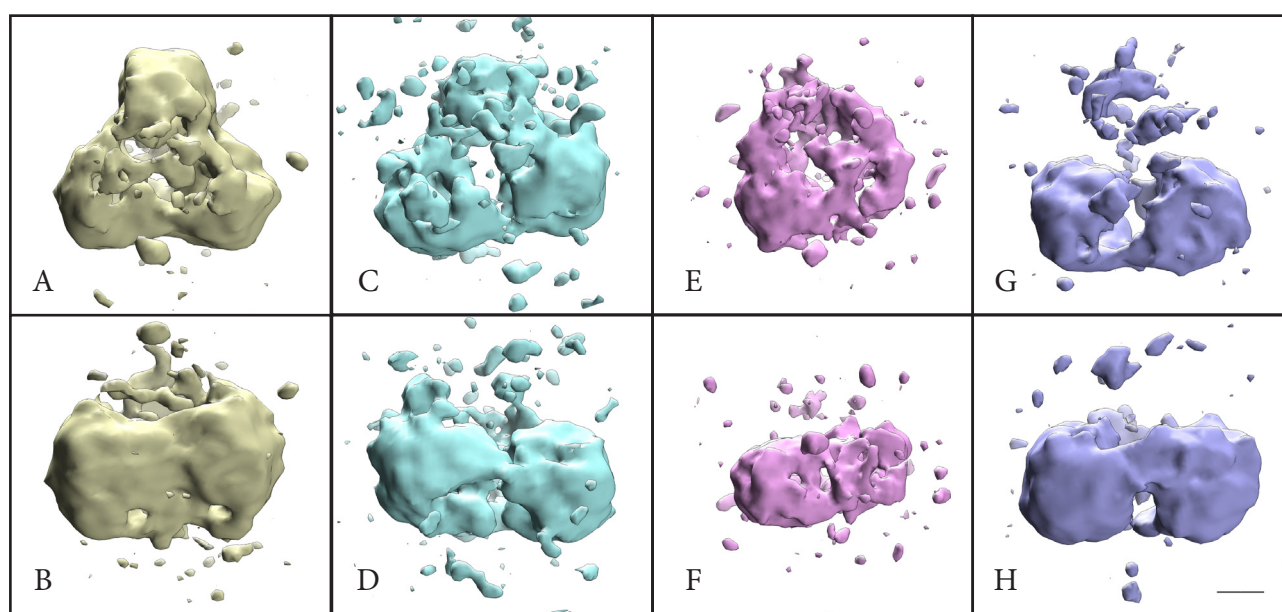
Figure 32 Results of the multiple 3-D refinement with EMAN1.9 using the PanHC cryo-EM dataset. Each reconstruction is shown from its top view (top row) and from its side view (bottom row). The first two reconstructions (A-D) were combined for a subsequent refinement in EMAN1.9. Scale bar: 50 Å.

dataset for further refinement of the structure in EMAN1.9 using C3 point group symmetry. The procedure reached stability after 20 iterations and resulted in a density map which shows three main densities that are each comparable in size to a hemocyanin hexamer (Figure 33). Consequently, it can be assumed that the density map resembles a 3x6-mer with a width of ~250 Å and a height of ~110 Å.

However, the density map is not sufficiently resolved to be structurally interpreted and molecular models of hemocyanin hexamers could not be docked into the structure by rigid-body fitting with sufficient accuracy. We knew that some of the micrographs used for particle selection showed a substantial amount of drift. Particles belonging to those micrographs were subsequently discarded, in order to test whether or not they could have led to the unsatisfactory result of the EMAN1.9 refinement. A total amount of 2,796 remaining particles were subjected to 50 iterations of refinement in EMAN1.9, using the reconstruction of the previous refinement as a reference and applying C3 symmetry. However, the result did not show a significant improvement when judged by eye (Figure 34).
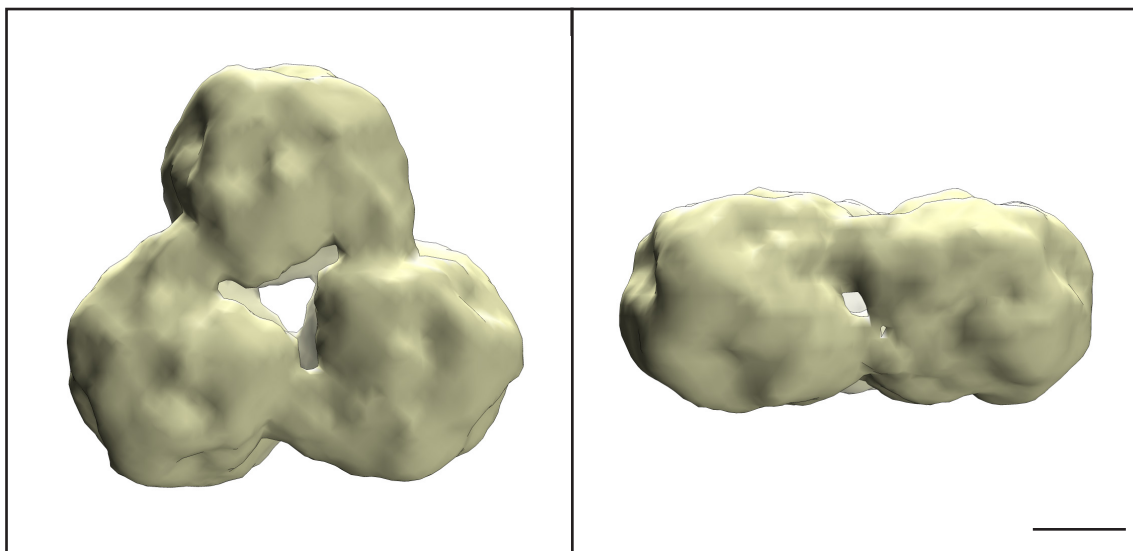
Figure 33 The final density map of the PanHc cryo-EM dataset after the EMAN1.9 refinement that followed the multiple 3-D refinement (see text for details). Each of the main masses is of similar size as a hexamer of arthropod hemocyanin. Scale bar: 50 Å.
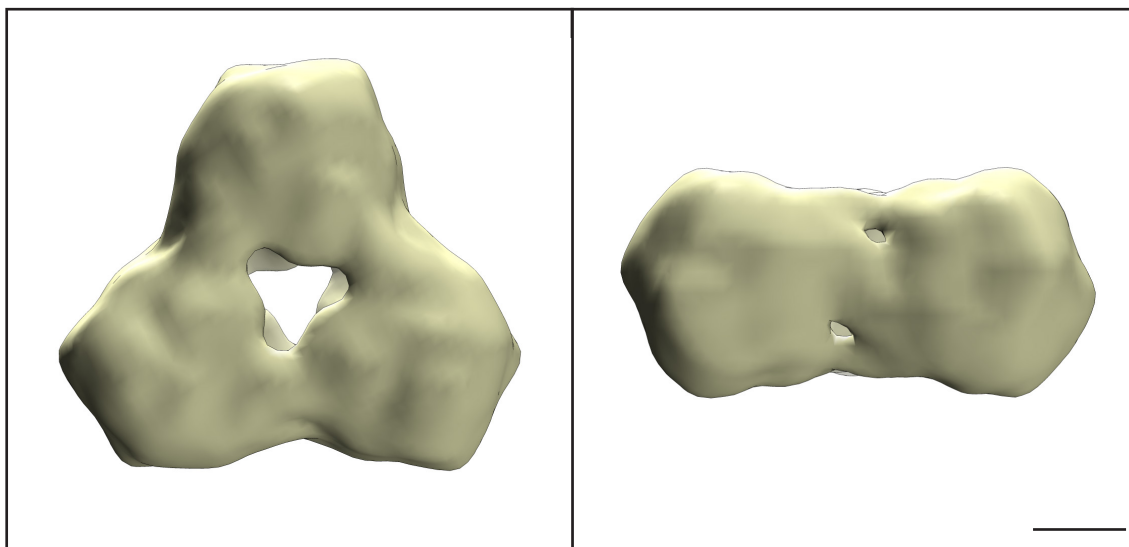
Figure 34 The final density map of the EMAN1.9 refinement of the PanHc cryo-EM dataset after the exclusion of drift micrographs. Each of the main masses is of similar size as a hexamer of arthropod hemocyanin. Scale bar: 50 Å.

1.2.3 NEGATIVE STAINING DATASET

Due to the problems encountered with the cryo-EM images, the available dataset of negatively stained PanHc was considered for re-assessment. As I mentioned above, this analysis did not produce an interpretable result in the past. However, the information that was gained from the analysis of the cryo-EM dataset, namely that a stable 3x6-mer seems to exist, permitted an alternative processing scheme for the negatively stained dataset. Iterative MSA-MRA cycles yielded class averages that resemble the averages described for the cryo-EM dataset (Figure 35). Class averages of unidentifiable densities, hexamer-sized densities, one rectangular shape and densities that seem too far apart to belong to one quaternary assembly are present. Nevertheless, many class averages clearly resemble the projections of a 3x6-mer. The rectangular shape present in the class averages shows two outer densities that are brighter than the two inner densities. As in the cryo-EM class averages, rod-shaped structures of equally and unequally distributed densities are visible.

On the basis of the cryo-EM analysis, only particles belonging to 3x6-mer class averages were used for further processing. A rough initial density map was calculated by angular reconstitution, using the method described by SAUR (2009). This initial reconstruction served as a reference for a subsequent refinement of 40 iterations in EMAN1.9, gradually increasing the point group symmetry from C1 and C3 to D3. The final reconstruction has a resolution of 25 Å (FSC 0.5 threshold) and clearly outlines the general structural features of three hemocyanin hexamers (Figure 36). The three main densities are connected by an inner and
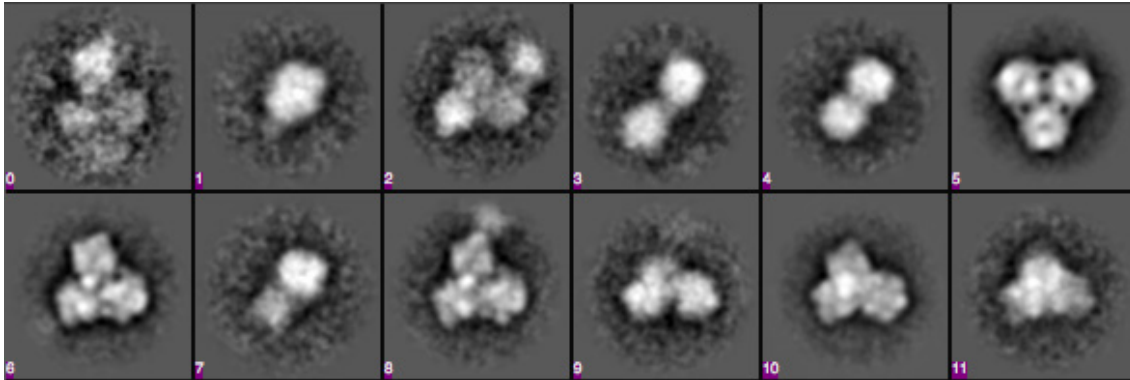
Figure 35 A selection of class averages resulting from 2-D analysis of negatively stained PanHc. Apart from unidentifiable structures (image 0), hexamer-sized densities (image 1), one rectangular sized structure in all class averages (image 2), densities that are too far apart for a typical quaternary assembly of arthropod hemocyanin (image 3) and structures that resemble a 2x6-mer in size (with equal brightness of the two main densities; image 4), most class averages were recognizable as projections of 3x6-mers (images 5-11).

an outer ring of inter-hexamer interfaces and amount for a width of 220 Å and a height of 110 Å. Hemocyanin hexamers, such as *Panulirus interruptus* hemocyanin (PDB code: 1hcy) could be readily placed into the structure by rigid-body fitting, confirming the identity of the main masses as hemocyanin hexamers. However, comparing the density map to the ScoHc reconstructions, indicated a necessary adjustment of the pixel size from 4.98 to 4.3 Å/pixel.

1.2.4 Pseudo-atomic model of PanHc

In order to structurally interpret the PanHc density map, homology models of PanHc subunits were created using the previously published homology model of ScoHcA as a template (Markl et al. 2009). A 3x6-meric assembly was created according to the phylogenetic analysis that was presented for ScoHc by Markl et al. (2009), positioning PanHc1 in the core of the structure and PanHc2 in the mantle by complying to a stoichiometry of 1:2 (Figure 37 A). The histidine loop that distinguishes *Polydesmus angustus* subunit 1 from the other Myriapoda (see Introduction) is positioned in six copies at the top and the bottom of the molecule (Figure 37 B). The interface nomenclature that will be used in the subsequent description of inter-hexamer interfaces is based on the terminology introduced by Markl et al. (2009). The hemocyanin of *Polydesmus angustus* shows two types of inter-hexamer interfaces, namely 3↔4 and 1↔2.

Interface 3↔4 is situated in the mantle of PanHc and constituted by PanHc2. A multitude of charged amino acids can be found in the vicinity of the interface (Figure 38). Two residues, namely D^{400} and K^{399}, of each subunit are located in very close proximity to each other and might form an ionic bridge.
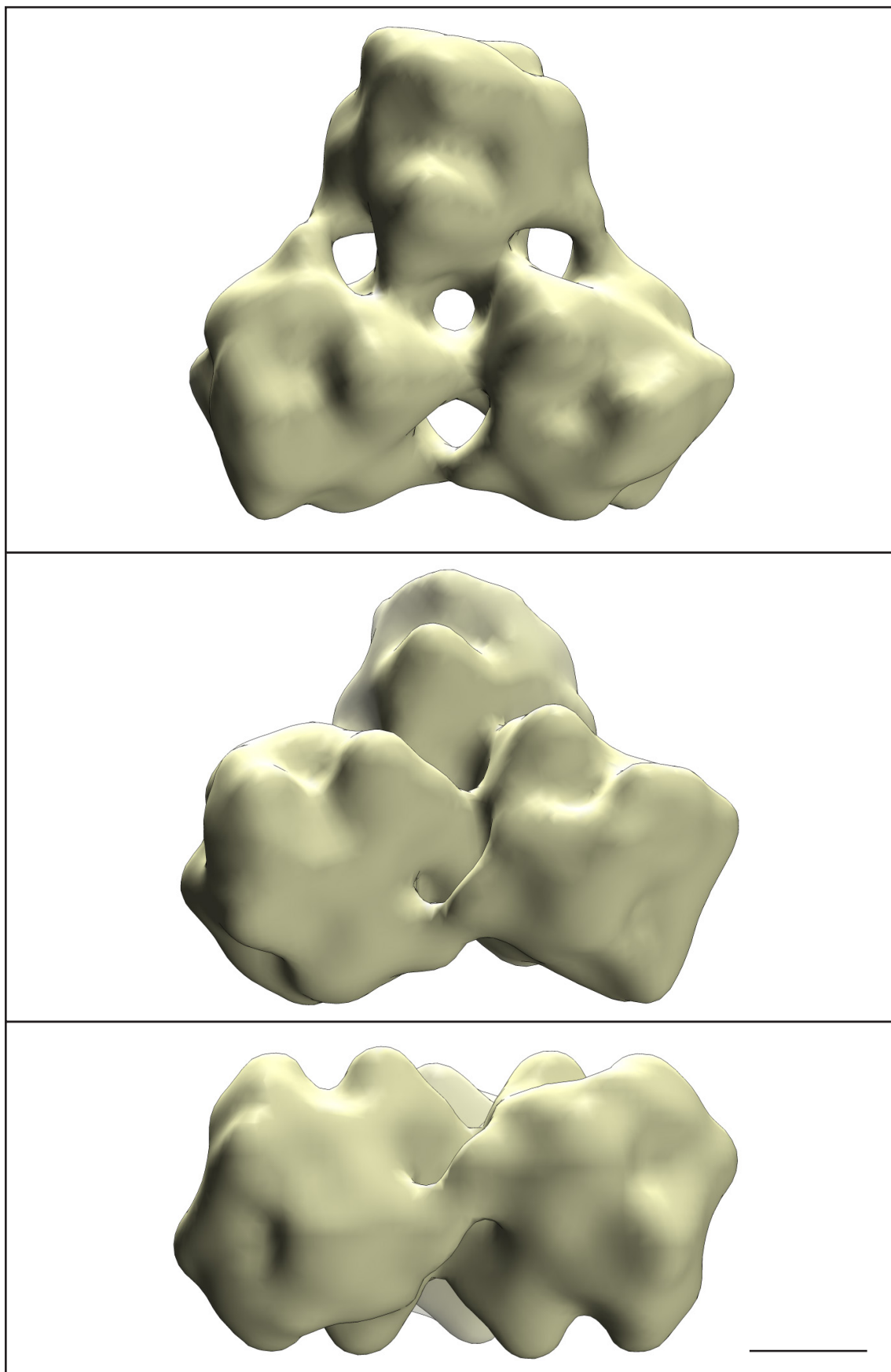
Figure 36 Final density map of negatively stained PanHc after refinement in EMAN1.9, displayed in top view (top), tilted view (center) and side view (bottom). Main structural features of hemocyanin hexamers are visible in the three main densities. Scale bar: 50 Å.
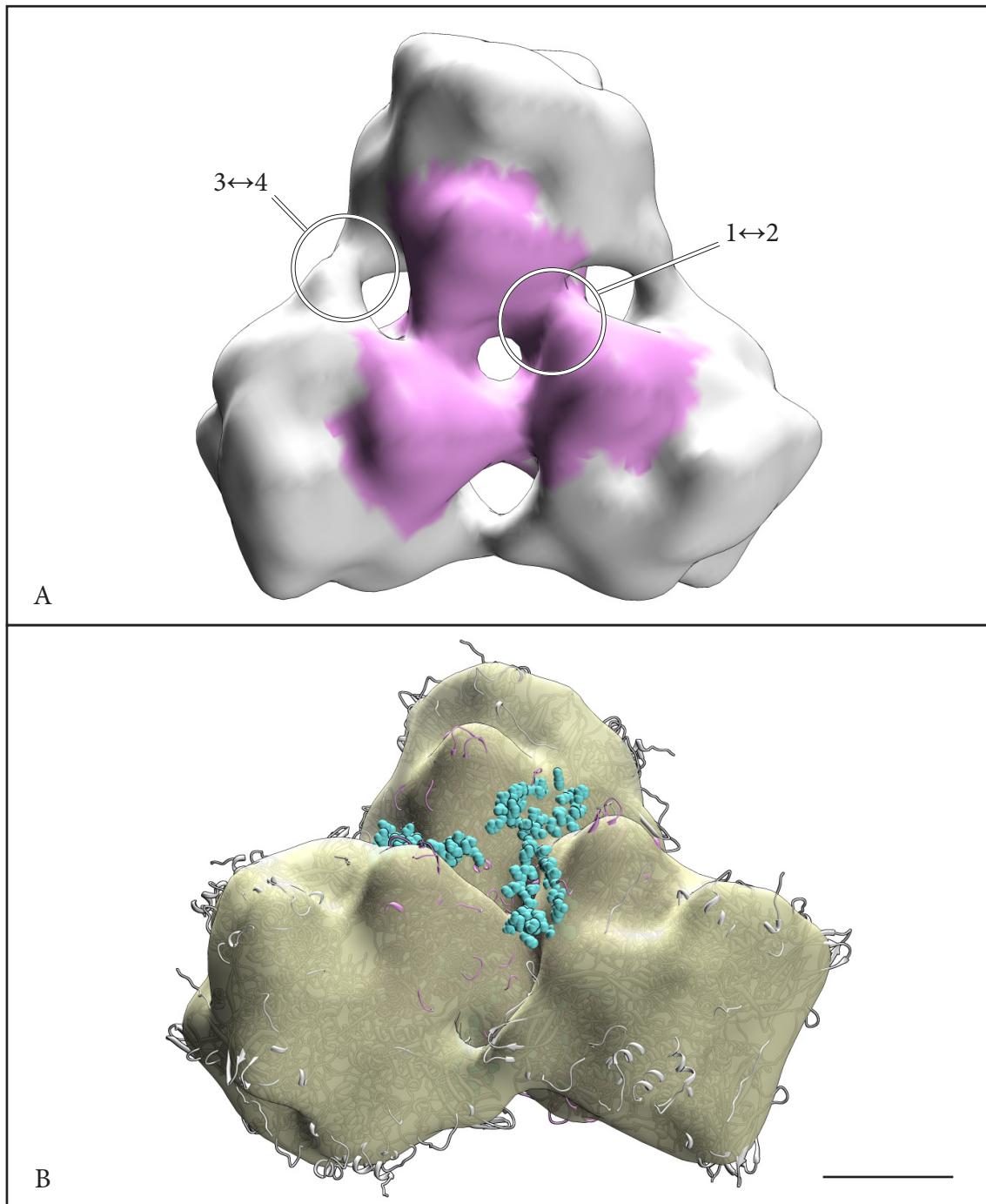
Figure 37 Final density map of negatively stained PanHc after refinement in EMAN1.9. A) Top view displaying the location of PanHc1 (purple), PanHc2 (gray), interface 3↔4 and interface 1↔2. B) Tilted view showing the location of the histidine loop found in PanHc1 (cyan). Scale bar: 50 Å.

Interface 1↔2 is the central connection that builds a 3x6-mer. The peculiarity of this interface is that one subunit of a hexamer is involved in the formation of two neighboring 1↔2 interfaces. In the contact region, mostly charged and polar amino acids can be found. For clarity reasons only charged amino acids are shown in Figure 39. Putative interactions

could exist between any of the amino acids in the interface that show close proximity and the resolution of the density map does not allow any further interpretation. Charged amino acids were chosen for display due to their proposed interaction with histidine residues for conformational changes (e.g. MARTIN ET AL. 2007), as will be elaborated more closely in the Discussion. Overall, eight histidines can be found in the interface, namely H^{144}, H^{469}, H^{665} and H^{668}, which exist in two copies each and are contributed to the interface by only two of the four monomers that constitute the interface.
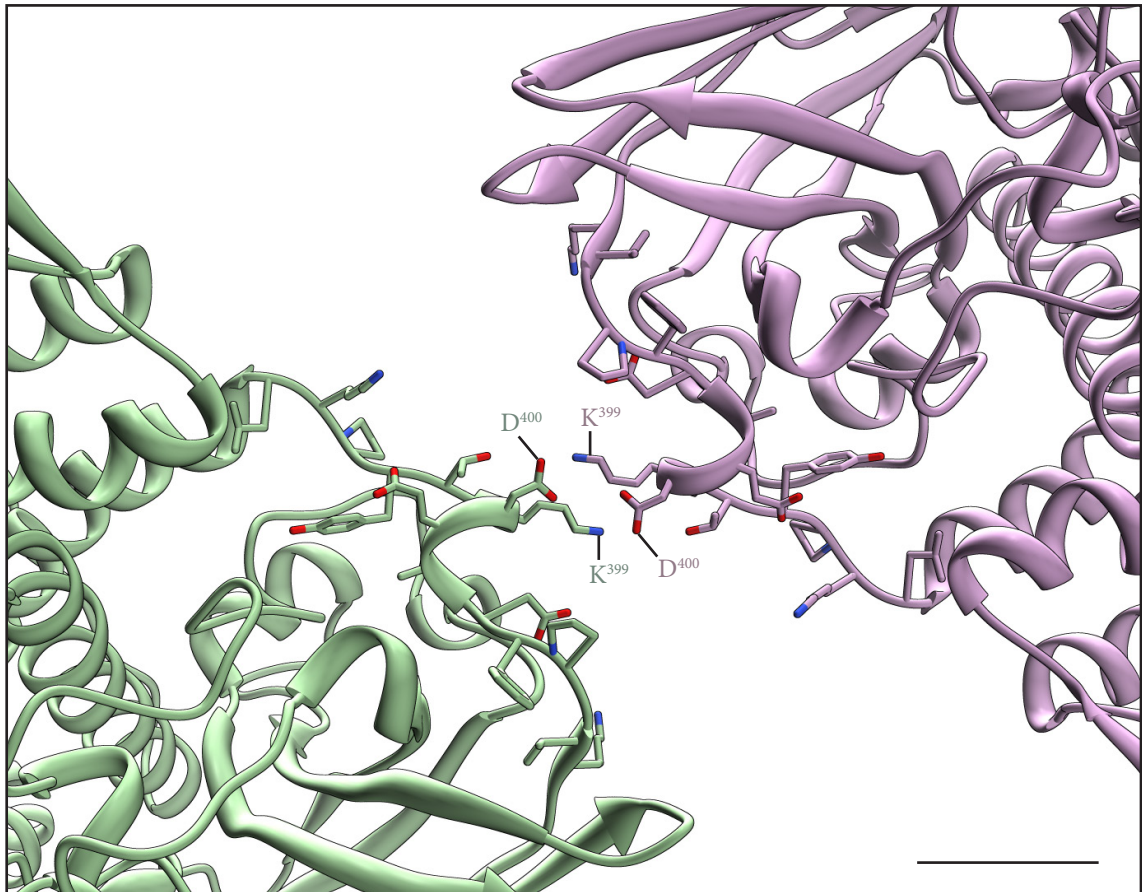
Figure 38 PanHc interface 3↔4. Two PanHc2 subunits provide charged amino acids for an inter-hexamer interaction which is presumably driven by an ionic interaction between D^{400} and K^{399}. Scale bar: 10 Å.

A

B

Interface amino acids:

D^1	D^{138}
R^2	K^{139}
K^{26}	H^{144}
K^{43}	K^{145}
E^{79}	D^{151}
E^{80}	D^{451}
K^{83}	D^{463}
	H^{469}
	H^{665}
	K^{667}
	H^{668}
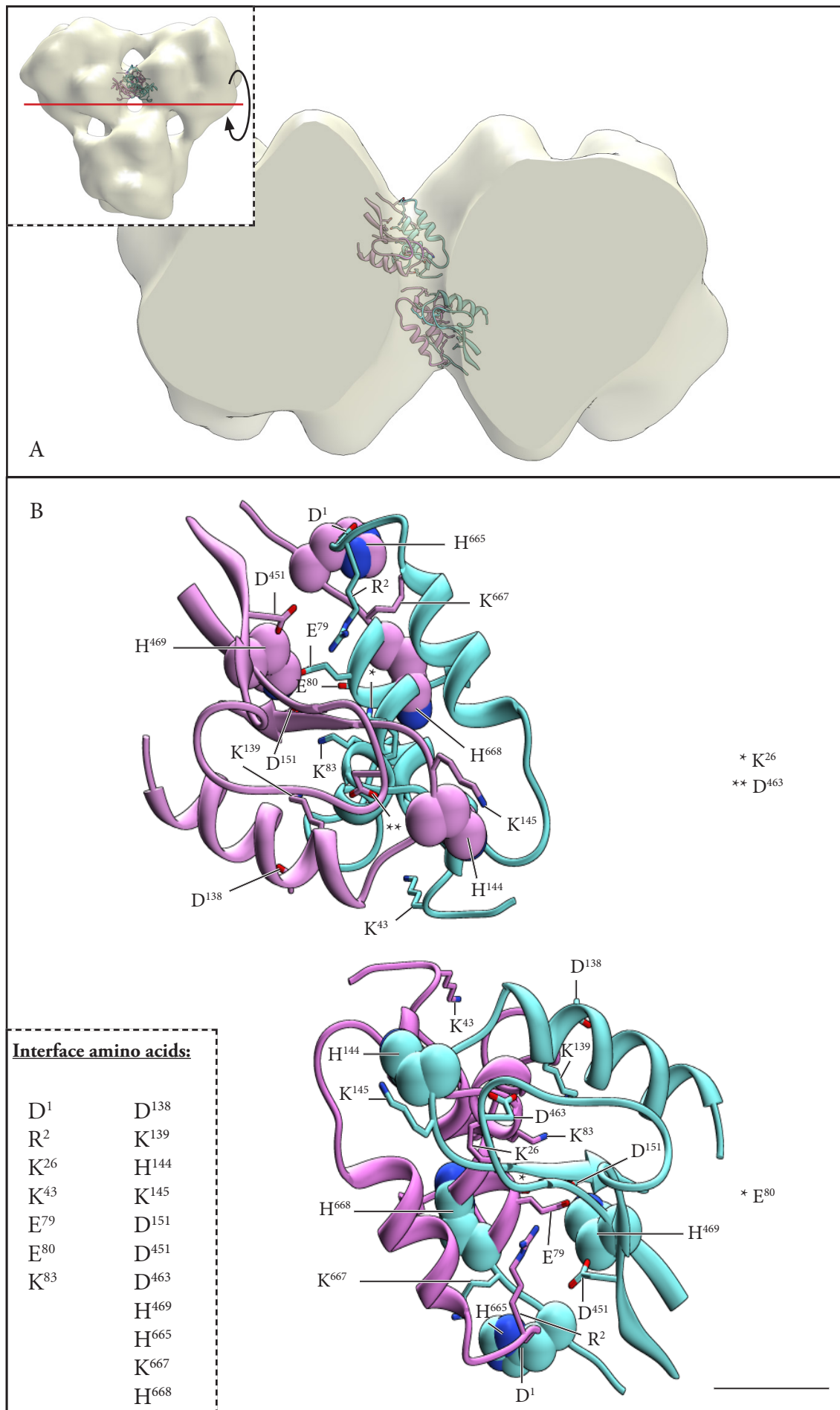
Figure 39 PanHc interface 1↔2. The red line through the top view of ScoHc indicates the rotation axis and the clipping plane which result in the image shown in A. B) Magnified view of the interface shown in A without the density map. Only charged amino acids and histidines are shown. Histidines are shown as spheres. Scale bar: 10 Å.

2 Vesicle-inducing protein in plastids 1

As already mentioned in section B5, all biochemical procedures (i.e. recombinant expressions, purifications, mutations, liposome preparations) were performed by ████████ of the Institute of Biochemistry at the JGU Mainz.

2.1 Wild type Vipp1

Purified Vipp1 was either stored in phosphate or HEPES buffer (for the composition of the two buffers see Appendix A). In order to determine which buffer was best suited for structural preservation of Vipp1 rings and single particle analysis, electron microscopic images of both samples were compared (Figure 40). It can be observed that the phosphate buffer sample shows a large amount of small densities which cannot be located in the HEPES buffer sample. Furthermore, the staining quality of the latter seems to be superior to the one of the phosphate buffer sample when judged by eye, rendering Vipp1 stored in HEPES buffer to be the samples of choice for subsequent analyses.

The heterogeneity of Vipp1 has already been described by Fuhrmann et al. (2009a) and the current sample confirmed the formation of different ring sizes, as well as the aggregation of the rings to rod-like structures and two-ring assemblies (Figure 41). The fact that relatively few side-views and even less tilted views are visible, complicates the 3-D reconstruction process. In a first two-dimensional analysis, refined class averages of Vipp1 show top-views
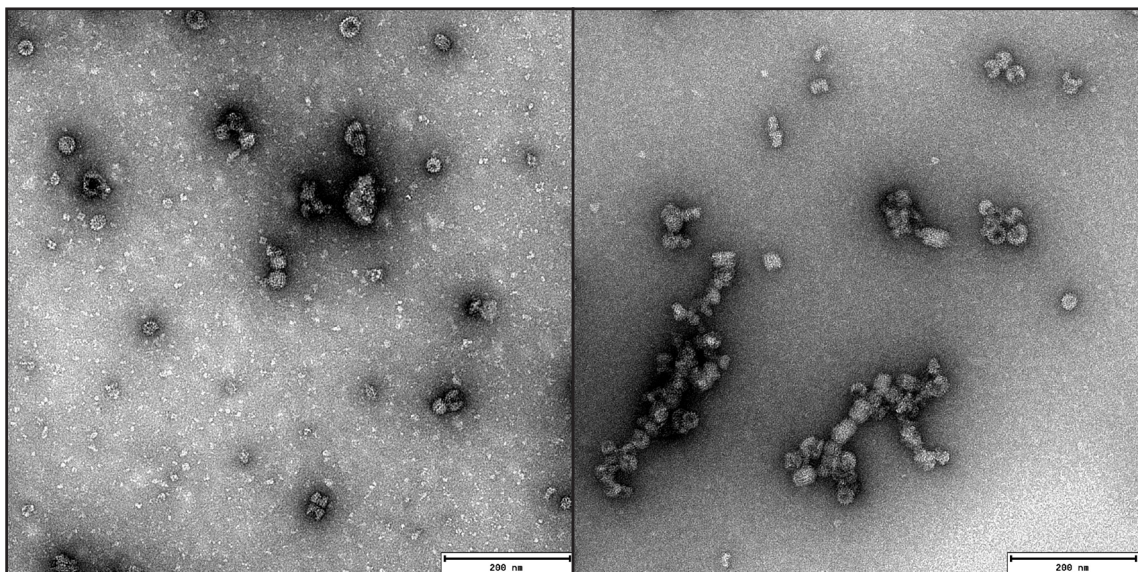
Figure 40 Electron-microscopic images of negatively stained Vipp1 in phosphate buffer (left) and HEPES buffer (right). Note the decrease in small background structures in the HEPES sample, allowing better staining.
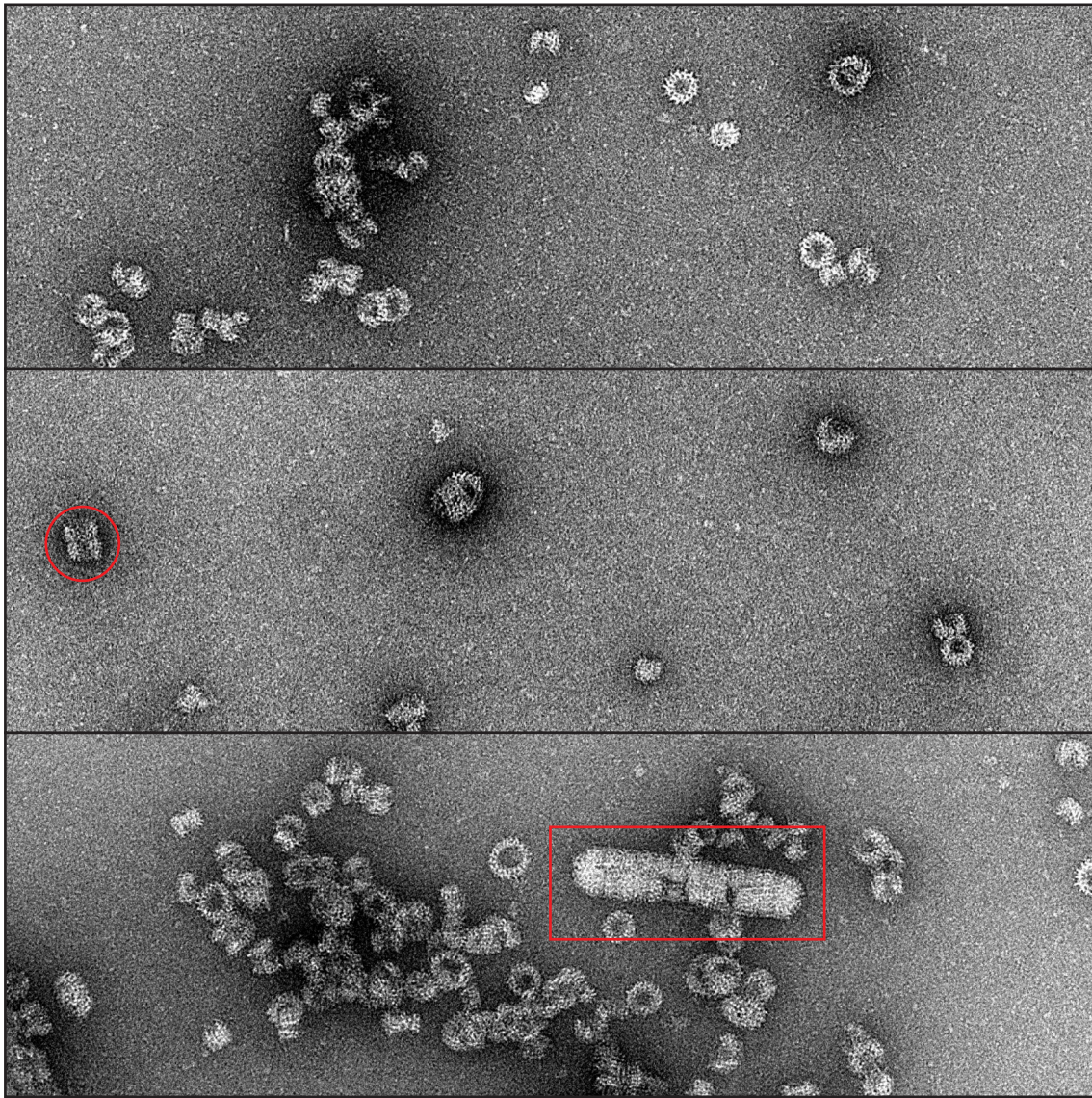
Figure 41 Electron-microscopic images of negatively stained Vipp1 showing mono-rings, double-rings (red circle), and rods (red rectangle).
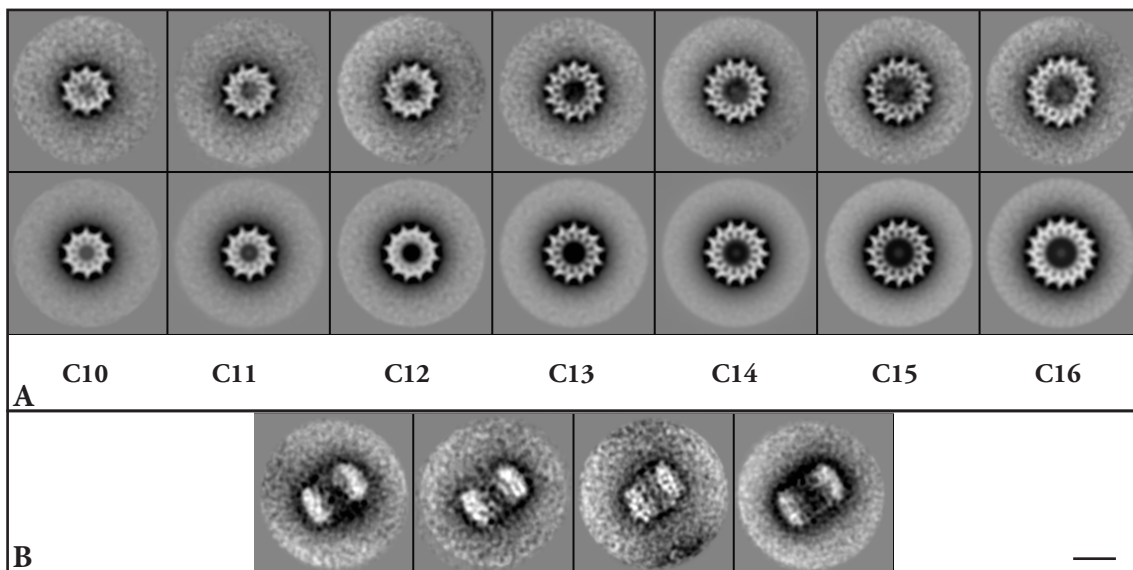
Figure 42 Representative class averages of Vipp1. A) Top-views of different ring sizes are shown without symmetrization (top) and with symmetrization according to the respective number of spikes (bottom). B) Typical side-views of Vipp1 also showing variable size. Scale bar: 200 Å.
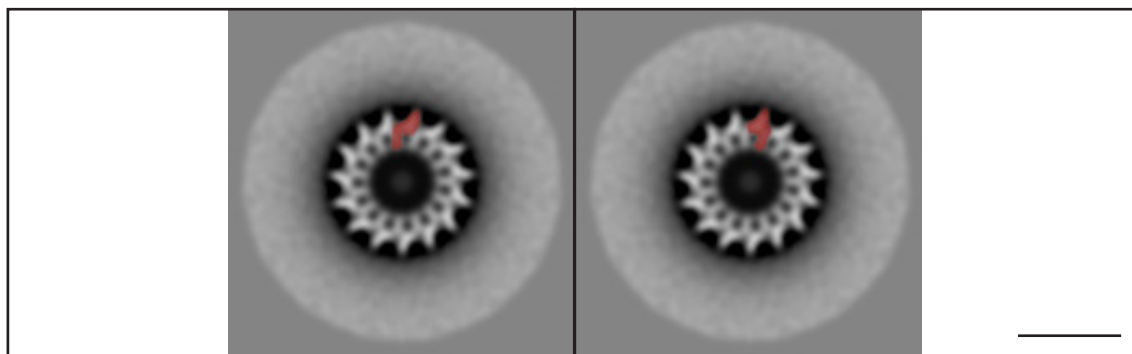
Figure 43 Two possibilities exist for describing one ring module, i.e. the basic building block of a ring (colored in red). Scale bar: 200 Å.

and side-views of different sizes (Figure 42). Rings bearing as few as seven and more than 20 spikes were also observed (data not shown). Class averages of top-views were symmetrized according to their respective number of spikes. The top-views clearly depict ring-like structures of varying diameters which range from ~230 Å for the 10-ring to ~330 Å for the 16-ring. Each ring seems to be constituted by a varying number of ring modules, which in turn appear to be defined by one outward protrusion and one inward rail (Figure 43). The side-view class averages define the rings as cylinder-shaped objects of ~190 Å in height (Figure 42 B). The spikes and rails seen for the top-views are partly depicted by the side-views as well. For this dataset, aggregations like the two-ring assemblies or the rods were not further processed. It has to be noted for both, however, that the assembly to these structures is not ring-size-specific.

The heterogeneity seen in the dataset led to the decision of using the multirefine procedure of EMAN1.9 for 3-D reconstruction. First trials of this procedure showed that using Gaussian spheres, cylinders or rotational averages (of top-view class averages that were stacked to a certain height) as references does not produce a good separation of different ring sizes. Consequently, class averages showing top-views were stacked to the height of Vipp1 (stacked 3Ds) and used as references for a first multirefine procedure (Figure 44 A). One of the resulting 3-D reconstructions was segmented and one ring module excised, which was then docked into the stacked 3D references and symmetrized according to the respective point-group symmetry (Figure 44 B). This procedure yielded simulated 3Ds, which were used as more accurate references in a second multirefine procedure with a newly acquired dataset comprising 1,710 particles (Figure 44 C). It has to be noted that although an improvement of ring-size separation could be observed by this procedure, the algorithm still failed to entirely separate the dataset correctly.
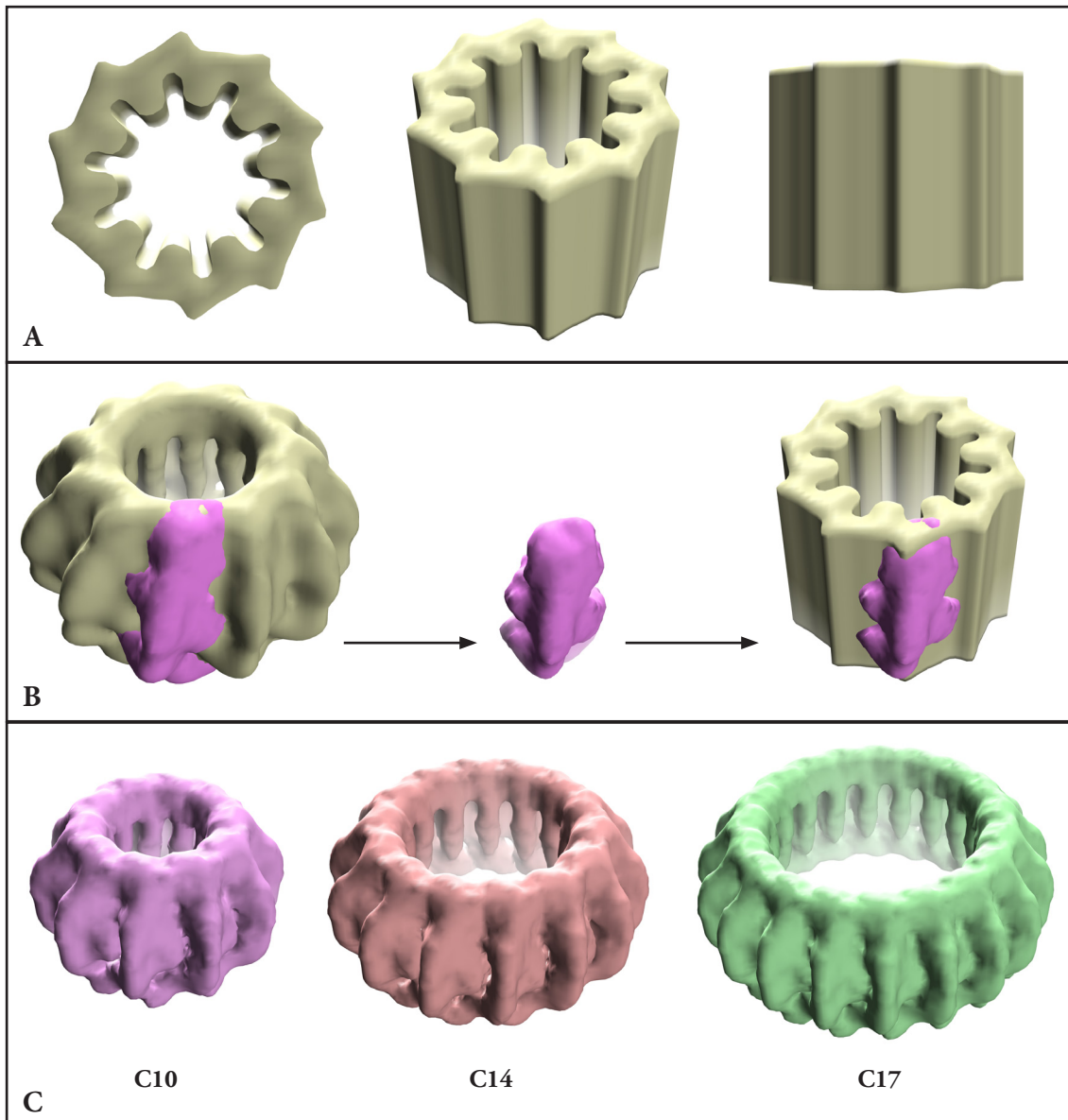
Figure 44 Creating Vipp1 references for the multirefine procedure. A) Top-view class averages of different ring sizes were symmetrized according to their respective number of spikes, stacked to the height of a Vipp1 molecule and used as references for a first multirefine (stacked 3-Ds). B) One of the resulting 3-D reconstructions was used to excise a ring module, which was aligned to each of the stacked 3-Ds. C) The aligned monomer was then symmetrized to simulated 3-Ds of different sizes.

For 11-, 16- and 17-rings, only few particles were assigned to the respective 3-D reconstruction, resulting in poorly defined structures. For 10-, 12-, 13-, 14-, 15-, and 18-rings, the resulting density maps demonstrate comparable structural assemblies among different ring sizes at resolutions between 20 and 30 Å (Figure 45). As already suggested by the 2-D analysis, Vipp1 rings are cylindrical assemblies with outward protruding spikes and inward protruding rails. The outer walls of the structures are slightly curved and the spikes are tilted to one direction with respect to the longitudinal axis of the cylinder. Since the tilt angles vary between
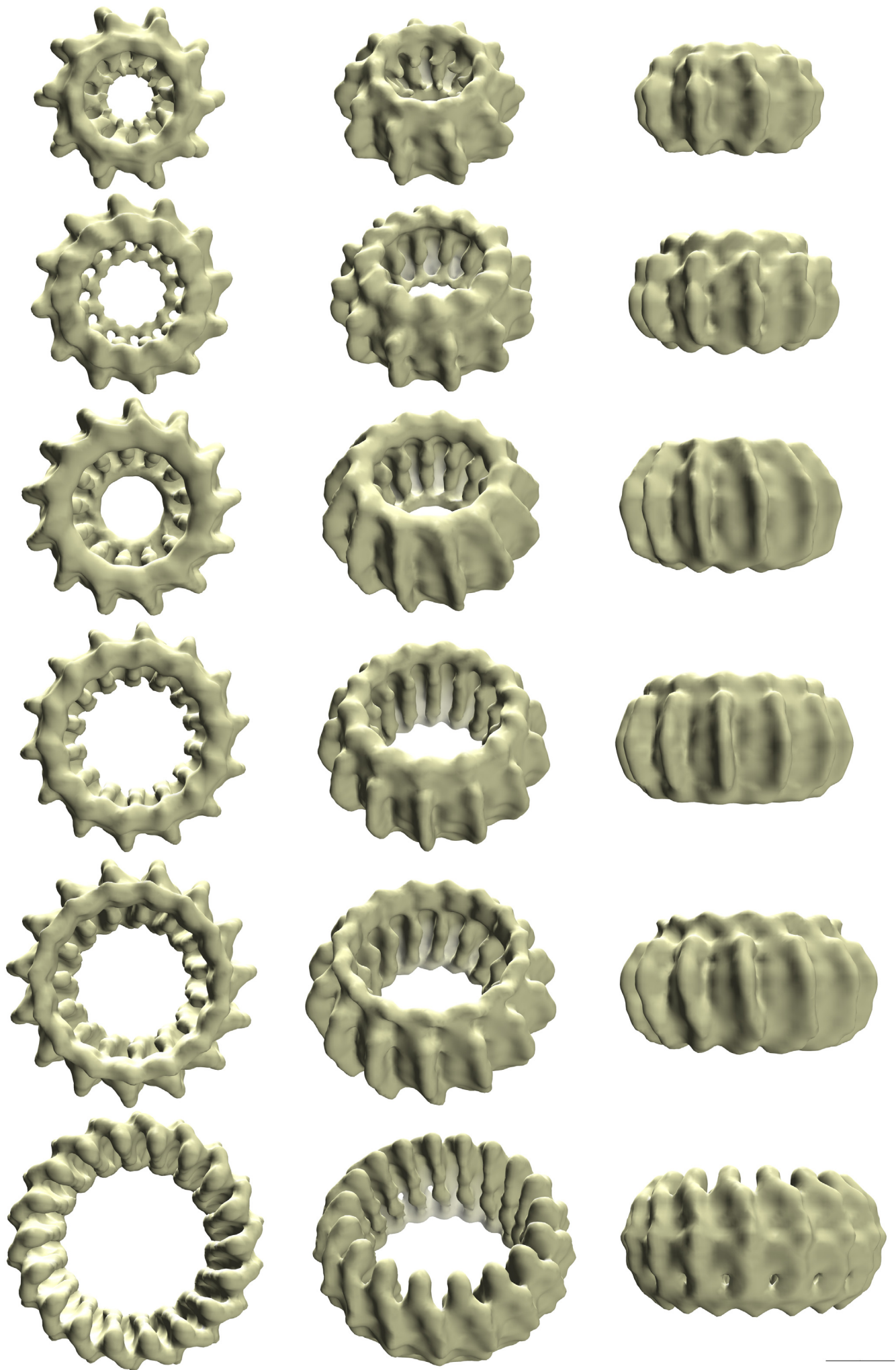
Figure 45 wtVipp1 density maps of different ring sizes (rows) in top (left), tilted (middle) and side views (right). Scale bar: 100 Å.

the density maps, no measurements of the former are shown. When depicting the structures from their top-view, it is apparent that one opening of the cylinder is wider than the other one. The wider opening will hereafter be defined as the top of the ring, whereas the narrower opening will be its bottom. The narrower opening is produced by a "ridge" that is very pronounced in smaller ring sizes and less pronounced in larger ones and which seems to originate from the rails (Figure 46). Dimensions of different ring sizes are displayed in Table 4 and show a similar height for all of the measurements. The slight differences can be accounted for by resolution- and measurement-dependent inaccuracies, as well as inconsistencies resulting from an incomplete ring-size separation. The diameters of the rings increase almost linearly with an increasing amount of ring modules (Figure 47). The rail length was measured from the top tip of the rail down to the bottom of the ring and shows similar values for five out of six density maps. The density map of the 18-ring exhibits no distinguishable rails and thus no measurement could be performed. Examining the ring-structures more closely, the spikes of all density maps show clearly pronounced ledges which show an equal length as the rails.

The attempt to improve the resolution of the density maps using the above mentioned 3-D reconstruction procedure with a combination of multiple datasets (6,171 particles) was not successful.
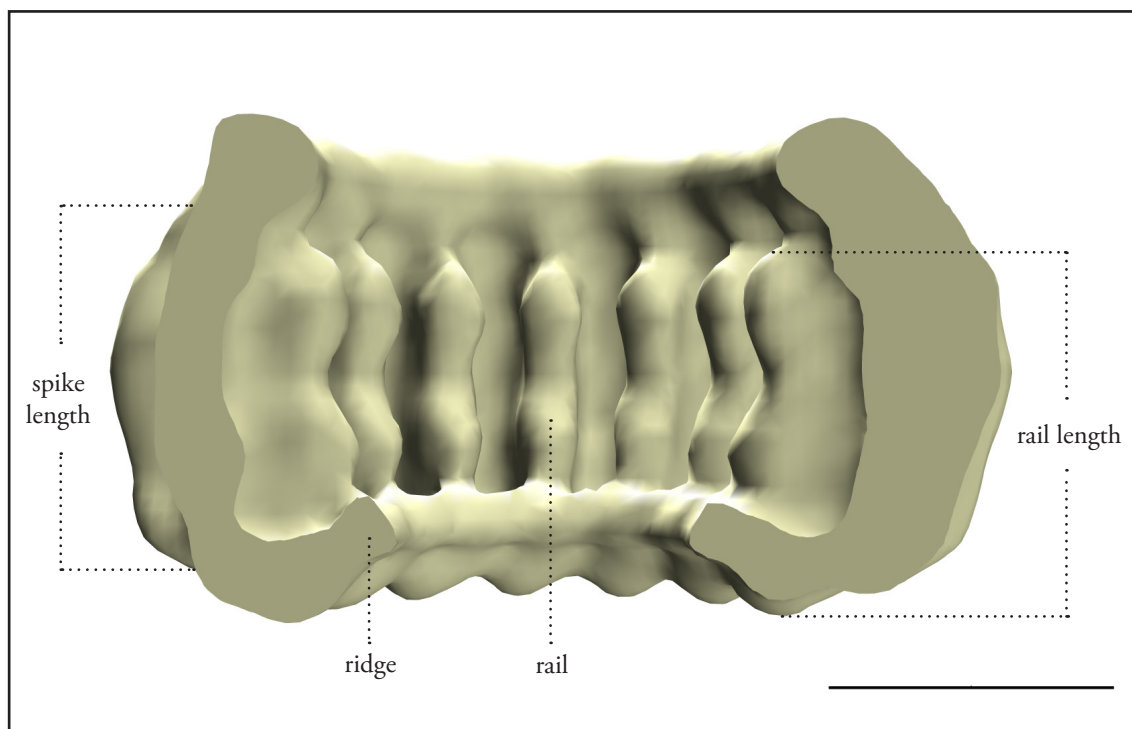
Figure 46 Internal view of a Vipp1 13-ring showing the curvature of the outer wall, the spikes, the rails and the ridge. Note that the distances between the horizontal lines indicating spike and rail lengths are equal for both structures. Scale bar: 100 Å.

Table 4 Dimensions of wtVipp1 density maps. The ring symmetry is equal to the number of spikes protruding from the rings. For C^{18}, no distinguished rail exists in the density map.

Ring Symmetry	Height (Å)	Diameter (Å)	Rail length (Å)
C10	131	243	100
C12	141	268	103
C13	150	279	107
C14	144	299	108
C15	152	320	105
C18	155	334	N/A
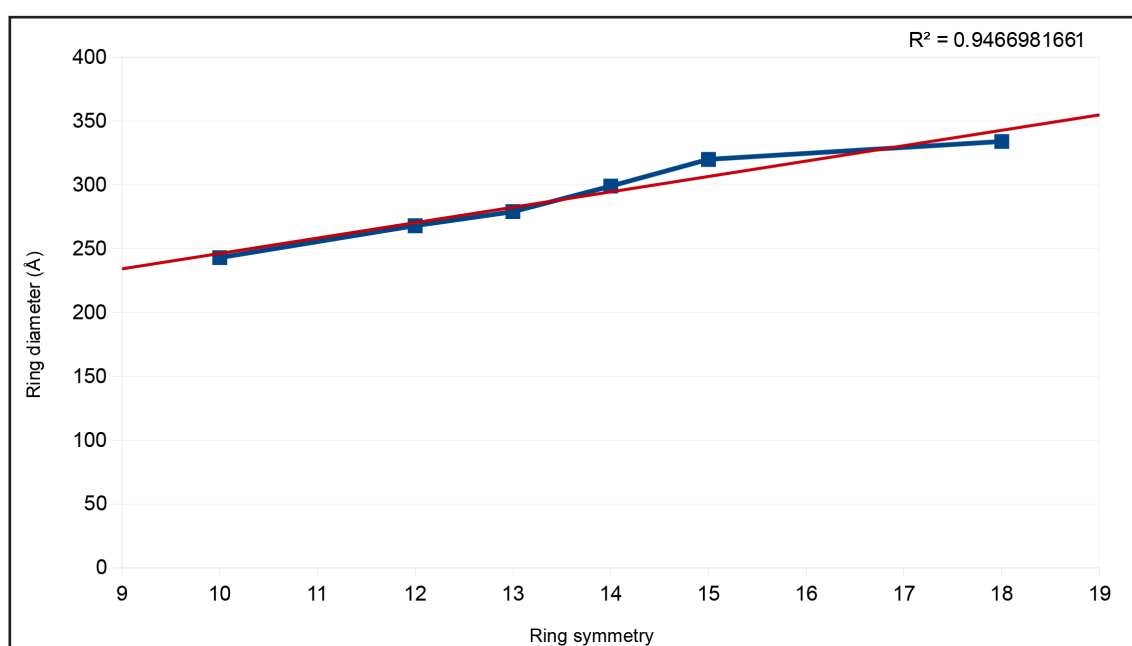
Figure 47 Vipp1 ring diameter in relation to the amount of ring modules (=ring symmetry). The red trend line highlights the almost linear increase in diameter with increasing ring symmetry.

2.2 MUTATIONAL ANALYSES

In order to identify the amino acids that are essential for the formation of such versatile cylindrical assemblies, the primary structure of the protein was altered by (i) introducing point mutations into different locations of the N-terminal region, (ii) eliminating half of the N-terminal helix α1 by deleting the first 10 amino acids (ΔH1a), or (iii) deleting the 52 C-terminal amino acids (ΔCT). The resulting samples of Vipp1 mutants were then negatively stained and visualized by transmission electron microscopy. Figure 48 gives an overview of the N-terminal mutations that were surveyed by negative staining TEM. However, due to the fact that only few of these mutations produced visible implications for the ring formation, only mutants ΔCT, N11 and N12 will be described in more detail. A representative image of each of the other mutations will be provided in Appendix D.

	M	G	L	F	D	R	L	G	R	V	V	R	A	N	L	N	D	L	V	S	K	A	E	D	P
														Helix α1											
N1				C																					
N2				C			A																		
N4				Y			Y																		
N5				Y			Y			Y															
N8						A																			
N11													A		A										
N12																					A	A	A		
N18										P	P														
S4				S			S			S															
S6				S			S			S	S														
S10				S			S			S							S		S						
13-4-1				Y			Y								Y										
13-4-2/3				Y			Y				Y				Y										
ΔH1a	x	x	x	x	x	x	x	x	x	x	x														
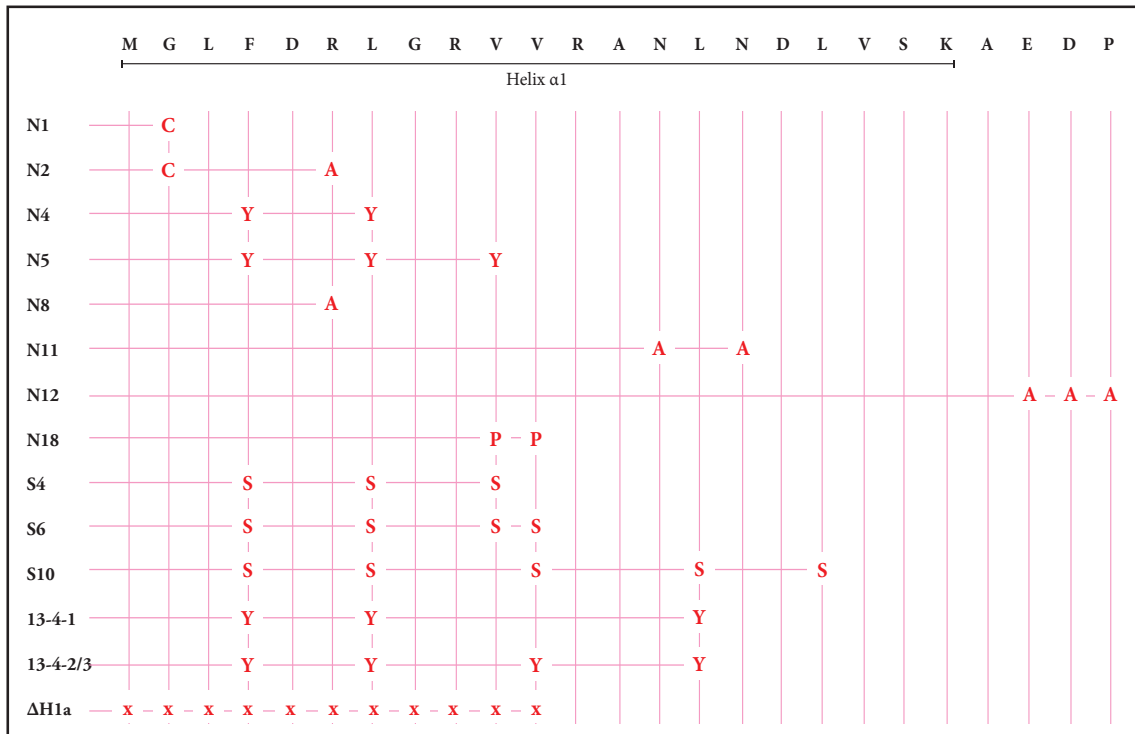
Figure 48 N-terminal mutations that were monitored with the transmission electron microscope. The N-terminal sequence of Vipp1 is shown in black above a marker indicating the length of the predicted Helix α1. The mutations are named in the right column. Red letters indicate the substitute amino acids for the respective position of the wild-type sequence. x=deletion
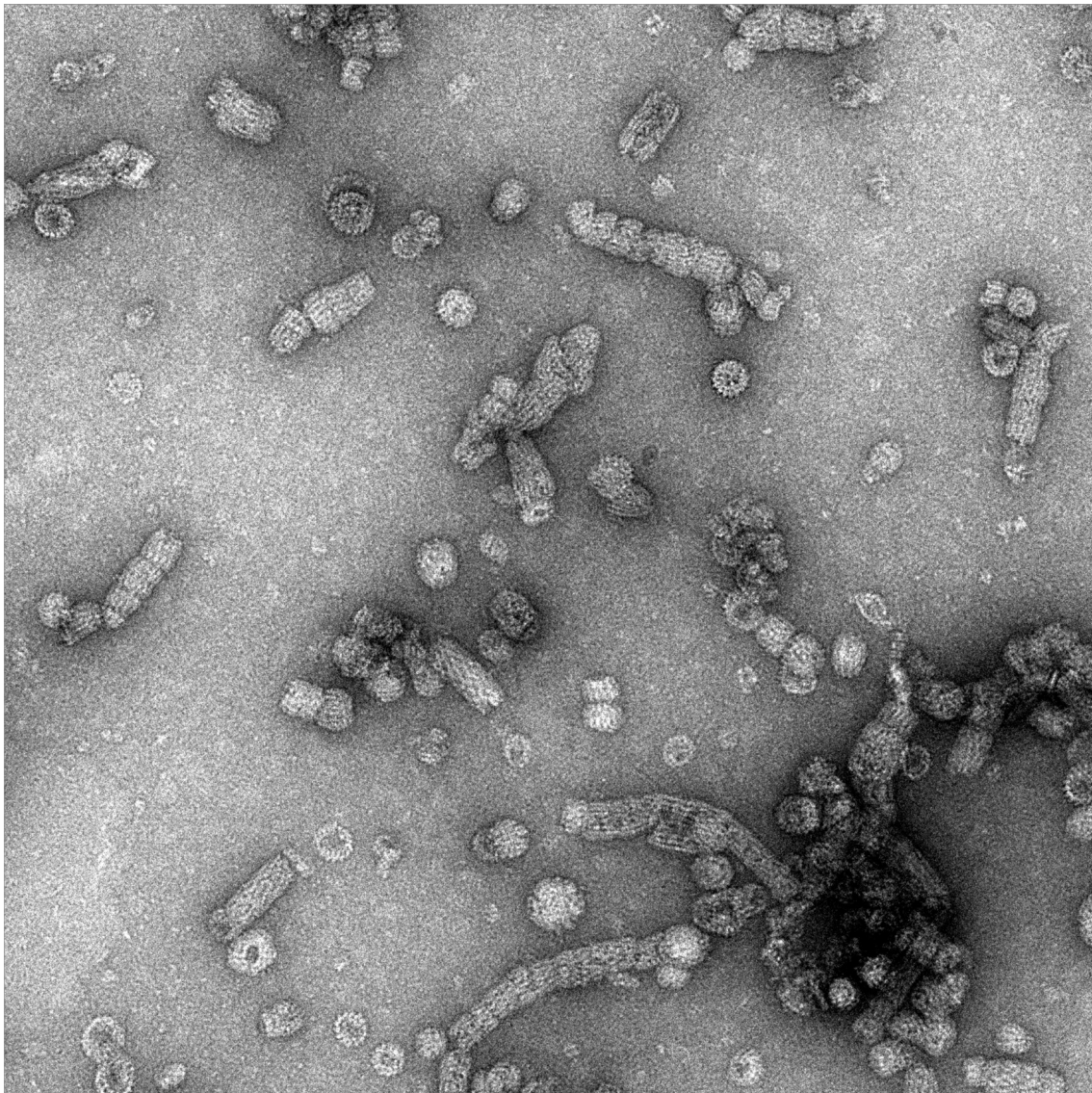
Figure 49 Typical electron-microscopic image of the Vipp1 ΔCT mutant, forming mono-rings and aggregates of double-rings and rods.

Deletion of the 52 C-terminal amino acids produced rings that resemble the ones seen for wild-type Vipp1 (Figure 49), although substantially more rings seem to aggregate into rod-like structures. Furthermore, double-rings and a smaller box-like structure are visible, both of which were already visible in wtVipp1, albeit the latter was not considered further in wtVipp1 samples. In spite of the seemingly normal assembly into rings of different sizes, two-dimensional analysis was performed with ΔCT due to the fact that the deleted region corresponds to the amount of C-terminal amino acids that PspA lacks in comparison to Vipp1. By selecting mono-rings, double-rings and the box-like structures, a dataset of 4,708 particles was created and subjected to class average refinement in iterative MRA-MSA cycles. Using the resulting class averages, the dataset was split into top-views of mono-rings, side-views of double-rings and views of box-like structures, followed by further MRA-MSA cycles. Refined class averages of the top-views show structures comparable to the ones presented for wt-Vipp1 (Figure 50 A; compare to Figure 42 A). Class averages of the double-rings clearly show that different ring sizes can interact with each other to create these assemblies (Figure 50 B). The contact between the rings seems to be created by their narrower opening. The box-shaped structures show four distinguishable masses that constitute the assembly (Figure 50 C). Different views of the particle suggest a diameter 175 Å. Aggregates of these peculiar structures can readily be observed in single images and are well defined in the class averages as rod-shaped structures of varying length.

As mentioned above, the N-terminal mutations generally did not show any implication for the formation of Vipp1 rings. However, mutation N12 showed a dramatic reduction of Vipp1 rings and the occurrence of bar-like structures (Figure 51). Electron microscopic images of the N12 mutant with subsequent selection of these bar-like structures yielded a dataset of 4,403 particles, which were subjected to class average refinement *via* iterative MRA-MSA cycles in IMAGIC-5. The bars differ in length between 70 Å and 140 Å and often show a zigzag-pattern (Figure 52). Some bars also show a certain curvature where one part, the upper or the lower one, is curved slightly more than the rest (see red dotted circle in Figure 52). Similar destructive impact on Vipp1 rings were observed for mutant N11, where the structural damage seems to be even greater due to the absence of bar-like structures. Instead, an increased protein background and very small structural fragments are visible (Figure 53). The small size of the resulting structures did not allow two-dimensional analysis.
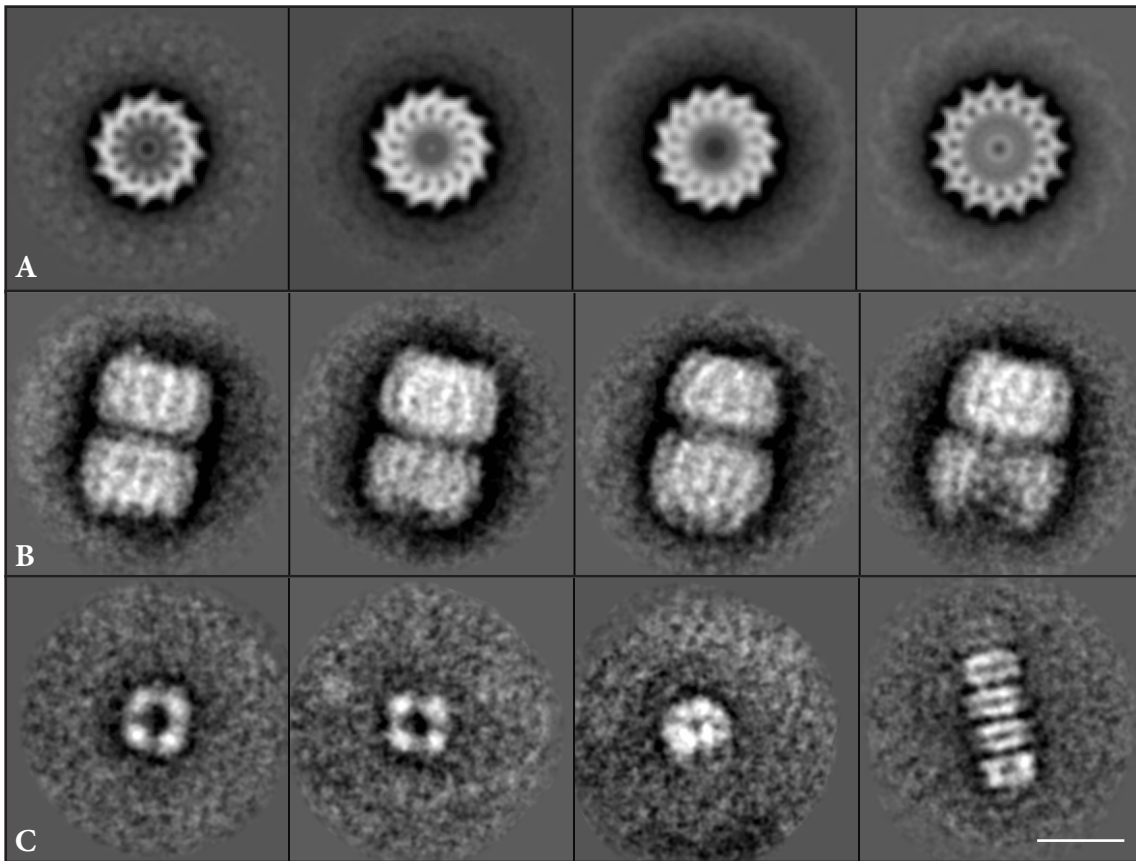
Figure 50 Class averages of ΔCT Vipp1. A) Top-views symmetrized according to their respective number of spikes. B) Double-rings showing the aggregation of different ring sizes. C) Box-shaped structure constituted of four distinct masses and putative rod-like aggregation of box-shaped particles. Scale bar: 200 Å.
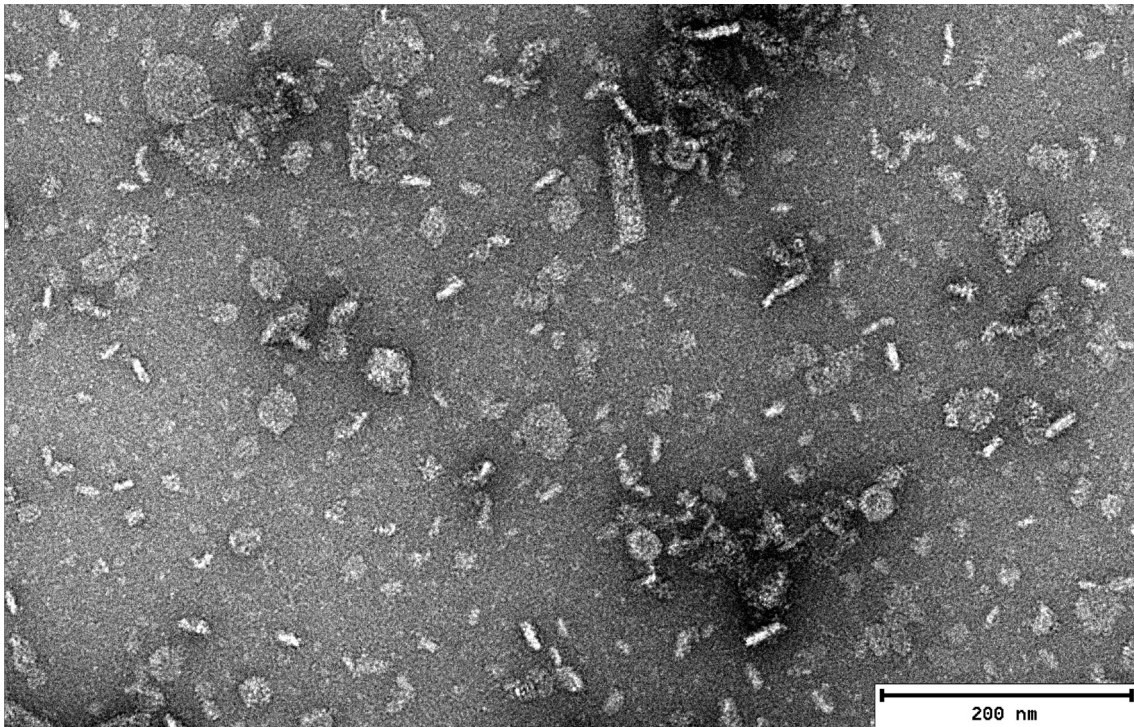
Figure 51 Typical micrograph of Vipp1 N12 mutant. The mutation led to a dramatic decrease in Vipp1 rings and an appearance of bar-like structures.
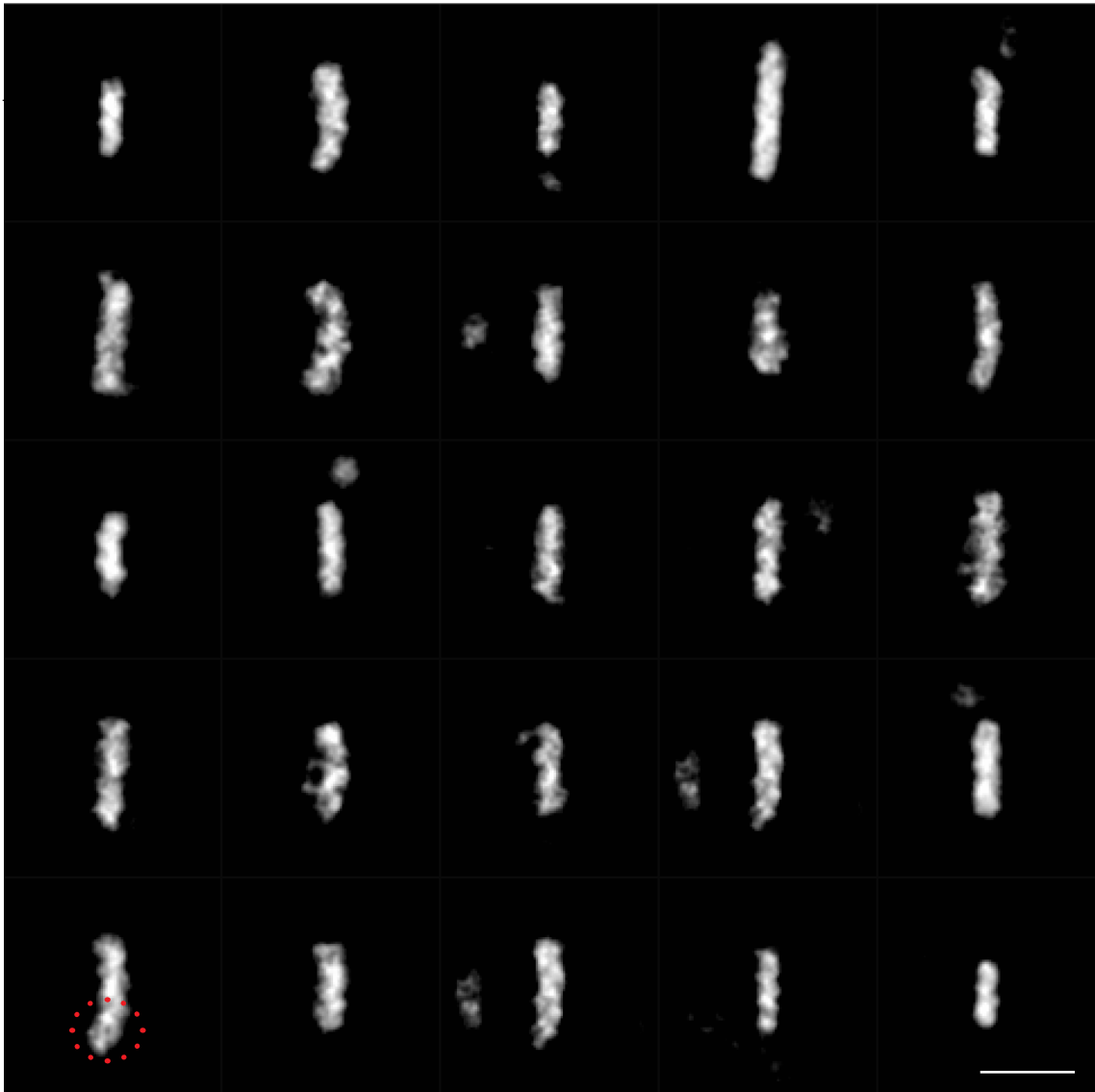
Figure 52 Representative class averages of the bar-like structures found in a sample of mutant N12. The bars are of different length and some show a certain curvature. The red dotted circle indicates an increase in curvature of part of the bar. Scale bar: 100 Å.
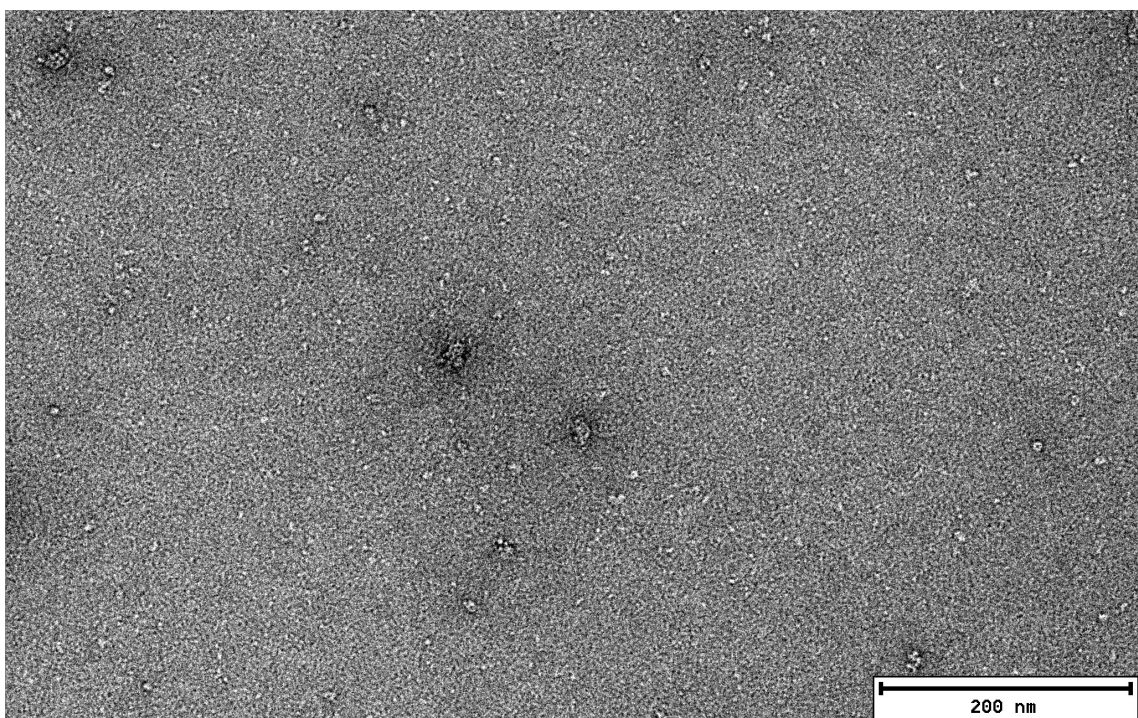
Figure 53 Typical electron micrograph of the Vipp1 N11 mutant. Vipp1 rings are very sparsely visible and an increment of small structures can be noted. No rods or bar-like structures can be found.

2.3 Vipp1-liposome interactions

In order to test whether Vipp1 rings are a structure only built to transport the protein, while dissolving for the actual interaction with lipids, different samples of lipids and Vipp1 were visualized with the transmission electron microscope. For these studies, 5 % ammonium molybdate with 0.1% trehalose at pH 7.1 was used due to the benefits of this stain when working with liposomes (see Harris 1986).

Liposomes of DOPG appeared as rounded structures, which often were much larger than the 500 Å that the sizing of the liposomes should have allowed. Inside these structures, small rounded vesicle-like structures were often seen (Figure 54 A). Some squares of the EM grid showed large bilayer "myelin bodies" (*ibid.*;Figure 54 B). Liposomes of SQDG or mixtures of SQDG+DOPG did not differ in appearance from the ones containing DOPG (Figure 54
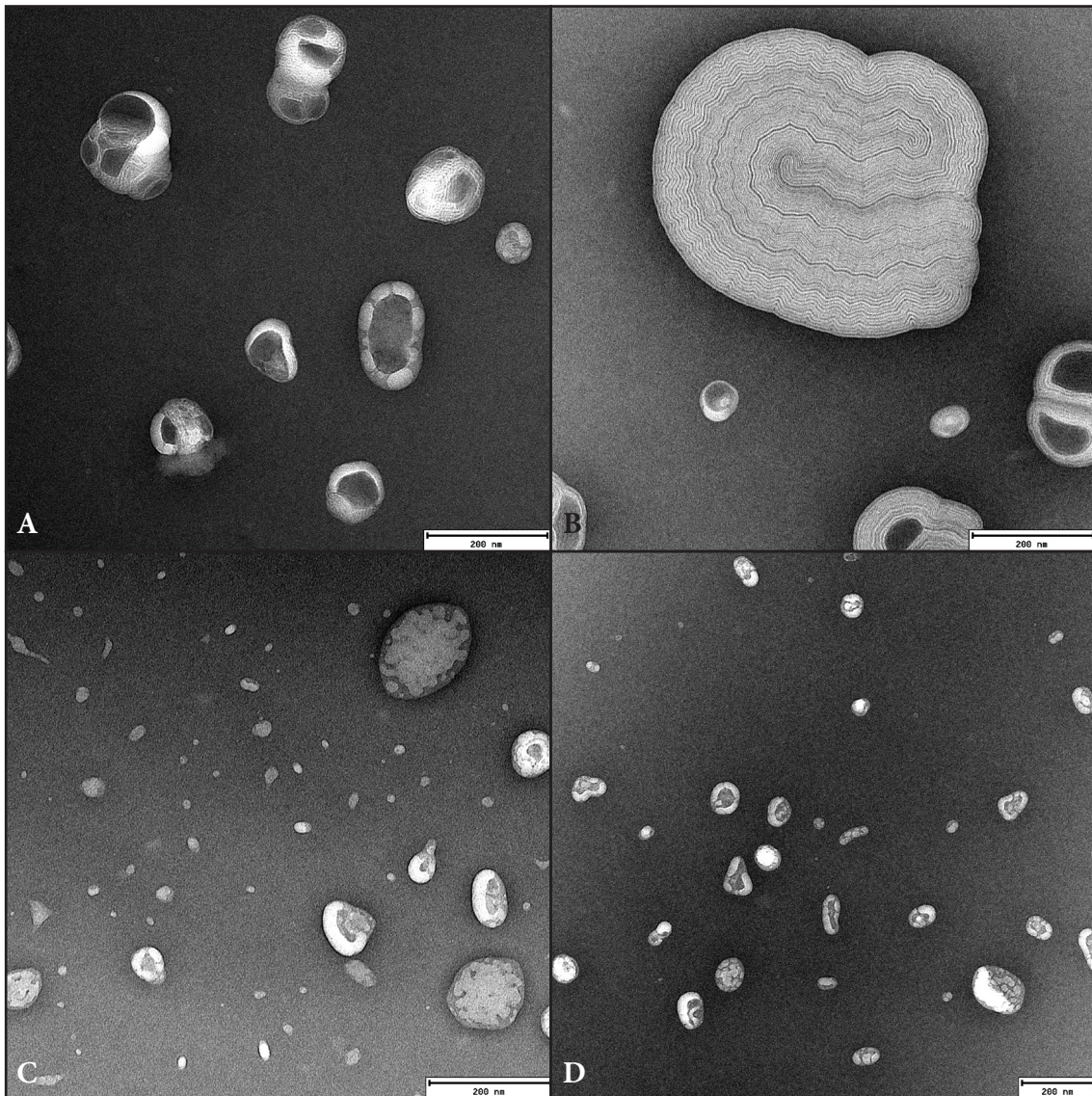
Figure 54 Liposomes containing different lipids stained with ammonium molybdate (5 %; 0.1 % trehalose; pH 7.4). A+B) DOPG liposomes should be sized to 50 nm and mono-layered, but show larger structures and inclusions (A). Some parts of the grid also displayed large myelin bodies (B). As already seen for the DOPG sample, SQDG liposomes (C) and SQDG+DOPG liposomes (D) show rounded inclusions and are partly larger than the expected size of 50 nm.

C+D). Even when Vipp1 was added to the SQDG, DOPG, or SQDG+DOPG samples, no visible change in liposome structure was observed. Furthermore, Vipp1 rings were dispersed among the sample and mostly detached from the liposomes (Figure 55). This did not change, when the Vipp1 concentration was increased by a factor of 10.

For an alternative lipid composition of the liposomes, DOPG was mixed with the main thylakoid membrane lipid MGDG. The images of the electron microscope show that the sizing procedure did indeed not affect the sizes of the liposomes and that large myelin bodies are formed (Figure 56 A+B). The combination of DOPG and MGDG seemingly produced more multi-layered vesicle than the other lipids. When Vipp1 was added to the lipid mixture, the overall appearance of the liposomes did not change. In contrast to previous samples, however, Vipp1 rings were largely associated with the liposomes (Figure 56 C+D). Even when increasing the Vipp1 concentration by a factor of 10, only few Vipp1 rings can be seen detached from the liposomes, whereas most rings show an association with them.
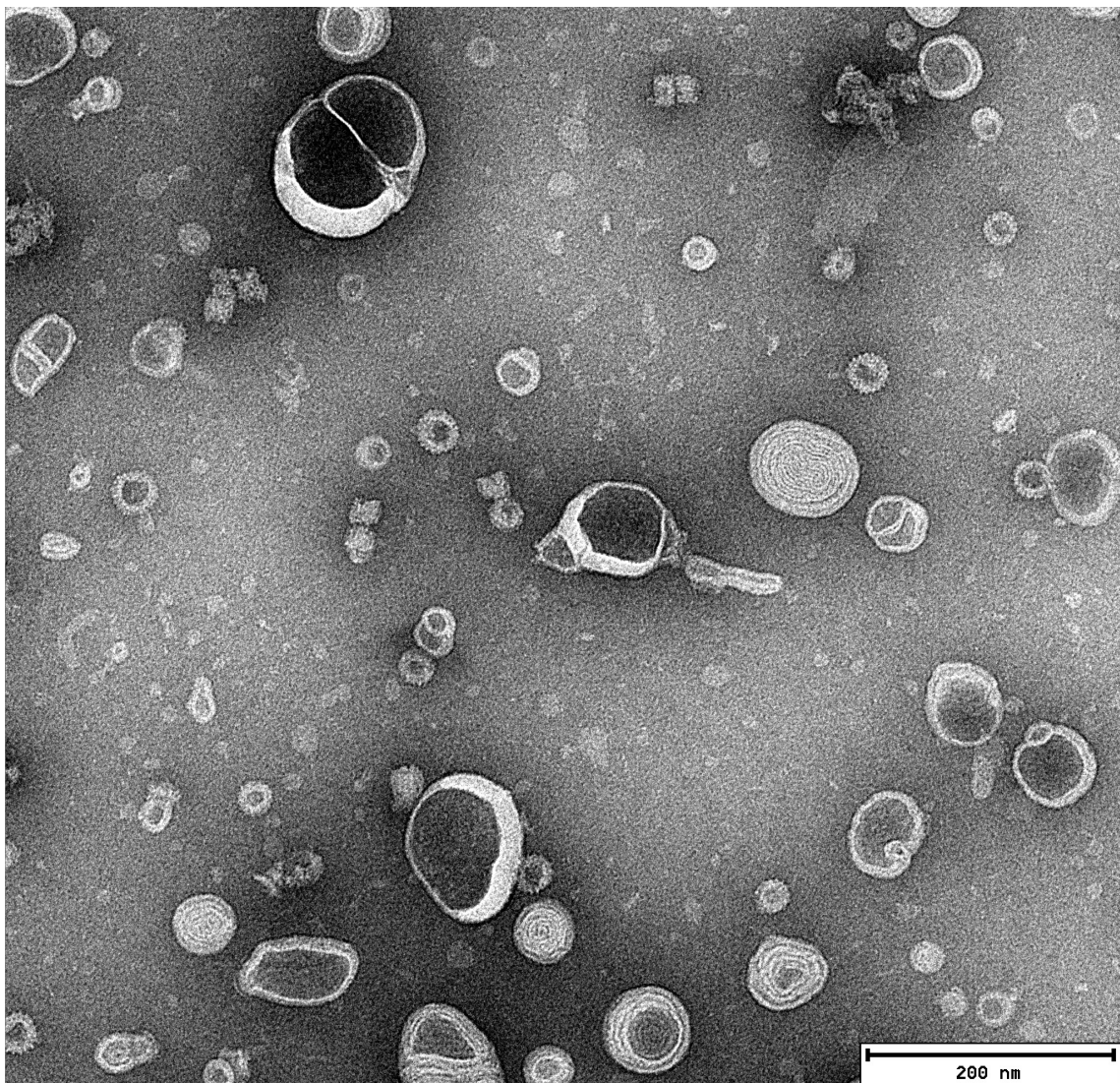
Figure 55 DOPG+Vipp1 stained with ammonium molybdate (5 %; 0.1 % trehalose; pH 7.4). Note that only a small amount of ring molecules is associated with liposomes, whereas most of the Vipp1 rings seem randomly dispersed. This was also seen for SQDG and DOPG+SQDG.
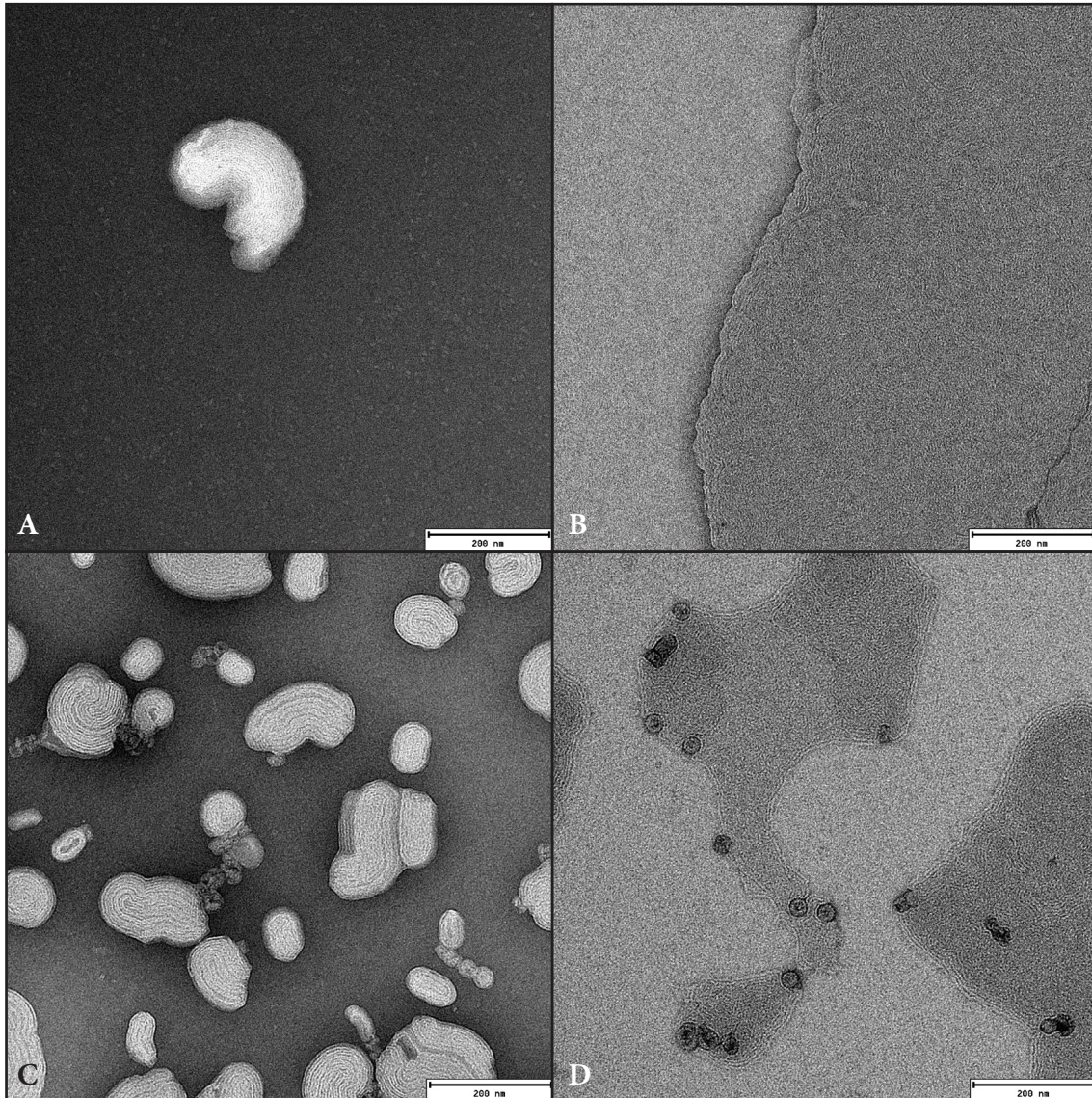
Figure 56 DOPG+MGDG and DOPG+MGDG+Vipp1 stained with ammonium molybdate (5 %; 0.1 % trehalose; pH 7.4). Most of the liposomes found were multi-layered and larger than the 50 nm to which they were sized (A). Furthermore, large myelin bodies could be found (B). Vipp1 can be seen closely associated to those lipids (C), or – on some parts of the grid – even seemingly incorporated into them (D).
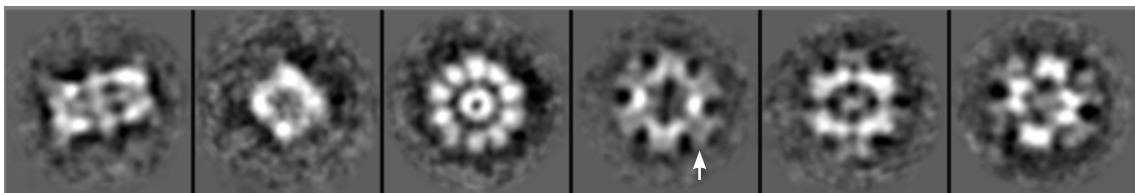
Figure 57 Representative class averages of the native BgAChBP dataset. The averages show a heterogeneous dataset, with the first two classes belonging to an unknown protein and the last four classes representing the rosette-like protein (i.e. BgAChBP). Note the protrusions emerging from the distal side of the chimney-like masses (white arrow).

3 *BIOMPHALARIA GLABRATA* ACETYLCHOLINE-BINDING PROTEIN

3.1 3-D ELECTRON MICROSCOPY OF HEMOLYMPH BgAChBP

Protein purification was conducted by ███████████. Hemolymph of *Biomphalaria glabrata* was centrifuged and the supernatant subjected to ion exchange chromatography, in order to yield the final sample which contained rosette-like protein (e.g. LIEB ET AL. 2006). Parallel to the biochemical experiments described in SAUR ET AL. (2012), I commenced three-dimensional analysis of the sample. A first batch of 52 micrographs had already been imaged and digitized by ███████████ and ███████████. Eight micrographs were discarded due to ice contamination and particle selection on the remaining ones by ███████████ ███████ yielded a dataset of 8,556 particles, which were subjected to MRA and MSA. The resulting class averages show high heterogeneity of the dataset, being comprised of highly symmetrical, rectangular and squared particles (Figure 57). Due to this apparent heterogeneity, the images were processed with the multirefine algorithm of EMAN1.9 (see B6.1.3 for details), which resulted in 3D reconstructions showing a dodecahedron-shaped particle, a cuboid and an amorphous mass, respectively (Figure 58). Given the fact that projections of the rosette-like protein could be explained by the dodecahedron-shaped structure, only the 4,248 images classified to the latter were used for the refinement dataset. The addition of a second batch of images yielded the final dataset, comprising 9,770 particles, which was then subjected to two different refinement strategies using EMAN1.9.
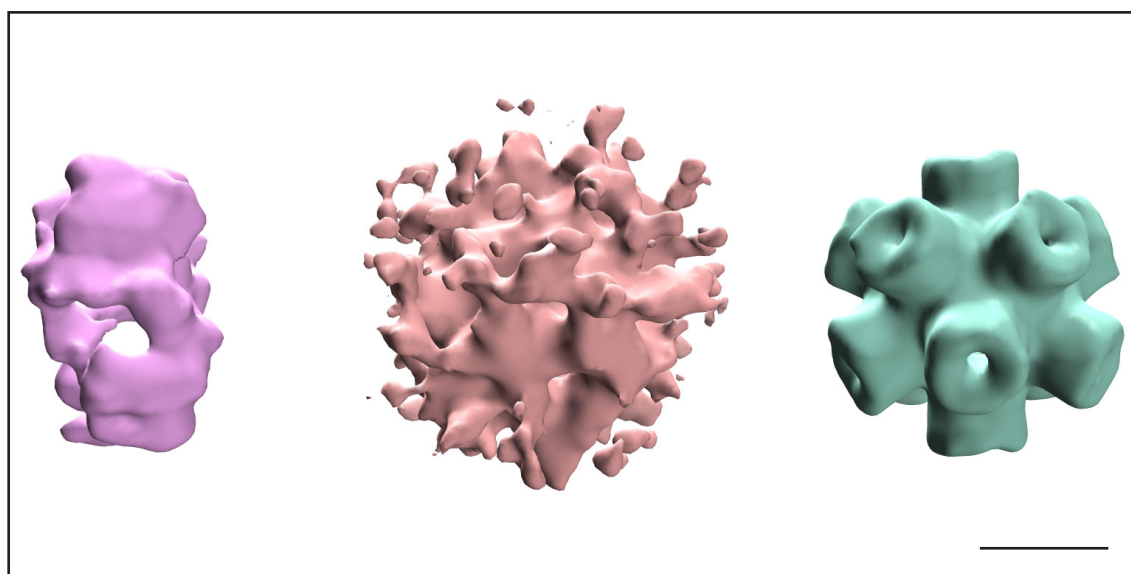
Figure 58 The 3D reconstructions resulting from the multirefine procedure. Images belonging to the cuboid structure (left) and the amorphous mass (middle) were discarded. Images belonging to the dodecahedral structure (right) were used for further refinement. Scale bar: 100 Å.

3.1.1 Traditional 3-D refinement

The definition of a "traditional refinement" will vary according to the person asked. In this context, I define this type of refinement as the one traditionally used in many laboratories, where the high resolution cut-off for a dataset is set according to the resolution of the last 3-D reconstruction. This cut-off tends to be approximately 1.5 times higher than the resolution measured (e.g. 5 Å for a resolution of 7.5 Å), whereas the 3-D reconstruction is generally low-pass filtered to the measured resolution. A peculiarity observed from the beginning of the reconstruction procedure, was the formation of finger-like protrusions around the five-fold symmetrical axes. These protrusions were present despite low-pass filtering the dataset to 20 Å in a control refinement (Figure 59). A closer examination of the class averages depicted in Figure 57 shows that they can already be seen by 2-D analysis. Due to the occurrence in the control refinement, the protrusions were included into the subsequent traditional refinement. After 85 iterations, the refinement reached a stable maximum measured by the FSC, showing a resolution of ~6 Å according to the 0.5 threshold (Figure 60). In order to account for the envelope function of the microscope, a negative temperature factor of 375.84 Å^2 was estimated for and imposed on the final reconstruction by EMBfactor, using the final FSC and a figure-of-merit weighting (Rosenthal and Henderson 2003; Fernández et al. 2008). The result was low-pass filtered to 6 Å which yielded the final density map that is depicted in Figure 61 and shows the general appearance of a pentagonal dodecahedron. The protein
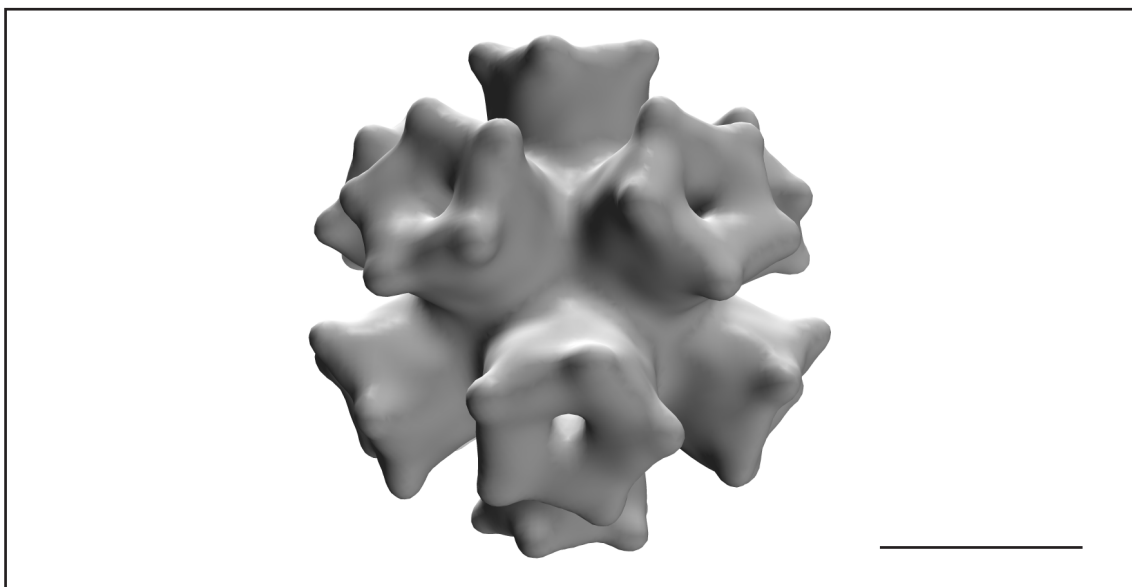
Figure 59 Result of a control refinement of native BgAChBP with a dataset low-pass filtered to 20 Å. Note that the protrusions described in the main text are also visible in this low-resolution structure. It must be emphasized that the density threshold for this structure is set to 1.7 times higher than the mass-correlated one. Scale bar: 100 Å.

consists of twelve pentameric chimney-like masses, which are joined at their bases. The outer diameter of the protein is 265 Å and each pentamer shows a height of ~85 Å and a width of ~105 Å (measured with protrusions). The center of the dodecahedron houses a cavity of ~100 Å in diameter that, judging by the density distribution of the reconstruction, seems to be empty. It is noticeable that the aforementioned protrusions have gained in size and are oriented almost parallel to the five-fold symmetrical axes. It is also peculiar that the surface of the density map is rather spiky.

Using a homology model of BgAChBP1 (see B6.1.4 for details), a pseudo-atomic model was created by placing 60 individual subunits into the 3-D density map *via* rigid-body fitting in UCSF Chimera (Figure 62). Judged by eye, the overall quality of the fitting seems to be excellent. A density map simulated on the basis of the pseudo-atomic model at 6 Å resolution shows a cross-correlation of 0.8 with the experimental density map. It is visible that the N-face (the side of the pentamer bearing the N-terminal α-helix) points towards the central cavity of the dodecahedron, whereas the C-face (the side of the pentamer bearing the C-terminus) points towards the solvent. The gate to the active site of the protein, constituted by loop C, is clearly visible on the outer surface of the 3D reconstruction, as is the single α-helix
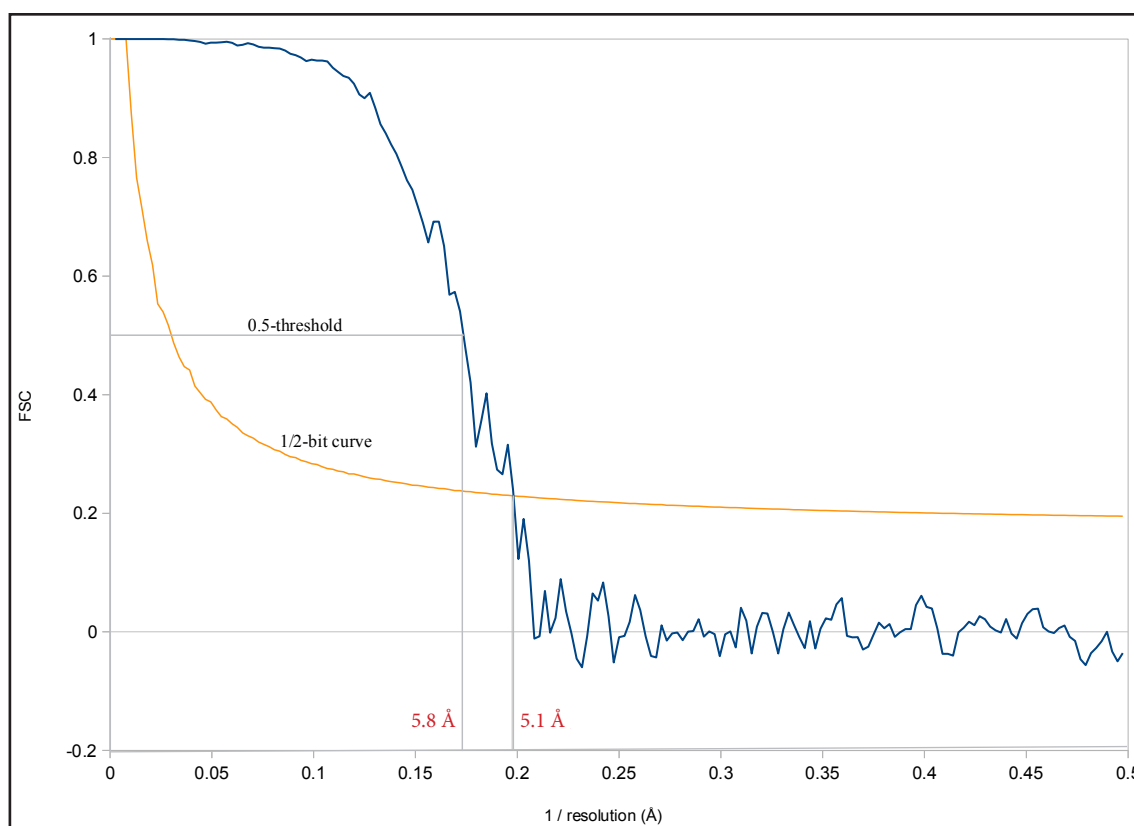
Figure 60 FSC of the final iteration of the traditional refinement.

5-fold axis

3-fold axis
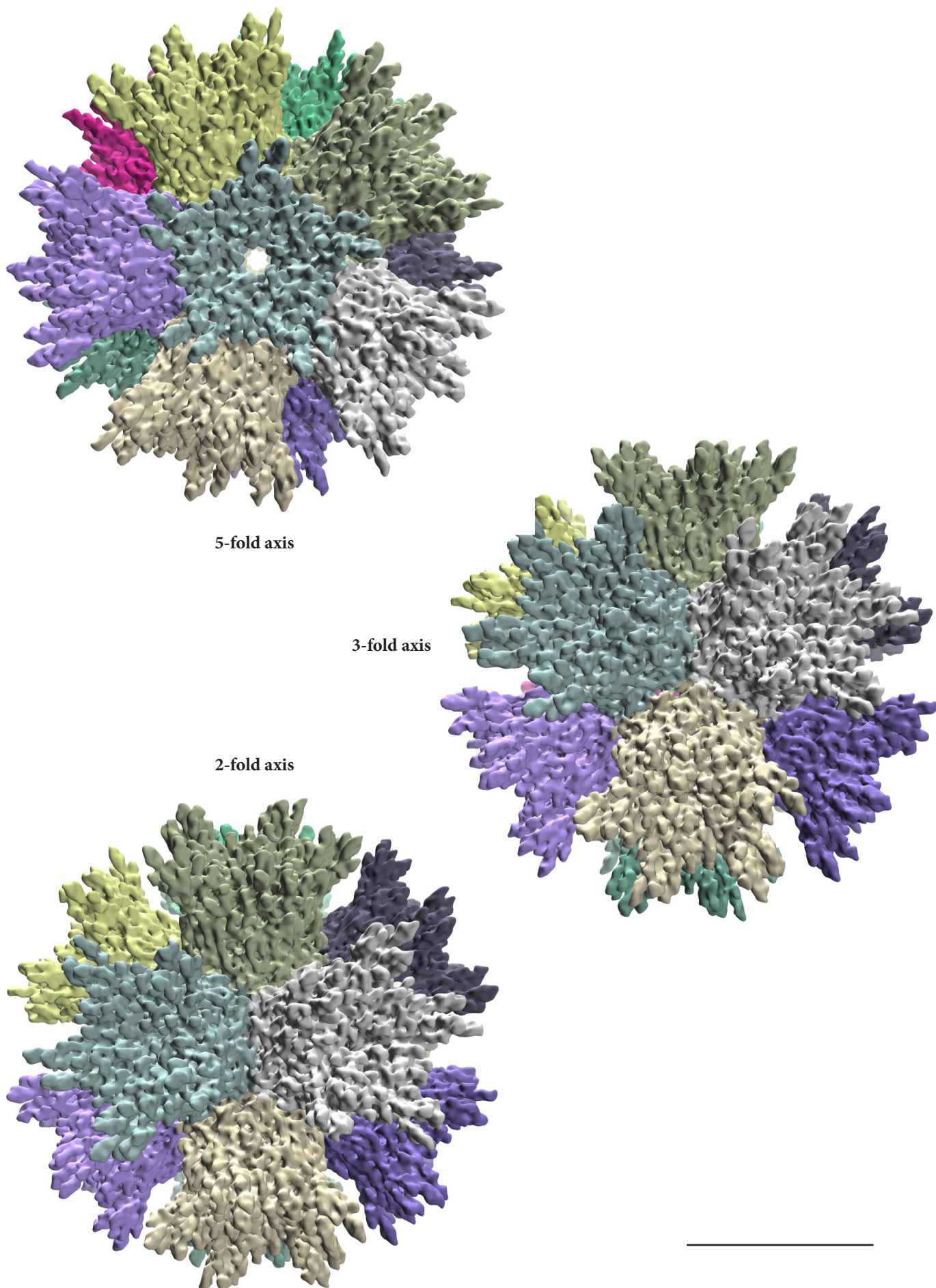
2-fold axis

Figure 61 Final 3-D reconstruction of the traditional refinement procedure. The density map is shown from the approximate views of one 5-fold, 3-fold and 2-fold axis, respectively. Each pentamer is highlighted in a different color. Scale bar: 100 Å.
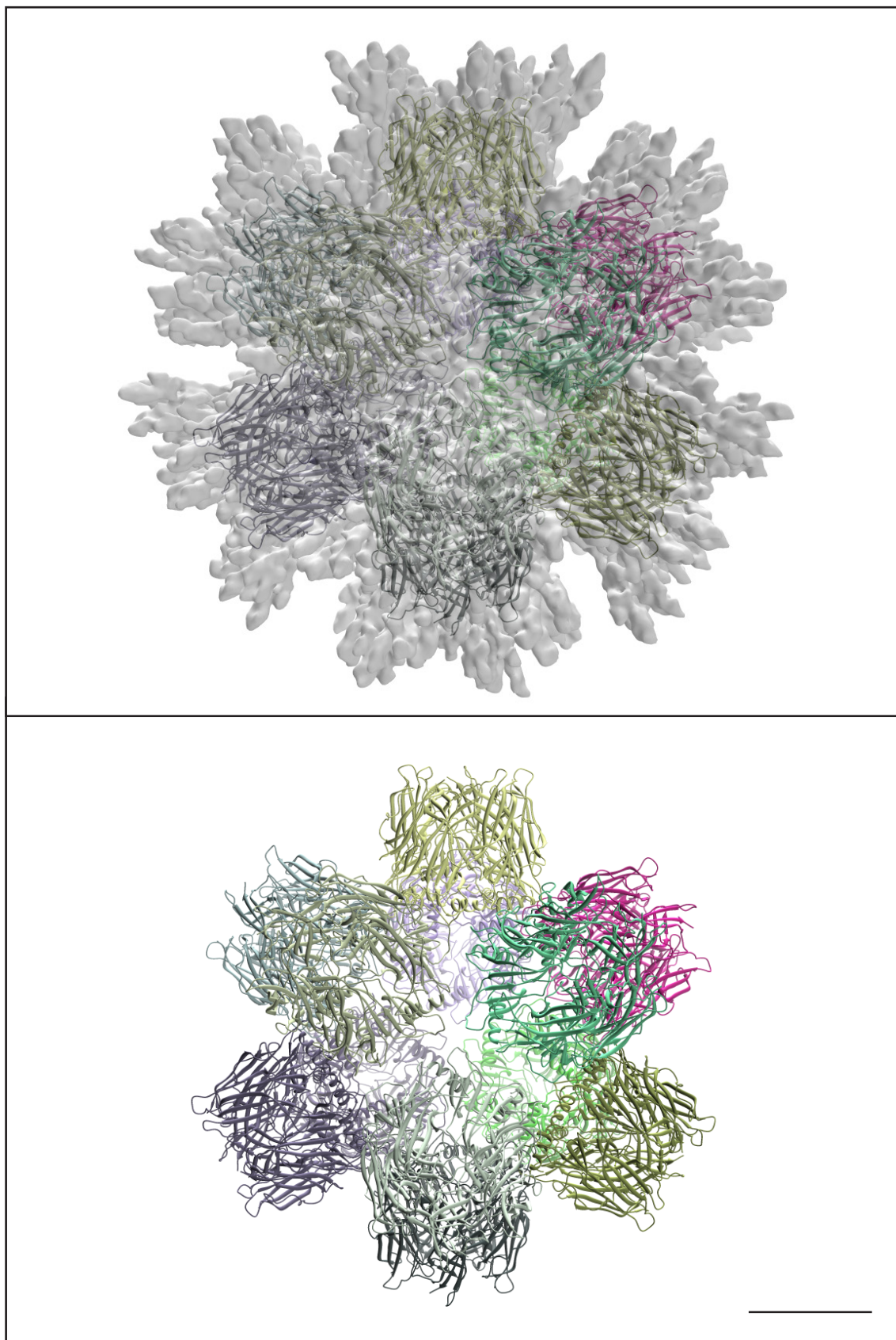
Figure 62 Homo-oligomeric pseudo-atomic model of BgAChBP1 using the density map of the traditional refinement as reference. The structures are shown approximately from one of their 3-fold axes. Scale bar: 50 Å.

present in this protein (Figure 63). The α-helix of the molecular model, however, would only fit well into the density map if flexible fitting would be applied.

The pseudo-atomic model seemed to substantiate the structural validity of the experimental density map. Nevertheless, prevailing doubts – primarily supported by the protrusions following the main symmetry axes and the spiky surface of the map – resulted in the decision to investigate a second type of refinement.

3.1.2 3-D REFINEMENT MINIMIZING NOISE BIAS

The low SNR experienced in cryo-EM images entails that multiple-reference alignments of images containing a lot of high-frequency information tend to partly align the noise, rather than the structural information, to the reference (see for example STEWART AND GRIGORIEFF 2004). To minimize this undesirable circumstance, I calculated a refinement that used different datasets for alignment and reconstruction. Both algorithms used band-pass filtered datasets, whose low-pass filters were set to 8 Å and 3 Å, respectively. Furthermore, the 3D reconstructions of each iteration were low-pass filtered to 10 Å prior to the re-projection of the references for the next iteration. As a final difference to the traditional refinement type, the protrusions emerging alongside the five-fold symmetry axes were carefully masked out after every iteration to prevent the accumulation of aligned noise in this region over the iterative cycles.

Stability of the FSC was reached within 15 iterations at a resolution of ~6 Å according
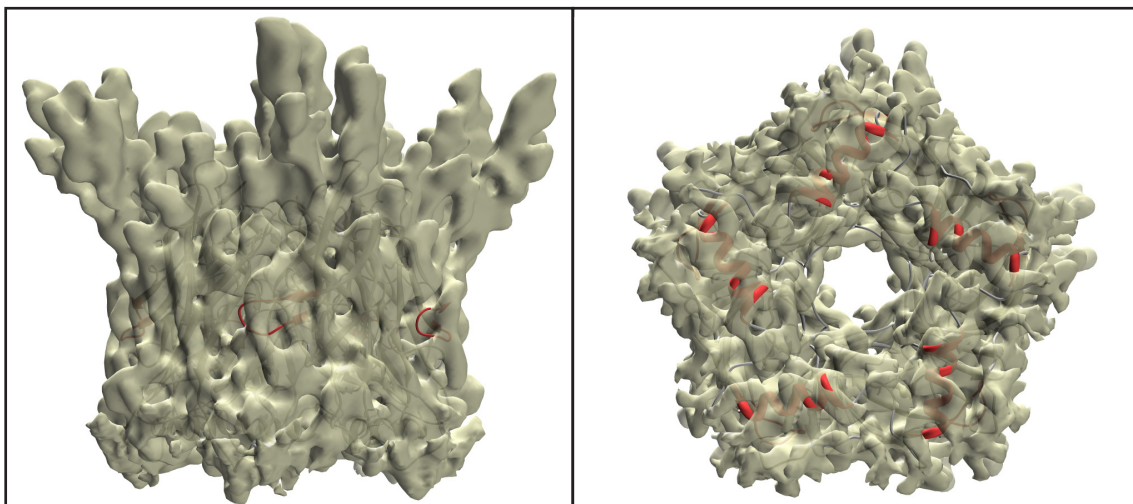
Figure 63 An excised pentamer of the BgAChBP reconstruction after traditional refinement, with the molecular model fit into it. Left: the gate to the ligand binding pocket constituted by loop C is clearly visible. Right: the density map is shown at lower than mass-correlated threshold to show the depiction of helix one by the surface representation. A barrel-like structure is visible, although for a good fit of the helix, flexible fitting would be necessary.

to the 0.5 threshold (Figure 64). Using the program bfactor, a negative temperature factor of 278.9 Å^2 was estimated and imposed on the final reconstruction, which was subsequently low-pass filtered to 6 Å. The resulting density map is depicted in Figure 65. The overall appearance of the structure is similar to the one obtained in the traditional refinement. It is conspicuous, however, that the surface of the density map seems much smoother than the previous one, i.e. the structure seems less noisy. It has to be noted that the aforementioned protrusions were still visible in the unfiltered and unsharpened reconstructions of each iteration, albeit much less pronounced and with a fuzzier appearance than the ones visible in the density map of the traditional refinement (Figure 66). Without the protrusions, the protein now has an outside diameter of ~220 Å while preserving the cavity diameter of ~100 Å described for the previous reconstruction. Each pentamer is ~60 Å in height and ~67 Å in width. Once more, a pseudo-atomic model was created by fitting 60 individual BgAChBP1 subunits into the structure. Due to the insubstantial difference to the previously shown model on a protein scale, it is not shown separately. A density map simulated on the basis of the pseudo-atomic model at 6 Å resolution, shows a cross-correlation of 0.89 with the experimental density map. The gate constituted by loop C and the α-helix (as barrel-like structure)
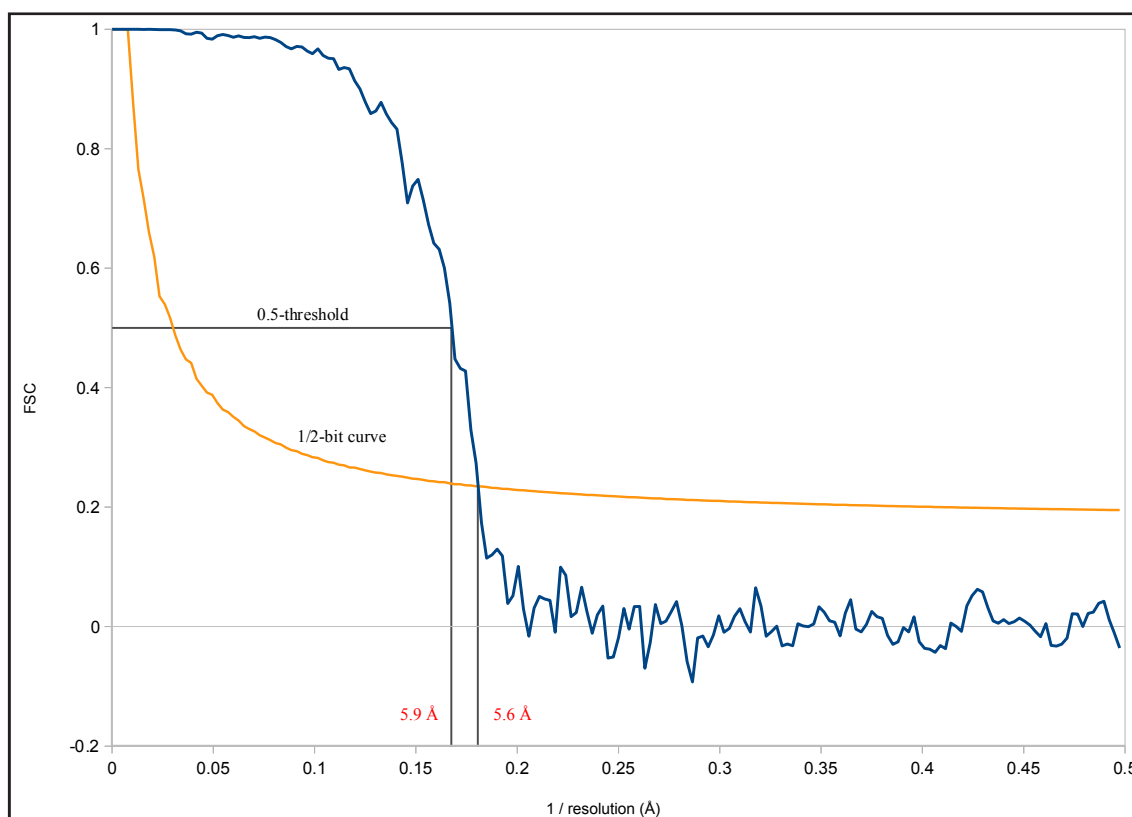
Figure 64 FSC of the final iteration of the refinement minimizing noise bias.
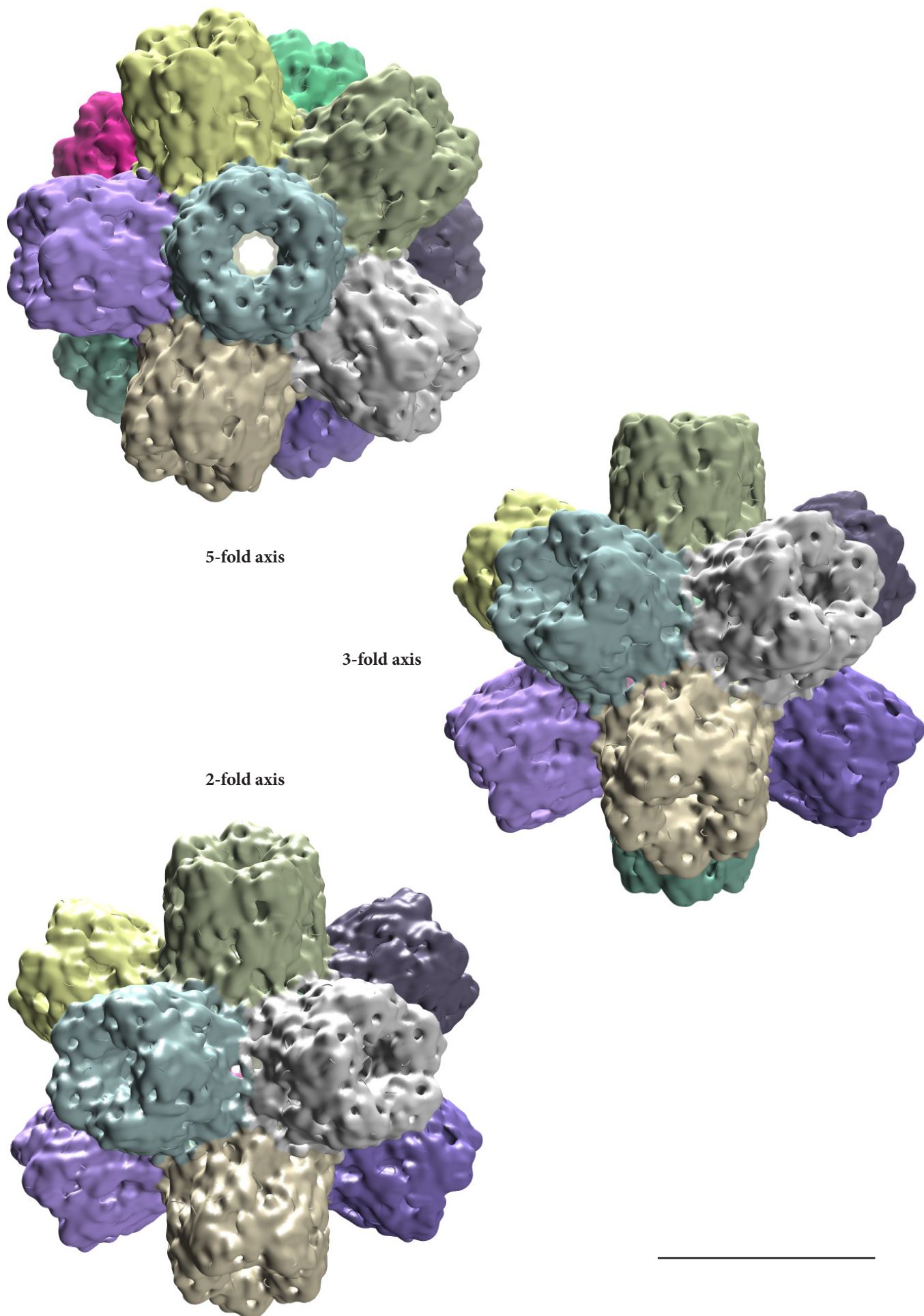
5-fold axis

3-fold axis

2-fold axis

Figure 65 Final 3-D reconstruction of the refinement minimizing noise bias. The density map is shown from the approximate views of one 5-fold, 3-fold and 2-fold axis, respectively. Each pentamer is highlighted in a different color. Scale bar: 100 Å.
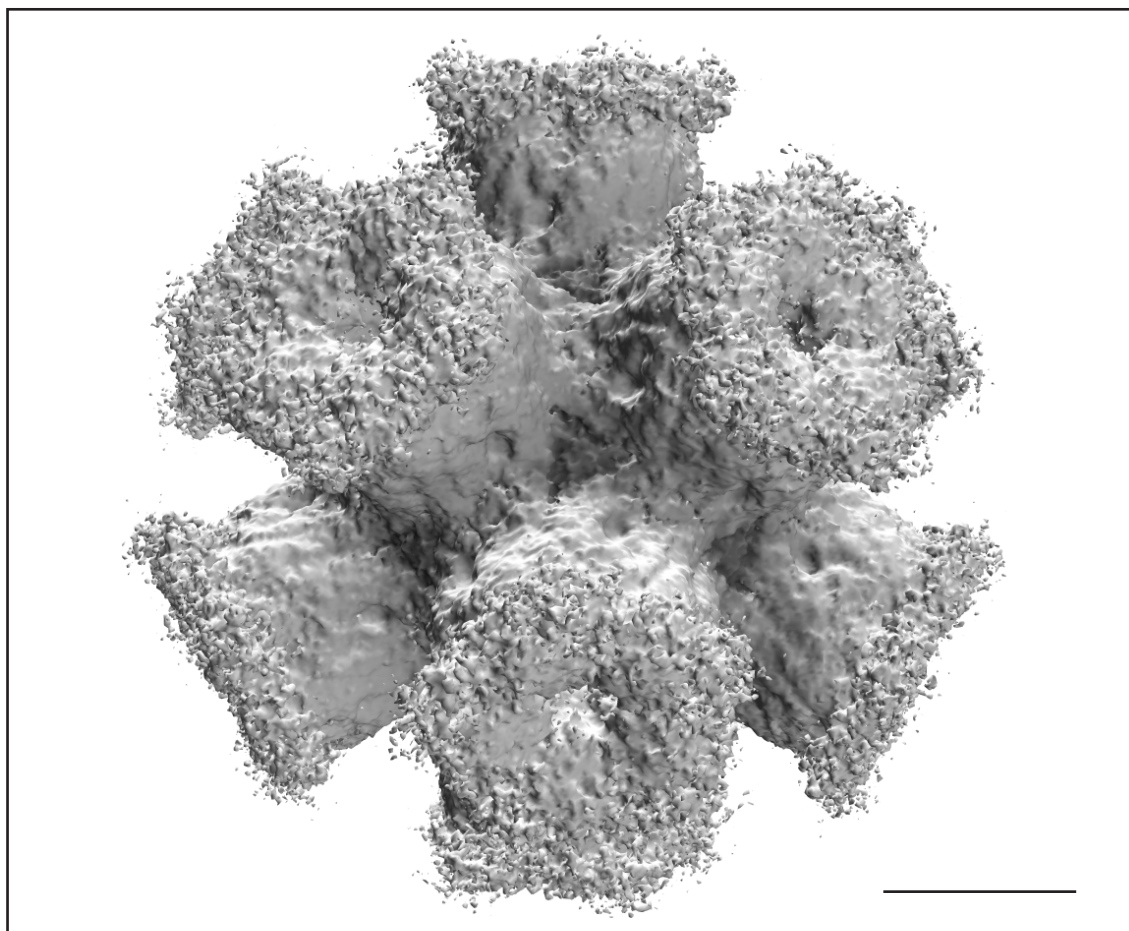
Figure 66 Final reconstruction of the refinement minimizing noise bias prior to low-pass filtering and sharpening. Note the fuzzy, noise-like structures surrounding the C-faces of the pentamers. The reconstruction is shown at a density threshold 1.5 times greater than the mass-correlated one. Scale bar: 50 Å

are clearly visible in the reconstruction, although the former, as already seen for the first refinement, sticks out a little bit from the density at a mass-correlated threshold (Figure 67). The α-helix of the molecular model would have to be flexibly fitted, in order to convincingly match the barrel-like structure of the density map.

The considerable differences seen between the results of the two refinement strategies will be discussed in chapter D4.1. The following description of the detailed structural topology will be based on the result of the second refinement, since it is more conservative and less prone to errors.

3.1.3 INTER-PENTAMER INTERFACES

The twelve pentamers of the dodecahedron are required to have contact zones designated to establish the quaternary assembly. Indeed, the 3-D reconstruction shows two discernible regions of bridging density, namely (i) at the 10 three-fold symmetrical axes between three
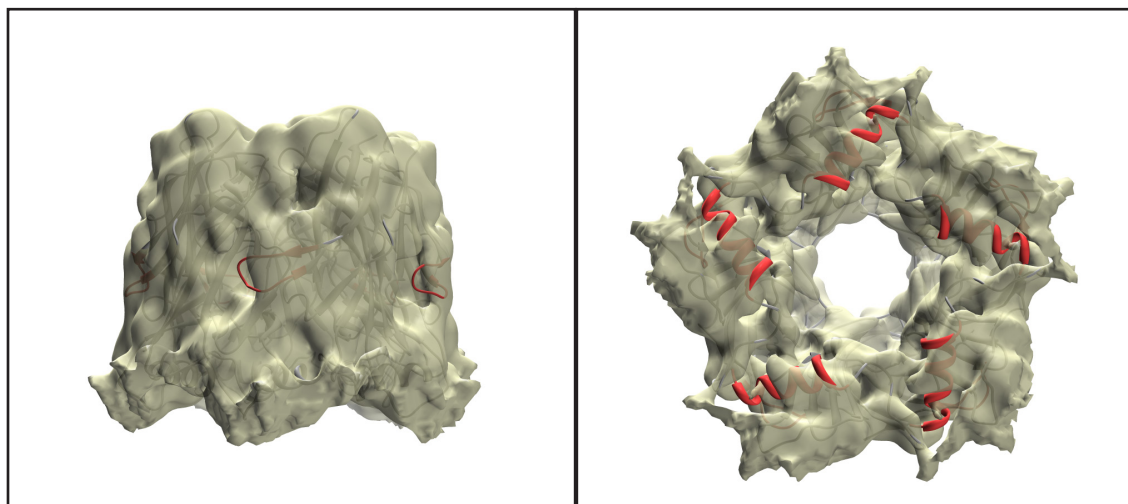
Figure 67 An excised pentamer of the BgAChBP reconstruction after the refinement minimizing noise bias, with the molecular model fit into it. Left: the gate to the ligand binding pocket constituted by loop C is clearly visible. Right: the density map is shown at lower than mass-correlated threshold to show the depiction of helix one by the surface representation. A barrel-like structure is visible, although for a good fit of the helix, flexible fitting would be necessary.

pentamers and (ii) at the 15 two-fold symmetrical axes between two pentamers. Evidence exists that dodecahedral BgAChBP is exclusively formed by BgAChBP1 (see C3.2 for details). Hence, the homo-oligomeric dodecahedral assembly of BgAChBP1 will be used for the description of the inter-pentamer contact zones. Since the transitions between the two interface types are fluent – due to their spatial proximity – a joint description of the interfaces is thought to be advantageous.

The contact zone is depicted in Figure 68. It is conspicuous that the phenylalanines F^{71} of three pentamers come together at the center of the interface, which is also represented by the density map showing a spherical mass at said position. The gathering of phenylalanines is enclosed by two rings of putative inter-pentamer contacts between charged amino acids, namely $R^3 \leftrightarrow E^{70}$ of helix α1 and the MIR loop, and $D^{25} \leftrightarrow R^{63}$ of loop L1 and the MIR loop, respectively. In the vicinity of these putative inter-pentamer contacts, an additional intra-pentamer contact might be established between R^2 (helix α1) and D^{26} (loop L1). Moreover, the close proximity of C^{16} (just behind helix α1) and C^{64} (MIR loop) indicates the formation of an intra-subunit disulfide bridge previously undescribed for an acetylcholine-binding protein. Due to reasons of clarity, it is not shown in Figure 68 that the three N^{69} and the three T^{72} residues are found above and below the plane of the phenylalanine cluster, respectively (see D4.2).

3.1.4 Intra-pentamer interfaces

A mixture of aliphatic, aromatic, charged and polar amino acids can be found at the contact
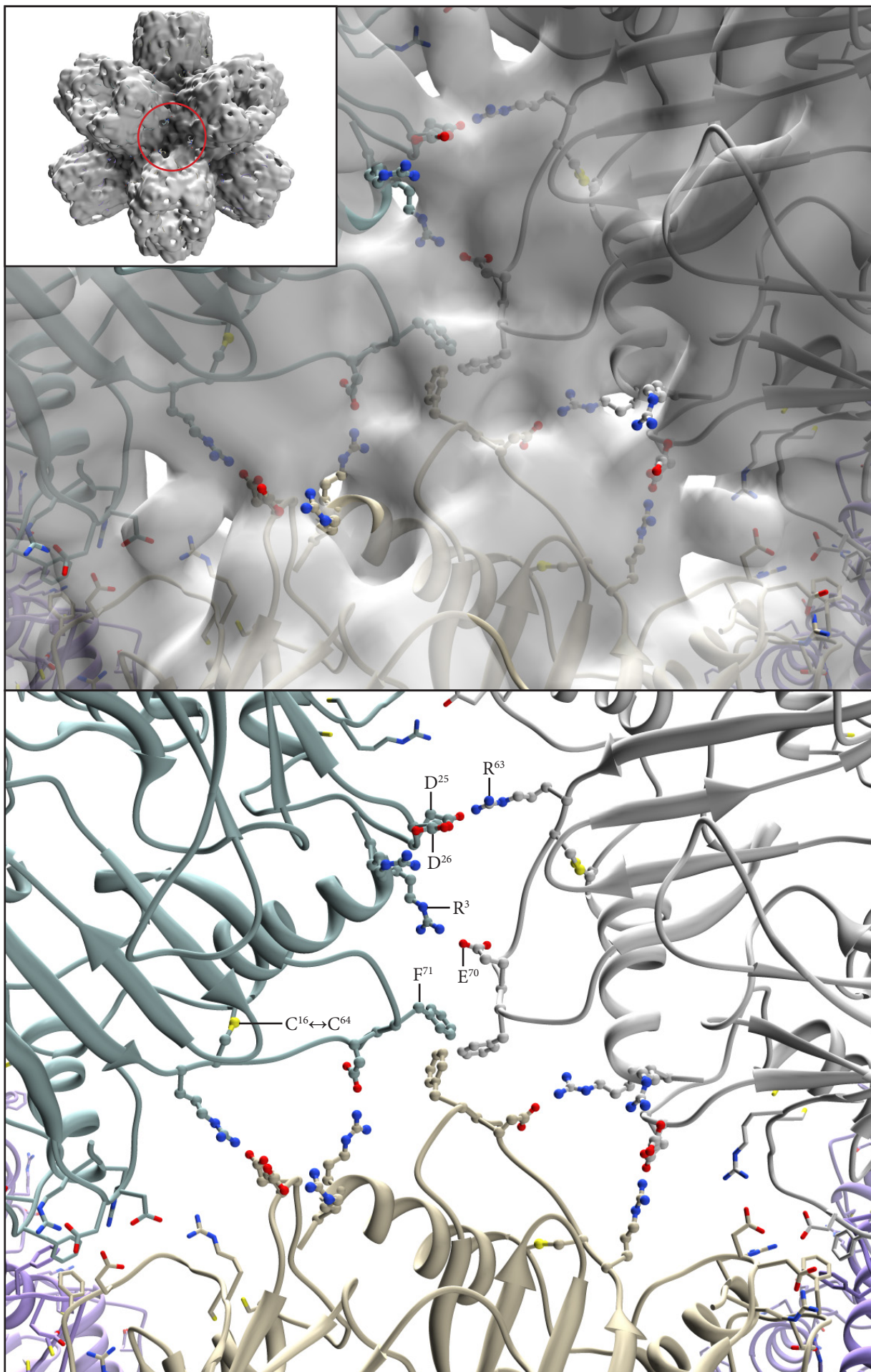
Figure 68 Inter-pentamer interface of a BgAChBP1 homo-oligomeric dodecahedron. Three phenylalanines (F^{71}) are enclosed by two rings of putative ionic contacts, namely $R^3 \leftrightarrow E^{70}$ and $D^{25} \leftrightarrow R^{63}$. Intriguing putative intra-pentamer contacts – $R^2 \leftrightarrow D^{26}$ and the novel disulfide bridge $C^{16} \leftrightarrow C^{64}$ – are also shown. Amino acids belonging to one interface are depicted by a ball-and-stick view. The general topological position of the interface inside the dodecahedron can be derived by the inset (red circle).

zone of two pentameric subunits. These amino acids and their general properties concerning contact formation are listed in Table 5. While hydrophilic and hydrophobic residues are almost equally distributed on the plus-side, the minus-side shows substantially more hydrophilic residues. A comparison of the amino acid distribution at this contact zone between BgAChBP1 and LsAChBP will be presented in the Discussion.

Table 5   Amino acids found at the intra-pentamer contact zone according to Brejc et al. (2001) and intriguing residues R^{42}, D^{48} and E^{46}. Coloring is according to the amino acid properties at physiological pH: positive (blue), negative (red), polar (O and N: purple; O: green), hydrophobic (black), Histidine (yellow).

Plus side	*Minus side*
S^{17}	S^{4}
L^{19}	R^{5}
N^{20}	L^{9}
I^{21}	Q^{10}
N^{47}	P^{80}
D^{88}	R^{171}
V^{95}	D^{103}
G^{96}	T^{105}
A^{97}	F^{41}
S^{125}	E^{99}
P^{127}	F^{101}
W^{146}	K^{121}
T^{147}	K^{39}
H^{148}	W^{55}
S^{149}	T^{78}
E^{152}	S^{102}
Y^{186}	R^{107}
P^{187}	D^{165}
R^{42}	Y^{166}
D^{48}	E^{46}

3.1.5 TOPOLOGY OF THE LIGAND-BINDING SITE

As seen in the reconstruction and in the pseudo-atomic model, AChBPs exhibit a ligand-binding pocket that is protected by a gate-like structure and which is common for the members of the pLGIC super-family. The amino acids contributing to the binding site were selected according to Brejc et al. (2001) and are shown in Figure 69. It is noticeable that the principal side mostly contributes aromatic residues, whereas the complementary side mostly contributes polar and charged residues.
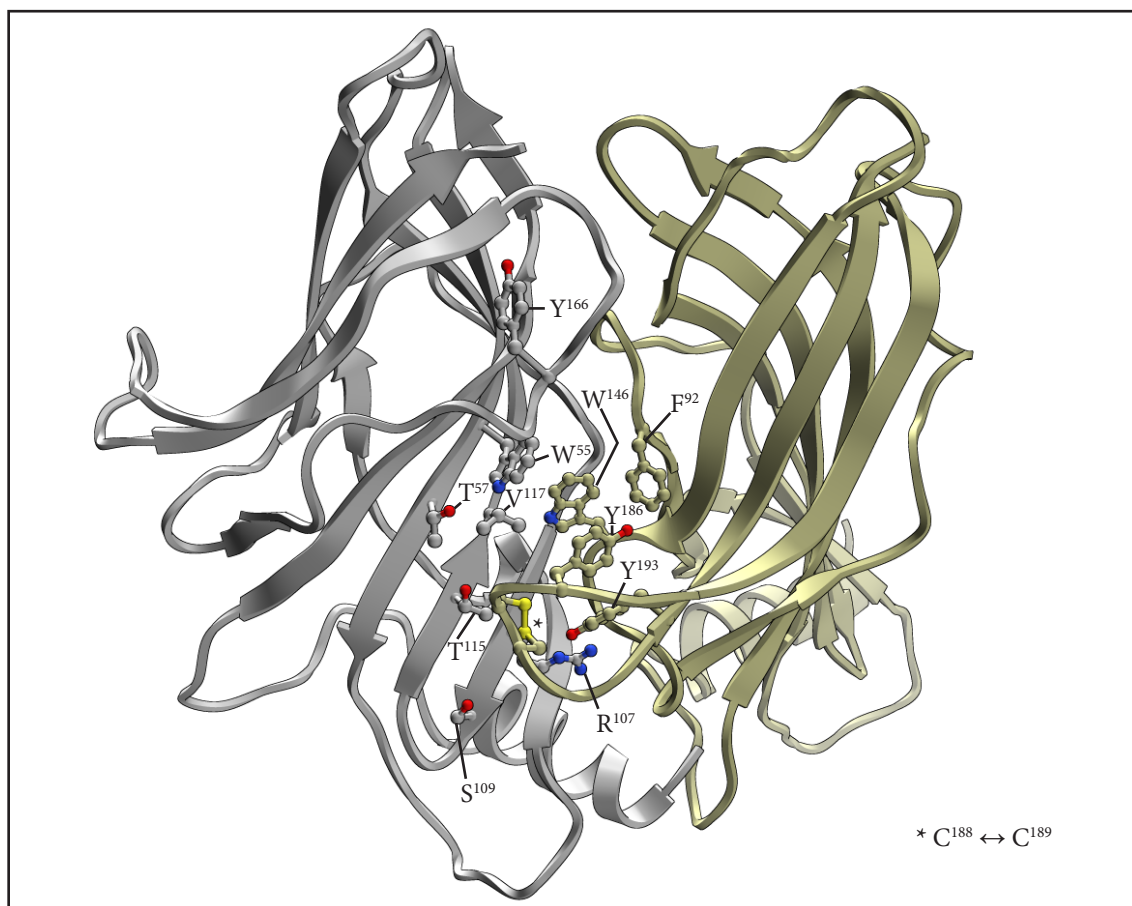
Figure 69 Amino acids of the ligand-binding site. Two monomers of BgAChBP1 are illustrated as a side-view tilted slightly to the top. The principal monomer is colored golden, the complementary monomer gray. Note the high amount of aromatic amino acids contributed by the principal monomer. The selection of amino acids followed BREJC ET AL. (2001).

3.2 RECOMBINANT BgAChBP 1

As mentioned in section B6.2, recombinantely expressed BgAChBP1 was provided by ██████████ and digitization, preprocessing and refinement was conducted under my supervision by ██████████████ as part of her Diploma thesis (BRAUKMANN 2012).

Negative staining

C-terminally His-tagged variants of the protein were chosen, since the N-terminus was likely to be of importance for inter-pentamer contact formation. Initial aggregation of the particles could be resolved by adding 50mM imidazole to the buffer (data not shown). A typical negatively stained image of recombinantely expressed BgAChBP1 after refolding can be seen in Figure 70. The characteristic projections of the rosette-like protein are immediately distinguishable, thus indicating that a successful assembly of twelve pentamers has taken place. It is also apparent, however, that the sample did not exclusively assemble into dodecahedra,

since various pentamer-like structures are also visible. In order to validate the structure of the seemingly dodecahedron-shaped particles, a multirefine procedure was calculated in ten iterations, using three Gaussian spheres with different Gaussian noise added to them as references. Particles assigned to the dodecahedron-shaped result of the multirefine procedure (Figure 71) were extracted from the dataset and initially refined using C1-symmetry in EMAN1.9 (see B6.2.3). The resulting 3-D reconstruction indicates icosahedral symmetry and is very similar to the low-resolution reconstructions obtained from native BgAChBP. The density map derived from this initial refinement was used as a reference for refinements with icosahedral symmetry in both EMAN1.9 and IMAGIC-5. Both resulting structures are of similar dimensions as native BgAChBP, but show different mass distributions at the inter-pentamer contact regions (Figure 72). The exact dimensions of the reconstructions were not measured due to the poor comparability between the cryo-EM dataset of native BgAChBP and this negatively stained dataset. A pseudo-atomic model, created using PDB entry 4aod, partly protrudes from both density maps at the N-face and the pentamer pores (Figure 72).
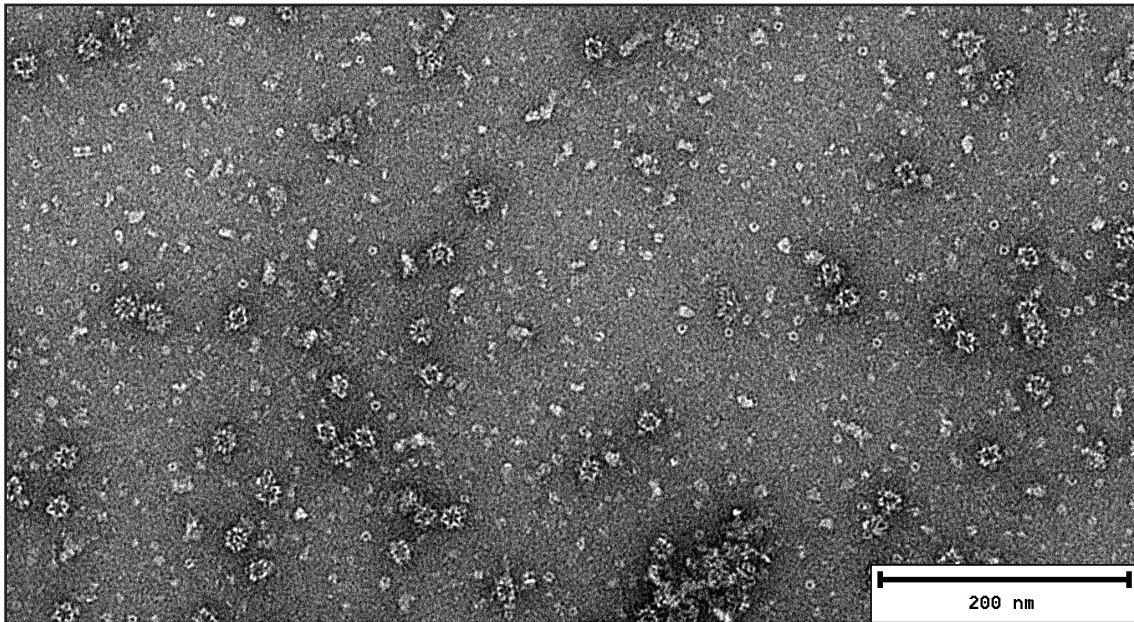
Figure 70 Electron-microscopic image of recombinant BgAChBP1. Besides pentamers, dodecahedral assemblies are visible.
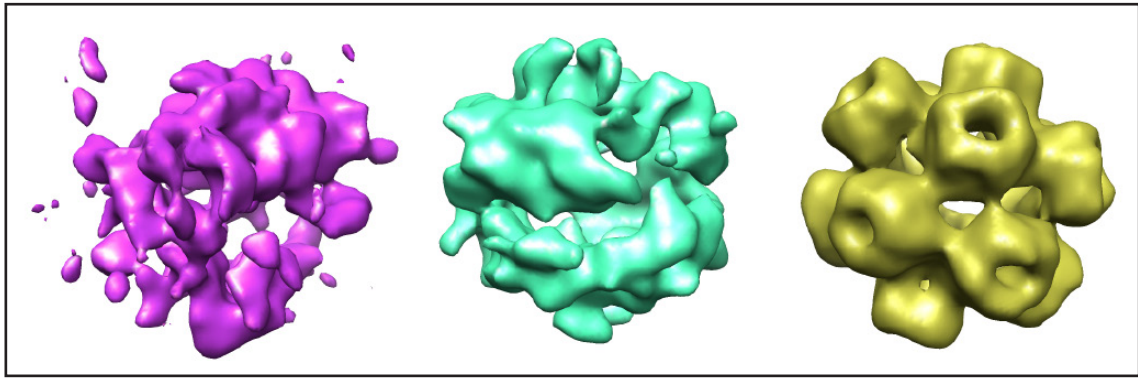
Figure 71 Density maps resulting from the multirefine procedure with negatively stained recombinant BgAChBP1. One structure clearly shows the general shape of a pentagonal dodecahedron (right). Only few particles were assigned to the other two reconstructions (reprinted from Braukmann 2012).
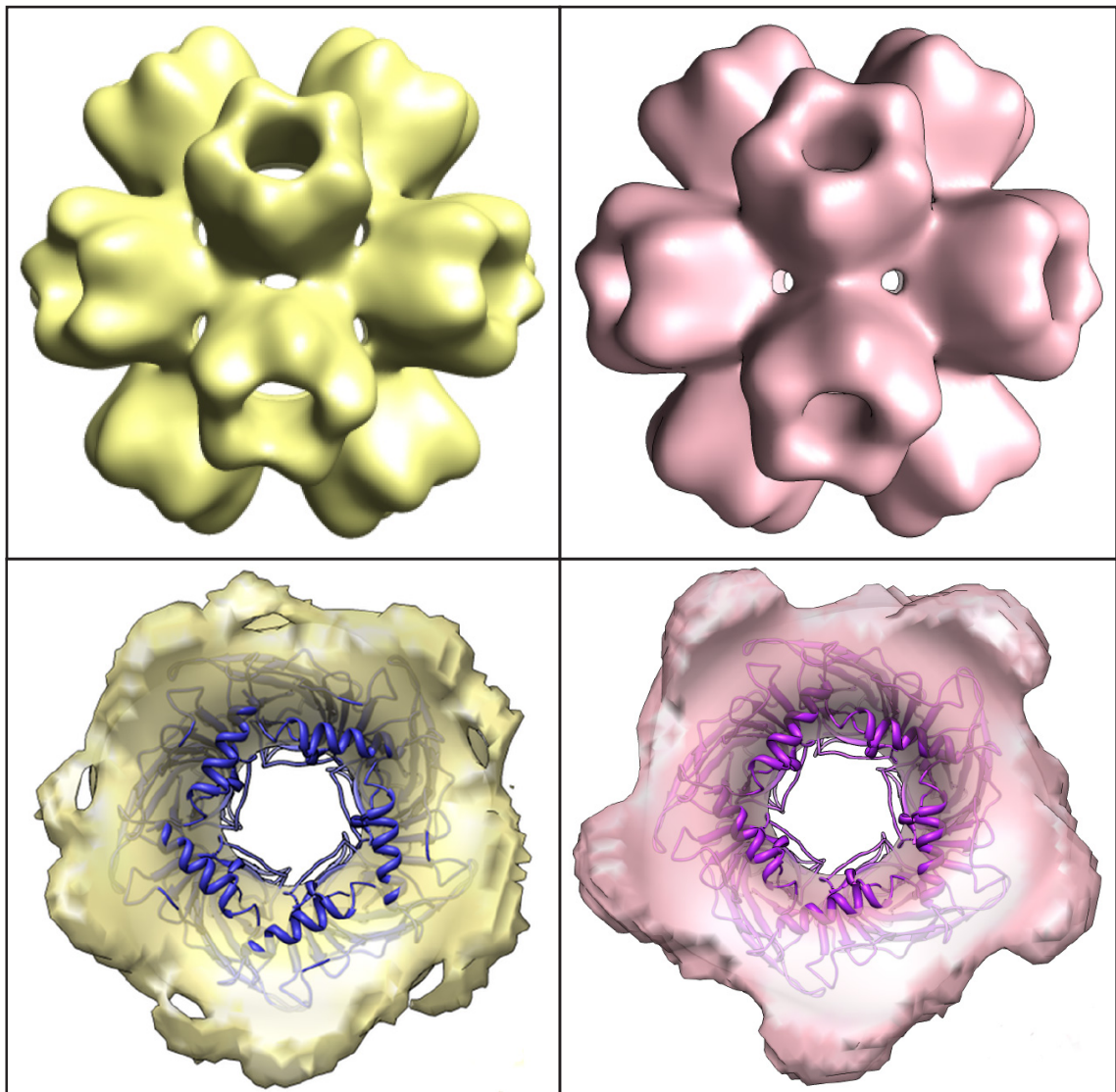
Figure 72 Results of a refinement of a negatively stained recombinant BgAChBP1 sample in EMAN1.9 (left) and IMAGIC-5 (right). Both reconstructions show a similar general appearance, but differences in bridging density between pentamers (top row). The homology model of BgAChBP1 docked into the structures by rigid-body fitting protrudes from the density at the N-face and inside the pentamer pore (bottom row; pentamers shown from the N-face; reprinted from Braukmann 2012).

Cryo-EM

To gain a better understanding of the recombinantely expressed BgAChBP1 structure, cryo-EM of the C-terminally His-tagged protein was pursued. A typical section of an electron micrograph obtained from the sample is shown in Figure 73. Hardly any particles can be found freely in the ice layer, whereas large protein aggregations are visible at the edges of the carbon film, indicating that the added imidazole did not result in the benefits experienced for the negatively stained datasets. In an attempt to reconstruct the non-ideal experimental data, a dataset comprising 2,952 particles was generated. Due to the concentration of the particles at the carbon edges – close to the micrograph edges which are excluded by CTFFIND3 – CTF estimation was very difficult and most micrographs did not show more than two or three Thon rings. After CTF correction, 2,200 particles remained for the calculation of a 3-D reconstruction. Refined class averages were obtained by MRA-MSA cycles and show no protrusions comparable to the ones observed for native BgAChBP (Figure 74). Refinements in EMAN1.9 and IMAGIC-5 were calculated as described in B6.2.4, using the 3-D reconstruction obtained from the negatively stained data as a reference. After stabilization of the FSC, the final reconstructions both show a resolution of 8 Å according to the 0.5 threshold.

The density map calculated with EMAN1.9 has a very spiky appearance and, as opposed to the density maps of native BgAChBP, does not unequivocally show the gate of the ligand-binding pocket (loop C; Figure 75). Some inter-pentamer bridging densities are visible at the 3-fold symmetry axes, whereas no mass can be seen at the 2-fold symmetry axes. The pentameric homology model of BgAChBP1 (PDB entry: 4aod) could be placed into the density map by rigid body fitting. Individual subunits, however, could not be docked into the structure. Helix α1 is not encased by density and lies entirely outside of the 3D-reconstruction.

The refinement of the BgAChBP1 dataset using IMAGIC-5 first converged at a resolution of 11Å (0.5 threshold). In order to increase the accuracy of the MRA – and thus the resolution of the 3-D reconstructions – two-dimensional masking was performed after every iteration by first calculating an initial MRA, then creating matching 2-D masks for every projection using the last density map and finally calculating a second MRA with the masked dataset. By following this procedure, the resolution could be improved to 8 Å (0.5 threshold). The final density map has similar dimensions as the density maps of native BgAChBP after refinement minimizing noise bias and BgAChBP1 after refinement in EMAN1.9. An intriguing aspect about the structure is the pronounced bridging density that can be seen at the 2-fold symmetrical axes (Figure 75). For helix α1, no corresponding density can be found.

Due to the inefficacy of the added imidazole for cryo-samples (see above), the His-tag
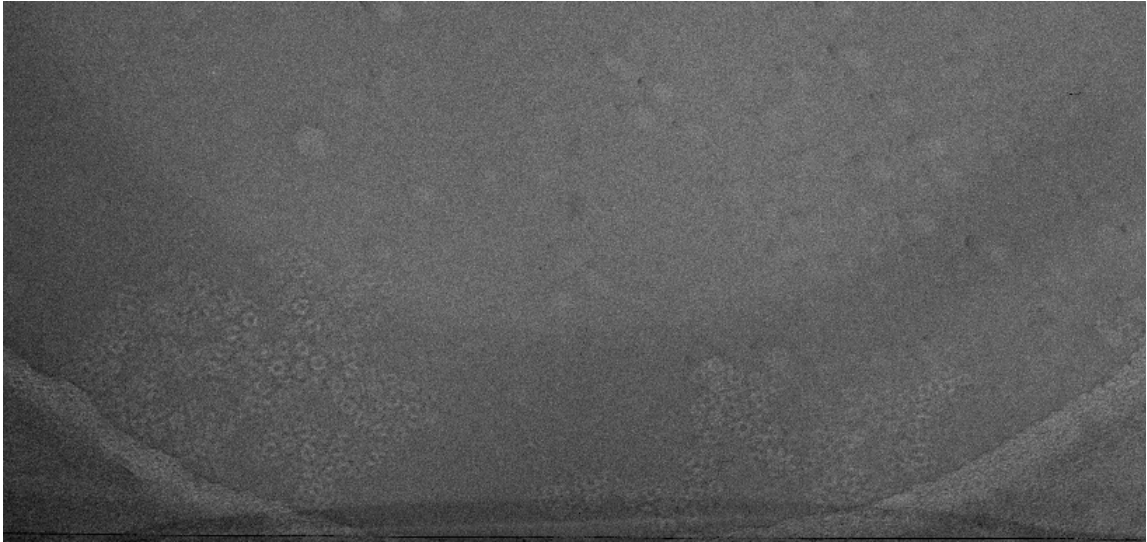
Figure 73 Cryo-EM image of recombinant BgAChBP1. Hardly any particles were found in the center of the ice layer; they mostly aggregate close to the carbon edges.
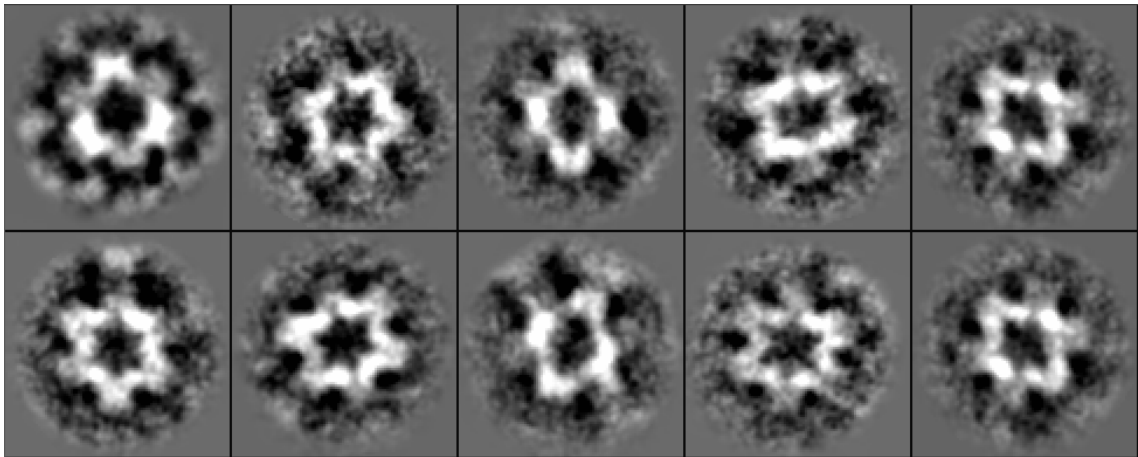
Figure 74 Class averages obtained from the cryo-EM dataset of recombinant BgAChBP1. Note that the protrusions described in Figure 57 are not visible.
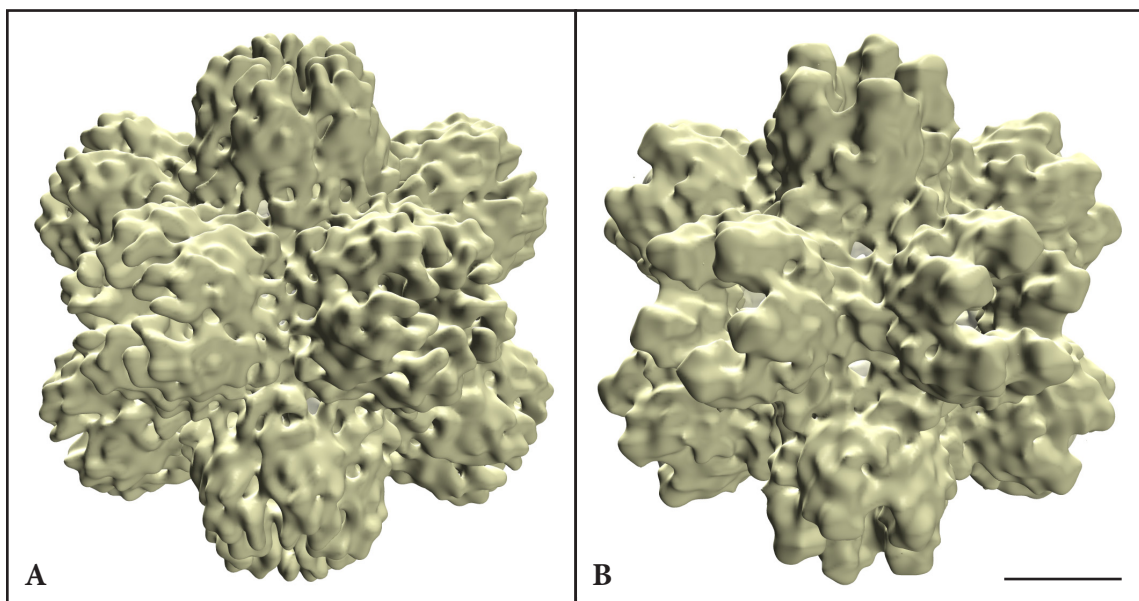
Figure 75 Density maps resulting from the refinement of the cryo-EM dataset of recombinant BgAChBP1. Both reconstructions are depicted along one of the two-fold symmetrical axes at mass-correlated threshold. A) EMAN1.9 refinement. The reconstruction has a spiky appearance and loop C cannot unequivocally be identified. Bridging densities between pentamers are visible at the two- and the three-fold symmetrical axes. B) IMAGIC-5 refinement. Loop C and the general outlines of the BgAChBP1 monomers are clearly distinguishable. Bridging masses can only be seen at the two-fold symmetrical axes. Scale bar: 50Å.

was attached at the N-terminus, where a TEV protease restriction site was added between the protein and the tag which enabled the cleavage of the His-tag after protein purification. The only additional N-terminal amino acid resulting from this procedure is a glycine residue which should not interfere with the dodecahedral assembly. The procedure was performed by ███████████ and is described in detail in MÖLLER (2012). Figure 76 shows the successful assembly of BgAChBP1 after removing the His-tag. Although the majority of the protein is still in a pentameric state, the occurrence of typical dodecahedral views is a promising result for future cryo-EM studies of BgAChBP1.
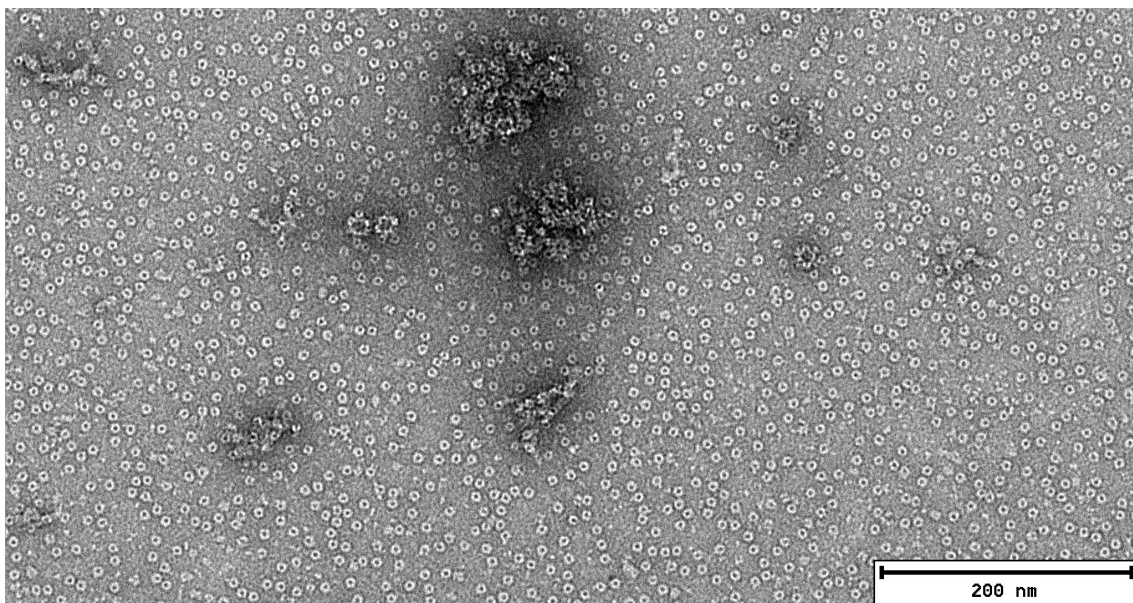
Figure 76 Recombinant BgAChBP1 after removing the N-terminal His-tag at the TEV protease restriction site. Most of the protein exhibits a pentameric assembly, but some fully assembled dodecahedra are also visible.

3.3 RECOMBINANT BgAChBP 2

Recombinantely expressed BgAChBP2 was provided by ███████████. In contrast to BgAChBP1, recombinantely expressed BgAChBP2 did not readily form dodecahedra (Figure 77). However, apart from the pentamer-like top-views that were present in the samples of native BgAChBP and BgAChBP1, elongated structures are discernible whose dimensions fit to side-views of a hypothetical dipentameric assembly (Figure 78). In order to elucidate the nature of this structure, negatively stained particles were imaged and a dataset comprising 2,109 of these elongated molecules was created. A refinement in EMAN1.9 assuming D5 point-group symmetry using a cylinder as a reference and only re-projecting references between alt-angles of 85-90°, yielded an elongated, pentagonal structure (Figure 79, top). Two pentameric assemblies of the homology model of BgAChBP2 (PDB entry: 4aoe) were placed into the density map *via* rigid-body fitting (Figure 79, center). Although in width the pentamers fit adequately, the density map seems to be too short, leading to an overlap of the two docked pentamers. Adapting the pixel size of the density map to 5 Å/pixel and a subsequent
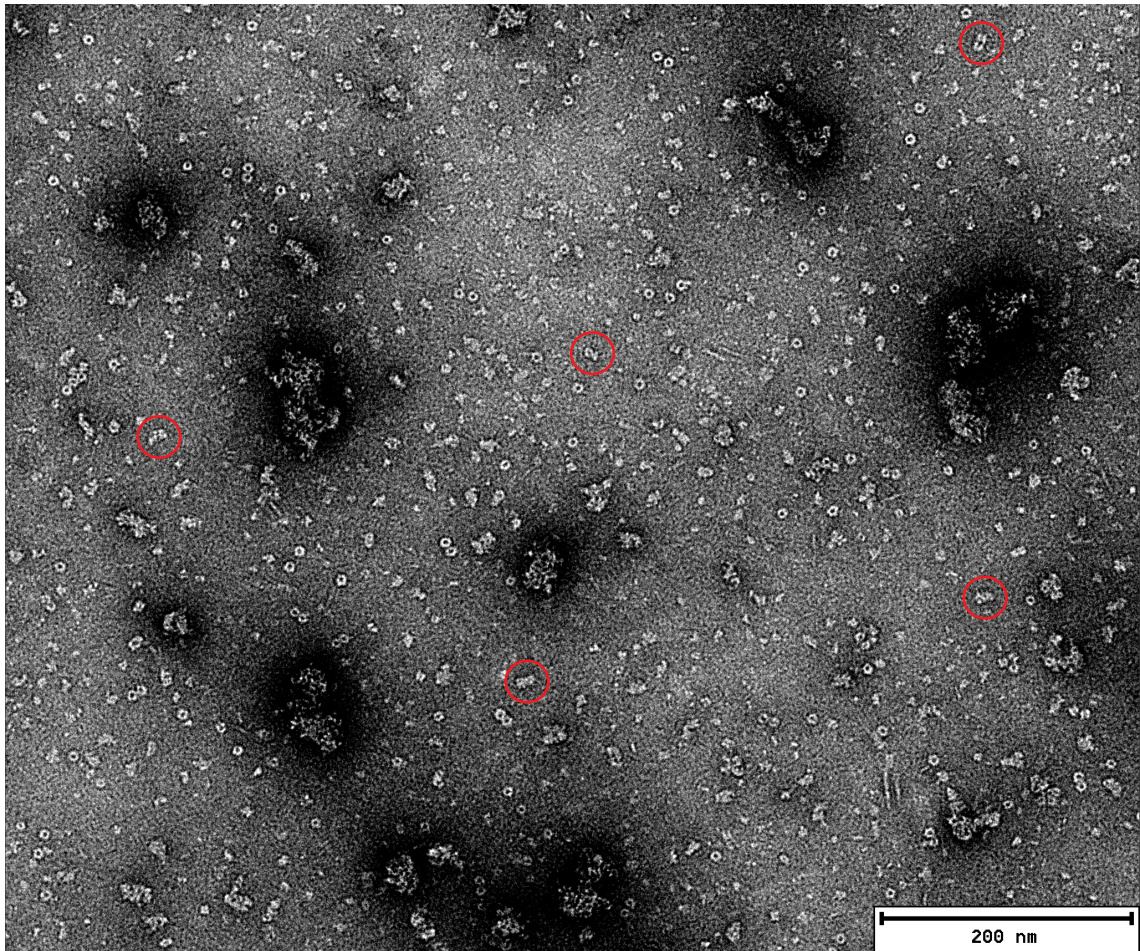
Figure 77 Recombinant BgAChBP2 did not form dodecahedral assemblies. However, elongated structures interpreted as dipentameric assemblies are visible (e.g. structures inside the red circles)
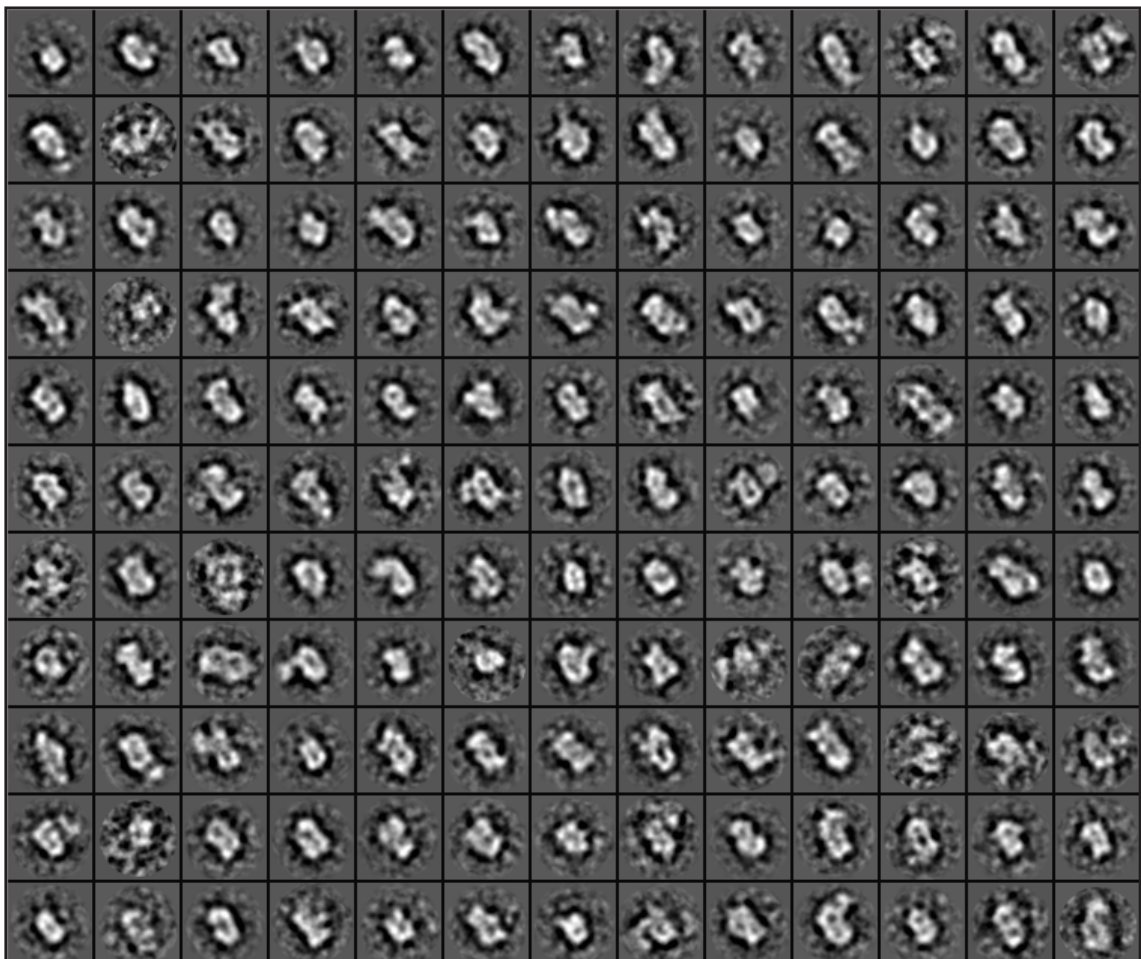
Figure 78 Refined class averages after four MRA-MSA cycles. Many elongated structures can be seen, which might be side-views of a dipentameric BgAChBP2.

refitting resulted in a good description of the models by the density map (Figure 79, bottom). It has to be noted that the change in pixel size alters the longitudinal shift of the models, but does not affect the rotation of the models along the longitudinal axis (5-fold symmetrical axis). Thus, apart from the increased distance between the models, the relative orientation between pentamers does not change from one pixel size to the other. A simulated density
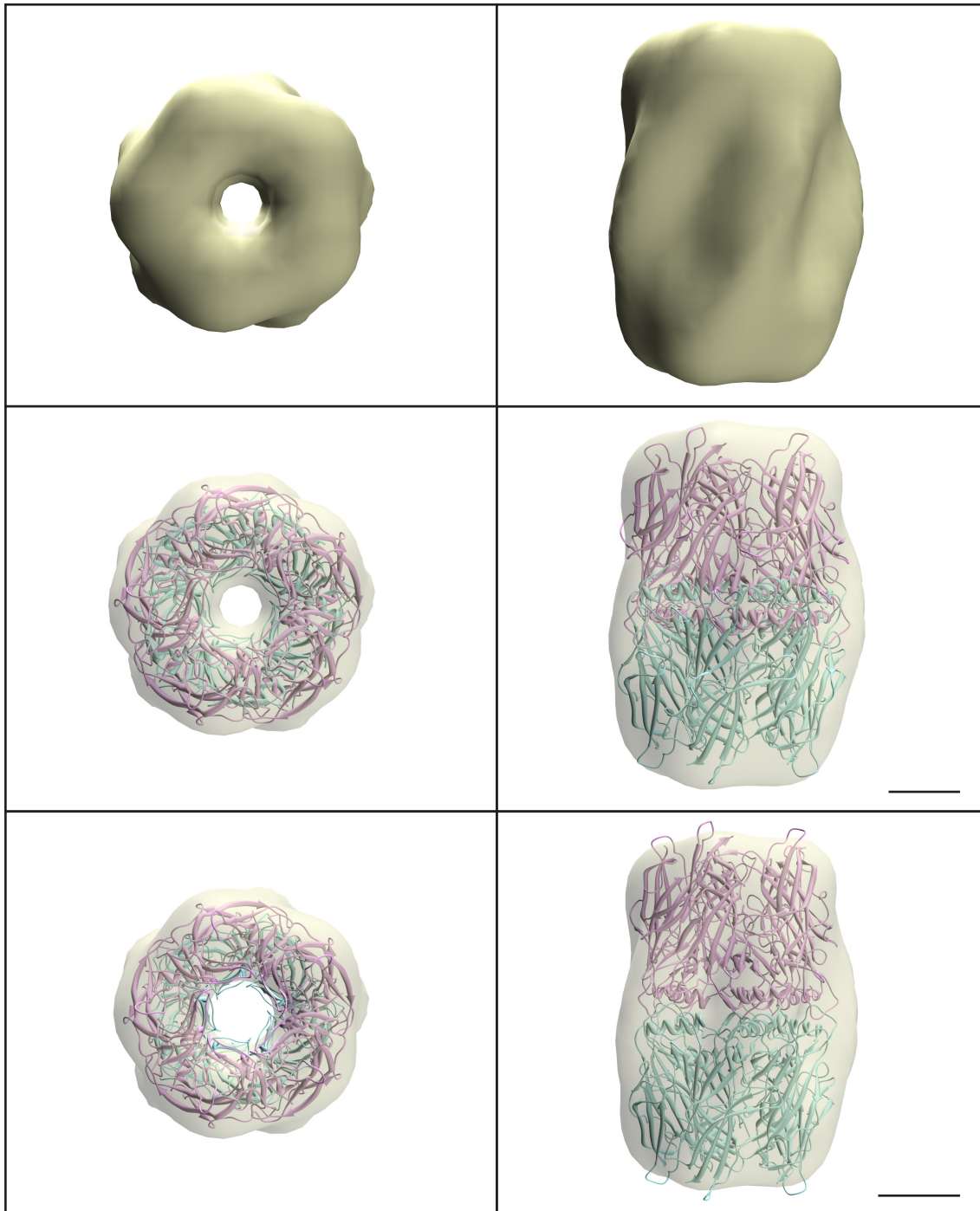
Figure 79 3-D reconstruction and pseudo-atomic model of BgAChBP2 at mass-correlated threshold. An elongated, dipentamer-shaped density map was obtained after refinement in EMAN1.9 (top row). Two copies of PDB entry 4aoe were placed into the structure by rigid-body fitting in UCSF Chimera and overlap at the inter-pentamer contact zone when choosing a pixel size of 4.36 Å/pixel (center). Displaying the density map at 5 Å/pixel and re-fitting the pseudo-atomic model, separates the pentamers without introducing any rotational difference. This results in a convincing molecular interface at the inter-pentamer contact (bottom; Figure 80). Scale bars: 25 Å

map of this dipentameric pseudo-atomic model low-pass filtered to 25 Å reaches a correlation of 0.94 with the experimental density map at 5 Å/pixel. Interestingly, the unpaired cysteine residues C^{71} of each pentamer result lying within a short distance of each other. Moreover, this putative inter-pentamer contact site shows a multitude of charged residues contributed by helix α1 and the MIR loop (Figure 80).
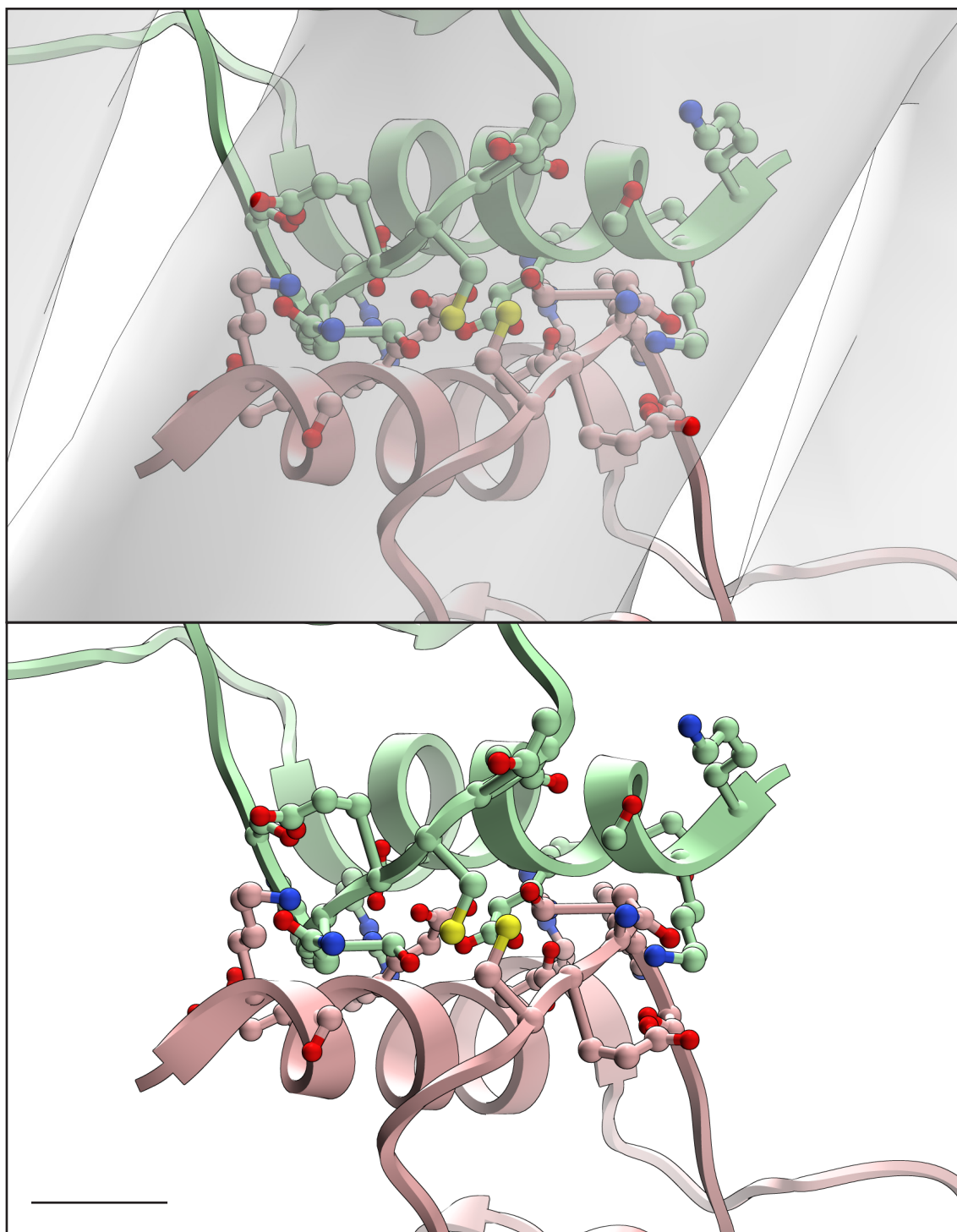
Figure 80 Inter-pentamer interface between two monomers of a dipentameric BgAChBP2. The contact site is constituted by helix α1 and the MIR loop and is also described by the density map at slightly below mass-correlated threshold (top). The five unpaired C^{71} of each pentamer are in close contact with those of the adjacent pentamer. Note that the contact site is rich in charged amino acids. Color-coding according to heteroatoms: C=color of model; N=blue; O=red. Scale bar: 5 Å

CHAPTER D: DISCUSSION

1 METHODOLOGY

In consideration of the presented results, it seems appropriate to more thoroughly analyze the algorithms used to process the data and their respective hazards towards error accumulation during the iterative cycles associated with 3-D reconstruction. Therefore, the general pitfalls of alignment algorithms will be presented with an analysis of their possible implications for the over-refinement of 3-D structures. Subsequently, the solutions of different software suites for circumventing these pitfalls will be discussed, combined with a critical description of the 3-D refinement strategies that were presented in the results for cryo-EM samples.

1.1 ALIGNMENT ERRORS AND OVER-REFINEMENT OF 3-D STRUCTURES

The necessity to align particle images for subsequent class averaging and 3-D reconstruction requires calculations that, according to JOYEUX AND PENCZEK (2002), resemble a special case of template matching algorithms. These alignment algorithms – particularly in the case of EM image analysis of biological macromolecules – are associated with problems based on the corruption of signal by noise components that disturb accurate alignment, and on interpolation errors that result from shifting particles by non-integer amounts or rotation (*ibid.*).

In general, all alignment algorithms in EM image processing try to minimize the squared difference between a pair of images (*ibid.*), a strategy that can be described by

$$\sum_{k \in [0, 0.5]} |F_X(k) - F_Y(k)|^2 \, ,$$

where k is the resolution in reciprocal space, and F_X and F_Y are the Fourier transforms of

images X and Y, respectively (Stewart and Grigorieff 2004). For correlation-based align-
ments, this minimization is equivalent to the maximization of the cross-correlation between
the images (*ibid.*). The alignment method described in section B3.7 uses linear cross-cor-
relation functions which, in real space, are described by

$$CCF(r_k) = f(r_i) * g(r_i) = \sum_{i \in [1,N]} f(r_i)\, g(r_i + r_k)$$

where r_i is each pixel of images f or g and r_k each pixel of the CCF. The Fourier space equiva-
lent of the convolution described above, is a multiplication given by

$$CCF(r_k) = F(r_i)\, G^*(r_i)\,,$$

with $G^*(r_i)$ being the complex conjugate Fourier transform of $g(r_i)$ (e.g. Frank 1975; saxton
and Frank 1977; Steinkilberg and Schramm 1980). Linear cross-correlation approaches,
however, have been widely discussed in the literature to be a major source for alignment er-
rors, since a linear weighting in favor of large amplitudes is the result of the procedure (e.g.
van Heel et al. 1992; Stewart and Grigorieff 2004; Scheres 2012a and b) . While this
fact would be desirable in noiseless images, large amplitudes in EM images for single particle
analysis do not necessarily imply a good signal at the respective position, but can occur due
to the contribution of noise. The weighting in favor of such noisy components can result in
the false alignment of the particle to one of the many local minima found in noisy images
(e.g. Scheres et al. 2005), which further leads to a summation of the errors introduced
by falsely aligned particles during the iterative 3-D reconstruction process (e.g. Joyeux and
Penczek 2002; Scheres 2012a; Stewart and Grigorieff 2004). In order to call attention
to the possible extent of noise bias, Stewart and Grigorieff (2004) impressively describe
the resemblance of a pure-noise dataset to their initial 3-D reference after refinement (Figure
81). This outcome can be explained by the fact that the generally used models that assume
independent, Gaussian distributed noise are not entirely correct. In EM images particular-
ly, noise correlates with a Gaussian distribution whose standard deviation is 1/[image edge
length in pixels] (*ibid.*).

   In an iterative refinement, the correlation of noise, the linear weighting of amplitudes and
the resulting misalignment of particles towards local minima can lead to so-called over-re-
fined 3-D structures that are mostly described by the noise of the images rather than their sig-
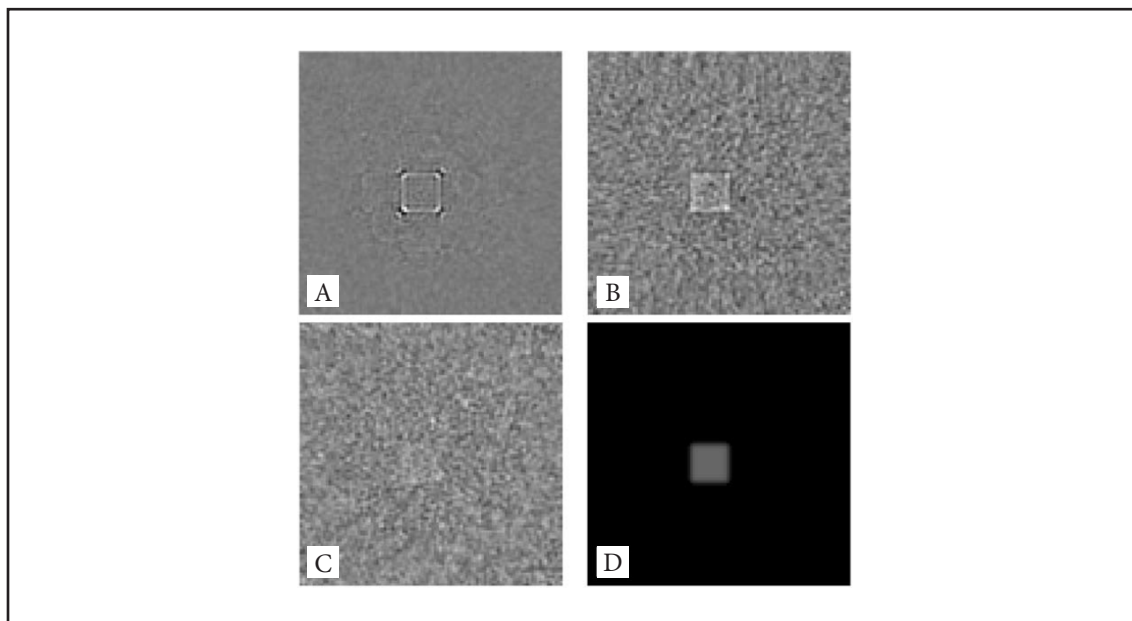
Figure 81 Example of noise alignment. Pure-noise images were aligned against a reference (D), either using phase residual (A), linear cross-correlation (B) or weighted cross-correlation (C). Note that the results for A and B clearly resemble the reference structure, although no structural information was present in the images (reprinted from Stewart and Grigorieff 2004 with permission from Elsevier).

nal (Figure 82; Scheres 2012a). Logically, falsely high cross-correlation values due to noise correlation and the consequential 3-D refinement against over-refined structures give rise to overly high resolution estimates when calculating an FSC (Stewart and Grigorieff 2004; Zhou 2008). For a traditional refinement strategy, a vicious circle begins due to the adaptation of high-resolution cut-off values according to the resolution measured by the FSC. The allowance of more high-frequency, noise-dominated components to the alignment, however, increases the noise bias in the reconstruction process and further increases resolution estimates of the FSC. Again, an adaptation of the high-resolution cut-off towards the allowance of higher spatial frequencies will be the result, and so forth.

If noise-bias can be mostly ruled out – may it be due to low-resolution cryo-EM refinements or 3-D structure determination of negatively stained particles – standard linear cross-correlation-based alignments work well. Although Joyeux and Penczek (2002) showed that the alignments using ACF (Frank 1975; van Heel et al. 1992), polar coordinates (Penczek et al. 1994) or singular value decomposition of sinograms (Bellon et al. 2001) are slightly more accurate than the direct alignment using FFT that is described above, to my knowledge, all of the programs mentioned in the subsequent sections use the latter as a similarity measure between projection and reference.

Even though linear correlation is not a problem when dealing with the above mentioned
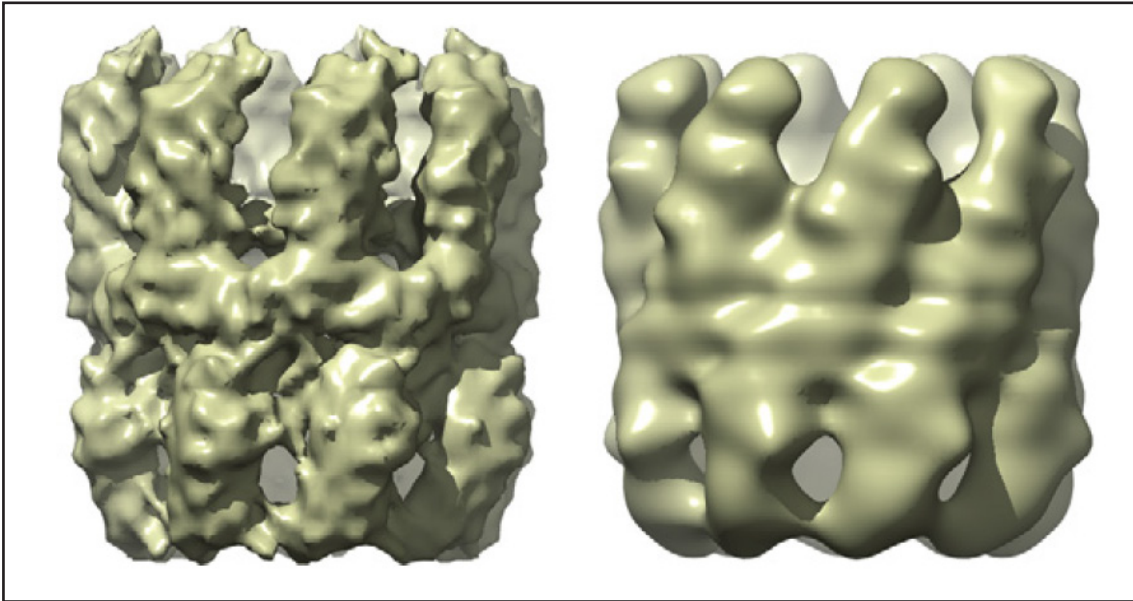
Figure 82 The effects of over-refinement. A structure with a measured resolution of 10 Å is shown on the left. Over-refinement can be already deduced by the very spiky appearance of the structure. In contrast, the density map on the right was calculated avoiding noise bias and showed a resolution of 16 Å. Comparing the reconstructions to the pseudo-atomic models of the protein resulted in a measured resolution of 30 Å for the left and 16 Å for the right structure. Hence, for high spatial frequencies, the left structure is mainly described by noise. (reprinted from Scheres 2012a with permission from Elsevier).

refinement types, medium- to high-resolution cryo-EM structure refinements should consider the pitfalls that calculations using such algorithms could have towards the structure determination. Hence, most of the cryo-EM structures presented in this work have been calculated using special refinement algorithms that are implemented in different software packages. The solutions to over-refinement pursued by each of these packages will be described in the following sections, concluding with a short description of promising future strategies concerning the alignment of particle images.

1.2 Refinement algorithms

1.2.1 EMAN1.9

Generally, the refinement algorithm of EMAN1.9 tries to minimize noise correlation by implementing a noise model during Wiener filtering. Downing and Glaeser (2008) describe the effects of a Wiener filter for CTF-correction and noise-suppression with the filter given by

$$w(s) = \frac{CTF(s)}{(CTF(s))^2 + \dfrac{1}{SNR}} \, ,$$
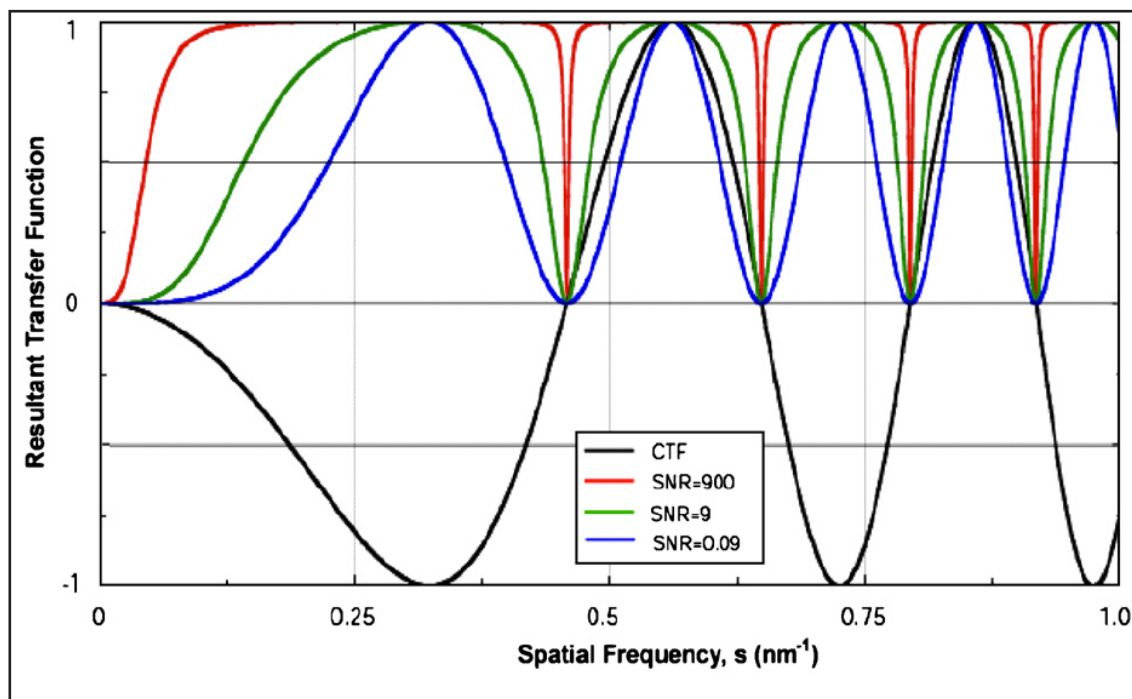
Figure 83 Effects of the Wiener filter on the contrast transfer function. For high SNR, amplitudes are almost completely restored. For very low SNR (blue curve; e.g. cryo-electron microscopy), small amplitudes are suppressed to prevent the over-amplification of noise. The blue curve is comparable to the result when multiplying by the CTF for CTF-correction (reprinted from Downing and Glaeser 2008 with permission from Elsevier).

where w(s) is the Wiener filter value, SNR the signal to noise ratio and CTF(s) the CTF value for each spatial frequency s (Figure 83). It is apparent that the filter suppresses small amplitudes, if a low SNR is present in the image, thus preventing over-amplification of noise. In comparison, just flipping the negative parts of the black curve – as is done when correcting for the CTF *via* phase flipping – would not result in any weighting. Multiplying by the CTF during CTF-correction, on the other hand, yields a result similar to the blue curve (*ibid.*). Consequently, the latter correction method should be preferred when no Wiener filter algorithm is to be used for CTF-correction.

Apart from the Wiener filter, EMAN contains a 'classiter' algorithm to prevent refinements to be trapped in local minima, which iteratively aligns members of a class against their class average (Ludtke et al. 1999). Furthermore, the user can activate the 'usefilt' option, which uses differently filtered datasets for alignment and reconstruction. The benefits of this option will be outlined below.

Concerning the datasets used in the context of this work, Wiener filtering was not conducted in any of the EMAN1.9 refinements, since the datasets were CTF-corrected *via* phase flipping prior to refinement. In consequence, the traditional refinement of the native BgAChBP dataset using EMAN1.9 implied a linear correlation alignment without any

weighting, thus being highly susceptible for over-refinement. Yet, the danger of converging to a local minimum was decreased due to the use of classiter.

A second strategy – applied for native BgAChBP, deoxy-ScoHc and oxy-ScoHc – used different datasets for alignment and reconstruction. As Henderson et al. (2011) pointed out, it is most likely that significant contributions for the alignment of particles are limited to spatial frequencies between 120 - 15 Å. Scheres (2012b) presents similar results, stating that spatial frequencies higher than 10 Å do not contribute to the alignment of the signal. In the light of these results, the question emerges why one should include frequencies higher than these values into the alignment procedure at all, especially in consideration of the importance of high frequencies in noise-bias and over-refinement. In combination with the 'usefilt' option, datasets with high-resolution cut-offs of 8 or even 12 Å were used for alignment, whereas datasets with high-resolution cut-offs of 3 Å or only normalized datasets were used for reconstruction. This ensured that the very high-frequency noise components could not interfere with the alignment steps, but were allowed into the 3-D reconstruction in order to include all relevant information. Low-pass filtering of the maps according to the measured resolution excluded frequencies not showing significant differences between signal and noise. Refinements calculated following this procedure can be expected to show significantly less tendencies for over-refinement and an over-all more consistent appearance of the density map.

1.2.2 FREALIGN

As a program explicitly written to refine structures that were already calculated in different program packages, FREALIGN incorporates many procedures to overcome noise-bias and the resulting over-refinement of noisy datasets.

As previously described, linear correlation coefficients are very noise-sensitive and impose a weighting in favor of large amplitudes, which in turn do not carry any information about the quality of the signal (Stewart and Grigorieff 2004). For these reasons, the authors first derived a Wiener-filter-weighted equation

$$R = \frac{1}{\sum_i N_i} \sum_i N_i CC_i^3 \, ,$$

with R being the term to maximize for alignment, N being the number of Fourier terms in

a resolution zone i and CC the average correlation coefficient (for detailed information concerning the derivation of the equation see Stewart and Grigorieff 2004). In order to minimize the correlation of noise between two images – which is given by a Gaussian distribution with the standard deviation 1/[image edge length in pixels] – the CC values are filtered with the standard deviation of the noise correlation (*ibid.*). As a third step, FREALIGN tries to minimize the probability of remaining in a local optimum. The iterative change of a particle's position is tracked as a gradient of the original position towards the new values. Since small positional changes of low frequencies lead to large changes in high frequencies, algorithms tend to remain in local optima and converge towards the high frequency values (*ibid.*). By filtering the convergence with the inverse of the spatial frequency, this bias is dramatically decreased (*ibid.*). As a final step, FREALIGN exploits the fact that the alignment of signal in one resolution zone implicates the alignment of signal in another resolution zone. This "convergence" of signal, as Stewart and Grigorieff (2004) call it, cannot be observed for noise, i.e. for each image the noise can be randomly in phase or out-of-phase. Consequently, by taking the absolute value of CC, anti-correlation is avoided and noise components will destructively interfere during the averaging process. This calculation is only possible because FREALIGN optimizes the correlation coefficient for each resolution zone, as opposed to the global maximization that is used in many other software suites. The final equation to be maximized during alignment would thus read

$$\tilde{R}'_f = \sum_i \left| CC^3_{f,i} \right| ,$$

where f is the noise filter described above.

After the alignment has calculated a weighted correlation coefficient for each image, the latter can be inserted into 3-D Fourier space for 3-D reconstruction. Each image is introduced as follows:

$$R_i = \frac{\sum_j w_j^2 b^2 c_j P_{i,j}}{f + \sum_j (w_j b c_j)^2}$$

with R_i as a sample i of the 3-D reconstruction in Fourier space, P_{ij} as a sample of the Fourier transform of the particle image j (prior to CTF-correction) contributing to R_i, w_j as a weight for image j that is derived from the weighted correlation coefficient and the average correlation coefficient of the dataset, b as a necessary transform constant from 2-D to 3-D Fourier

space, c_j as the CTF of image j at point i, and f as a filter constant (GRIGORIEFF 2007). The constant f is set to 10 % of the sum in the denominator averaged over the whole dataset. When comparing the above formula with the one given for the Wiener filter (see D1.2.1), it is apparent that FREALIGN performs a Wiener-filter-like operation by substituting 1/ SNR by the constant f. It is thus to assume that CTF-correction in FREALIGN will be superior to mere phase flipping.

After the analysis of the algorithm used in FREALIGN, it is noticeable that the program undergoes severe efforts to prevent or minimize noise-bias, the convergence to local minima, and an over-refinement of the 3-D structure to be determined. The traditionally used, and compared to linear correlation coefficients almost equally biased, phase residual has been replaced by an extensive armada of weighting and optimization algorithms (GRIGORIEFF 1998; STEWART AND GRIGORIEFF 2004).

Finally, a short remark concerning the resolution determination in FREALIGN has to be made. In contrast to the generally used 0.5 threshold, Grigorieff states that FREALIGN values compare best to the resolution of other software packages, when the 0.143 threshold is used in FREALIGN (personal communication). This might be due to the previously described weighting algorithms which could result in a more accurate estimate of the resolution at a lower threshold. This assumption would be in accordance with the observations of over-estimated resolutions in over-refined density maps (e.g. STEWART AND GRIGORIEFF 2004).

Additionally to the measures implemented by FREALIGN, my refinement strategy for the oxy-ScoHc and deoxy-ScoHc datasets calculated with this program only allowed resolutions lower than 13 Å to be used for alignment. The resulting reconstructions can consequently be expected to show only a small degree, if any, of over-refinement.

1.2.3 RELION

The newest software used for some of the refinements in this work explores new algorithms for the analysis of cryo-EM data. The two main implementations that let RELION be one of the most interesting recent developments concerning noise bias in cryo-EM refinements, are the implementations of (i) a statistical refinement framework that is robust against noise bias and (ii) the calculation of gold-standard FSCs, both of which will be described in more detail below.

Instead of maximizing the different steps of original refinement procedures (such as align-

ment, classification and reconstruction) separately, RELION calculates and maximizes one single probability function, resulting in minimum user interaction (Scheres 2012b). For the implementation into the field of cryo-electron microscopy, a novel expectation-maximization algorithm was developed that calculates *maximum a posteriori* estimates following Bayesian statistics. A detailed description of the algorithm can be found in Scheres (2012a and b). Apart from the experimental data, expectation-maximization algorithms need priors to calculate the probability of the newly created model (Kucukelbir et al. 2012). In RELION, this prior is a Gaussian distribution of the Fourier components of the reference structure, i.e. a smoothed reference structure (Scheres 2012a). The introduction of smoothness into a cryo-EM refinement is one of the reasons for the sturdiness of the algorithm against over-refinement: in contrast to other refinement algorithms, images are not assigned to discrete orientations when they are aligned against the references. Rather, probability-weighted integrals are calculated over all possible orientations, i.e. projections will contribute with multiple Gaussian-distributed orientations to the 3-D reconstruction, the sharpness of this distribution being determined by the noise model of the data (*ibid.*). This said, it is logical that the Gaussian distribution acts as an adaptive low-pass filter (Wahba 1980 *cited in* Kucukelbir et al. 2012) and that the inclusion of a noise model results in a "Bayesian interpretation of the [...] Wiener filter" (Scheres 2012a). For example, if the data is very noisy, the distribution of a projection over a range of orientations will be high, effectively smoothing (i.e. low-pass filtering) the resulting reconstruction. The contrary holds true for low noise levels. The noise model is updated in every iteration and – in contrast to the Wiener constants of EMAN and FREALIGN – is given for each resolution shell, making it possible to describe non-white, colored noise (*ibid.*).

As a second important implementation, RELION calculates gold-standard FSCs (Henderson et al. 2012) by separately refining two halves of the dataset. Hence, noise correlation between the halves of a dataset – a known problem when refining the whole dataset – is entirely avoided, making the resolution estimate using an FSC more reliable (Grigorieff 2000). The SNR calculated between the two reconstructions of every iteration is then used for the previously mentioned adaptive filter and given by

$$SNR^{MAP}(v) = \frac{FSC(v)}{1 - FSC(v)} \, ,$$

where v is the spatial frequency (Scheres 2012b). This calculation essentially uses the results

obtained by FRANK AND AL-ALI (1975), which confirm the possibility of calculating the SNR by using the cross-correlation between two images of one dataset:

$$SNR = \frac{CC}{1-CC}.$$

Ultimately, the two halfsets are combined to a final reconstruction with improved signal and resolution. The fear of not reaching the same resolution when separately refining two halfsets, compared to refining the whole dataset, is disapproved by SCHERES AND CHEN (2012) in several experiments.

Due to the aforementioned properties of the refinement algorithm that is implemented into RELION, the resulting 3-D reconstructions – similarly to the ones calculated in FRE-ALIGN – can be expected to show minimal, if any, over-refinement.

1.2.4 OTHERS

In light of the problems arising from linear cross-correlation, the necessity for alternative algorithms is immanent. Apart from RELION, promising statistical algorithms include the use of maximum likelihood (e.g. Sigworth 1998; SCHERES ET AL. 2005; SIGWORTH ET AL. 2010) which, for two-dimensional analysis, has already proven to be very valuable, as it is very insensitive to model bias and noise bias and converges very quickly (STEWART AND GRIGORIEFF 2007). However, its computationally highly demanding nature currently only permits 3-D refinements with angular step sizes of 10°, too much for obtaining high-resolution structures (SCHERES 2012B). In the future, other alignment strategies might be feasible when considering, for example, BONNET (2004), who predicts the usage of artificial intelligence and pattern recognition algorithms for image processing.

For the time being, the efforts pursued by the 3-D EM community to solve the problem of noise bias and over-refinement evoked the creation of several promising approaches, including the ones discussed above. In conclusion, the 'usefilt' option in EMAN can help to decrease over-refinement as a basic and simple method, although it is doubtful that it entirely prevents it. In the light of the increasing awareness of the problem in the community, I introduced FREALIGN to our laboratory. The extensive efforts for the suppression of over-refinement should yield superior structures when compared to the refinement strategy with prior CTF-correction that I pursued in EMAN. Finally, when RELION was published recently, I

implemented it for extensive testing, especially using the ScoHc datasets. The algorithm and the examples of refined structures presented by SCHERES (2012B) are compelling arguments for using this software for 3-D refinements of cryo-EM data. Due to the lack of time, I was not able to confirm all of the cryo-EM reconstructions presented in this work by re-calculating refinements in RELION, although this would be my recommendation.

2 Myriapod hemocyanin

2.1 *Scutigera coleoptrata* hemocyanin

The analysis of oxy-ScoHc and deoxy-ScoHc showed similar visual differences between the two conformational states when comparing three different refinement algorithms. Furthermore, these differences were confirmed by a 3-D reconstruction processed independently at Brandeis University, Boston, USA. All 3-D reconstructions have in common that the upper/lower deoxy-ScoHc 3x6-mer seemingly rotates counter-clockwise/clockwise against oxy-ScoHc when considering the top view of the molecule. It was shown, however, that indeed each hexamer is rotated by ~2° instead of a rotation of the whole 3x6-mer. Yet, as expected, FREALIGN and RELION did calculate density maps that – judged by the resolution estimates – appeared less resolved and lacked outer densities. This effect – originating from the implementation of a more sophisticated CTF correction as opposed to the phase-flipping used for the EMAN1.9 or previously calculated IMAGIC-5 reconstructions – should have been corrected for by imposing a B-factor (bfac reconstructions). However, the resulting maps were not visually comparable to the results of the phase-flipped data. Due to the fact that FREALIGN does not have the option to only phase-flip the data, this visual comparability was achieved by re-calculating the last 3-D reconstructions of the FREALIGN and RELION refinements while correcting for the CTF using an amplitude contrast of 70 % (amp70 reconstructions). The utilization of amp70 reconstructions, however, requires a critical analysis concerning their validity within the scope of this study.

2.1.1 *Validity of amp70 reconstructions*

It was shown that the visual difference between bfac and amp70 reconstructions did not have an effect on the amount of hexamer rotation and little effect on the position of the rotation axis when comparing oxy- and deoxy-ScoHc. To further confirm that amp70 reconstructions did not have a negative impact on the measurements, root mean square deviations (RMSD), rotations and axial shifts between hexamers fitted into different deoxy-ScoHc reconstructions were compared (Table 6). In RELION, the 3-D reconstruction of the final deoxy-ScoHc refinement iteration was re-calculated by either correcting for the CTF using only phase-flipping or omitting the first CTF peak for correction (both are selectable RELION options). Both reconstructions visually resembled the amp70 reconstructions and the reconstruction obtained in EMAN1.9. Table 6 clearly shows that the values between amp70 and other

Table 6   Comparison of the orientational changes introduced when docking a ScoHc hexamer into different deoxy-ScoHc reconstructions by rigid-body fitting. It is apparent that the differences between amp70 reconstructions and phase-flipped ones is not larger than the differences between different phase-flipped reconstructions.

RMSD					
	RELION ommit first peak	RELION phase flip	RELION amp70	EMAN	FREALIGN amp70
RELION ommit first peak	---	0.191	0.179	0.826	1.042
RELION phase flip	0.191	---	0.214	0.812	1.035
RELION amp70	0.179	0.214	---	0.815	1.041
EMAN	0.826	0.812	0.815	---	0.654
FREALIGN amp70	1.042	1.035	1.041	0.654	---
Rotation					
RELION ommit first peak	---	0.07	0.05	0.18	0.3
RELION phase flip	0.07	---	0.05	0.25	0.31
RELION amp70	0.05	0.05	---	0.23	0.28
EMAN	0.18	0.25	0.23	---	0.3
FREALIGN amp70	0.3	0.31	0.28	0.3	---
Axial shift					
RELION ommit first peak	---	-0.18	0.01	0.54	0.85
RELION phase flip	-0.18	---	-0.1	0.31	0.85
RELION amp70	0.01	-0.1	---	0.48	0.8
EMAN	0.54	0.31	0.48	---	0.52
FREALIGN amp70	0.85	0.85	0.8	0.52	---

reconstructions are not substantially different to the values obtained when comparing two phase-flipped reconstructions. Similar results were found for oxy-ScoHc (results not shown). Thus, the validity of using amp70 reconstructions for the scope of this study is substantiated by (i) similar rotations between oxy-ScoHc and deoxy-ScoHc hexamers when comparing bfac and amp70 reconstructions (Table 7) and (ii) similar values measured when comparing hexamers fitted into phase-flipped and amp70 reconstructions of the same conformational

Table 7 Comparison of the orientational changes of oxy- and deoxy-ScoHc hexamers fitted into reconstructions calculated with different software packages. RMSD: root mean square deviation.

	RMSD (Å)	Rotation (°)	Axial shift (Å)
EMAN1.9			
oxy-ScoHc ↔ deoxy-ScoHc	2.947	2.16	1.48
RELION			
oxy-ScoHc (bfac) ↔ deoxy-ScoHc (bfac)	2.681	2.08	1.38
oxy-ScoHc (amp70) ↔ deoxy-ScoHc (amp70)	2.856	2.17	1.55
FREALIGN			
oxy-ScoHc (bfac) ↔ deoxy-ScoHc (bfac)	2.844	2.29	0.82
oxy-ScoHc (amp70) ↔ deoxy-ScoHc (amp70)	2.433	2.11	0.84
oxy-ScoHc (Brandeis) ↔ deoxy-ScoHc (bfac)	4.207	2.67	2.6
IMAGIC (previously calculated)			
oxy-ScoHc ↔ deoxy-ScoHc	2.837	2.28	1.4

state (Table 6). It must be emphasized, however, that this procedure should be restricted to the current comparison between software packages, whereas concerning future reconstructions, it is strongly advised to use the more sophisticated Wiener-filter-like CTF-correction in combination with B-factor sharpening for more accurate results.

2.1.2 Pinpointing the structures inducing hexamer rotation

Visual inspection of the conformational changes that emerged in the 3-D reconstructions of oxy-ScoHc and deoxy-ScoHc first suggested a torsion between the two 3x6-mers of the molecule. Yet, the analysis of the rotations between the respective pseudo-atomic models led to a new understanding of the allosteric movements found in ScoHc. When comparing the rotation axes between hexamers fitted into oxy-ScoHc and deoxy-ScoHc reconstructions of different software packages, a strong convergence towards the region of interface 1↔2 is visible (Figure 84). Previous reconstructions that were calculated in IMAGIC-5 were also
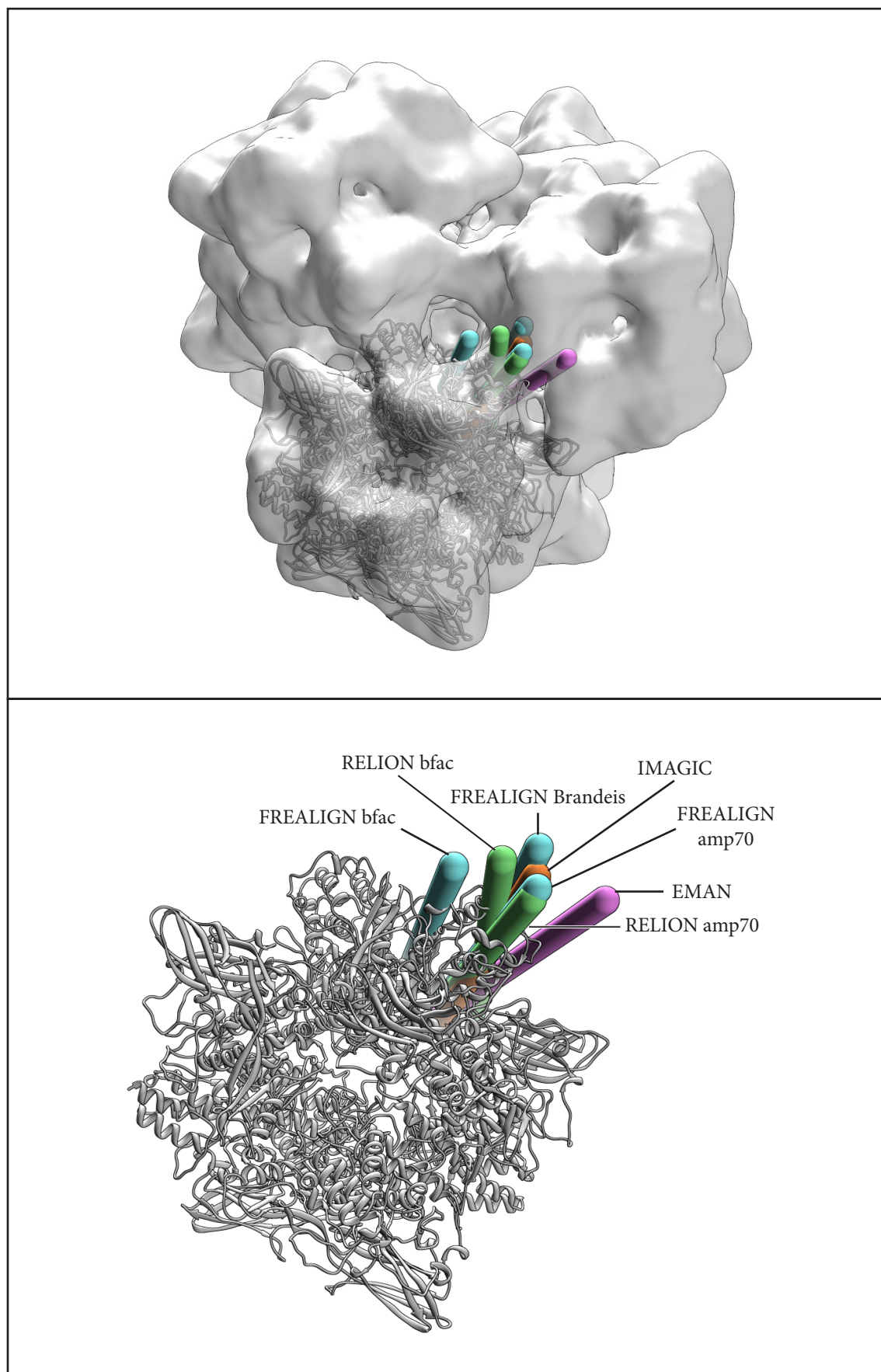
Figure 84 Rotation axes between hexamers of oxy-ScoHc and deoxy-ScoHc that were fitted into density maps refined with IMAGIC-5 (orange), EMAN1.9 (magenta), FREALIGN (cyan) and RELION (green). The axes are all confined to one region of the hexamer.

included into the analysis and show similar results. The position of the rotation axes strongly suggests an involvement of interface 1↔2 in the allosteric movements that are necessary for achieving the highly cooperative oxygen binding measurable for ScoHc. Structural determinants arguing in favor of interface 1↔2 for the involvement in the conformational variability will be discussed in section D2.3.

2.2 *POLYDESMUS ANGUSTUS* HEMOCYANIN

The two- and three-dimensional analyses of PanHc by single particle analysis showed that a stable 3x6-mer exists in both negatively stained and cryo-EM samples. However, although it could be shown that the vast majority of the 2-D class averages obtained by MSA are projections of a 3x6-mer, indications of a possible heterogeneity of the sample remained. Apart from the clearly discernible 1x6-mers present in the sample, for instance, putative 2x6-mers were visible as rod-like structures that show a similar brightness between two main densities (see Figure 35). Given their size, these structures could well be side views of a 3x6-mer. However, in a 3x6-mer side view that shows two distinguishable densities, one of the densities has to be brighter than the other, given the fact that two of the three hexamers would overlap in the projection image. Consequently, it might be possible that PanHc is also present as stable 2x6-mer. An alternative consideration involves the alignment and classification algorithms that were used for single particle analysis. It is conspicuous that the four different 3-D classes of the multirefine algorithm contain almost equal amounts of particles, which had typically not been the case with other proteins that I processed using this algorithm. Moreover, although one of the 3-D classes resembles a 2x6-mer, it is apparently comprised of a heterogeneous set of class members, judged by the faint density of a third hexamer that is visible in the 3-D reconstruction (see Figure 32 G/H). These observations suggest that the alignment algorithms might have difficulties to accurately determine the orientations of the particles. The latter is reinforced by the fact that MSA class averages of the alleged 2x6-mers
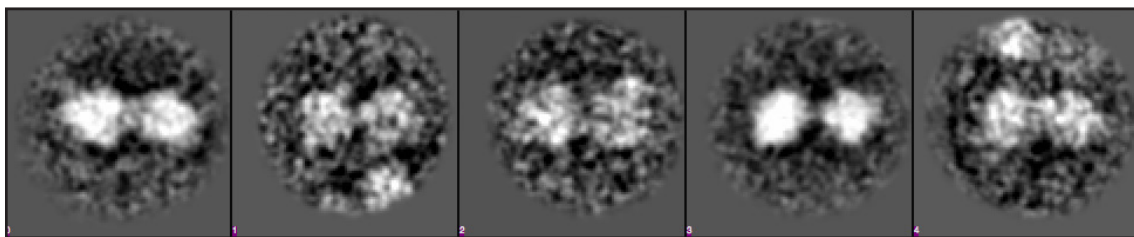
Figure 85 Class averages of alleged 2x6-mers that appeared after carefully selecting the negatively stained dataset for 3x6-mer projections.

contain class members that are clearly side views of 3x6-mers (data not shown). Furthermore, these class averages are still found in the negatively stained dataset for which only 3x6-mer projections were selected (Figure 85). In conclusion, the presence of stable 2x6-mers cannot be confirmed, but evidence suggests that alleged 2x6-mer class averages are the result of misalignment and/or misclassification.

Considering a 6x6-mer assembly, class averages of rectangular structures might indicate its presence. But also alleged top views of 3x6-mers could in theory be projections of a 6x6-mer. The sparse distribution of rectangular views, however, lets a stable 6x6-mer seem very unlikely for the analyzed samples. Yet, it was previously shown by size-exclusion chromatography that the hemocyanin of the diplopod *Spirostreptus* sp. (SpiHc) probably shows a 6x6-meric quaternary structure with the tendency of degrading into 3x6-mers (Jaenicke et al. 1999). The same authors show an electron microscopic image presumably displaying 6x6-mer views of the hemocyanin (Figure 86). Although some particles might indeed be 6x6-mers, the analysis of the 3x6-meric PanHc identifies most of the alleged 6x6-mer particles as 3x6-mers. The contradicting results between size-exclusion chromatography and electron microscopy led to the decision of comparing a simulated PanHc 6x6-mer with the known 6x6-mer of *Scutigera coleoptrata*. As can be seen, the PanHc 3x6-mer is very similar to a ScoHc 3x6-mer, which suggests that – in a 6x6-meric assembly – the second PanHc 3x6-mer should be positioned similar as the second ScoHc 3x6-mer (Figure 87). A similar assembly as ScoHc, however, would inevitably define the top views of the PanHc EM datasets as 3x6-mers, due to the fact that top views of ScoHc show a different appearance (Figure 88 A). Alternatively, the second
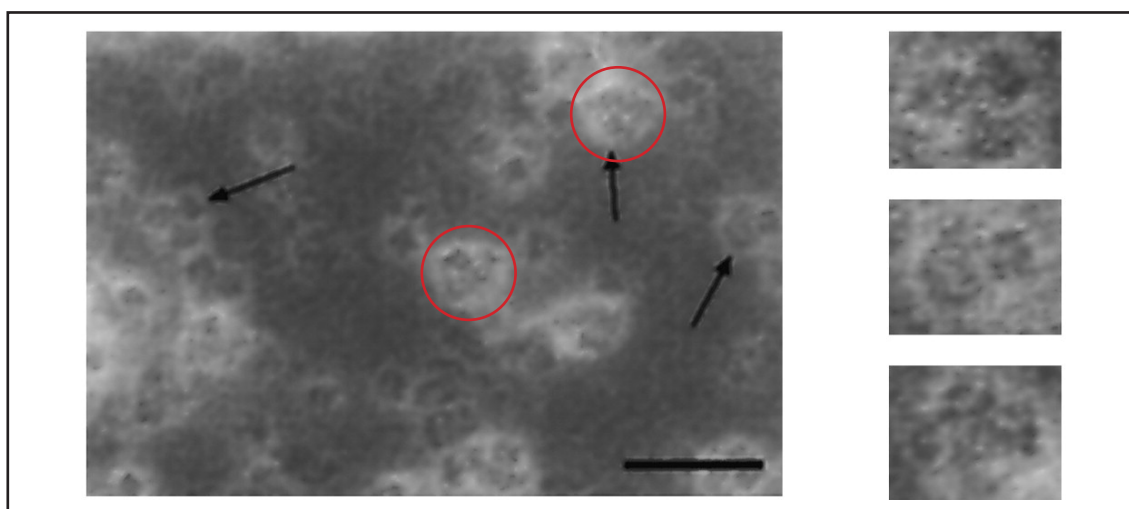
Figure 86 Electron microscopic images of *Spirostreptus* sp. hemocyanin as shown by Jaenicke et al. (1999; Fig. 1B). The particles marked by a red circle might indeed represent 6x6-mer assemblies, whereas the highlighted particles on the right most probably show 3x6-mers.
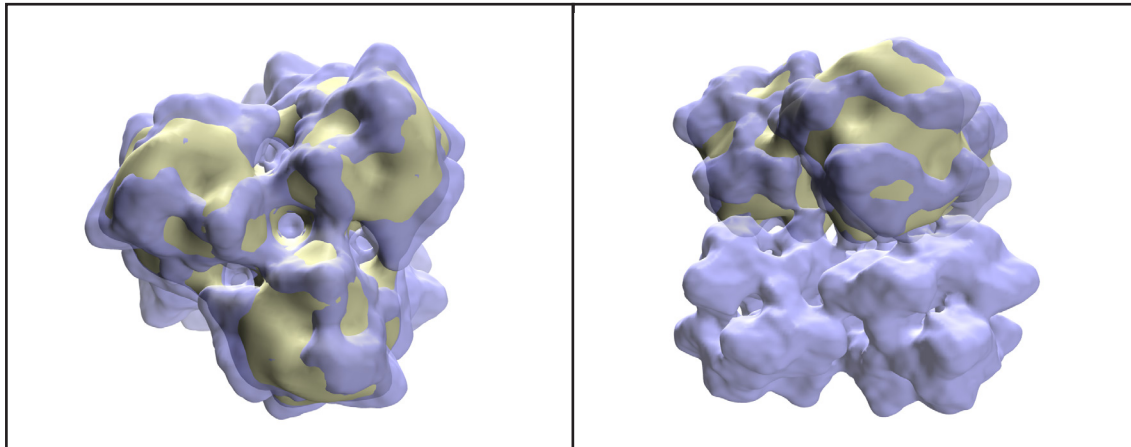
Figure 87 Comparison between PanHc (yellow) and ScoHc (purple) in top view (left) and side view (right).

PanHc 3x6-mer could be positioned underneath the first one without any torsion between the two 3x6-mers, which could yield top views as seen in the EM datasets. Yet, this assembly would not be in accordance with the sparsely observed rectangular structures, since they show a shift of two 3x6-mers against each other (Figure 88 B). Consequently, a 6x6-meric PanHc is currently only supported by size-exclusion chromatography aof SpiHc. In view of these considerations, a comparison between the inter-3x6-mer interfaces of ScoHc and the simulated PanHc 6x6-mer seems immanent. The simulated 6x6-mers were built according to the ScoHc quaternary structure. Ideally, SpiHc would have been included in the comparison of all interfaces, but no sequence for subunit 2 has been published to date. As was shown by MARKL ET AL. (2009), interface 3↔3 of PanHc does not essentially differ from ScoHc in more than the lack of ScoHc H^{151}. Similarly, interface 1_A↔1_E does contain equivalent residues when comparing PanHc and ScoHc. Histidines H^{571} and H^{593} found in ScoHc are not present in PanHc, but might be substituted by H^{493} and H^{609}, which also lie in the vicinity of the interface. In SpiHc, no substitution for H^{493} can be found, although the rest of the interface remains similar to PanHc. Finally, interface 1_A↔1_D shows relatively subtle differences between ScoHc and SpiHc which are characterized by the substitution of ScoHc K^{405} by SpiHc E^{403}, the lack of ScoHc H^{407} in SpiHc and the addition of SpiHc H^{401} and H^{408}. In PanHc, however, this interface is entirely interrupted by the additional histidine loop that distinguishes PanHc from the other myriapodal hemocyanins (Figure 89). This loop might disturb the formation of a 6x6-mer, as it unites 162 (27*6) additional amino acids in the center of the molecule. Alternatively, the loop could be a source of interaction between 3x6-mers, given the fact that it bears positively charged histidine and lysine residues, as well as negatively charged aspartate residues.
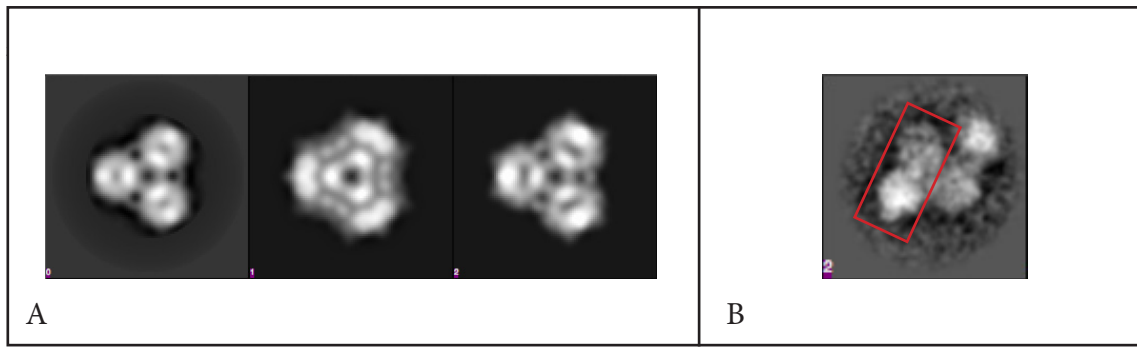
Figure 88 2-D analysis of PanHc. A) Comparison between top view re-projections of a PanHc 3x6-mer (left), a ScoHc 6x6-mer (middle) and a simulated ScoHc 3x6-mer (right). All reconstructions were low-pass filtered to 25 Å. B) Rectangular view of a putative 6x6-mer occasionally found in PanHc datasets. One 3x6-mer is highlighted in red.
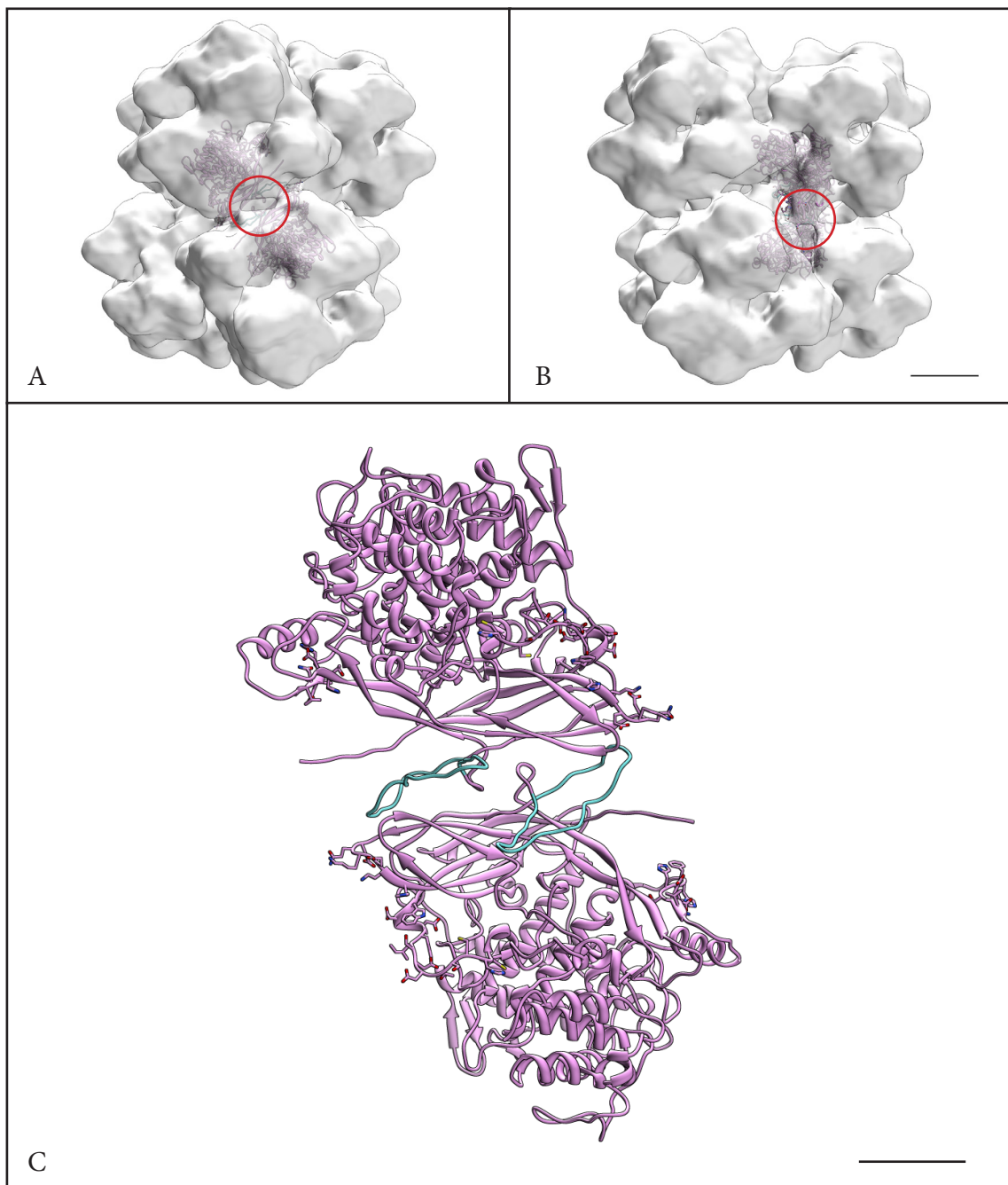
Figure 89 Simulated interface $1_A \leftrightarrow 1_D$ of PanHc. A) Position of the interface indicated by a ScoHc density map as reference. B) Same as A), but turned by -90° around the three-fold symmetry axis of ScoHc. C) Close-up view of the interface from the inside of the molecule. The additional histidine loop is colored in cyan. Scale bars: 50 Å (B), 20 Å (C).

In summary, there is a possibility that SpiHc and PanHc hemocyanins exist as a stable 3x6-mer and interact loosely as a 6x6-mer, which could be explained by the results obtained by size-exclusion chromatography and electron microscopy. Furthermore, they are the first arthropod hemocyanins showing an odd number of hexamers described to date (excluding 1x6-mers). It must be also considered that sample preparation for electron microscopy could have resulted in a degradation of 6x6-mers to 3x6-mers. However, a stable 3x6-mer would provide, as will be discussed in the next section, an intelligible explanation for the differences in cooperativity that can be measured between ScoHc and SpiHc. Moreover, the occurrence of odd-numbered multi-hexameric assemblies – in contrast to the more frequently encountered even-numbered assemblies – can be explained by the independent evolution of multi-hexameric hemocyanins among arthropods (for a review see Burmester 2002).

2.3 Cooperativity-driving mechanisms in Myriapod hemocyanin

Multi-hexameric hemocyanins of arthropods typically show high cooperativity in oxygen binding. For instance, the hemocyanin of the horse-shoe crab *Limulus polyphemus* is comprised of eight hexamers which collectively display a Hill coefficient of ~9 (Martin et al. 2007). The six hexamers of ScoHc amount for a Hill coefficient of up to 10.7, although at the pH values of the present study they are more likely to approximate 9 (Mangum et al. 1985). At such high cooperativity, it can be assumed that the oxygenated state corresponds to the R-state of allosteric proteins and the deoxygenated state corresponds to the T-state (Decker et al. 2007). However, the concerted model of allostery, or MWC model (Monod et al. 1965), is not sufficient to explain the differential effects of allosteric effectors on arthropod hemocyanins that are larger than one hexamer (Decker et al. 2007). A nested extension of this model was proposed by Decker et al. (1986) that circumvents this problem. Single particle analysis, as it was used in the course of this study, is very unlikely to allow any insights into these allosteric models in ScoHc. Firstly, the resolutions achieved until now are not sufficient to recognize small conformational movements of one subunit. Secondly, the symmetry-related averaging of all six hexamers, in order to achieve high resolutions, would only allow for the reproduction of intra-hexameric differences between subunits. Finally, the latter would only be seen if all hexamers were in an equal state. The only solution to this problem would be to pursue a 3-D analysis without imposing point group symmetry, albeit it is questionable that the alignment algorithms could distinguish between the minute differences inside the hexamers.

This said, single particle analysis was able to show global conformational movements that are present between oxy-ScoHc and deoxy-ScoHc. Measurements between the pseudo-atomic models of the two states could show that the hexamers of ScoHc rotate around axes which are in the direct vicinity of interface 1↔2 (the rotations found for the density maps calculated by different software packages can be seen in Table 7). Interestingly, this interface is almost entirely formed by amino acids that are situated in domain 1 of the hemocyanin subunits. HAZES ET AL. (1993) could show a rotation of ~8° of domain 1 against the other domains, when comparing a crystal structure of *Panulirus interruptus* hemocyanin (VOLBEDA AND HOL 1989) with a crystal structure of *Limulus polyphemus* hemocyanin of supposedly different conformational states. Furthermore, the authors could show that domain 1 has a direct influence on the active site *via* F^{49} (ScoHc F^{58}; see Figure 90 for an illustration of interface 1↔2 in ScoHc). Concerning the 2° rotation measured between oxy-ScoHc and deoxy-ScoHc hexamers, it must be considered that the pseudo-atomic model was created by rigid-body fitting. Consequently, it might well be that some regions inside a hexamer move substantially more than seen in the current pseudo-atomic model and that an 8° rotation of domain 1 ultimately translates into the 2° rotation of a whole hexamer observable in the 3-D reconstructions. One could argue that the different buffer and imaging conditions of the samples compared in this study might have biased the results. Indeed, future work should keep these conditions as stable as possible. However, it is highly doubtful that the different buffer conditions led to any bias at all. Despite the fact that myriapod hemolymph is slightly alkaline (pH 8; XYLANDER 2009), ScoHc has been shown to be highly cooperative at pH values between 7 and 8 (see above; MANGUM ET AL. 1985). Therefore, it can be assumed that ScoHc is fully oxygenated under atmospheric conditions and fully deoxygenated in a 100 % nitrogenic atmosphere. As to the different imaging conditions, a third, independently processed dataset confirmed the conformational state of oxy-ScoHc and substantiated the current model of conformational movement. At the current resolutions, the possible changes in interface 1↔2 upon rotation should not be over-interpreted, as the data does not allow any sustainable conclusions. Nevertheless, the identification of interface 1↔2 as the major source of conformational movement will allow the conception of precisely targeted experiments that focus on this region in ScoHc.

In PanHc, the lack of a stable 6x6-mer could be in accordance with the almost absent cooperativity in SpiHc. However, oxygen binding studies have not been performed on PanHc. Using SpiHc oxygen binding values for the analysis of PanHc only based on phylogenetic
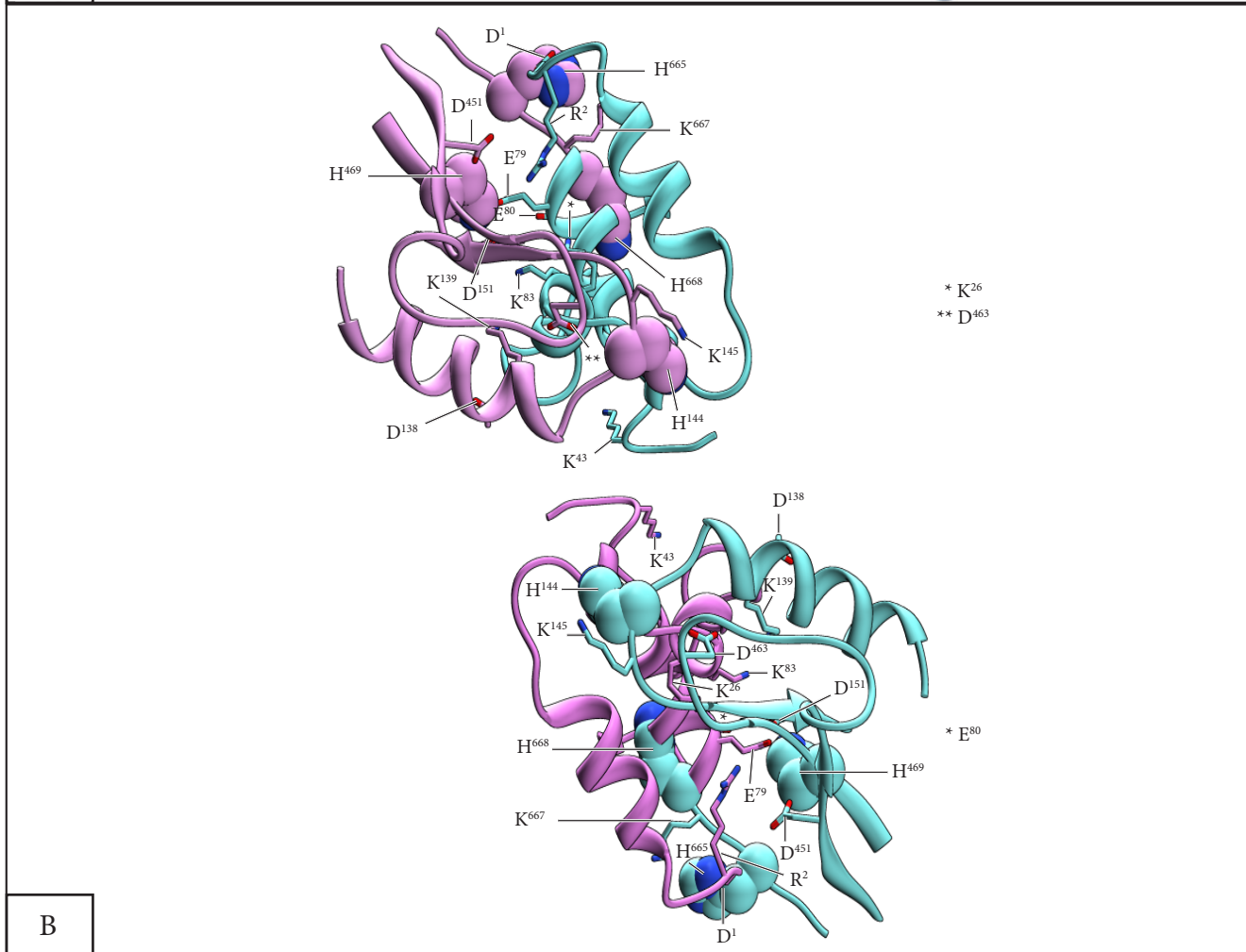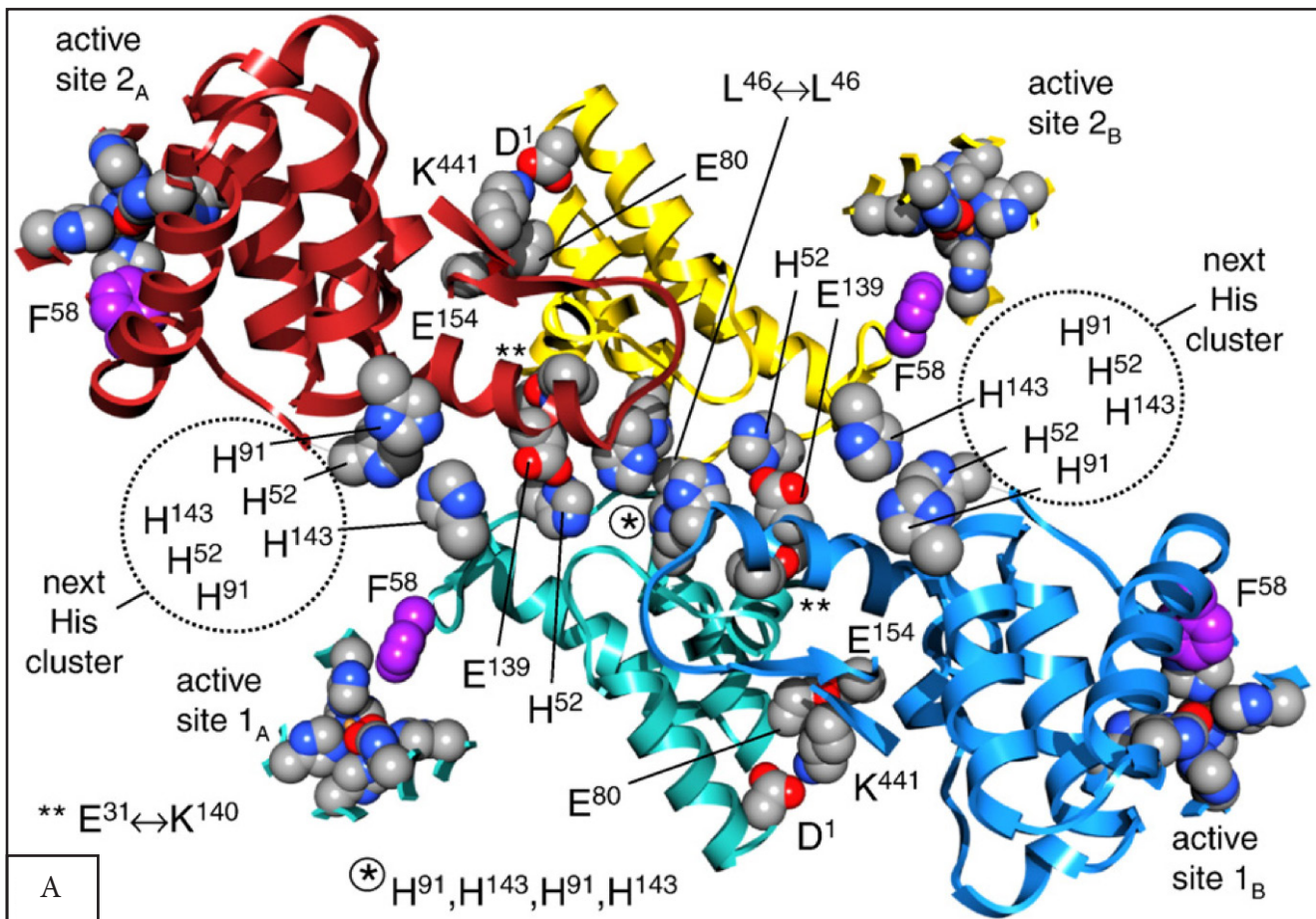
Figure 90 Comparison of interface 1↔2 in ScoHc and PanHc. A) ScoHc shows several charged amino acids and a central histidine cluster (H^{52}, H^{91}, H^{143}). Note the connection of the interface to the active sight, especially via F^{58} (reprinted from MARKL ET AL. 2009 with permission from Elsevier). B) PanHc does not contain a central histidine cluster. Two other histidines (H^{665} and H^{668}) are widely dispersed over the interface (H^{668} is also present in ScoHc).

proximity seems too speculative. Savel-Niemann et al. (1988) showed that a hemocyanin half-molecule (2x6-mer) of the tarantula *Eurypelma californicum* (EcaHc) was only able to achieve 50 % of the cooperativity measurable for an entire 4x6-mer. It seems consistent that the rotations between the two ScoHc 3x6-mers, which amount for an accumulative shift of ~8-10 Å between peripheral amino acids of different 3x6-mers and which are induced by a rotation of individual hexamers, are the main reason for cooperativity in ScoHc. Furthermore, it can be assumed that 18 subunits in a 3x6-mer would not reach the cooperativity of 36 subunits in a 6x6-mer. Nevertheless, Savel-Niemann et al. (1988) also described that the oxygen affinity of the EcaHc half-molecule did not change in comparison to the 4x6-mer. In SpiHc, however, the affinity is increased almost 10-fold to P_{50} = 4.7 Torr when compared to ~55 Torr in ScoHc (both at pH 7.5; Mangum et al. 1985; Jaenicke et al. 1999). Yet, the occurrence of SpiHc as a stable 3x6-mer and only loosely interacting 6x6-mer has still to be proven. Consequently, the altered affinity might only be an adaptation to the habitats of the animals without large differences in the quaternary structures. It was shown in section D2.2 that the most pronounced difference between inter-3x6-mer interfaces of ScoHc and simulated interfaces of PanHc, was the distribution of histidine residues. Besides, histidines have been suggested to be important agents in conformational changes (Martin et al. 2007). A comparison of the allosterically important interface 1↔2 in ScoHc and PanHc reveals that the histidine cluster found in ScoHc (H^{52}, H^{91}, H^{143}; Markl et al. 2009) is only comprised of H^{144} in PanHc (Figure 90). It has to be noted that the greater distance between the two PanHc H^{144} in Figure 90 compared to ScoHc results from the rigid-body fitting of the pseudo-atomic model into the lower-resolution density map of PanHc. Although H^{665} and H^{668} are also contributing to the PanHc interface, they are dispersed among it. Furthermore, H^{668} has a correspondent histidine in ScoHc (H^{638}). Therefore, the possibility that oxygen binding properties of ScoHc and PanHc are partly regulated *via* different dispersions of histidines in the respective inter-hexamer interfaces, seems very conclusive. Nevertheless, in my opinion, the transmission of allosteric information through the large histidine loop of a loosely interacting PanHc 6x6-mer seems unlikely, as a large unstable loop should not be able to provide the necessary rigidity for the initiation of allosteric movements. A stable 3x6-mer could thus theoretically allow a certain amount of cooperativity which, in the case of PanHc, might not be given due to the lack of inter-hexamer interface histidines. These results will have to be substantiated in future studies which should focus on answering the question of how many hexamers are combined in the physiologically active PanHc quaternary structure.

2.4 Concluding remarks

The hemocyanins of the chilopod *Scutigera coleoptrata* and a representative of diplopods (*Spirostreptus* sp.) show substantially different oxygen binding properties. Both types of hemocyanin were thought to function as a 6x6-meric assembly for which the mechanisms driving the differences in oxygen binding were poorly understood. By means of single particle analysis, I was able to three-dimensionally reconstruct an oxy-ScoHc and a deoxy-ScoHc sample with different software packages and compare putative conformational movements. The results of all software packages yielded similar results that show a 2° rotation of each hexamer along an axis that is close to interface 1↔2. Between two hexamers of different 3x6-mer half-molecules, an overall rotation of 4° against each other is the result of oxygen binding, which translates into a shift of approximately 8-10 Å for peripheral amino acids. Due to the rigid-body fitting used in this study, it might be possible that parts of the hexamer rotate more than 2°. In fact, domain 1, which carries most of the amino acids that form interface 1↔2, has been shown to rotate by ~8° against the other domains (Hazes et al. 1993). A third, independently processed refinement confirmed the results of the oxy-ScoHc structure. The results allow future studies to focus on the region that comprises interface 1↔2, in order to fully understand the mechanistic implications of oxygen binding in ScoHc. A further insight into myriapod hemocyanin cooperativity was achieved by analyzing PanHc, which showed a stable 3x6-mer in the electron microscopic images. The pseudo-atomic model created on the basis of the 3-D reconstruction revealed a divergent distribution of histidines in inter-hexamer interfaces when compared to ScoHc. Furthermore, the characteristic large histidine loop of PanHc is located at an inter-3x6-mer interface, possibly obstructing the formation of a stable 6x6-mer. An electron microscopic analysis of SpiHc, which was conducted by Jaenicke et al. (1999), also predominantly showed projections belonging to a 3x6-mer. However, size-exclusion chromatography showed a mixture of 6x6- and 3x6-mers (*ibid.*). The structure of SpiHc 6x6-mers will have to be analyzed more closely in future studies, as two 3x6-mers might only loosely interact with each other. Furthermore, oxygen binding studies will have to be performed for PanHc. However, the occurrence of a stable 3x6-mer in PanHc and SpiHc and different histidine patterns in inter-hexamer interfaces might conclusively explain the highly different oxygen binding properties of ScoHc and SpiHc.

3 Vesicle-inducing protein in plastids 1

3.1 Wild type Vipp1

Single particle analysis allowed to visualize Vipp1 in three dimensions for the first time. All 3-D reconstructions calculated are very similar when judged by eye, showing outer spikes, inner rails, an equal height and – some of them – a ridge at the bottom inside the ring (Figure 46). The ridge can be mainly found in smaller ring sizes, which might be due to increasing contact of individual ring modules when ring sizes become smaller. Yet, this hypothesis can not be proven on the basis of the current data. Other possibilities for the formation of a ridge might be artifacts due to symmetrization, incomplete penetration of the stain and incomplete separation of different ring sizes, or genuine structural variability. Both possibilities, artifact formation and structural variability, would be substantiated by the very smooth ring-like appearance of the ridge which does not show clear structural details. The latter would imply an arbitrary assembly or disassembly of the ridge, possibly related to the protein function. The diameter of the rings appears to be in a linear relationship with the amount of ring modules, thus enabling the prediction of ring sizes. The latter might be beneficial for the determination of size-exclusion chromatography pore sizes during the purification process. The equal length of spikes and rails shown in Figure 46 indicates that both structures are constituted by similar structural features of Vipp1. In consequence, each ring module is defined as being comprised of two or a multiple of two Vipp1 monomers. The most likely alternative will be discussed in section D3.4.

3.2 Mutational analyses

3.2.1 C-terminal deletion

The ability of Vipp1 to oligomerize without its typical C-terminal extension confirms the results obtained by Aseeva et al. (2004) and Liu et al. (2007). None of the authors, however, specifically describe an increased occurrence of rods and double-rings in samples with C-terminally truncated Vipp1, indicating that the increase experienced in this work might have resulted from different sample concentrations as noted by Aseeva et al. (2004). Yet, the effect of sample concentration on rod formation has, to my knowledge, thus far only been postulated but not demonstrated. The C-terminal region might be important for Vipp1 protein function (Aseeva et al. 2007) and is presumably located on the surface of the protein

complex (ASEEVA ET AL. 2004). Thus, the formation of rods might be the result of an induced conformational change which is favored by the cleavage of the C-terminus. This hypothesis could be supported by speculations expressing the importance of Vipp1 rod-like structures for lipid trafficking (LIU ET AL. 2007; VOTHKNECHT ET AL. 2012).

Two-dimensional analysis of top-views and double-rings did not show apparent structural differences between ΔCT-Vipp1 and wtVipp1. Given a molecular mass of approximately 5.3 kDa for the cleaved section of a monomer, a 10-ring would loose 212 kDa (40 x 5.3 kDa) by the truncation of Vipp1. A protein density of this magnitude should be clearly discernible by negative staining TEM. However, due to the fact that missing densities are distributed among the whole assembly, it is comprehensible that ΔCT-Vipp1 and wtVipp1 appear to be indistinguishable in 2-D analysis. Three-dimensional analysis might provide further information as to the exact location of the C-terminal region, but is largely hindered by the aggregation of Vipp1 to rods. The box-like structures found in the sample do not resemble anything structurally described in the context of Vipp1 research thus far and when comparing these structures with Vipp1, no obvious connection can be established. Consequently, the identity and function of the box-like structures remains to be determined.

3.2.2 N-TERMINAL MUTATIONS

OTTERS ET AL. (2012) described the inability of Vipp1 to form rings upon deletion of helix α1. The current mutational analyses confirm these results and specify the interactions further. For instance, the oligomerization to rings is apparently neither compromised by numerous mutations in the first half of helix α1 nor upon deletion of the latter. In contrast, certain mutations in the second half of helix α1 lead to a significant decrease of Vipp1 oligomers. Furthermore, it could be shown for the first time that amino acids located outside of α-helical regions are essential for oligomerization as well.

The N12 mutant of Vipp1 displays peculiar bar-like structures which are of different lengths and show a zigzag pattern. The latter could be attributed to the previously described propensity of forming coiled coils (e.g. ASEEVA ET AL. 2004; BULTEMA ET AL. 2010). The different lengths cannot be interpreted as easily. Shorter bars might represent tilted views, whereas longer ones represent side views. Some class averages display the typical curvature of the Vipp1 rings that was visible in the 3-D analysis. Others are straight, possibly due to a 90° tilt along their longitudinal axis compared to the curved ones. Generally, it can be assumed that the bar-like structures are either dimeric or tetrameric Vipp1. The exact determination

of the oligomeric state, however, will require biochemical analyses.

The second N-terminal mutant showing structural effects, N11, contains little amounts of bar-like structures and increased amounts of small background density. This might indicate that oligomerization even to a dimeric or tetrameric state is entirely abolished. However, when observing the sample under the electron microscope, the sample concentration appeared to be very low. Hence, EM analysis of this mutation should be repeated to assure the results.

3.3 VIPP1-LIPOSOME INTERACTIONS

Vipp1 has been reported to be associated with the inner envelope and thylakoid membranes (e.g. LI ET AL. 1994). Furthermore, numerous studies suggest that oligomerization of Vipp1 and PspA beyond the tetrameric state is essential for the interaction of the protein with membranes and, thus, for protein function (e.g. JOLY ET AL. 2009; Aseeva et al. 2004; ASEEVA ET AL. 2007; OTTERS ET AL. 2012). For PspA, it has been suggested that the protein is able to form clathrin-like networks (STANDAR ET AL. 2008). However, the majority of the structures interpreted as network can be attributed to the striations created by slightly tilted spikes and rails of PspA rings. Consequently, these results will be regarded as misinterpretation and not considered for Vipp1.

The results obtained by negative staining electron microscopy indicate divergent affinities of Vipp1 to the lipids surveyed in this study (see Figure 54 - Figure 56). Apparently, the protein only binds to MGDG while being freely distributed in samples that did not contain the latter. Nevertheless, considering the alleged function of the protein as lipid transporter, Vipp1 rings not bound to the large lipid vesicles do not unequivocally prove the absence of lipid on the rings. It has to be considered that free rings might have already bound to the vesicles, extracted lipids and detached again. Moreover, the SQDG used in this study was found to show abnormal biochemical behavior and had to be replaced. Consequently, samples containing SQDG have to be discarded for interpretation. The different behavior of Vipp1 rings in combination with DOPG or DOPG+MGDG might also be the result of the different lipid composition. Finally, although the concentration of Vipp1 rings in protein samples did not seem to significantly decrease in comparison to the protein-lipid samples, the probability that Vipp1 rings partly disassemble to attach to the vesicles as monomeric, dimeric or tetrameric forms cannot be excluded either. Using immuno-gold labeling of the protein in combination with electron microscopy could assess this question. However, given

the previously demonstrated significance of Vipp1 rings for lipid binding and functionality, the disassembly of the former seems unlikely.

Judged by the electron microscopic images, the sizing procedure was unsuccessful for most lipids using the extrusion technique. Alternatively, lipids could be sonicated or produced in a French pressure cell to produce small, unilamellar liposomes (e.g. Papahadjopoulos and Miller 1967; Suurkuusk et al. 1976; Hamilton et al. 1980; Harris 1986). It has to be emphasized, however, that the unsuccessful sizing procedure might be attributed to negative staining, as it has been described to produce strong artifacts in liposome studies (e.g. Harris 1986).

In summary, the liposome studies conducted in this work have to be considered as preliminary results, since preparation techniques might have introduced significant changes to the lipids. Given previous work, it is unlikely that Vipp1 rings will disassemble for lipid binding and the results presented above do not show a decline of Vipp1 rings when adding lipids. In order to obtain more significant results, sonication of lipids should be tested for the creation of small, unilamellar vesicles and cryo-EM should be considered as more physiological alternative to negative staining.

3.4 Vipp1 – a hypothetical model

No crystal structure of Vipp1 or any closely related protein is available, which greatly limits structural interpretation at the current resolution. However, combining results of the present 3-D reconstructions, mutational analyses and published research, I was able to create a model that outlines the general spatial distribution of the Vipp1 structural elements. Predictions of the secondary structure elements of both Vipp1 and PspA have been described previously (e.g. Bultema et al. 2010) and are shown in Figure 91 as a reference point for the subsequent hypothetical model. Generally, PspA and Vipp1 show 5 and 6 α-helices, respectively, although other authors predict a slightly different secondary structure (Otters et al. 2012).

The definition of the ring module is the first step of the structural analysis, considering that it has been shown in Figure 43 that two possibilities exist to describe one such module. The most probable alternative was identified by displaying the density maps at a high contour level, where only few density remains visible. This method indicated that more structural interaction exists between spikes and rails of the first alternative compared to the second one. Rails and spikes show an equal average length of ~110 Å. Given the predominantly α-helical nature of Vipp1, the translation of the measured lengths into amino acid quantities was
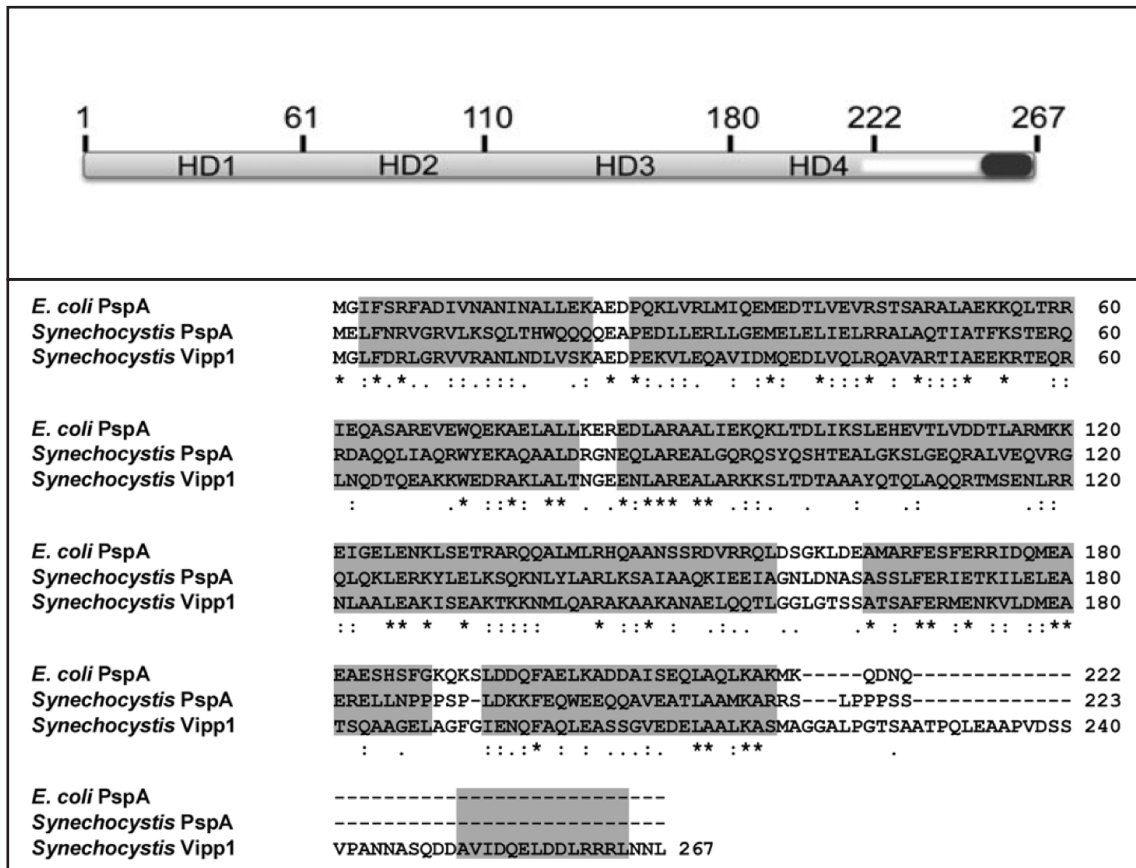
Figure 91 PspA and Vipp1 sequence structure. Top: Domain structure of PspA (HD1 to HD4) with the elongation seen in Vipp1 from amino acids 222 to 267 (reprinted from Vothknecht et al. 2012 by permission of Oxford University Press). Bottom: Multiple sequence alignment of PspA and Vipp1 showing predicted α-helical regions shaded in gray (reprinted from Bultema et al. 2010).

considered. Having a pitch of 5.4 Å and bearing 3.6 amino acids per turn, an α-helix with a length of 110 Å would translate into an amount of ~73 amino acids (e.g. Hyman et al. 1981). Indeed, helix α3 of both Vipp1 and PspA is comprised of 73 amino acids. Being an uninterrupted helix, α3 is thus a viable candidate for the most basal (towards the center of the ring) and the most apical (towards the outside of the ring) rail and spike densities, respectively, which show the mentioned length of ~110 Å.

It has been shown that ring modules are most likely comprised of a tetrameric assembly of a dimer of dimers (Fuhrmann et al. 2009a), which implies that each spike and each rail must contain a dimeric assembly of Vipp1. Since spatial restraints imposed by the density map would not allow a side-by-side assembly of a dimer, I propose a model describing two intertwined monomers. Furthermore, I assume a coiled coil interaction between helix α3 of two monomers, based on a coiled coil prediction for helix α3 using the NPS@ web interface (Combet et al. 2000) and the previously described propensity for coiled coils of Vipp1 (e.g. Fuhrmann et al. 2009a). Given their importance in nature (e.g. Oakley and Hollenbeck

2001), I chose an anti-parallel coiled coil, although it has to be emphasized that a parallel assembly would result in a similar spatial distribution of the secondary structure elements of Vipp1. The resulting anti-parallel coiled coil was placed at the most basal density of a rail, in order to continue the assembly of a hypothetical Vipp1 dimer (Figure 92 A).

A short, three amino acid linker and spatial constrains given by the density map restrict helix α2 to be situated apical of α3 (Figure 92 B). The same linker length imposes a position of helix α1 close to α2. Moreover, at a higher contour level, the density map shows bridging density between ring modules, which is situated at the longitudinal center of the modules. Finally, the mutational studies presented above and by OTTERS ET AL. (2012) have shown the importance of helix α1 for ring formation. In consideration of these facts, I placed helix α1 close to α2, bent by 180° in order to lie near the longitudinal center of a ring module and close to the bridging density between neighboring modules that resembles the inter-tetramer interface (Figure 92 C). An anti-parallel assembly of the α3 coiled coil would imply that the two helices α1 point to opposing directions. However, while the first α1 helix would lie close to the inter-tetramer contact zone, the second one would point away from it. Given the flexibility between individual helices due to the linker, I turned the second α1 helix by 180° against α2 which led to an approximation of the two α1 helices, both of them being situated at the inter-tetramer interface (Figure 92 C).

Spatial restrictions imposed by the density map and a linker length of seven amino acids between α3 and α4 indicate that the latter is positioned apical of α3, α1 and α2, facing the intra-tetramer interface (Figure 92 D). Helix α5 could remain in line with α4 or bend 180° to run anti-parallel to the latter, considering a linker length of 4 amino acids and the spatial constrains imposed by the density map. Interestingly, the H4 region of PspA (which encompasses approximately 1/3 of α4 and the entire α5 helix) has been suggested to mediate intra-tetramer contact (VOTHKNECHT ET AL. 2012 based on the findings of JOLY ET AL. 2009) which substantiates the hypothetical model. Due to the interaction between the α4 and α5 helices of four monomers, I decided to position helix α5 as an extension of α4 – rather than using an anti-parallel orientation – which, when observing the entire tetrameric assembly, more closely resembles the fact that only the last 8 amino acids of α4 are incorporated in the H4 region (Figure 92 E).

The results obtained by previous studies (ASEEVA ET AL. 2004; OTTERS ET AL. 2012) and the increased formation of rod-like assemblies in the ΔCT-Vipp1 sample imply that helix α6 is located near the surface of the protein. According to the current model, unfilled densities
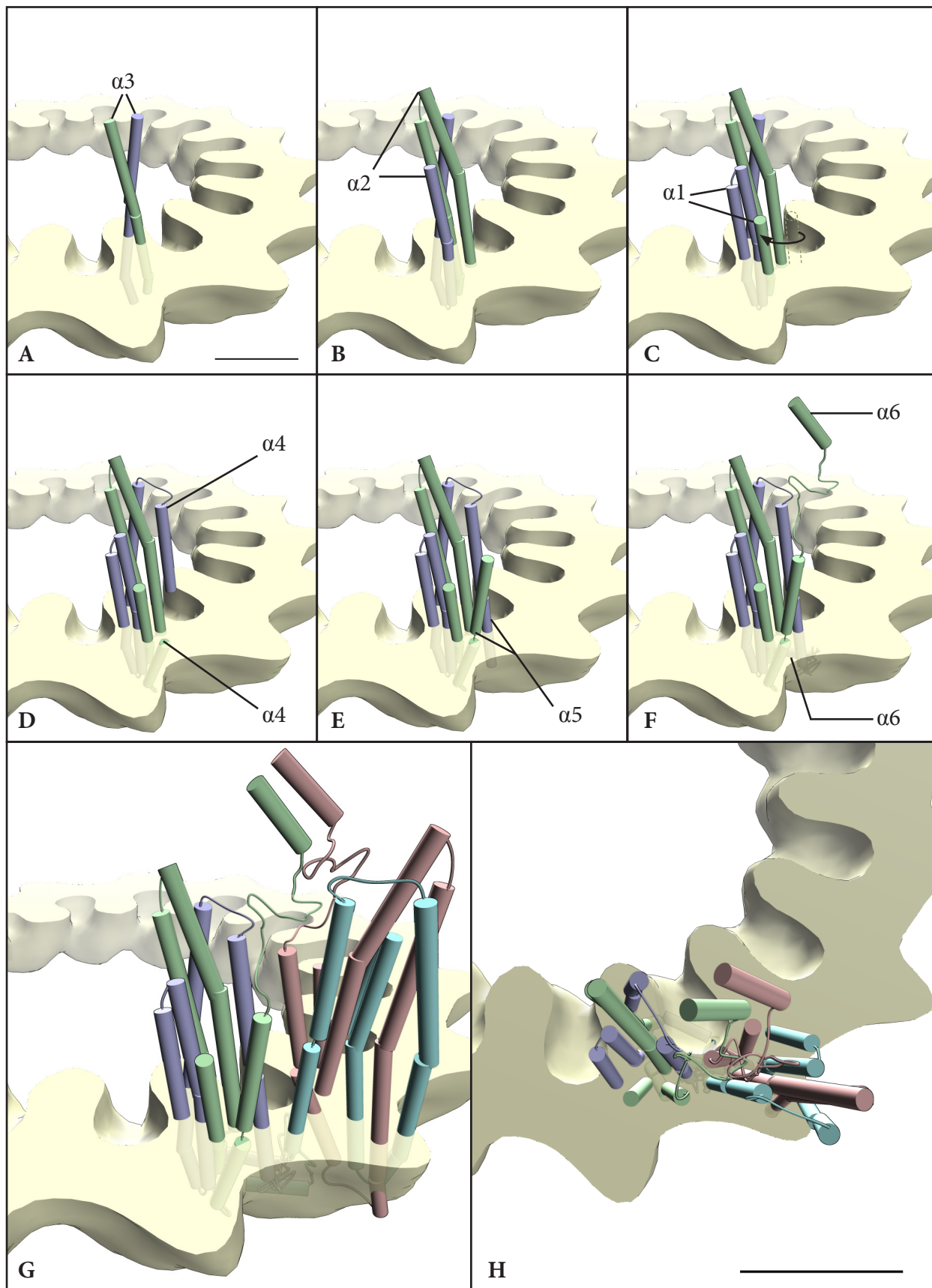
Figure 92 Hypothetical molecular model of a Vipp1 tetramer on the basis of the structural and mutational analyses and published data. One dimer is created by placing an antiparallel coiled coil of helices α3 into the density map (A) – the latter is clipped for clarity reasons – followed by helices α2 (B), α1 (C), α4 (D), α5 (E) and α6 (F). The dashed helix in C) denotes the original position of the green α1. The tetramer of a ring module is assembled by duplicating the dimer, rotating it by approximately 160° along its longitudinal axis and placing it into the spike by consecutive slight rotations and shifts (G). Note the convergence of helices α4 and α5 at the intra-tetramer interface. Helices α1 are all located near the inter-tetramer interface (H). Helices are depicted as pipes due to the hypothetical character of this model. For details see section D3.4. Scale bars: 50 Å.

above and below the latter might accommodate the four α6 helices (Figure 92 F). This assumption is substantiated by the lower height of PspA rings, which lack the C-terminal extension (Hankamer et al. 2004). Given a linker length of 30 amino acids between α5 and α6, the exact location of the α6 helices could be very variable. In line with these considerations, it might be possible that α6 is the source of the ridges seen in some reconstructions and that these ridges are formed as a part of the protein's function either at the top or the bottom of the protein. The observation of different structures for equal ring sizes throughout this project would concur with this hypothesis. Furthermore, the fact that ridges only occur at the user-defined bottom of the molecule is likely to be an artifact imposed by user interaction, since the ridges were always placed at the bottom manually before correcting the handedness of the reconstruction according to the pointing direction of the spikes. Hence, ridges at the top of the molecule might be possible as well. Therefore, Vipp1 might exist in different conformational states, where either no ridge is present or a constriction of the ring is achieved by the formation of a ridge at the top or the bottom of the ring. Similar conformational changes have been observed, for example, in GroEL which constricts the entry of its active site in the absence of substrate and ATP (e.g. Bigotti and Clarke 2008). The hypothesis of a conformationally variable constriction of Vipp1 rings is congruent with the excision of membrane portions by vesicle constriction, although the latter is a hypothesis previously associated with the different Vipp1 ring sizes (Fuhrmann et al. 2009a). Furthermore, conformational variability would explain the poor separation of ring sizes achieved by the multirefine algorithm.

After assembling the dimeric Vipp1, a tetramer can be formed by manually placing a copy of the dimer in the spike opposing the rail chosen for model building. In this case, the spike dimer copy was turned by approximately 160° along its longitudinal axis and placed into the spike by consecutive slight rotations and shifts. It has to be noted that the 30 amino acid linker and helix α6 were positioned separately for each monomer. Upon tetramer formation, the putative interaction between helices α4 and α5 of all four monomers becomes apparent. Furthermore, when imposing the symmetry of the density map upon the tetramer, the inter-pentamer interfaces show a putative contact between the four helices α1 of two tetramers. A simulated 15 Å density map of a 15-ring Vipp1 assembly that was calculated using the model shows expected differences, but coincides strikingly well with the density map obtained by single particle analysis (Figure 93). Re-projections of simulated density maps and "real" density maps of different ring sizes show very similar top views and slightly different side views (Figure 94). This similarity, however, is not surprising, as the hypothetical model was built to fit the density map.
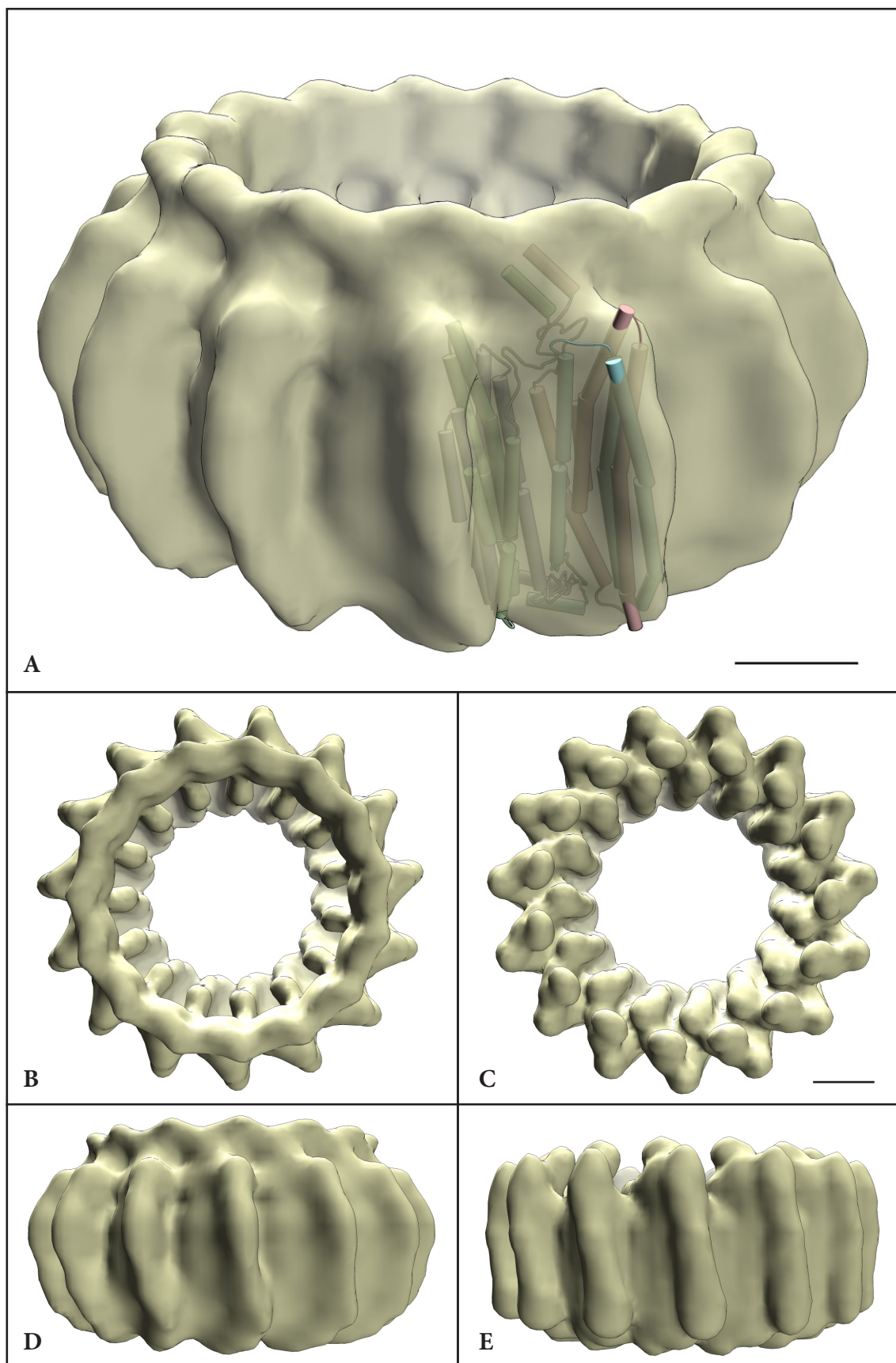
Figure 93 Fitting of the tetramer and comparison between calculated and simulated density map. The hypothetical tetramer fills the density of a ring module (A). B-E) Comparison of the "real" map (B+D) and the simulated map (C+E) of a 15-ring in top view and side view. Scale bars: 50 Å.

It has to be emphasized that this hypothetical model was created to show the spatial dimensions of the Vipp1 components inside the ring using the structural information available for Vipp1 and PspA. Yet, the exact distribution of the individual helices is still unknown and might be misinterpreted in the model. However, the interaction of α1 at the inter-tetramer and of α4 and α5 at the intra-tetramer interfaces, as well as the corresponding lengths between the density maps and helix α3 support the present model.

3.5 CONCLUDING REMARKS

By means of single particle analysis, I was able to confirm the variability of Vipp1 ring sizes and three-dimensionally reconstruct six differently sized Vipp1 rings from negatively stained images to a resolution between 20 to 30 Å. The reconstructions allow more accurate measurements of the molecule than had been possible by the previously published 2-D analyses. I could show that the width of rings increases linearly with the number of ring modules, enabling the prediction of ring width. Furthermore, the height of the molecules was determined to be approximately 150 Å which is only 2/3 of previously measured 220 Å (FUHRMANN ET AL. 2009A). Measurements conducted on 2-D class averages during this work also showed substantially higher values than the measurements derived from the density maps (~190 Å). Mutational analyses showed that the first half of helix α1 is not essential for ring formation, whereas mutations of certain amino acids on the second half and between helices α1 and α2 greatly diminished Vipp1 ring assemblies. Furthermore, 2-D analysis of C-terminally truncated Vipp1 images might indicate that the C-terminus controls the formation of rod-like structures. Preliminary results of electron microscopic images of Vipp1 and different lipids indicate that Vipp1 binds to lipids in the oligomeric ring assembly. As discussed above, however, the results will have to be substantiated by repeating the experiments and considering different preparative methods.

On the basis of the 3-D reconstructions, mutational analyses and published results, I created a hypothetical model of a Vipp1 tetrameric assembly. The model is based on a dimer of dimers according to FUHRMANN ET AL. (2009A). The coincidence of rail and spike lengths with the length of helix α3 indicates that this helix is situated in the most basal and apical positions of rails and spikes, respectively. The convergence of helix α1 of two tetramers at the inter-tetramer interface is in accordance with the mutational analyses and recently published biochemical results (OTTERS ET AL. 2012). The close proximity of helices α4 and α5 of one tetramer follows the results presented by JOLY ET AL. (2009). Due to the inclusion of 3-D
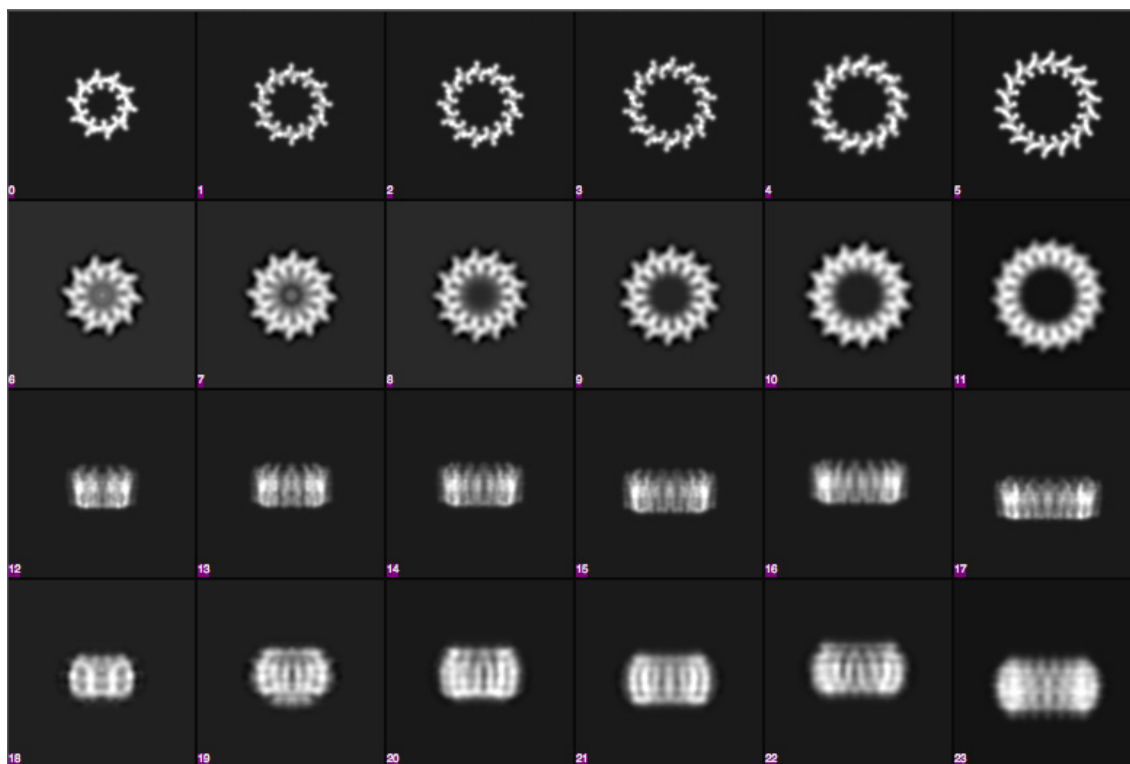
Figure 94 Comparison between re-projections of simulated and "real" density maps. Top view and side view re-projections of the simulated density maps are displayed in rows one and three, top view and side view re-projections of the "real" density maps are displayed in rows two and four, respectively. For a comparison of the above images with class averages resulting from 2-D analysis, see Figure 42.

structural data, this model should provide a more accurate description of the spatial distribution inside the Vipp1 tetramer than the model previously published by BULTEMA ET AL. (2010) and might enable the design of site-specific peptide antibodies that would increase the accuracy of Vipp1 detection for *in vivo* experiments. Immuno-gold labeling of different Vipp1 domains could further substantiate the model in future studies. Cryo-EM of Vipp1 will only be feasible, if rings can be narrowed down to one or two different sizes. If conformational variability exists, the separation of different states would complicate the acquisition of a sufficiently large dataset.

4 *Biomphalaria glabrata* acetylcholine-binding Protein

4.1 Evaluation of the 3-D reconstructions

The substantial differences between the density maps resulting from the traditional refinement and the refinement minimizing noise bias are striking (Figure 61 and Figure 65). The traditional refinement density map has a spiky over-all appearance, which indicates that the low-pass filter applied to the final reconstruction was set too high. However, since it was set to the resolution measured at the 0.5 threshold, it has to be assumed that the calculated FSC shows over-optimistic values. In combination with the fact that, for this refinement type, a final low-pass filter of 3 Å was used for the dataset during iterative, linear-correlation-based refinement, it is a reasonable assumption that the final reconstruction shows significant amounts of over-refinement. In contrast, the refinement minimizing noise bias used a dataset low-pass filtered to 8 Å for the alignment of particles throughout the entire process. Nevertheless, this value seems too high, considering that the spatial frequencies contributing to the alignment will most likely not exceed 10 Å (e.g. Scheres 2012b). The reason why 8 Å were chosen lies in the nature of the Gaussian low-pass filter, which implicates that the amplitudes of spatial frequencies are increasingly dampened towards higher frequencies, starting at the lowest one. In order to not substantially decrease the amplitudes at 10 Å, the slightly lower value of 8 Å was chosen. In light of these considerations, small amounts of noise bias might still be present after this refinement. In retrospective, a randomization of phases for spatial frequencies higher than 10 Å would be a superior approach to ensure that amplitudes at 10 Å are fully available for alignment while information beyond that threshold is randomized, ultimately minimizing noise bias when calculating a 'usefilt' refinement (see D1.2.1) in EMAN1.9 with prior CTF-correction.

Considering this, the aforementioned protrusions of the 3-D reconstructions (see C3.1.1 and C3.1.2) could be assumed to largely emerge as product of aligned noise. However, their appearance in the second refinement and a control refinement, as well as their early appearance in the 2-D analysis of the particle images and in the vicinity of a putative N-glycosylation site, suggests that N-glycane chains might also contribute to these densities (Figure 95). Indeed, 20 % of the molecular mass of BgAChBP is attributable to glycosylation (Saur et al. 2012). The absence of the protrusions in the structural analyses of recombinant BgAChBP1 further substantiates that the protrusions of the traditionally calculated density map result from a combination of noise bias and glycosylation of BgAChBP. Due to the pre-
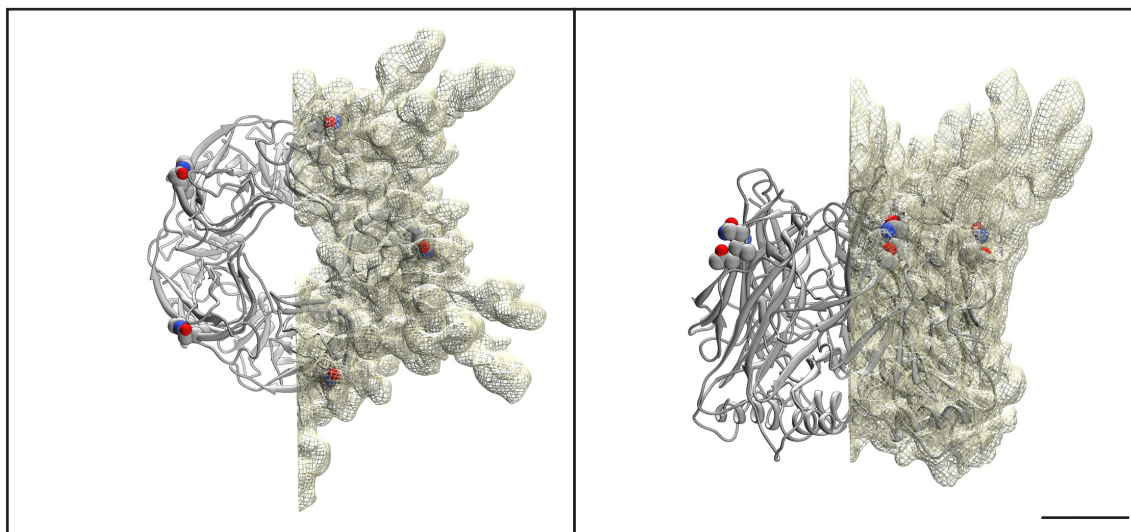
Figure 95 BgAChBP1 pentameric pseudo-atomic model, overlayed with a reconstructed pentamer of the traditional refinement. The putative N-glycosylation site (^{176}NAT^{178}) is colored by heteroatom and shows close proximity to the protrusions of the 3-D reconstruction that are not filled by the pseudo-atomic model. Note that the β-sheet slope comprising the site points towards the protrusions. Scale bar: 20 Å.

sumed flexibility of the sugar moieties, structural analysis is not feasible. Yet, the statistical overlap of the different conformations could result in undefined density as seen in the 2-D analyses, the control refinement and, lastly, the refinement minimizing noise bias.

As a final consideration, it was mentioned in the Results that helix α1 would have to be flexibly fitted to be entirely incorporated into the 3-D reconstruction. The need for this alteration of the pseudo-atomic model does not reflect a flawed density map. It is rather an expression of the flexibility of this helix that has already been described in the literature (e.g. CELIE ET AL. 2005; RUCKTOOA ET AL. 2009).

4.2 INTER-PENTAMER INTERFACES

The novel quaternary structure found for BgAChBP implicates the appearance of previously undescribed contact zones. As mentioned above, the inter-pentamer interfaces show an intriguing combination of amino acids which possibly interact with each other. The most peculiar of these interactions is the convergence of three phenylalanines (F^{71}) from three different pentamers which could theoretically form a hydrophobic pocket. A survey of the PDB databank, conducted by LANZAROTTI ET AL. (2011), showed that aromatic trimers in proteins are surprisingly common, with 44 % of all analyzed unique structures bearing at least one such trimer. Aromatic clusters are thought to be important for protein folding (e.g. SERRANO ET AL. 1991), protein-ligand interactions (e.g. AUERBACH 2013) and protein stability (e.g. BURLEY AND PETSKO 1985), amongst others. Especially the latter aspect is important for the

analysis of the inter-pentamer interface exhibited in dodecahedral BgAChBP. It was found that 80 % of a protein's aromatic residues are involved in networks that enhance protein stability, of which phenylalanine is the most abundant residue (Burley and Petsko 1985; Chourasia et al. 2001). Aromatic trimers have been shown to mostly adapt the lowest energy conformation already described for benzene clusters with an edge-to-face orientation (e.g. Oikawa et al. 1985; Singh and Thornton 1985; Burley and Petsko 1985; Lanzarotti et al. 2011). A similar orientation can be seen in our loop-refined model. The stabilization of the BgAChBP dodecahedron is a compelling option, although bond energies among aromatic residues tend to be much weaker than other bonding types (Burley and Petsko 1985). Disregarding the increased stability by the accumulation of 12 equal aromatic interfaces in a dodecahedron, each aromatic trimer might be individually stabilized by the ionic environment present in its surrounding. Polar residues (N^{69} and T^{72}), situated below and above the plane of the phenylalanine trimer, could shield the latter from the solvent (Figure 96 A). Furthermore, the aforementioned residues R^2, R^3, D^{25}, D^{26}, E^{70} and R^{73} could engage in ionic bonding (Sheinerman et al. 2000), which would result in the formation of two hydrophilic rings surrounding the trimer. Taken together, these polar and charged amino acids could form a hydrophilic cage-like structure that would strengthen the hydrophobic interaction between the three phenylalanines (Figure 96 B+C).

The 3-D reconstruction obtained of dodecahedral BgAChBP cannot unequivocally prove the convergence of the three phenylalanines, as the bond distance for these trimeric clusters is described to lie between 4.5 and 7 Å (Burley and Petsko 1985; Lanzarotti 2011) and because a phenylalanine ring has a diameter of just under 3 Å, both of which are not readily resolvable at 6 Å resolution. Nevertheless, the indirect convergence of the three F^{71} by rigid-body fitting and the spherical mass shown by the reconstruction at the position of convergence (see C3.1.3), affirm the postulated aromatic interaction.

The description of the inter-pentamer interfaces of dodecahedral BgAChBP also included cysteines C^{16} and C^{64}, which were characterized as a novel disulfide bond previously undescribed for AChBPs or AChRs. A disulfide bond in this position would tether the F^{71}-bearing MIR-loop to helix α1, thus decreasing the flexibility of the loop in favor of a more rigid positioning of F^{71}. Until know, however, the disulfide bridge has been a structural interpretation based on the close proximity of C^{16} and C^{64}. It would therefore be advisable to substantiate this bridge using biochemical analyses as described, for example, by Gorman et al. (2002) and Pérez et al. (2010).
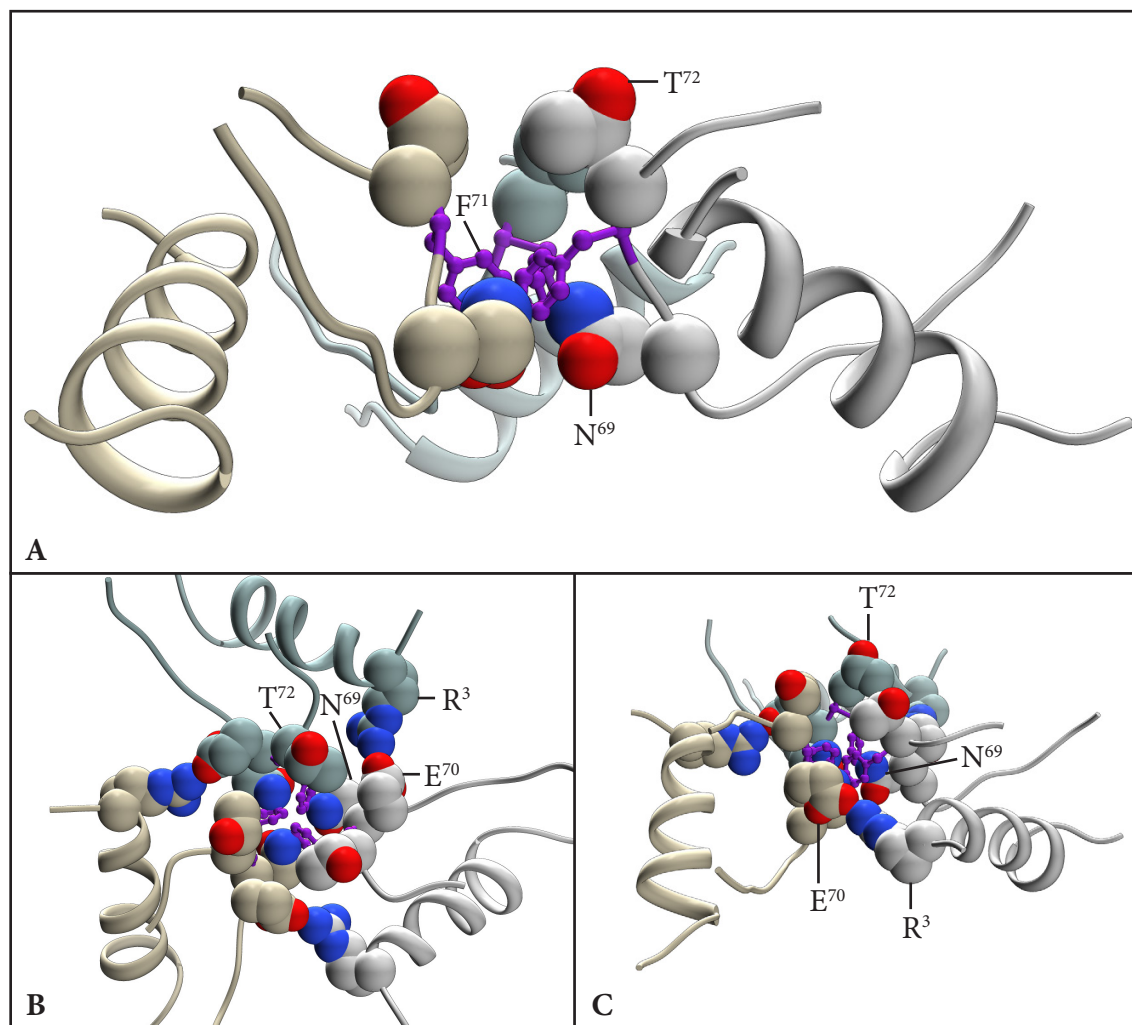
Figure 96 Aromatic trimer of the inter-pentamer interface surrounded by polar and charged amino acids. A) N^{69} and T^{72} are situated below and above the plane of the aromatic trimer. B) The first ring of charged amino acids surrounds the aromatic trimer and, together with N^{69} and T^{72}, builds a cage surrounding it (top view). C) same as B) in tilted view. The second ring of charged amino acids (R^2, D^{25}, D^{26}, R^{63}) is not shown. Phenylalanines are shown in purple.

In Saur et al. (2012), we describe the putative dodecahedral homo-oligomer constituted by BgAChBP type 2, in which the inter-pentamer interface is comprised of three converging cysteines (C^{71}) instead of F^{71}. Given the plausibility of an aromatic interface, the recent advances in the 3-D reconstruction of BgAChBP2 (see C3.3), the discovery of rectangular structures (as seen for recombinant BgAChBP2) in native BgAChBP samples and my unsuccessful literature research on cysteine trimers in proteins, this homo-oligomeric dodecahedral assembly of BgAChBP2 was intentionally omitted in this work. Following the same lines of evidence, a hetero-oligomeric assembly of BgAChBP1 and BgAChBP2 seems improbable. However, the results do not unequivocally disprove such an assembly.

4.3 INTRA-PENTAMER INTERFACES

Residues responsible for the preservation of a pentameric assembly are reportedly not well conserved among the members of pLGICs (BREJC ET AL. 2001). The variability between amino acids belonging to the plus or minus sides of BgAChBP1, BgAChBP2, *Lymnaea stagnalis* AChBP (LsAChBP) and the *Torpedo marmorata* AChR alpha subunit (TmAChR-α) is emphasized in Table 8. Although these contact regions have been shown to potentially be hydrophobic or charged, the four selected subunits show the tendency to build polar intra-pentamer interfaces (e.g. BREJC ET AL. 2001). While the amino acids of both BgAChBP subunits mostly coincide well with each other, their biochemical properties entirely contradict those of LsAChBP and TmAChR-α in several locations. Overall, the results obtained for BgAChBP are in concert with the observations of previous studies that described high variability and the importance of shape complementarity in this contact region (*ibid.*).

Table 8   Amino acids found at the intra-pentamer contact zone according to BREJC ET AL. (2001) and intriguing residues R^{42}, D^{48} and E^{46} (SAUR ET AL. 2012). Coloring is according to the amino acid properties at physiological pH: positive (blue), negative (red), polar (O and N: purple; O: green), hydrophobic (black), histidine (yellow). The rows between plus and minus site do not correspond structurally. Bg, *Biomphalaria glabrata*; Ls, *Lymnaea stagnalis*; Tm, *Torpedo marmorata*.

Plus side				Minus side			
BgACh-BP1	BgACh-BP2	LsAChBP	TmAChR-α	BgACh-BP1	BgACh-BP-2	LsAChBP	TmAChR-α
S^{17}	N^{17}	R^{15}	N^{16}	S^4	S^4	D^2	H^3
L^{19}	N^{19}	D^{17}	V^{18}	R^5	R^5	R^3	E^4
N^{20}	N^{20}	V^{18}	I^{17}	L^9	V^9	L^7	V^8
I^{21}	I^{21}	I^{19}	R^{18}	Q^{10}	K^{10}	Y^8	A^9
N^{47}	T^{47}	T^{45}	N^{47}	P^{80}	P^{80}	P^{77}	P^{81}
D^{88}	H^{88}	D^{85}	D^{89}	R^{171}	R^{171}	R^{170}	E^{175}
V^{95}	I^{95}	I^{92}	A^{96}	D^{103}	D^{103}	P^{100}	M^{105}
G^{96}	G^{96}	S^{93}	D^{97}	T^{105}	L^{105}	L^{102}	K^{107}
A^{97}	K^{97}	K^{94}	G^{98}	F^{41}	Y^{41}	L^{39}	I^{41}
S^{125}	S^{125}	S^{122}	Y^{127}	E^{99}	G^{99}	E^{96}	F^{100}
P^{127}	A^{127}	D^{124}	E^{129}	F^{101}	H^{101}	L^{98}	V^{103}
W^{146}	W^{146}	W^{143}	W^{149}	K^{121}	T^{121}	R^{118}	I^{123}
T^{147}	T^{147}	T^{144}	T^{150}	K^{39}	D^{39}	N^{37}	Q^{39}
H^{148}	Y^{148}	H^{145}	Y^{151}	W^{55}	W^{55}	W^{53}	R^{55}
S^{149}	D^{149}	H^{146}	D^{152}	T^{78}	T^{78}	S^{75}	R^{79}
E^{152}	D^{152}	E^{149}	K^{155}	S^{102}	S^{102}	T^{99}	H^{104}
Y^{186}	Y^{186}	Y^{185}	Y^{190}	R^{107}	R^{107}	R^{104}	L^{109}
P^{187}	P^{187}	S^{186}	T^{191}	D^{165}	T^{165}	Y^{164}	T^{169}
R^{42}	T^{42}	E^{40}	N^{42}	Y^{166}	Y^{166}	F^{165}	F^{170}
D^{48}	D^{48}	N^{46}	Q^{48}	E^{46}	D^{46}	I^{44}	V^{46}

4.4 Ligand-binding site

Comparing the ligand-binding pocket of BgAChBP1 and BgAChBP2 to other AChBPs and to the alpha subunit of *Torpedo marmorata* AChR, it becomes apparent that the principle face residues are very conserved, whereas the complementary face residues are variable (Table 9). This result is concurrent with recent publications (e.g. Brejc et al. 2001; Rucktooa et al. 2009; Blum et al. 2013). However, it is also apparent that both BgAChBP1 and BgAChBP2 show a substitution of very conserved tyrosines by phenylalanines at different positions. The substitution of Y^{89} with F^{89} in α7 AChR (position 92 in BgAChBP1) severely decreased the affinity for acetylcholine (Sine et al. 1994).

Generally, the principal face binds ligands *via* cation-π-interactions between charged nitrogens of the ligand and aromatic residues of the ligand-binding site. For aqueous solutions, this type of bonding was found to be stronger than salt bridges, which is the strongest non-covalent interaction type in air (Gallivan and Dougherty 2000; Lu et al. 2013a). Consequently, cation-π-interactions are wide-spread in biological processes, as for example the self-assembly of bio-macromolecules, molecular sensing, signal transduction, and the attachment of molluskan byssal threads to a substrate (e.g. Lin et al. 2007, Lu et al. 2013a, Lu et al. 2013b). The strength of these interactions is also found in AChBPs, were certain ligands bind with affinities in the nano-molar range (e.g. Celie et al. 2004; Hansen et al. 2004). Various other interactions were described for the ligand-binding site, such as dipole cation, hydrogen bond and van der Waals interactions (e.g. Smit et al. 2001; Rucktooa et al. 2009; Arias 2012). In general, ligands of AChBPs and AChRs must minimally provide a hydrogen bond acceptor (for the complementary face) and a charged nitrogen (for the principal face), which can be protonated or quaternary (Romanelli and Gualtieri 2003; Blum et al. 2013), although ligands lacking a nitrogen atom have also been described (e.g. cembr-

Table 9 Comparison of ligand-binding pocket amino acids between BgAChBP1 (Bg1), BgACHBP2 (Bg2), *Lymnaea stagnalis* AChBP (Ls), *Aplysia californica* AChBP (Ac), *Bulinus truncatus* AChBP (Bt) and *Torpedo marmorata* AChR subunit α (Tm). Amino acids were chosen according to Brejc et al. (2001). Strictly conserved amino acids are colored in red, mostly conserved ones (at least 4 of 6 equal) in green.

Principle face						Complementary face					
Bg1	*Bg2*	*Ls*	*Ac*	*Bt*	*Tm*	*Bg1*	*Bg2*	*Ls*	*Ac*	*Bt*	*Tm*
F^{92}	Y^{92}	Y^{89}	Y^{91}	Y^{88}	Y^{93}	W^{55}	W^{55}	W^{53}	Y^{53}	W^{51}	R^{55}
W^{146}	W^{146}	W^{143}	W^{145}	W^{142}	W^{149}	T^{57}	V^{57}	Q^{55}	Q^{55}	Q^{53}	R^{57}
Y^{186}	Y^{186}	Y^{185}	Y^{186}	Y^{184}	Y^{190}	R^{107}	R^{107}	R^{104}	V^{106}	V^{103}	L^{109}
$C^{188-189}$	$C^{188-189}$	$C^{187-188}$	$C^{188-189}$	$C^{186-187}$	$C^{192-193}$	S^{109}	Y^{109}	V^{106}	T^{108}	S^{105}	D^{111}
Y^{193}	F^{193}	Y^{192}	Y^{193}	Y^{191}	Y^{198}	T^{115}	T^{115}	L^{112}	M^{114}	I^{111}	M^{117}
-	-	-	-	-	-	V^{117}	V^{117}	M^{114}	I^{116}	V^{113}	T^{119}
-	-	-	-	-	-	Y^{166}	Y^{166}	F^{165}	Y^{166}	Y^{164}	F^{170}

anoids and lophotoxins; see Bourne et al. 2010). While the principle face tightly binds the ligand, the complementary face was shown to be important not only for the stabilization of the bond, but also for substrate affinity and selectivity (Rucktooa et al. 2009; Blum et al. 2013). Furthermore, studies indicate that residues outside of the ligand-binding site might be relevant for ligand binding as well (Celie et al. 2005). Lastly, ligands have been shown to switch their selectivity from $\alpha_2\beta_3$ subunits towards α7 subunits by introducing a single point-mutation (Hogg et al. 1999; Luo et al. 1999).

Resuming to the substitution of tyrosines by phenylalanines in both types of BgAChBP, the implications for the ligand-binding properties of BgAChBP cannot be foreseen in consideration of the complexity of the ligand-protein interaction described above. Predictions of ligand-protein interactions would at least require *in silico* docking studies, although the latter have been shown to predict interactions that were disproved biochemically (Blum et al. 2013). Biochemical analysis might involve methods such as affinity labeling, mutagenesis and the substituted cysteine accessibility method (Sixma and Smit 2003).

Conformational changes in AChBP upon ligand binding have been described as locally confined to loop C, interpreting the AChBP conformation as a desensitized or a resting state of AChR (Sixma and Smit 2003; Hansen et al. 2005). The conformational change of loop C is described as a closed state upon agonist binding, whereas binding an antagonist results in an open and a resting state in an intermediate state (e.g. Hansen et al. 2005; Mukhtasimova et al. 2009). Comparing the density map of the refinement minimizing noise bias with different AChBP subunits indicates a position between the intermediate and the closed state (Figure 97). Moreover, the occupation state of the ligand-binding site cannot be determined at 6 Å resolution, since density from the neighboring amino acids is projected into the binding pocket, thus concealing any putative ligand. In consequence, a better resolved structure would be essential to determine the conformational state of loop C and the occupation of the ligand-binding-site.

4.5 Recombinant BgAChBP1

Negative staining

The results of the refinements in IMAGIC-5 and EMAN1.9 affirmed the assumption that BgAChBP1 is able to assemble into dodecahedral structures. The differently distributed interface densities are most likely to be the result of different low-pass filtering, which we found was able to create an IMAGIC-5-type density map from an EMAN1.9-type one. Due to
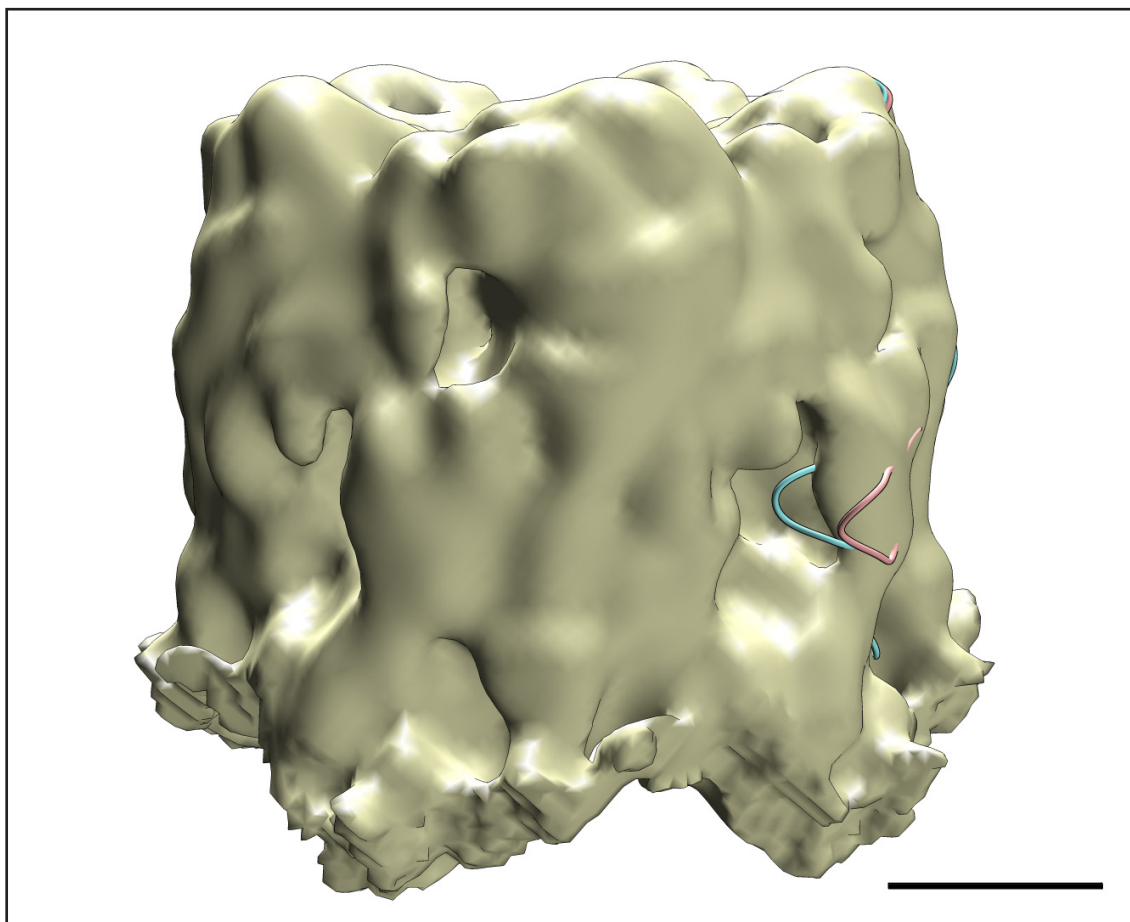
Figure 97 NatBgAChBP pentamer excised from the final density map of the refinement minimizing noise bias, with two fitted AcAChBP monomers. The conformational position of loop C appears to be in an intermediate state between apo-AcAChBP (pink; PDB code 2BYN) and closed state HEPES-AcAChBP (blue; PDB code 2BR7). Scale bar: 20 Å.

the limited resolution and the possible artifacts of negative staining, differences in interface density distribution should generally not be interpreted. Missing densities at the locations of helix α1 can be attributed to the incomplete penetration of stain into the protein, which is a common problem in negative staining EM (e.g. BREMER ET AL. 1992; HARRIS AND HORNE 1994). Consequently, the image processing of the negatively stained dataset of BgAChBP1 can be seen as a successful proof of the dodecahedral assembly.

Cryo-EM

The cryo-EM micrographs clearly showed a reappearance of the aggregations already seen in the first negatively stained images and that could be previously prevented by adding 50 mM imidazole to the sample. Arguably, the aggregations might have been prevented by an increased imidazole concentration. However, due to the fact that the sample should remain in a close-to-physiological buffer environment, we decided not to pursue this strategy. We as-

sumed that the aggregations were evoked by an interaction between the C-terminal His-tags of a subunit, since a dodecahedral assembly displays 360 histidines on its surface and protein aggregation due to the interaction of His-tags is a known problem (e.g. Hansen et al. 2004; Carson et al. 2007). As described in section C3.2, the introduction of a cleavable His-tag resulted in positive preliminary results and will most likely reduce the aggregation of single particles.

The density maps obtained after refining the cryo-EM dataset in EMAN1.9 and IMAGIC-5 display considerable differences that can be only attributed to image processing, since the same dataset was used for both refinements (see B6.2; Braukmann 2012). The density map calculated by EMAN1.9 has a very spiky appearance which, as discussed above, indicates over-refinement of the structure. Furthermore, loop C must be visible at a claimed resolution of 8 Å. The general lack of AChBP-specific structural features is a strong indication that the refinement of this structure was not successful. In contrast, the density map calculated by IMAGIC-5 clearly shows loop C and the general outline of a BgAChBP monomer. Yet, the inter-pentamer contact zone is structured significantly different than the one described for native BgAChBP. Moreover, the reconstruction does not show any attributable density for helix α 1 which, at a claimed resolution of 8 Å, should be visible as a barrel-like structure similar to the ones characterized for native BgAChBP. Hence, both refinements of BgAChBP1 at their current states are highly questionable. In my opinion, the difficulties experienced with this dataset are manifold. Firstly, due to the concentration of the particles at the micrograph edges, Thon rings were sparsely visible during CTF-correction, introducing errors into the dataset due to inaccuracies in CTF-determination. Secondly, the aggregation of the proteins has led to an erroneous alignment of the dataset, an assumption validated by the sudden increase in resolution as a consequence of 2-D masking for alignment in IMAGIC-5. Thirdly, the dataset comprised only 2200 particles, which is very little even considering the icosahedral symmetry of the particle. The erroneous determination of particle orientations has an especially high detrimental impact on icosahedral structures, as applying a 60-fold symmetry might lead to a summation of alignment errors. Hence, for now, the difference of the inter-pentamer contact region between native BgAChBP and BgAChBP1 calculated by IMAGIC-5 will be not considered further, as it might be an artifact resulting from the unfavorable circumstances described above. The possibility that the inconsistencies of the inter-pentamer contact region may be attributed to different low-pass filtering values – as described for the negatively stained dataset – can be generally discarded regarding the cryo-

EM dataset, since this effect was only experienced while using low-frequency filter values. Similarly, possible reference bias has been excluded by calculating control refinements using different references.

4.6 Recombinant BgAChBP2

The inability of BgAChBP2 to assemble into dodecahedral quaternary structures, the occurrence of elongated structures of dipentameric length and the observation of these elongated structures in native BgAChBP samples, substantiated our assumption that both types of BgAChBP were situated in different quaternary assemblies with potentially diverging functionalities. The subsequently obtained 3-D reconstruction is in concert with the hypothesis presented in Saur et al. (2012) that BgAChBP2 could assemble into a dipentamer by forming five disulfide bridges between the ten C^{71} residues. In addition to the putative disulfide bridges, the contact zones between the ten alpha-helices are rich in possibilities for the formation of salt bridges. The necessary adjustment of the pixel size could be the result of an "impure" dataset. The difficulty to distinguish tilted views from side-views might have led to the inclusion of the former into the dataset. Due to the confinement of tilt angles during refinement, any tilted views present in the dataset were inserted into the 3-D reconstruction as side views, potentially leading to a shortening of the reconstruction due to their similar width but shorter length when compared to side views.

For future structural studies, the selection of larger datasets could allow the incorporation of tilted views, as the alignment and reconstruction should be more stable. The inability to incorporate top views into the datasets – due to the possibility of selecting pentameric structures – implies the necessity of improving sample preparation to increase the sparsely observed side views. With a molecular mass of 250 kDa cryo-EM may be considered, although it must be emphasized that the small size of the protein combined with a preferential selection of side views (which show less density in micrographs than top views) might result in substantial difficulties concerning image acquisition and image processing.

4.7 Possible biological functions of BgAChBP

As mentioned in the Introduction, several different biological functions of BgAChBP have been proposed. The simultaneous occurrence of two AChBP-like molecules has been described for *Bulinus truncatus* and the abalone *Haliotis discus hannai* (Celie et al. 2005; Huang et al. 2009). The former, however, have not been sufficiently characterized for fur-

ther consideration. The latter share 41 % sequence identity and a conservation of the characteristic residues of pentameric LGICs. Yet, spatial and ontogenetic expression analyses, as well as sequence comparisons to known AChBPs and the ACCBP of *Pinctada fucata* showed that the two proteins most likely have different functions *in vivo* (*ibid.*). It was finally hypothesized that the first protein – which more closely resembles AChBP – could have a neuronal function, whereas the second protein – which more closely resembles ACCBP – could have a function related to shell nacre growth (*ibid.*). The proteins were hence called HdhAChBP and HdhACCBP (*ibid.*).

The two types of BgAChBP could theoretically relate similarly. They share 52 % sequence identity and show a phylogenetic position and branching similar to HdhAChBP and HdhACCBP (Saur et al. 2012). The sequence of BgAChBP1 bears one putative N-glycosylation site and does not contain an N-terminal cysteine for dimerization into dipentameric assemblies. Furthermore, HdhAChBP also contains residue F^{71} which was discussed to be important for dodecahedral assembly (see D4.2). However, the putative disulfide bridge $C^{16} \leftrightarrow C^{64}$ is not present in HdhAChBP. In contrast, BgAChBP2 does contain an unpaired cysteine which is not positioned N-terminally – as in other ACCBPs – but located at the N-face, and which has been shown to lie in a spatially favorable position for disulfide bonding in a dipentameric quaternary structure by means of single particle analysis. However, BgAChBP2 only bears one putative N-glycosylation site, which contrasts results of multiple such sites in other ACCBPs. The ability of native BgAChBP to bind amorphous calcium carbonate showed a high-affinity and a low-affinity component (Saur et al. 2012). The latter was attributed to BgAChBP2, due to the appearance of subunit dimers in non-reduced SDS-PAGE of this sample, thus defining the high-affinity component as BgAChBP1. These results would contradict the hypothesis that BgAChBP1 resembles AChBP in functionality. If both types of BgAChBP would show an ACCBP-like functionality, the occurrence of two different types could be explained by a different spatial distribution inside the animal. Hemolymph is known to be supersaturated with $CaCO_3$ (Huang et al. 2009). Hence, in order to maintain stability and prevent the $CaCO_3$ from precipitation, ACCBP-like factors must be distributed in the hemolymph as well. High concentrations of ACCBP inside the hemolymph, however, would further increase the colloid osmotic pressure inside the animal, which could be unfavorable for an organism living in a freshwater environment. It has been mentioned that the benefit of large quaternary assemblies might be a lower colloid osmotic pressure in comparison to an equal mass of loosely distributed subunits (e.g. Mangum and Johansen 1975; Mangum

1980; Markl et al. 1991). The dodecahedral assembly would be in accordance with this claim. Moreover, the high affinity of BgAChBP1 for amorphous calcium carbonate might implicate a lower essential protein concentration, additionally lowering the colloid osmotic pressure. In contrast, the colloid osmotic pressure in the extra-pallial fluid (EPF) is not of importance, being a region outside the animal. Hence, the second type of BgAChBP could be distributed in the EPF. Its lower affinity for amorphous calcium carbonate could be of importance during shell nacre formation, considering Ma et al. (2007) who showed that ACCBP is not incorporated into normally grown nacre lamellae. However, the low concentration of BgAChBP in the hemolymph (5 % according to Lieb et al. 2006), does not support the argument concerning the colloid osmotic pressure. Furthermore, electron microscopic analyses of the hemolymph and the EPF have shown that the BgAChBP quaternary assemblies do not differ between he two samples.

Other than exhibiting AChBP- or ACCBP-like functions, BgAChBP could be important for the inactivation of phycotoxins, as it has been shown that AChBP binds Gymnodimines and Spirolides – diatom toxins – in picomolar affinities (Bourne et al. 2010). Feeding strategies of pulmonate snails indeed show a substantial ingestion of diatoms, which may carry toxic products (e.g. Thomas et al. 1985; Madsen 1992; Brown and Piscopo 2011). Moreover, pulmonate diet may also include blue-green algae which, however, are typically avoided by the animals (Madsen 1992). Nevertheless, freshwater ponds may be subjected to cyanobacterial blooms accompanied by an extensive production of toxins, such as the cholinergic agonist anatoxin (e.g. Romanelli and Gualtieri 2003). A dodecahedron is the most efficient spatial distribution of pentamers, thus leading to a maximized number of ligand-binding sites on a minimized surface area.

In summary, the biological functions of the two BgAChBP types are yet to be investigated. However, the advantages of a dodecahedral assembly were highlighted concerning the optimized presentation of active sites. A difference in functionality between BgAChBP1 and BgAChBP2 might be indicated by the substitution of the strictly conserved residue D^{88} with H^{88} in BgAChBP2, which was identified to be important in receptor activation (Lee and Sine 2004), and the substitutions of active site BgAChBP1 F^{92} and Y^{193} by BgAChBP2 Y^{92} and F^{193}.

4.8 Concluding Remarks

Based on cryo-EM images of native BgAChBP, I calculated a 6 Å 3-D reconstruction showing a previously undescribed dodecahedral assembly of AChBP. The assembly of a pseudo-atomic model enabled the interpretation of the inter-pentamer interfaces and the analysis of the intra-pentamer contact zones and the ligand-binding site compared to other AChBPs.

Single particle analysis of recombinantely expressed BgAChBP1 and BgAChBP2 led to preliminary results that show a dodecahedral assembly of the former and a dipentameric assembly of the latter, indicating divergent functions of the AChBP types. Cryo-EM analysis of BgAChBP1 was accompanied by aggregation of particles which imposed strong limitations to the 3-D reconstructions. However, the assembly to dodecahedral particles could be demonstrated at physiological pH values. The exchange of the C-terminal His-tag for an N-terminal TEV-protease-cleavable His-tag shows promising results and is currently optimized in our laboratory. Once a satisfactory sample preparation is achieved, the exploitation of the inherent icosahedral symmetry should allow high-resolution reconstructions reaching near-atomic resolution, if imaging conditions and image processing algorithms are chosen correctly. A confirmation of the $C^{16}\leftrightarrow C^{64}$ disulfide bridge by methods such as mass spectrometry (e.g. Sun et al. 1996) is required, since the occurrence of the former is currently a hypothesis based on the structural proximity of the cysteines. Additionally, an examination of the hypotheses concerning the importance of $C^{16}\leftrightarrow C^{64}$ and F^{71} for the formation of inter-pentamer interfaces in a dodecahedron could be achieved by site-directed mutagenesis. Cryo-EM of BgAChBP2 is possible in theory. However, for a 250 kDa complex, I would rather recommend to optimize sample preparation for X-ray crystallographic studies. Again, mutagenesis could provide indications whether C^{71} is essential for the formation of dipentameric assemblies or not, and consequently provide even stronger indications for the hypothesis of two homo-oligomeric BgAChBP1 and BgAChBP2 quaternary structures.

As to the functional relevance of BgAChBP1 and BgAChBP2 in *Biomphalaria glabrata*, published hypotheses concerning the general function of AChBPs and ACCBPs have been analyzed and discussed, especially in consideration of the advantages a dodecahedral assembly might provide concerning the optimized spatial presentation of active sites. Ultimately, the biological function of the two BgAChBP types requires agonist and antagonist binding studies, as well as spatial and temporal expression analyses. Due to the absence of post-translational glycosylation in *E. coli*, analysis of ACCBP activity might only be feasible by initiating eukaryotic expression.

The insights into BgAChBP provided in this work might be relevant for the structural analysis of AChBP and AChR ligand-binding pockets, as they establish the possibility to visualize different protein-ligand interactions of AChBP/AChR in a physiological environment for the first time. Furthermore, putative industrial relevance concerning the protein family's affinity to toxins should be examined, as the dodecahedron with its 60 densely packed active sites might provide a suitable framework for toxin inactivation or detection. The detection of toxins by AChBPs has already been proposed by Bourne et al. (2010) as a substitution for a currently used mouse bioassay.

Summary

Three-dimensional electron microscopy (3-D EM) provides a framework for the analysis of large protein quaternary structures. The advantage over the generally higher resolving method of X-ray crystallography is the embedding of the proteins in their physiological environment. However, results of the two methods can be combined to obtain superior structural information. In this work, three different protein types – (i) Myriapod hemocyanin, (ii) vesicle-inducing protein in plastids 1 (Vipp1) and (iii) acetylcholine-binding protein (AChBP) – were structurally analyzed by 2-D and 3-D EM and, where possible, functionally interpreted.

Myriapod hemocyanins have been previously shown to be 6x6-meric assemblies that, in case of *Scutigera coleoptrata* hemocyanin (ScoHc), show two 3x6-mer planes whith a staggering angle of approximately 60°. Here, previously observed structural differences between oxy- and deoxy-ScoHc could be substantiated. A 4° rotation between hexamers of two different 3x6-mer planes was measured, which originates at the most central inter-hexamer interface. Further information about allosteric behaviour in myriapod hemocyanin was gained by analyzing *Polydesmus angustus* hemocyanin (PanHc), which shows a stable 3x6-mer and divergent histidine patterns in the inter-hexamer interfaces when compared to ScoHc. Both findings would conclusively explain the very different oxygen binding properties of chilopod and diplopod hemocyanin.

Vipp1 is a protein found in cyanobacteria and higher plants which is essential for thylakoid membrane function and forms highly variable ring-shaped structures. In the course of this study, the first 3-D analysis of Vipp1 was conducted and yielded reconstructions of six differently sized Vipp1 rings from negatively stained images at resolutions between 20 to 30 Å. Furthermore, mutational analyses identified specific N-terminal amino acids that are essential for ring formation. On the basis of these analyses and previously published results, a hypothetical model of the Vipp1 tertiary and quaternary structure was generated.

AChBP is a water-soluble protein in the hemolymph of mollusks. It is a structural and functional homologue of the ligand-binding domain of nicotinic acetylcholine receptors. For the freshwater snail *Biomphalaria glabrata*, we previously described two types of AChBP (BgAChBP1 and BgAChBP2). In this work, a 6 Å 3-D reconstruction of native BgAChBP is presented, which shows a dodecahedral assembly that is unprecedented for an AChBP. Single particle analysis of recombinantely expressed BgAChBP types led to preliminary results showing a dodecahedral assembly of BgAChBP1 and a dipentameric assembly of BgAChBP2. This indicates   divergent biological functions of the two types.

Zusammenfassung

Die dreidimensionale Elektronenmikroskopie (3-D EM) ermöglicht die strukturelle Analyse großer Proteinquartärstrukturen. Gegenüber der grundsätzlich höher auflösenden Röntgenkristallographie bietet diese Methode den Vorteil, die zu untersuchenden Proteinkomplexe unter physiologischen Pufferbedingungen abbilden zu können. Beide Methoden können kombiniert werden, um verbesserte Strukturinformationen zu erhalten. In der vorliegenden Arbeit wurden drei unterschiedliche Proteintypen – (i) Myriapodenhämocyanin, (ii) vesikelinduzierendes Protein in Plastiden 1 (Vipp1) und (iii) Acetylcholin-bindendes Protein (AChBP) – mittels 2-D und 3-D EM untersucht und nach Möglichkeit einer funktionellen Analyse unterzogen.

Myriapodenhämocyanin wurde als 6x6-mere Struktur beschrieben, die – im Falle des *Scutigera coleoptrata* Hämocyanins (ScoHc) – aus zwei um ca. 60° gegeneinander verdrehten 3x6-meren Ebenen besteht. Diese Arbeit konnte zuvor beschriebene strukturelle Unterschiede zwischen oxy- und deoxy-ScoHc bestätigen und zeigt eine 4°-Rotation zwischen zwei Hexameren unterschiedlicher 3x6-mer-Ebenen auf, welche durch eine zentral gelegene histidinreiche, inter-hexamere Kontaktstelle eingeleitet zu werden scheint. Die strukturelle Analyse des *Polydesmus angustus* Hämocyanins (PanHc) lieferte weitere Hinweise auf die allosterische Regulation in Myriapodenhämocyaninen, da dieses als stabiles 3x6-mer vorliegt und, verglichen mit ScoHc, Unterschiede in der Histidinverteilung innerhalb der inter-hexameren Kontaktstellen aufweist.

Vipp1 wird in Cyanobakterien und höheren Pflanzen exprimiert, ist essentiell für den Aufbau der Thylakoidmembranen und bildet sehr heterogene, ringartige Strukturen aus. Dreidimensionale Analysen von Vipp1 wurden bisher allerdings nicht durchgeführt. Im Zuge dieser Arbeit war es möglich, mittels 3-D EM aus negativ-kontrastierten Proben sechs unterschiedlich große Vipp1-Ringe mit Auflösungen zwischen 20 und 30 Å zu rekonstruieren. Desweiteren konnten mittels Mutagenesestudien spezifische N-terminale Aminosäuren identifiziert werden, die für die Ringbildung essentiell zu sein scheinen. Auf der Grundlage dieser und zuvor publizierter Ergebnisse, wurde ein hypothetisches Modell der Tertiär- und Quartärstruktur von Vipp1 erstellt.

AChBP ist ein wasserlösliches Protein in der Hämolymphe von Mollusken. Es ist strukturell und funktionell homolog zu der Ligandenbindedomäne der nikotinischen Acetylcholinrezeptoren. Die Süßwasserschnecke *Biomphalaria glabrata* exprimiert zwei verschiedene AChBP-Typen (BgAChBP1 und BgAChBP2). In dieser Arbeit wurde eine 3-D Rekonstruktion des nativen BgAChBPs mit einer Auflösung von 6 Å erstellt, welche die für AChBP zuvor nicht bekannte Struktur eines pentagonalen Dodekaeders zeigt. Einzelpartikelanalyse von rekombinant exprimiertem BgAChBP1 und BgAChBP2 ergab eine Dodekaederstruktur für BgAChBP1 und eine Dipentamerstruktur für BgAChBP2. Dies deutet auf eine unterschiedliche biologische Funktion der beiden Isoformen hin.

Bibliography

Adrian, M., Dubochet, J., Lepault, J. & McDowall, A.W., 1984. Cryo-electron microscopy of viruses. *Nature*, 308(5954), pp.32–36.

Arias, H.R., 2012. Molecular Interactions between Ligands and Nicotinic Acetylcholine Receptors Revealed by Studies with Acetylcholine Binding Proteins. *Journal of Thermodynamics & Catalysis*, 03(04).

Aseeva, E., Ossenbühl, F., Eichacker, L.A., Wanner, G., Soll, J.U.R. & Vothknecht, U.C., 2004. Complex formation of Vipp1 depends on its alpha-helical PspA-like domain. *J Biol Chem*, 279(34), pp.35535–35541.

Aseeva, E., Ossenbühl, F., Sippel, C., et al., 2007. Vipp1 is required for basic thylakoid membrane formation but not for the assembly of thylakoid protein complexes. *Plant physiology and biochemistry : PPB / Société française de physiologie végétale*, 45(2), pp.119–128.

Auerbach, A., 2013. The Energy and Work of a Ligand-Gated Ion Channel. *Journal of molecular biology*, pp.1–15. Available at: http://www.sciencedirect.com/science/article/pii/S0022283613000430.

Bai, X.-C., Fernandez, I.S., McMullan, G. & Scheres, S.H., 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife*, 2, p.e00461.

Baldwin, P.R. & Penczek, P.A., 2007. The Transform Class in SPARX and EMAN2. *Journal of Structural Biology*, 157(1), pp.250–261.

Banks, G., Kemenes, I., Schofield, M., O'Shea, M. & Korneev, S.A., 2009. Acetylcholine binding protein of mollusks is unlikely to act as a regulator of cholinergic neurotransmission at neurite-neurite synaptic sites in vivo. *Journal of Structural Biology*, 23(9), pp.3030–3036.

Beintema, J.J., Stam, W.T., Hazes, B. & Smidt, M.P., 1994. Evolution of Arthropod Hemocyanins and Insect Storage Proteins (Hexamerins). *Molecular biology and evolution*, 11(3), pp.493–503.

Bellon, P.L., Cantele, F. & Lanzavecchia, S., 2001. Correspondence analysis of sinogram lines. Sinogram trajectories in factor space replace raw images in the orientation of projections of macromolecular assemblies. *Ultramicroscopy*, 87(4), pp.187–197.

Benning, C., 2009. Mechanisms of lipid transport involved in organelle biogenesis in plant cells. *Annual review of cell and developmental biology*, 25, pp.71–91.

Bigotti, M.G. & Clarke, A.R., 2008. Chaperonins: The hunt for the Group II mechanism. *Archives of Biochemistry and Biophysics*, 474(2), pp.331–339.

Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C. & Sarma, V.R., 1965. Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. *Nature*, 206(4986), pp.757–761.

Blum, A.P., Van Arnam, E.B., German, L.A., Lester, H.A. & Dougherty, D.A., 2013. Binding Interactions with the Complementary Subunit of Nicotinic Receptors. *Journal of Biological Chemistry*, 288(10), pp.6991–6997.

Bock, H.-H., 2008. Origins and extensions of the-means algorithm in cluster analysis. *Journal Électronique d'Histoire des Probabilités et de la Statistique* [electronic only], 4(2).

Bonnet, N., 2004. Some trends in microscope image processing. *Micron*, 35(8), pp.635–653.

Borland, L. & van Heel, M., 1990. Classification of image data in conjugate representation spaces. *J opt Soc Am*, A7, pp.601–610.

Bourne, Y., Radic, Z., Ar a oz, R.O.M., et al., 2010. Structural determinants in phycotoxins and AChBP conferring high affinity binding and nicotinic AChR antagonism. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13), pp.6076–6081.

Bouzat, C., Gumilar, F., Spitzmaul, G., et al., 2004. Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel. *Nature*, 430(7002), pp.896–900.

Braukmann, S., 2012. *3D-Elektronenmikroskopie des rekombinanten Acetylcholin-Bindeproteins aus der Tellerschnecke Biomphalaria glabrata*. JGU Mainz, Germany.

Brejc, K., van Dijk, W.J., Klaassen, R.V., et al., 2001. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*, 411(6835), pp.269–276. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11357122.

Bremer, A., Henn, C., Engel, A., Baumeister, W. & Aebi, U., 1992. Has negative staining still a place in biomacromolecular electron microscopy? *Ultramicroscopy*, 46(1), pp.85–111.

Brenner, S. & Horne, R.W., 1959. A negative staining method for high resolution electron microscopy of viruses. *Biochimica et biophysica acta*, 34, pp.103–110.

Brissette, J.L., Russel, M., Weiner, L. & Model, P., 1990. Phage shock protein, a stress protein of *Escherichia coli. Proceedings of the National Academy of Sciences*, 87(3), pp.862–866.

Brown, E.R. & Piscopo, S., 2011. Ion channels in key marine invertebrates; their diversity and potential for applications in biotechnology. *Biotechnology Advances*, 29(5), pp.457–467.

Brusca, R.C. & Brusca, G.J., 1990. Invertebrates. *Sinauer, Sunderland, Mass*, 264, pp.435–482.

Budinger, T.F. & Glaeser, R.M., 1976. Measurement of focus and spherical aberration of an electron microscope objective lens. *Ultramicroscopy*, 2, pp.31–41.

Bultema, J.B., Fuhrmann, E., Boekema, E.J. & Schneider, D., 2010. Vipp1 and PspA: Related but not twins. *Communicative & integrative biology*, 3(2), pp.162–165.

Burley, S.K. & Petsko, G.A., 1985. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science*, 229(4), pp.23–28.

Burmester, T., 2001. Molecular Evolution of the Arthropod Hemocyanin Superfamily. *Molecular biology and evolution*, 18(2), pp.184–195.

Burmester, T., 2002. Origin and evolution of arthropod hemocyanins and related proteins. *J. Comp. Physiol. [B]*, 172, pp.95–117.

Carson, M., Johnson, D.H., McDonald, H., Brouillette, C. & DeLucas, L.J., 2007. His-tag impact on structure. *Acta Crystallographica Section D*, 63(3), pp.295–301.

Caspar, D.L., 1966. An analogue for negative staining. *Journal of molecular biology*, 15(1), pp.365–371.

Celie, P.H.N., Klaassen, R.V., van Rossum-Fikkert, S.E., et al., 2005. Crystal Structure of Acetylcholine-binding Protein from *Bulinus truncatus* Reveals the Conserved Structural Scaffold and Sites of Variation in Nicotinic Acetylcholine Receptors. *Journal of Biological Chemistry*, 280(28), pp.26457–26466.

Celie, P.H.N., van Rossum-Fikkert, S.E., van Dijk, W.J., Brejc, K., Smit, A.B. & Sixma, T.K., 2004. Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron*, 41(6), pp.907–914.

Chourasia, M., Sastry, G.M. & Sastry, G.N., 2011. Aromatic-Aromatic Interactions Database, A(2)ID: an analysis of aromatic π-networks in proteins. *International journal of biological macromolecules*, 48(4), pp.540–552.

Combet, C., Blanchet, C., Geourjon, C. & Deléage, G., 2000. NPS@: network protein sequence analysis. *Trends Biochem Sci*, 25(3), pp.147–150.

Conway, J.F. & Steven, A.C., 1999. Methods for Reconstructing Density Maps of "Single" Particles from Cryoelectron Micrographs to Subnanometer Resolution. *Journal of Structural Biology*, 118, pp.106–118.

Corringer, P.-J., Poitevin, F., Prevost, M.S., Sauguet, L., Delarue, M. & Changeux, J.-P., 2012. Structure and Pharmacology of Pentameric Receptor Channels: From Bacteria to Brain. *Structure*, 20(6), pp.941–956.

Crowther, R.A., 1971. Procedures for Three-Dimensional Reconstruction of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 261(837), pp.221–230.

Crowther, R.A., DeRosier, D.J. & Klug, A., 1970. The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 317(1530), pp.319–340.

Danev, R. & Nagayama, K., 2001. Transmission electron microscopy with Zernike phase plate. *Ultramicroscopy*, 88(4), pp.243–252.

Davis, I.W., Leaver-Fay, A., Chen, V.B., et al., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research*, 35(Web Server issue), pp.W375–83.

De Rosier, D.J. & Klug, A., 1968. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*, 217(5124), pp.130–134.

Decker, H., 1986. Nesting - An extension of the allosteric model and its application to tarantula hemocyanin. In B. Linzen, ed. *Invertebrate oxygen carriers*. Berlin, Heidelberg, New York: Springer, pp. 383–388.

Decker, H., Hellmann, N., Jaenicke, E., Lieb, B., Meissner, U. & Markl, J., 2007. Minireview: Recent progress in hemocyanin research. *Integr Comp Biol*, 47(4), pp.631–644.

DeLisa, M.P., Lee, P., Palmer, T. & Georgiou, G., 2004. Phage Shock Protein PspA of *Escherichia coli* Relieves Saturation of Protein Export via the Tat Pathway. *Journal of Bacteriology*, 186(2), pp.366–373.

Downing, K.H. & Glaeser, R.M., 2008. Restoration of weak phase-contrast images recorded with a high degree of defocus: the "twin image" problem associated with CTF correction. *Ultramicroscopy*, 108(9), pp.921–928.

Dubochet, J., Adrian, M., Chang, J.-J., et al., 1988. Cryo-electron microscopy of vitrified specimens. *Quarterly reviews of biophysics*, 21(2), pp.129–228.

DeRosier, D.J. & Klug, A., 1972. Structure of the tubular variants of the head of bacteriophage T4 (Polyheads): I. Arrangement of subunits in some classes of polyheads. *Journal of molecular biology*, 65(3), pp.469–488.

Erni, R., 2010. Aberration-Corrected Imaging in Transmission Electron Microscopy, London: Imperial College Press.

Fahrenbach, W.H., 1970. The cyanoblast: hemocyanin formation in *Limulus polyphemus*. *Journal of Cell Biology*, 44(2), pp.445–453.

Fernandez Moran, H., Luque, D., Caston, J.R. & Carrascosa, J.L., 2008. Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of Structural Biology*, 164(1), pp.170–175.

Fiser, A., Do, R.K. & Sali, A., 2000. Modeling of loops in protein structures. *Protein science: a publication of the Protein Society*, 9(9), pp.1753–1773.

Frank, J., 1975. Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy*, 1(2), pp.159–162.

Frank, J. & Al-Ali, L., 1975. Signal-to-noise ratio of electron micrographs obtained by cross correlation. *Nature*, 256, pp.376–379.

Frank, J., Goldfarb, W., Eisenberg, D. & Baker, T.S., 1978. Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy*, 3(3), pp.283–290.

Freedman, T.B., Loehr, J.S. & Loehr, T.M., 1976. A resonance Raman study of the copper protein,

hemocyanin. New evidence for the structure of the oxygen-binding site. *J Am Chem Soc*, 98(10), pp.2809–2815.

Fuhrmann, E., Bultema, J.B., Kahmann, U., Rupprecht, E., Boekema, E.J. & Schneider, D., 2009a. The Vesicle-inducing Protein 1 from *Synechocystis* sp. PCC 6803 Organizes into Diverse Higher-Ordered Ring Structures. *Molecular Biology of the Cell*, 20, pp.4620–4628.

Fuhrmann, E., Gathmann, S., Rupprecht, E., Golecki, J. & Schneider, D., 2009b. Thylakoid membrane reduction affects the photosystem stoichiometry in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant physiology*, 149(2), pp.735–744.

Gallivan, J.P. & Dougherty, D.A., 2000. A Computational Study of Cation–π Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering. *J Am Chem Soc*, 122(5), pp.870–874.

Gaykema, W., Hol, W., Vereijken, J.M., Soeter, N.M., Bak, H.J. & Beintema, J.J., 1984. 3.2 Å structure of the copper-containing, oxygen-carrying protein *Panulirus interruptus* haemocyanin. *Nature*, 309(5963), pp.23–29.

Gebauer, W. & Markl, J., 1999. Identification of four distinct subunit types in the unique 6x6 hemocyanin of the centipede *Scutigera coleoptrata*. *Naturwissenschaften*, 86(9), pp.445–447.

Giribet, G. & Edgecombe, G.D., 2012. Reevaluating the Arthropod Tree of Life. *Annu. Rev. Entomol.*, 57(1), pp.167–186.

Gonen, T., Sliz, P., Kistler, J., Cheng, Y. & Walz, T., 2004. Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature*, 429(6988), pp.193–197.

Gorman, J.J., Wallis, T.P. & Pitt, J.J., 2002. Protein disulfide bond determination by mass spectrometry. *Mass Spectrometry Reviews*, 21(3), pp.183–216.

Grant, T., 2007. *Advances in single particle electron microscopy*. Imperial College, London, UK.

Grigorieff, N., 2007. FREALIGN: High-resolution refinement of single particle structures. *Journal of Structural Biology*, 157, pp.117–125.

Grigorieff, N., 2000. Resolution measurement in structures derived from single particles. *Acta Crystallographica Section D*, 56(10), pp.1270–1277.

Grigorieff, N., 1998. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. *Journal of molecular biology*, 277(5), pp.1033–1046.

Hall, C.E., 1955. Electron densitometry of stained virus particles. *The Journal of biophysical and biochemical cytology*, 1(1), pp.1–12.

Hamilton, R.L., Goerke, J., Guo, L.S., Williams, M.C. & Havel, R.J., 1980. Unilamellar liposomes made with the French pressure cell: a simple preparative and semiquantitative technique. *Journal of lipid research*, 21(8), pp.981–992.

Hankamer, B.D., Elderkin, S.L., Buck, M. & Nield, J., 2004. Organization of the AAA(+) adaptor protein PspA is an oligomeric ring. *J Biol Chem*, 279(10), pp.8862–8866.

Hansen, S.B., Sulzenbacher, G., Huxford, T., Marchot, P., Taylor, P. & Bourne, Y., 2005. Structures of *Aplysia* AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *The EMBO Journal*, 24(20), pp.3635–3646.

Hansen, S.B., Talley, T.T., Radic, Z. & Taylor, P., 2004. Structural and ligand recognition characteristics of an acetylcholine-binding protein from *Aplysia californica*. *The Journal of biological chemistry*, 279(23), pp.24197–24202.

Harauz, G. & Ottensmeyer, F.P., 1984. Direct three-dimensional reconstruction for macromolecular complexes from electron micrographs. *Ultramicroscopy*, 12(4), pp.309–319.

Harauz, G. & van Heel, M., 1986. Exact filters for general geometry three dimensional reconstruction. *Optik*, 73, pp.146–156.

Harauz, G., Borland, L., Bahr, G.F., Zeitler, E. & Heel, M., 1987. Three-dimensional reconstruction of a human metaphase chromosome from electron micrographs. *Chromosoma*, 95(5), pp.366–374.

Harris, J.R., 1986. A comparative negative staining study of aqueous suspensions of sphingomyelin. *Micron and Microscopica Acta*, 17(2), pp.175–200.

Harris, J.R. & Agutter, P., 1970. A negative staining study of human erythrocyte ghosts and rat liver nuclear membranes. *Journal of ultrastructure research*, 33(3), pp.219–232.

Harris, J.R. & Horne, R.W., 1994. Negative Staining - a Brief Assessment of Current Technical Benefits, Limitations and Future Possibilities. *Micron*, 25(1), pp.5–13.

Hazes, B., Kalk, K.H., Hol, W.G.J., et al., 1993. Crystal structure of deoxygenated *Limulus polyphemus* subunit II hemocyanin at 2.18 Å resolution: Clues for a mechanism for allosteric regulation. Protein Science, 2(4), pp.597–619. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21844650.

Henderson, R. & Unwin, P.N., 1975. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*, 257(5521), pp.28–32.

Henderson, R., Chen, S., Chen, J.Z., et al., 2011. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *Journal of molecular biology*, 413(5), pp.1028–1046.

Henderson, R., Sali, A., Baker, M.L., et al., 2012. Outcome of the First Electron Microscopy Validation Task Force Meeting. *Structure*, 20(2), pp.205–214.

Heymann, J.B., Chagoyen, M. & Belnap, D.M., 2005. Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *Journal of Structural Biology*, 151(2), pp.196–207.

Hogg, R.C., Miranda, L.P., Craik, D.J., Lewis, R.J., Alewood, P.F. & Adams, D.J., 1999. Single amino acid substitutions in alpha-conotoxin PnIA shift selectivity for subtypes of the mammalian neuronal nicotinic acetylcholine receptor. *Journal of biological chemistry*, 274(51), pp.36559–36564.

Huang, F., Fulda, S., Hagemann, M. & Norling, B., 2006. Proteomic screening of salt-stress-induced changes in plasma membranes of *Synechocystis* sp. strain PCC 6803. *Proteomics*, 6(3), pp.910–920.

Huang, J., Wang, H., Cui, Y., et al., 2009. Identification and comparison of amorphous calcium carbonate-binding protein and acetylcholine-binding protein in the abalone, *Haliotis discus hannai*. *Marine biotechnology*, 11(5), pp.596–607.

Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J. & Kim, W., 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature*, 413(6852), pp.154–157.

Hyman, J.M., McLaughlin, D.W. & Scott, A.C., 1981. On Davydov's alpha-helix solitons. *Physica D: Nonlinear Phenomena*, 3(1-2), pp.23–44.

Ilan, E., Hammel, I., David, M.M. & Daniel, E., 1986. Erythrocruorin from the aquatic snail *Helisoma trivolvis*. Quaternary structure and arrangement of subunits. *Biochemistry*, 25(21), pp.6551–6554.

Im, C.S. & Grossman, A.R., 2002. Identification and regulation of high light-induced genes in *Chlamydomonas reinhardtii*. *The Plant Journal*, 30(3), pp.301–313.

Jaenicke, E., Decker, H., Gebauer, W., Markl, J. & Burmester, T., 1999. Identification, Structure, and Properties of Hemocyanins from Diplopod Myriapoda. *Journal of biological chemistry,* 274(41), pp.29071–29074.

Joly, N., Burrows, P.C., Engl, C., Jovanovic, G. & Buck, M., 2009. A Lower-Order Oligomer Form of Phage Shock Protein A (PspA) Stably Associates with the Hexameric AAA+ Transcription Activator Protein PspF for Negative Regulation. *Journal of molecular biology*, 394(4), pp.764–775.

Joyeux, L. & Penczek, P.A., 2002. Efficiency of 2D alignment methods. *Ultramicroscopy*, 92(2), pp.33–46.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. & Phillips, D.C., 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610), pp.662–666.

Kleerebezem, M. & Tommassen, J., 1993. Expression of the pspA gene stimulates efficient protein export in *Escherichia coli*. *Molecular Microbiology*, 7(6), pp.947–956.

Kobayashi, R., Suzuki, T. & Yoshida, M., 2007. *Escherichia coli* phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. *Molecular Microbiology*, 66(1), pp.100–109.

Kroll, D., Meierhoff, K., Bechtold, N., et al., 2001. VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. *PNAS*, 98(7), pp.4238–4242.

Kucukelbir, A., Sigworth, F.J. & Tagare, H.D., 2012. A Bayesian adaptive basis algorithm for single particle reconstruction. *Journal of Structural Biology*, 179(1), pp.56–67.

Kusche, K. & Burmester, T., 2001. Diplopod hemocyanin sequence and the phylogenetic position of the Myriapoda. *Molecular biology and evolution*, 18(8), pp.1566–1573.

Kusche, K., Hembach, A., Hagner-Holler, S., Gebauer, W. & Burmester, T., 2003. Complete subunit sequences, structure and evolution of the 6 x 6-mer hemocyanin from the common house centipede, *Scutigera coleoptrata*. *Eur. J. Biochem.*, 270, pp.2860–2868.

Lander, G.C., Stagg, S.M., Voss, N.R., et al., 2009. Appion: An integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology*, 166(1), pp.95–102.

Langley, J.N., 1905. On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. *The Journal of physiology*, 33(4-5), pp.374–413.

Lanzarotti, E., Biekofsky, R.R., Estrin, D.A., Marti, M.A. & Turjanski, A.G., 2011. Aromatic–Aromatic Interactions in Proteins: Beyond the Dimer. *Journal of Chemical Information and Modeling*, 51(7), pp.1623–1633.

Larkin, M.A., Blackshields, G., Brown, N.P., et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp.2947–2948.

Lee, W.Y. & Sine, S.M., 2004. Invariant aspartic Acid in muscle nicotinic receptor contributes selectively to the kinetics of agonist binding. *The Journal of general physiology*, 124(5), pp.555–567.

Li, H.M., Kaneko, Y. & Keegstra, K., 1994. Molecular cloning of a chloroplastic protein associated with both the envelope and thylakoid membranes. *Plant Molecular Biology*, 25(4), pp.619–632.

Lieb, B., Dimitrova, K., Kang, H.-S., et al., 2006. Red blood with blue-blood ancestry: intriguing structure of a snail hemoglobin. *PNAS*, 103(32), pp.12011–12016.

Lin, Q., Gourdon, D., Sun, C., et al., 2007. Adhesion mechanisms of the mussel foot proteins mfp-1 and mfp-3. *PNAS*, 104(10), pp.3782–3786.

Liu, C., Willmund, F., Golecki, J.R., et al., 2007. The chloroplast HSP70B-CDJ2-CGE1 chaperones catalyse assembly and disassembly of VIPP1 oligomers in *Chlamydomonas*. *The Plant Journal*, 50(2), pp.265–277.

Liu, C., Willmund, F., Whitelegge, J.P., et al., 2005. J-domain protein CDJ2 and HSP70B are a plastidic chaperone pair that interacts with vesicle-inducing protein in plastids 1. *Molecular Biology of the Cell*, 16(3), pp.1165–1177.

Lu, Q., Danner, E., Waite, J.H., Israelachvili, J.N., Zeng, H. & Hwang, D.S., 2013a. Adhesion of mussel foot proteins to different substrate surfaces. *Journal of The Royal Society Interface*, 10(79).

Lu, Q., Oh, D.X., Lee, Y., Jho, Y., Hwang, D.S. & Zeng, H., 2013b. Nanomechanics of Cation–π Interactions in Aqueous Solution. *Angewandte Chemie-International Edition*, 52(14), pp.3944–3948.

Ludtke, S.J., Baldwin, P.R. & Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology*, 128(1), pp.82–97.

Luo, S., Nguyen, T.A., Cartier, G.E., Olivera, B.M., Yoshikami, D. & McIntosh, J.M., 1999. Single-residue alteration in alpha-conotoxin PnIA switches its nAChR subtype selectivity. *Biochemistry*, 38(44), pp.14542–14548.

Ma, Z., Huang, J., Sun, J., et al., 2007. A novel extrapallial fluid protein controls the morphology of nacre lamellae in the pearl oyster, *Pinctada fucata*. *The Journal of biological chemistry*, 282(32), pp.23253–23263.

Madsen, H., 1992. Food selection by freshwater snails in the Gezira irrigation canals, Sudan. *Hydrobiologia*, 228(3), pp.203–217.

Magnus, K.A., Hazes, B., Ton-That, H., et al., 1994. Crystallographic analysis of oxygenated and deoxygenated states of arthropod hemocyanin shows unusual differences. *Proteins*, 19(4), pp.302–309.

Majeran, W., Cai, Y., Sun, Q. & van Wijk, K.J., 2005. Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. *The Plant cell*, 17(11), pp.3111–3140.

Majorovits, E., Barton, B., Schultheiss, K., Pérez-Willard, F., Gerthsen, D. & Schröder, R.R., 2007. Optimizing phase contrast in transmission electron microscopy with an electrostatic (Boersch) phase plate. *Ultramicroscopy*, 107(2–3), pp.213–226.

Mangum, C.P., 1980. Respiratory Function of the Hemocyanins. *Integr Comp Biol*, 20(1), pp.19–38.

Mangum, C.P. & Johansen, K., 1975. The colloid osmotic pressures of invertebrate body fluids. *Journal of Experimental Biology*, 63(3), pp.661–671.

Mangum, C.P., Scott, J.L., Black, R.E., Miller, K.I. & van Holde, K.E., 1985. Centipedal hemocyanin: its structure and its implications for arthropod phylogeny. *PNAS*, 82(11), pp.3721–3725.

Markl, J., 2013. Evolution of molluscan hemocyanin structures. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. In press.

Markl, J. & Decker, H., 1992. Molecular structure of the Arthropod Hemocyanins. In C. H. P. Mangum, ed. Blood and Tissue Oxygen Carriers. Advances in Comparative and Environmental Physiology. Berlin: Springer Verlag, pp. 325–376.

Markl, J., Hofer, A., Bauer, G., et al., 1979. Subunit heterogeneity in arthropod hemocyanins: II. Crustacea. *Journal of comparative physiology*, 133(3), pp.167–175.

Markl, J., Moeller, A., Martin, A.G., Rheinbay, J., Gebauer, W. & Depoix, F., 2009. 10 Å cryoEM structure and molecular model of the Myriapod (*Scutigera*) 6x6mer hemocyanin: understanding a giant oxygen transport protein. *Journal of molecular biology*, 392(2), pp.362–380.

Markl, J., Savel-Niemann, A., Wegener-Strake, A., et al., 1991. The role of two distinct subunit types in the architecture of keyhole limpet hemocyanin (KLH). *Naturwissenschaften*, 78(11), pp.512–514.

Martin, A.G., Depoix, F., Stohr, M., et al., 2007. *Limulus polyphemus* hemocyanin: 10 Å cryo-EM structure, sequence analysis, molecular modelling and rigid-body fitting reveal the interfaces between the eight hexamers. *Journal of molecular biology*, 366(4), pp.1332–1350.

McCormack, T., Petrovich, R.M., Mercier, K.A., et al., 2010. Identification and functional characterization of a novel acetylcholine-binding protein from the marine annelid *Capitella teleta*. *Biochemistry*, 49(10), pp.2279–2287.

Mindell, J.A. & Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology*, 142(3), pp.334–347.

Miyazawa, A., Fujiyoshi, Y. & Unwin, N., 2003. Structure and gating mechanism of the acetylcholine receptor pore. *Nature*, 423(6943), pp.949–955.

Monod, J., Wyman, J. & Changeux, J.P., 1965. On the nature of allosteric transitions: a plausible model. *J Mol Biol*, 12, pp.88–118.

Möller, V., 2012. Charakterisierung und Expression zweier multimerer Hämolympheproteine der Schnecke *Biomphalaria glabrata*. PhD thesis. JGU Mainz, Germany.

Mukhtasimova, N., Lee, W.Y., Wang, H.-L. & Sine, S.M., 2009. Detection and trapping of intermediate states priming nicotinic receptor channel opening. *Nature*, 459(7245), pp.451–454.

Neumann, R., 2013. *3D Strukturanalyse von Hämocyanin aus Polydesmus angustus*. Bachelor thesis. JGU Mainz, Germany.

Oakley, M.G. & Hollenbeck, J.J., 2001. The design of antiparallel coiled coils. *Curr Opin Struct Biol*, 11(4), pp.450–457.

Ohi, M., Li, Y., Cheng, Y. & Walz, T., 2004. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biological procedures online*, 6, pp.23–34.

Oikawa, S., Tsuda, M., Kato, H. & Urabe, T., 1985. Growth mechanism of benzene clusters and crystalline benzene. *Acta Crystallographica Section B: Structural Science*, 41(6), pp.437–445.

Otters, S., Braun, P., Hubner, J., Wanner, G., Vothknecht, U.C. & Chigri, F., 2012. The first alpha-helical domain of the vesicle-inducing protein in plastids 1 promotes oligomerization and lipid binding. *Planta*, 237(2), pp.529–540.

Papahadjopoulos, D. & Miller, N., 1967. Phospholipid model membranes. I. Structural characteristics of hydrated liquid crystals. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 135(4), pp.624–638.

Paterson, D. & Nordberg, A., 2000. Neuronal nicotinic receptors in the human brain. *Progress in Neurobiology*, 61(1), pp.75–111.

Penczek, P.A.P., Grassucci, R.A.R. & Frank, J.J., 1994. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3), pp.251–270.

Pettersen, E.F., Goddard, T.D., Huang, C.C., et al., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), pp.1605–1612.

Pérez, V.I., Pierce, A., de Waal, E.M., et al., 2010. Chapter 8 - Detection and Quantification of Protein Disulfides in Biological Tissues: A Fluorescence-Based Proteomic Approach. In E. C. A. L. Packer, ed. *Thiol Redox Transitions in Cell Signaling, Part A: Chemistry and Biochemistry of Low Molecular Weight and Protein Thiols*. Methods in Enzymology. Academic Press, pp. 161–177.

Pintilie, G., Zhang, J., Chiu, W. & Gossard, D., 2009. Identifying components in 3D density maps of protein nanomachines by multi-scale segmentation. In Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH IS - SN - VO . pp. 44–47.

Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W. & Gossard, D.C., 2010. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *Journal of Structural Biology*, 170(3), pp.427–438.

RA, C., 2004. Viruses and the Development of Quantitative Biological Electron Microscopy. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)*, 56(5), pp.239–248.

Rajulu, G.S., 1969. Presence of haemocyanin in the blood of a centipede *Scutigera longicornis*. *Curr Sci Bangalore*, 7, pp. 168-169.

Ramundo, S., Rahire, M., Schaad, O. & Rochaix, J.-D., 2013. Repression of essential chloroplast genes reveals new signaling pathways and regulatory feedback loops in *chlamydomonas*. *The Plant cell*, 25(1), pp.167–186.

Romanelli, M.N. & Gualtieri, F., 2003. Cholinergic nicotinic receptors: competitive ligands, allosteric modulators, and their potential applications. *Medicinal research reviews*, 23(4), pp.393–426.

Rosenthal, P.B. & Henderson, R., 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of molecular biology*, 333(4), pp.721–745. Available at: http://linkinghub.elsevier.com/retrieve/pii/ S0022283603010222.

Rucktooa, P., Smit, A.B. & Sixma, T.K., 2009. Insight in nAChR subtype selectivity from AChBP crystal structures. *Biochemical Pharmacology*, 78(7), pp.777–787.

Sali, A. & Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3), pp.779–815.

Saur, M., 2009. *Scutigera coleoptrata hemocyanin: three-dimensional reconstruction and comparison of two physiological states*. Diploma thesis. Mainz: JGU Mainz, Germany.

Saur, M., Moeller, V., Kapetanopoulos, K., et al., 2012. Acetylcholine-binding protein in the hemolymph of the planorbid snail *Biomphalaria glabrata* is a pentagonal dodecahedron (60 subunits). *PloS one*, 7(8), p.e43685.

Saxton, W.O. & Frank, J., 1977. Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy*, 2(2-3), pp.219–227.

Schatz, M., 1992. *Invariante Klassifizierung elektronenmikroskopischer Aufnahmen von eiseingebetteten biologischen Makromolekülen*. Egelsbach, Köln, New York: Verlag Hänsel-Hohenhauser.

Scheres, S.H.W., 2012a. A Bayesian view on cryo-EM structure determination. *Journal of molecular biology*, 415(2), pp.406–418.

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

Scheres, S.H.W., Valle, M., Nu n ez, R., et al., 2005. Maximum-likelihood multi-reference refinement for electron microscopy images. *Journal of molecular biology*, 348(1), pp.139–149.

Scheres, S.H.W. & Chen, S.S., 2012. Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, 9(9), pp.853–854.

Serrano, L., Bycroft, M. & Fersht, A.R., 1991. Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *Journal of molecular biology*, 218(2), pp.465–475.

Sheinerman, F.B., Norel, R. & Honig, B., 2000. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol*, 10(2), pp.153–159.

Sierwald, P. & Bond, J.E., 2007. Current Status of the Myriapod Class Diplopoda (Millipedes): Taxonomic Diversity and Phylogeny. Annual Review of Entomology, 52(1), pp.401–420.

Sigworth, F.J., 1998. A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*, 122(3), pp.328–339. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9774537.

Sigworth, F.J., Doerschuk, P.C., Carazo, J.-M. & Scheres, S.H., 2010. Chapter Ten - An Introduction to Maximum-Likelihood Methods in Cryo-EM. *Methods Enzymol.*, 482, pp.263–294.

Sine, S.M., Quiram, P., Papanikolaou, F., Kreienkamp, H.J. & Taylor, P., 1994. Conserved tyrosines in the alpha subunit of the nicotinic acetylcholine receptor stabilize quaternary ammonium groups of agonists and curariform antagonists. *The Journal of biological chemistry*, 269(12), pp.8808–8816.

Singh, J. & Thornton, J.M., 1985. The interaction between phenylalanine rings in proteins. *FEBS Letters*, 191(1), pp.1–6.

Sixma, T.K. & Smit, A.B., 2003. Acetylcholine binding protein (AChBP): a secreted glial protein that provides a high-resolution model for the extracellular domain of pentameric ligand-gated ion channels. *Annual review of biophysics and biomolecular structure*, 32(1), pp.311–334.

Smit, A.B., Syed, N.I., Schaap, D., et al., 2001. A glia-derived acetylcholine-binding protein that modulates synaptic transmission. *Nature*, 411(6835), pp.252–255. Available at: http://www.nature.com/nature/journal/v411/n6835/abs/411261a0.html.

Standar, K., Mehner, D., Osadnik, H., et al., 2008. PspA can form large scaffolds in *Escherichia coli*. *FEBS Lett*, 582(25-26), pp.3585–3589.

Steinkilberg, M. & Schramm, H.J., 1980. Eine verbesserte Drehkorrelationsmethode für die Strukturbestimmung biologischer Makromoleküle durch Mittelung elektronenmikroskopischer Bilder. *Hoppe-Seyler´s Zeitschrift für physiologische Chemie*, 361(2), pp.1363–1370.

Stewart, A. & Grigorieff, N., 2004. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy*, 102(1), pp.67–84.

Sun, Y., Bauer, M.D., Keough, T.W. & Lacey, M.P., 1996. Disulfide Bond Location in Proteins. 61, pp.185–210.

Suurkuusk, J., Lentz, B.R., Barenholz, Y., Biltonen, R.L. & Thompson, T.E., 1976. A calorimetric and fluorescent probe study of the gel-liquid crystalline phase transition in small, single-lamellar dipalmitoylphosphatidylcholine vesicles. *Biochemistry*, 15(7), pp.1393–1401.

Suzuki, I., Kanesaki, Y., Mikami, K., Kanehisa, M. & Murata, N., 2001. Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*. *Molecular Microbiology*, 40(1), pp.235–244.

Tang, G., Peng, L., Baldwin, P.R., et al., 2007. EMAN2: an extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1), pp.38–46.

Taylor, K.A. & Glaeser, R.M., 1974. Electron diffraction of frozen, hydrated protein crystals. *Science (New York, N.Y.)*, 186(4168), pp.1036–1037.

Terwilliger, N.B., Terwilliger, R.C. & Schabtach, E., 1976. The quaternary structure of a molluscan (*Helisoma trivolvis*) extracellular hemoglobin. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 453(1), pp.101–110.

Thomas, J.D., Nwanko, D.I. & Sterry, P.R., 1985. The Feeding Strategies of Juvenile and Adult *Biomphalaria glabrata* (Say) under Simulated Natural Conditions and their Relevance to Ecological Theory and Snail Control. In Proceedings of the Royal Society of London. Series B. pp. 177–209.

Tzartos, S.J., Cung, M.T., Demange, P., et al., 1991. The main immunogenic region (MIR) of the nicotinic acetylcholine receptor and the anti-MIR antibodies. *Molecular Neurobiology*, 5(1), pp.1–29.

Undheim, E.A.B. & King, G.F., 2011. On the venom system of centipedes (Chilopoda), a neglected group of venomous animals. *Toxicon*, 57(4), pp.512–524.

Unwin, N., 2005. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *Journal of molecular biology*, 346(4), pp.967–989.

Unwin, P.N.T., 1970. An Electrostatic Phase Plate for the Electron Microscope. *Berichte der Bunsengesellschaft für physikalische Chemie*, 74(11), pp.1137–1141.

van Bruggen, E.F.J., 1983. *An electron microscopists view of the quaternary structure of arthropodan and molluscan hemocyanins*, Harwood, London: Life Chem. Rep. Suppl. 1.

van Heel, M., 1987. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2), pp.111–123.

van Heel, M., 1989. Classification of very large electron microscopical image data sets. *Optik*, 82, pp.114–126.

van Heel, M. & Frank, J., 1981. Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, 6(2), pp.187–194.

van Heel, M. & Stöffler-Meilicke, M., 1985. Characteristic views of *E. coli* and *B. stearothermophilus* 30S ribosomal subunits in the electron microscope. *The EMBO Journal*, 4(9), pp.2389–2395.

van Heel, M., Gowen, B., Matadeen, R., et al., 2000. Single-particle electron cryo-microscopy: towards atomic resolution. *Journal of Structural Biology*, 33(4), pp.307–369.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. & Schatz, M., 1996. A new generation of the IMAGIC image processing system. *Journal of Structural Biology*, 116(1), pp.17–24.

van Heel, M., Portugal, R. & Schatz, M., 2009. Multivariate Statistical Analysis in Single Particle (Cryo) Electron Microscopy. In A. Verkley & E. Orlova, eds. *An electronic text book: Electron microscopy in Life Science*. 3D-EM Network of Excellence, pp. 1–47.

van Heel, M., Schatz, M. & Orlova, E.V., 1992. Correlation functions revisited. *Ultramicroscopy*, 45, pp.307–316.

van Holde, K.E. & Miller, K.I., 1995. Hemocyanins. In F. M. R. C B Anfinsen John T Edsall and David S Eisenberg, ed. Advances in Protein Chemistry. Advances in Protein Chemistry. San Diego: Academic Press, pp. 1–81.

Volbeda, A. & Hol, W.G., 1989. Crystal structure of hexameric haemocyanin from *Panulirus interruptus* refined at 3.2 Å resolution. *Journal of molecular biology*, 209(2), pp.249–279.

Vothknecht, U.C., Otters, S., Hennig, R. & Schneider, D., 2012. Vipp1: a very important protein in plastids?! *Journal of experimental botany*, 63(4), pp.1699–1712.

Wade, R.H., 1992. A brief look at imaging and contrast transfer. *Ultramicroscopy*, 46(1-4), pp.145–156.

Wahba, G., 1980. *Spline bases, regularization, and generalized cross-validation for solving approximation problems with large quantities of noisy data*, Academic Press, New York.

Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), pp.236–244.

Westphal, S., Heins, L., Soll, J. & Vothknecht, U.C., 2001a. Vipp1 deletion mutant of *Synechocystis*: a connection between bacterial phage shock and thylakoid biogenesis? *Proceedings of the National Academy of Sciences*, 98(7), pp.4243–4248.

Westphal, S., Soll, J. & Vothknecht, U.C., 2001b. A vesicle transport system inside chloroplasts. *FEBS Letters*, 506(3), pp.257–261.

Wood, E.J. & Mosby, L.J., 1975. Physicochemical properties of *Planorbis corneus* erythrocruorin. *The Biochemical journal*, 149(2), pp.437–445.

Xylander, W.E., 2009. Physico-chemical properties of haemolymph of Chilopoda and Diplopoda (Myriapoda, Arthropoda): protein content, pH, osmolarity. *Soil organisms*, 81, pp.431–439.

Yu, X., Ge, P., Jiang, J., Atanasov, I. & Zhou, Z.H., 2011. Atomic Model of CPV Reveals the Mechanism Used by This Single-Shelled Virus to Economically Carry Out Functions Conserved in Multishelled Reoviruses. *Structure*, 19(5), pp.652–661.

Zernike, F., 1955. How I discovered phase contrast. *Science (New York, N.Y.)*, 121(3141), pp.345–349.

Zhang, J., Xue, F., Liu, Y., Yang, H. & Wang, X., 2013. The Structural Mechanism of the Cys-Loop Receptor Desensitization. *Molecular Neurobiology*, pp.1–12.

Zhang, L. & Sakamoto, W., 2013. Possible function of VIPP1 in thylakoids: Protection but not formation? *Plant signaling & behavior*, 8(2), p.e22860.

Zhang, L., Tong, H., Garewal, M. & Ren, G., 2012. Optimized negative-staining electron microscopy for lipoprotein studies. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1830(1), pp.2150–2159.

Zhou, Z.H., 2008. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr Opin Struct Biol*, 18(2), pp.218–228.

Danksagung

Ich danke...

... Prof. Dr. ████████ für das mir entgegengebrachte Vertrauen, viele fruchtbare Diskussionen, Hilfe bei jeglichen Problemen und die kritische Durchsicht meiner Arbeit.

... Prof. Dr. ████████ für die bereitwillige Übernahme der zweiten Berichterstattung meiner Arbeit und der sehr angenehmen Kooperation im Vipp1-Projekt.

... Dr. ████████ für die geduldige Einführung in die cryo-EM und EM-Aufnahmen von *Scutigera coleptrata* Hämocyanin und Acetylcholin-bindendem Protein.

... Prof. Dr.████████ und seiner Arbeitsgruppe (im Besonderen ████████, ████████ und ██████) für die Bereitstellung sämtlicher Daten ihrer oxy-ScoHc-Berechnungen.

... Prof. Dr. ████████ und seiner Arbeitsgruppe für die Bereitstellung des aufgereinigten *Polydesmus angustus* Hämocyanins.

... ████████ für die angenehme Zusammenarbeit bei der Analyse des *Polydesmus angustus* Hämocyanins

... ████████ für die tolle Zusammenarbeit im Vipp1-Projekt und ergebnisreiche Diskussionen während zahlreicher Mittagspausen

... ████████, Dr. ████████, ████████, Dr. ████████ und Dr. ██████ für die produktive Zusammenarbeit beim BgAChBP-Projekt.

... der gesamten AG Markl für eine wahnsinnig tolle Atmosphäre und insbesondere der 3-D Gruppe, die den Arbeitsplatz zu mehr machten als nur einen Platz zum Arbeiten.

... ██ und ██ für Korrekturvorschläge und Hilfe bei Namensgebungen.

... meiner Familie, dass sie immer für mich da war.

Appendices

- ScoHc Tris buffer A: 50 mM Tris/HCl
                        150 mM NaCl
                        pH 7.4

- ScoHc Tris buffer B: 50 mM Tris/HCl
                        5 mM $CaCl_2$
                        5 mM $MgCl_2$
                        pH 7.6

- ScoHc Tris buffer C: 50 mM Tris/HCl
                        5 mM $CaCl_2$
                        5 mM $MgCl_2$
                        pH 7.4

- Vipp1 phosphate buffer: 50 mM Na_2HPO_4 (84,5 % v/v)
                          50 mM NaH_2PO_4 (15,5 % v/v)
                          pH 7.6

- Vipp1 HEPES buffer: 20 mM HEPES/NaOH
                        pH 7.6

- BgAChBP buffer: 20 mM Tris/HCl
                        50 mM imidazole
                        pH 8.0

Appendix B: Sequence and PDB codes

	Database	Accession number
Limulus polyphemus **hemocyanin subunit 2**	Uniprot	P04253
Panulirus interruptus **hemocyanin**	PDB	1hcy
Panulirus interruptus **hemocyanin chain A**	Uniprot	P04254
Polydesmus angustus **hemocyanin subunit 1**	Uniprot	I7HPY6
Polydesmus angustus **hemocyanin subunit 2**	Uniprot	I7HZS1
Spirostreptus **sp. hemocyanin subunit 1**	Unitprot	Q9BHJ9
Synechocystis **sp. strain PCC6803 vesicle-inducing protein in plastids 1**	Uniprot	F7USL8
Aplysia californica **AChBP**	Uniprot	Q8WSF8
Aplysia californica **AChBP in HEPES buffer**	PDB	2br7
Aplysia californica **AChBP in apo state**	PDB	2byn
Biomphalaria glabrata **AChBP1**	PDB	4aod
Biomphalaria glabrata **AChBP2**	PDB	4aoe
Bulinus truncatus **AChBP**	PDB	2bj0
Haliotis discus hannai **AChBP1**	Genbank	ABU51880.1
Haliotis discus hannai **AChBP2**	Genbank	ABU51880.1
Lymnaea stagnalis **AChBP**	Uniprot	P58154
Torpedo marmorata **AChR subunit** α	Uniprot	P02711
Torpedo marmorata **AChR**	PDB	2bg9

APPENDIX C: REFINEMENT VARIABLES

EMAN1.9 deoxy-ScoHc:

General refinement command

refine 1 ang=15 mask=100 pad=384 hard=80 classkeep=2 classiter=5 sym=D3 refine proc=128 filt3d=25 median fscls usefilt xfiles=2,2700,99 amask=40,0.8,14

Variables altered during refinement

Iteration	*ang*	*hard*	*classkeep*	*classiter*	*filt3D*
1	15	80	2	5	25
2	10	60	0.9	5	25
3	10	50	0.8	5	34.13
4	10	40	0.7	5	34.13
5-10	8	40	0.7	5	34.13
11-20	5	40	0.7	5	42.6
21-30	3	40	0.7	3	51.2
31-25	2	40	0.7	3	51.2

EMAN 1.9 oxy-ScoHc:

General refinement command

refine 10 ang=5 mask=100 pad=384 hard=40 classkeep=0.9 classiter=5 sym=D3 refine proc=128 filt3d=42.6 median fscls usefilt xfiles=2,2700,99 amask=40,0.8,14

Variables altered during refinement

Iteration	*ang*	*hard*	*classkeep*	*classiter*	*filt3D*
10	10	40	0.9	5	42.6
20	3	40	0.9	3	25
30	2	40	0.9	3	51.2

FREALIGN deoxy-ScoHc:

Parameters kept stable throughout refinement:

FMAG=F; FDEF=F; FASTIG=F; FPART=F; IEWALD=0; FBEAUT=T; FFILT=F; FBFAC=F; FMATCH=T; IFSC=0; FSTAT=F; IBLOW=1; RO=170; RI=0; PSIZE=2; WGH=0.07; XSTD=0; PBC=5; DANG=20; ITMAX=100; IPMAX=10; MASK 1 1 1 1 1; ASYM=D3; RELMAG=1; DSTEP=15; TARGET=10; CS=1.2; AKV=300; TX=0; TY=0; RREC=6; RMAX1=300; DFSIG=200; RBFACT=0

Variables altered during refinement

Iteration	IFLAG	BOFF	THRESH	RMAX2
1	1	62	65	15
2	4	60.1	65	15
3-5	4	0	60	15
6-7	4	0	65	13
8-30	1	0	64	13

FREALIGN oxy-ScoHc:

Parameters kept stable throughout refinement:

FMAG=F; FDEF=F; FASTIG=F; FPART=F; IEWALD=0; FBEAUT=T; FFILT=F; FBFAC=F; FMATCH=T; IFSC=0; FSTAT=F; IBLOW=1; RO=170; RI=0; PSIZE=2; WGH=0.07; XSTD=0; PBC=5; DANG=20; ITMAX=100; IPMAX=10; MASK 1 1 1 1 1; ASYM=D3; RELMAG=1; DSTEP=11.8; TARGET=10; CS=1.2; AKV=300; TX=0; TY=0; RREC=6; RMAX1=300; DFSIG=200; RBFACT=0

<u>Variables altered during refinement</u>

Iteration	*IFLAG*	*BOFF*	*THRESH*	*RMAX2*
1	1	62	65	15
2-3	4	62	65	15
4	4	61.7	65	15
5	4	60.7	65	15
6	4	60.7	65	13
7	4	60.7	65	13
8	1	60.7	65	13
9	1	61.5	64	13
10-30	1	61.1	64	13

RELION deoxy-ScoHc:

Ref. map is on absolute greyscale? == No

Initial low-pass filter (A): == 30

Particle mask diameter (A): == 300

Pixel size (A): == 2

Reference mask (optional): ==

Do CTF-correction? == Yes

Have data been phase-flipped? == No

Ignore CTFs until first peak? == No

Has reference been CTF-corrected? == No

Symmetry group: == D

Symmetry number: == 3

Angular sampling interval: == 30 degrees

Offset search range (pix): == 5

Offset search step (pix): == 1

Local searches from auto-sampling: == 1.8 degrees

RELION oxy-ScoHc:

Ref. map is on absolute greyscale? == No

Initial low-pass filter (A): == 30

Particle mask diameter (A): == 300

Pixel size (A): == 2

Reference mask (optional): ==

Do CTF-correction? == Yes

Have data been phase-flipped? == No

Ignore CTFs until first peak? == No

Has reference been CTF-corrected? == No

Symmetry group: == D

Symmetry number: == 3

Angular sampling interval: == 30 degrees

Offset search range (pix): == 5
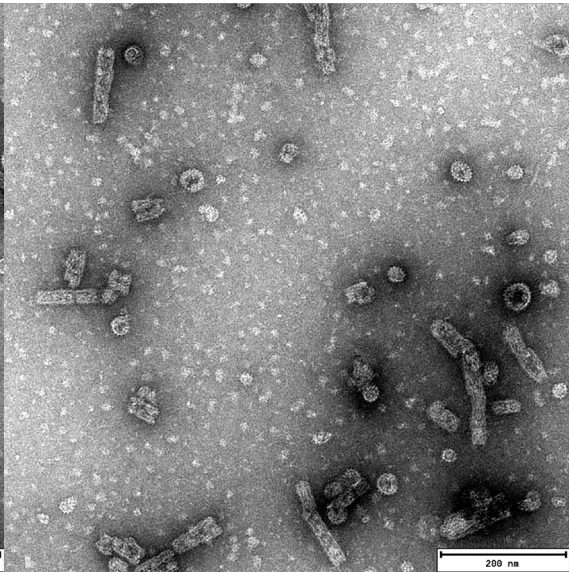
Offset search step (pix): == 1

Local searches from auto-sampling: == 1.8 degrees

Appendix D: Sample images of Vipp1 mutants

Mutant N1
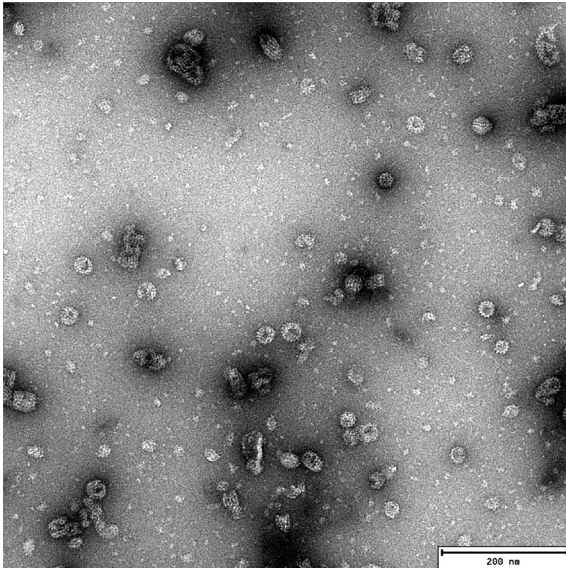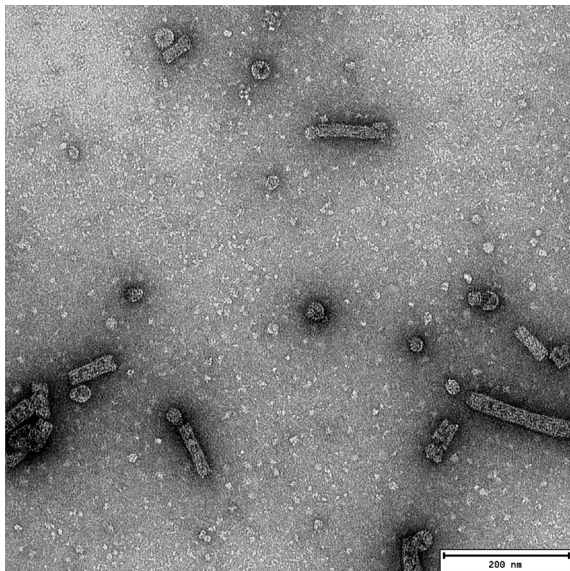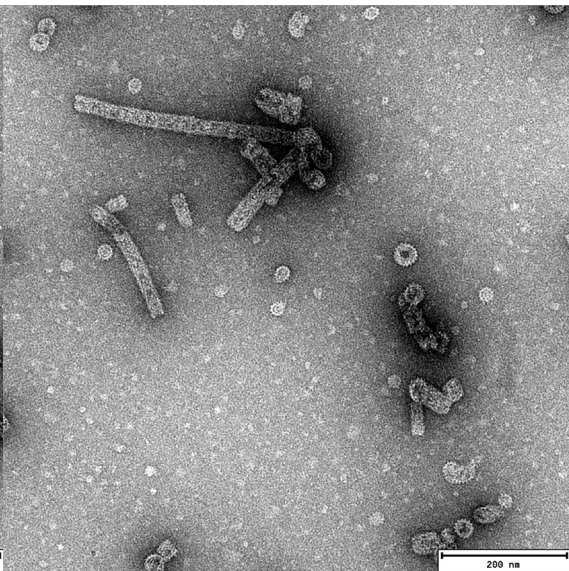
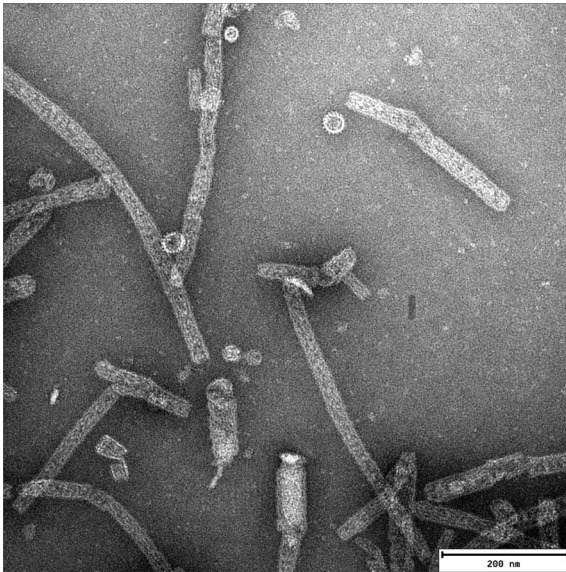**Mutant N2**
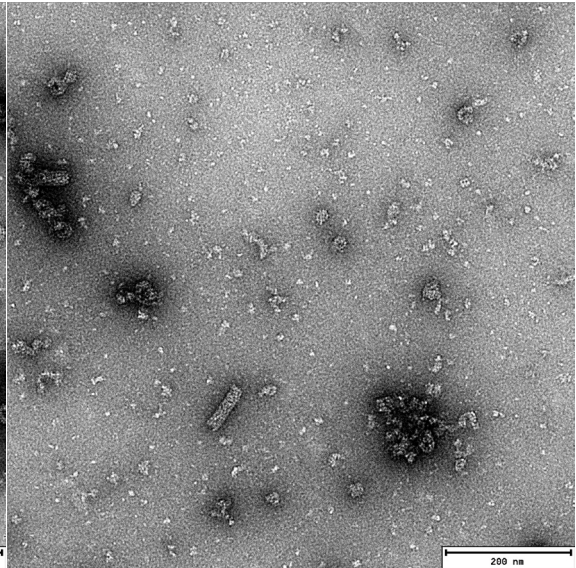
Mutant N4
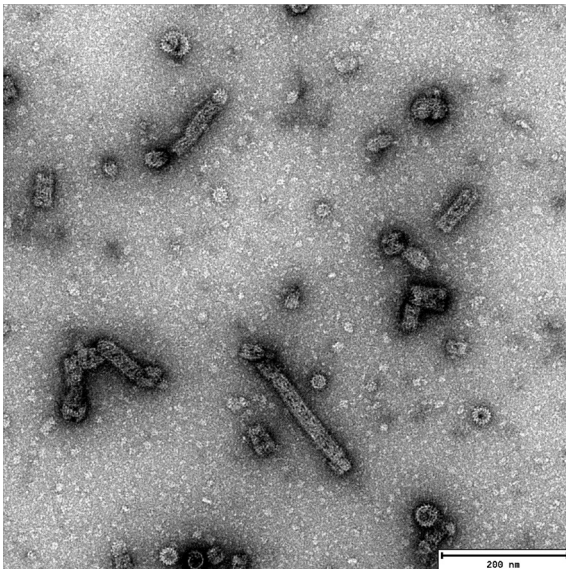
Mutant N5
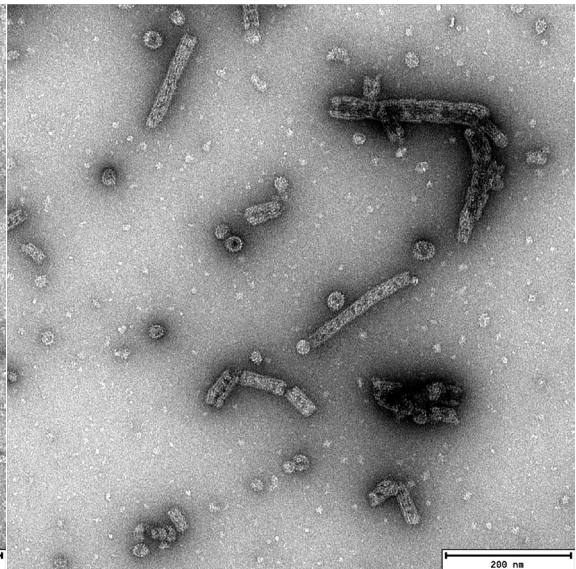
Mutant N8
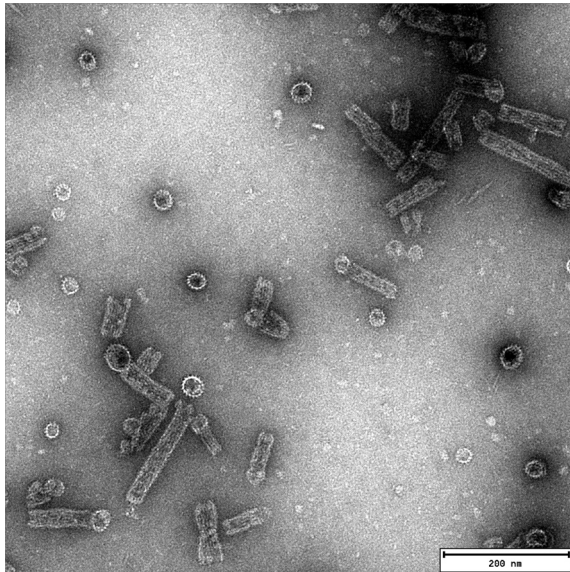
Mutant N18
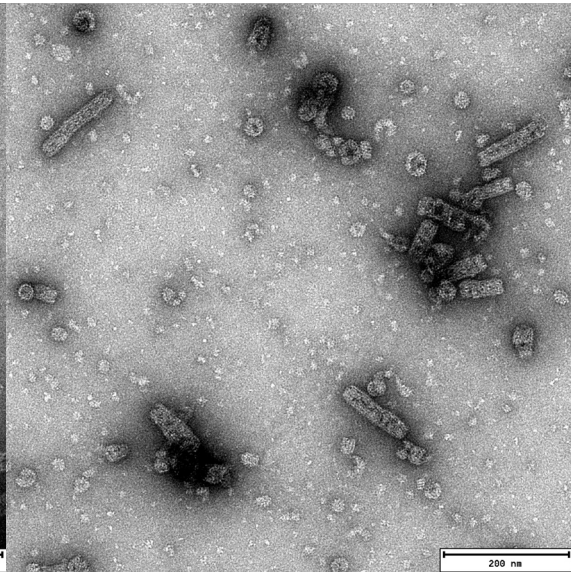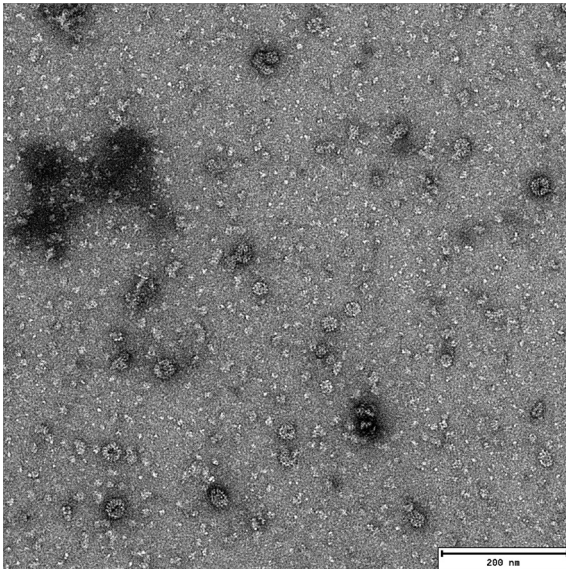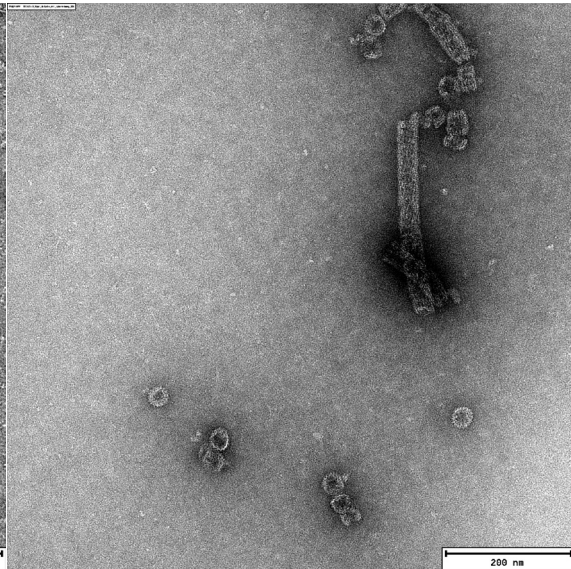

Mutant S4

Mutant S6

Mutant S10

Mutant 13-4-1

Mutant 13-4-2/3

Mutant ΔH1a

Erklärung und Copyright

Erklärung

Ich versichere hiermit, die vorliegende Arbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben.

Mainz, den 16.05.2013   _____

Michael Saur

Copyright

CURRICULUM VITAE | MICHAEL SAUR

EDUCATION

2010 - PRESENT
PhD-student and research associate in the department of structural biology at the Johannes-Gutenberg-University (JGU) Mainz with focus on 3-D electron microscopy, primarily on single particle analysis; estimated graduation: June 2013

2010
Completion of the teaching degree for secondary schools in Biology and Physical Education (Grade: 1.5; highest grade:1.0; lowest grade: 6.0)

2009
Completion of the German Diploma in Biology with distiction (Grade: 1.0); Topic of thesis: "*Scutigera coleoptrata* hemocyanin: three-dimensional reconstruction and comparison of two physiological states"

2008
Acceptance for a scholarship at the "Studienstiftung des deutschen Volkes", one of the most renown German scholarship providers

2007
Initiation of a second degree for the acquisition of a German Diploma in Biology (Majors: Zoology, Botany, Microbiology)

2006
Study-abroad year at James Cook University Townsville, AUS, in the specialties Marine- and Microbiology

2003
Initiation of higher education at JGU Mainz, GER, for a teaching degree at secondary schools in Biology and Physical Education

2002 - 2003
Military Service in the German Airforce

2002
Higher education entrance qualification at Martin-von-Cochem-Gymnasium in Cochem, GER

SPECIAL KNOWLEDGE & teaching activity

Bilingual education (Spanish-German)

Assistant in marine biology excursions and universitary courses of animal physiology

Tutor for Diploma and Master theses in our electron microscopy laboratory

Conception and implementation of practicals in 3-D EM for undergraduate students

WORKSHOPS & CONFERENCES

Brazil School for Single Particle Cryo Electron Microscopy workshop 2010, Engenheiro Passos, BRA

Gordon Research Conference on three-dimensional electron microscopy 2011, New England, USA

Horizons in Molecular Biology conference 2011, Göttingen, GER

SPnet PhD conference for single particle analysis 2011, Frankfurt a. M., GER

Interdisciplinary symposium on 3-D microscopy 2012, Les Diablerets, CH

Gordon Research Conference on three-dimensional electron microscopy 2012, Les Diablerets, CH

EMBO workshop for 3-D electron microscopy and image processing 2012, Heidelberg, Germany

PUBLICATIONS, POSTERS & PRESENTATIONS

"*Scutigera coleptrata* hemocyanin: three-dimensional reconstruction and comparison of two physiological states": presentation (Brazil School for Single Particle Cryo Electron Microscopy)

"3D-EM of ring-like Vipp1 multimers, a protein involved in thylakoid membrane formation": poster (GRC on three-dimensional electron microscopy 2011, Horizons in Molecular Biology 2011, SPnet PhD conference)

"Acetylcholine-binding in the hemolymph of the planorbid snail *Biomphalaria glabrata* is a pentagonal dodecahedron (60 subunits)": published article (PLOSOne 2012, 7(8)); presentation (EMBO workshop); poster (GRC on three-dimensional electron microscopy 2012, Horizons in Molecular Biology 2012)

CURRICULUM VITAE | MICHAEL SAUR

PRIMARY SCIENTIFIC METHODS

Transmission electron microscopy (TEM) on FEI microscopes (expert)

2-D immuno-gel electrophoresis (intermediate)

Preparation of TEM samples (negative staining / cryo) (expert)

High-performance liquid chromatography (intermediate)

Single Particle Analysis (expert) and Electron tomography (intermediate)

Cell culture of *Chlamydomonas reinhardtii* (intermediate)

Polyacrylamid gel electrophoresis (SDS / native) (intermediate)

Recombinant expression in vector plasmid systems (basics)

PRIMARY SOFTWARE & PROGRAMMING

Office (intermediate)

EMAN and EMAN2 (expert)

Adobe InDesign (intermediate)

FREALIGN (expert)

Adobe Illustrator (intermediate)

IMOD (intermediate)

Adobe Photoshop (basics)

Python programming (basics)

Bash scripting (expert)

RELION (intermediate)

IMAGIC-5 (expert)

UCSF Chimera (expert)

FURTHER KNOWLEDGE INTERESTS & SOCIAL COMMITMENT

Languages: German (Mother tongue)

English (fluent; TOEFL, computer-based, in 2005, score:277)

Spanish (Mother tongue)

French (basics)

Italian (reading basics)

Portuguese (reading basics)

Drums and Percussion:
Member of the Coverrock-Band „Brotlos" and the association for German folk music in Dohr, Germany; Participation in several productions of the JGU musical group (Musical Inc. e.V.)

Sports

Mainz, May 16th 2013

Michael Saur