JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

3D cryo-electron microscopy and molecular model of two arthropod hemocyanins

Dissertation zur Erlangung des Grades

“Doktor der Naturwissenschaften”

am Fachbereich Biologie

der Johannes Gutenberg Universität Mainz

Vorgelegt von

Wolf Christoph Kühne

geb. am 27. Februar 1982 in Frankfurt am Main

Mainz, 2016

Dekan: [redacted]

1. Berichterstatter: [redacted]

2. Berichterstatter: [redacted]

Tag der Prüfung:

I Table of contents

A Introduction 1

B Material & Methods 11

C Results & Discussion 35

II List of Figures

III List of Tables

A Introduction

The transformation of the Earth's anoxic state to oxic conditions some 2.5 billion years ago by the evolution of oxygenic photosynthesis presented a turning point for life on Earth (Lindahl, 2008; Review: Decker & van Holde, 2011). As most of the Earth's inhabitants at that time were obligate anaerobes, they died due to the toxicity of free dioxygen. However, some organisms were able to adapt to the oxygenic atmosphere by either retreating to anaerobic environments or by developing mechanisms to control the generation of the toxic by-products (e.g. reactive oxygen species). Moreover, these organisms evolved to use oxygen, which presents one of the most capable electron acceptors on Earth (Lindahl, 2008). The highly increased efficiency of ATP production by oxidative phosphorylation opened up new developmental possibilities and allowed the evolution of aerobic life forms from the first multicellular organisms to complex vertebrates and plants (Lindahl, 2008; Decker & van Holde, 2011). Their metabolism strictly depends on the availability of dioxygen and its constant supply is crucial for the survival of their cells.

In many animals with circulation systems, oxygen is bound reversibly by oxygen binding proteins in the respiratory tract (gills or lungs) and delivered to the various tissues where the oxygen is released and diffuses into the cells. The most common oxygen binding proteins are hemoglobins, hemocyanins and hemerythrins.

Hemoglobins are the most widely used oxygen transport proteins. They are present in most vertebrates as well as in the tissues of certain invertebrates like some arthropods, molluscs and annelids. The protein contains a heme group with an iron atom coordinated in its center (Fig. 1A). This iron atom binds the dioxygen molecule in a well-protected cavity (Decker & van Holde, 2011). Oxygen is therefore not free to react and generate reactive oxygen species during transport. Human hemoglobin is a rather small tetrameric protein of ~64 kDa that is present at high concentrations of 200 to 300 million molecules per erythrocyte / red blood cell. As each hemoglobin is able to carry four molecules of oxygen, one per heme group, the total oxygen capacity of blood is increased by a 50-fold quantity compared to dissolved oxygen in blood. However, the shape and size of hemoglobins can vary considerably. Many annelids, especially marine polychaetes, contain giant, multi-subunit complexes of more than 3,5 MDa, which are

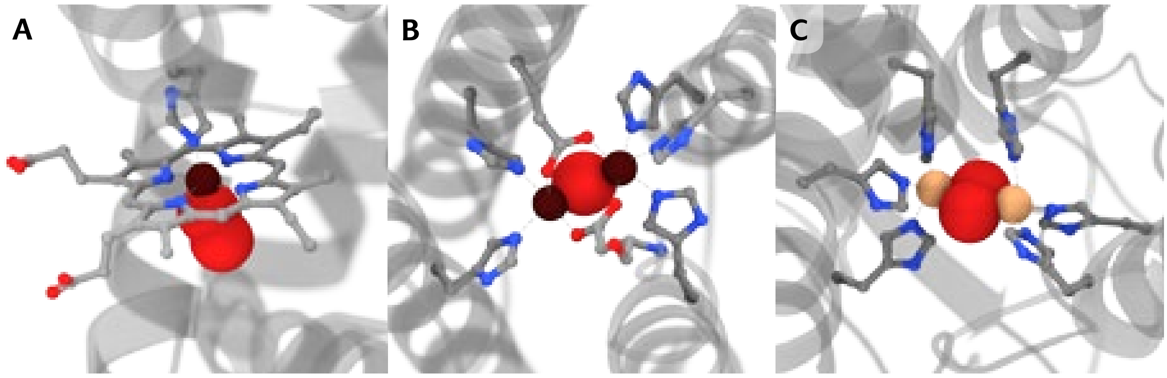


Fig. 1: Active sites of three common respiratory proteins. A: Hemoglobin B: Hemerythrin C: Hemocyanin Iron depicted dark red at 0,5 x atomic radius, copper colored light orange, oxygen red and nitrogen blue. Figure based on Solem, 2010.

also termed erythrocruorins (Royer et al., 2000; Pallavicini et al., 2001). These complexes are dissolved freely in the blood plasma. In *Lumbricus terrestris* it could be shown that the complex is composed of 144 oxygen-binding hemoglobin subunits and 36 nonhemoglobin linker subunits (Royer et al., 2000). While the hemoglobin subunits arrange in dodecameric substructures, the linker chains form twelve triple stranded coiled-coil α-helices towards the center of the complex/trimeric complexes (double-layered hexagonal shape with overall molecular D6 symmetry).

Hemerythrin, another oxygen transport protein, is found in several marine invertebrates (e.g. sipunculids, priapulid worms and brachiopods; French et al., 2007; Bailly et al.,2008). Similar to hemoglobin, the binding site for dioxygen consists of a pair of iron atoms (Fig. 1B). However, unlike hemoglobin, hemerythrin does not contain a heme. The iron atoms are bridged by a characteristic set of conserved glutamate, aspartate and histidine residues within the protein (French et al., 2007). So far, octomeric and trimeric complexes have been identified (Satake et al., 1990; Fuseya et al., 1989).

Hemocyanins are the second most common oxygen transport proteins found in nature. As this study is focused on hemocyanins, they will be characterized here in more detail.

1 Hemocyanin

Hemocyanins are giant extracellular proteins of many molluscs and arthropods and circulate freely in the hemolymph. The active site is composed of two copper proteins instead of iron and contains no heme (Fig. 1C). The copper ions are complexed by three histidines each. Due to this copper type-3 center, which is also present in tyrosinases, catecholoxidases and phenoloxidases, it is assumed that the proteins evolved from ancient tyrosinase-like proteins (Burmester & Scheller, 1996; Markl, 2013). Binding of oxygen to the copper ions leads to the oxidation of Cu(I) to Cu(II) which bind oxygen as O_2^{2-}. The oxygenated form of the complex is the reason for the deep blue color of the hemolymph (van Holde & Miller, 1995). Due to their similarities in the active site including the mode of oxygen binding, molluscsan and arthropod hemocyanin is clearly related. However, apart from copper-dependent oxygen binding, the molecular structures of both proteins differs considerably and they are considered as two distinct protein superfamilies (Fig. 2; Markl & Decker, 1992; van Holde et al., 2001; Markl, 2013).

Molluscan hemocyanins are composed of large polypeptide chains of 350-400 kDa each (Fig. 2). Each polypeptide contains seven or eight sequential globular "functional units" (FUs; termed FU-a to FU-g,h) which bind oxygen (van Holde et al., 2001). These FUs are connected by linker peptides of 10-15 amino acids and assemble to form cylindrical decamers, didecamers or multidecamers (Reviews: van Holde & Miller, 1995 and Markl, 2013).

1.1 Arthropod hemocyanins

Arthropod hemocyanins assemble into a single or multiples of hexamers (Fig. 3 – 1x6mer, 2x6mers, 4x6mers or 8x6mers). The monomers are roughly bean-shaped with a molecular mass of ~75 kDa (Reviews: Markl & Decker, 1992; van Holde & Decker, 2001; Decker et al., 2007). In arthropods, each subunit contains a single oxygen binding site.

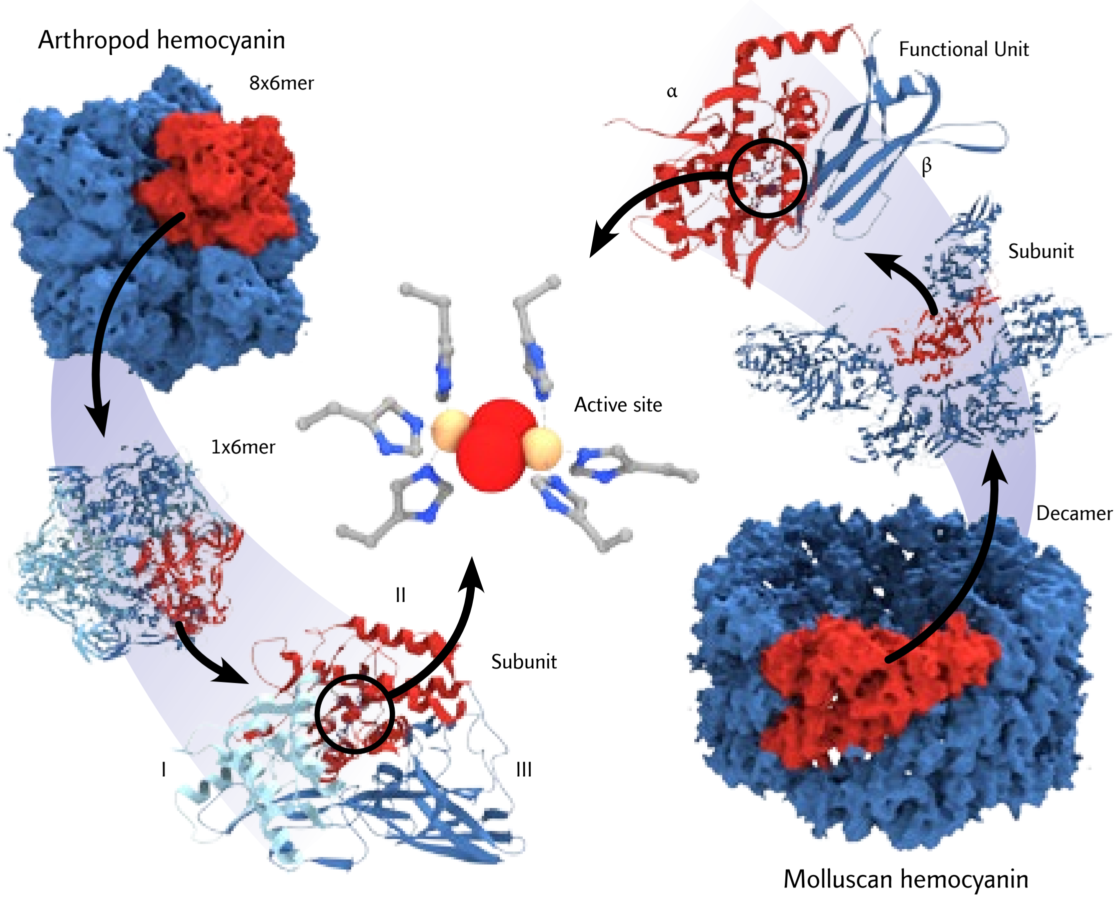


Fig. 2: Structural levels of arthropod and molluscan hemocyanins. Domain designations are indicated at the subunit/functional unit level. Modified from Decker et al., 2007.

The arthropod hemocyanin superfamily, aside from hemocyanin, comprises four additional subclasses which share sequence similarities but differ widely in function: phenolaxidases, pseudo-hemocyanins, hexamerins and hexamerin receptors (Burmester & Scheller, 1996; Burmester, 2001 & 2002). Phenoloxidases (comprising tyrosinases and catecholoxidases) are necessary for multiple biological functions such as browning, wound healing, primary immune defense and sclerotization (Decker et al., 2001). They are ubiquitous essential enzymes (Decker et al., 2007). Pseudo-hemocyanins, hexamerins and hexamerin receptors have lost the ability to bind oxygen (Burmester, 2002). Pseudo-hemocyanins might have a storage function, similar to hexamerins, an insect storage protein (Burmester, 1999).

The 1x6mer hemocyanin from *Panulirus interruptus* was the first hemocyanin structure solved by x-ray crystallography (Fig. 4; Gaykema et al., 1984) which also revealed the domain structure of the subunits (Gaykema et al., 1986). Domain #1 comprises five or six α-helices (helix α1.2 is missing in Chelicerate hemocyanin; Magnus et al., 1994). This configuration protects the active site contained in domain #2. A typical cluster of four anti-parallel helices (Presnell & Cohen, 1989) contains two cooper ions (Cu(A) and Cu(B)) which are coordinated by six highly conserved histidine residues. Domain #3 is dominated by β-barrels, parts of which reach back towards domain #1. This arrangement stabilizes the tertiary structure of the whole subunit (Markl & Decker, 1992).

1.1.1 Hemocyanin in Crustacea

The use of hemocyanin for oxygen transport has been shown for two crustacean taxa, Malacostraca and Remipedia. Other taxa might use hemoglobins or lack oxygen-binding proteins (Markl & Decker, 1992; Ertas et al., 2009; Scherbaum, 2010). Typical hemocyanin oligomer variants found in Malacostraca are 1x6mer and 2x6mers (Fig. 3 A and E). Atypical are the 4x6mer variants found in Thalassinidae; especially *Callianassa californiensis* has been intensively studied (Miller et al., 1976, 1977, 1981; Roxby et al., 1974). In Callianassidae, 4x6mer hemocyanins were also observed. In particular the structure of the hemocyanin of *Callianassa truncata* is a target of this study.

1.1.2 Subunit heterogeneity

The subunits of the decapoda were differentiated into three classes (α, β and γ) by means of immunological markers (Markl, 1986; Markl et al., 1986) and molecular phylogenies (Burmester, 2002; Hagner-Holler et al., 2005; Scherbaum et al., 2010).

Subunit heterogeneity has been shown to be essential for the formation of oligomer hemocyanin structures (Markl & Decker, 1992; van Holde & Miller, 1995). The 4x6mer of the Callianassidae stands out within crustacea as commonly found assembly are 1x6 and 2x6mers. For the *Callianassa truncata* 4x6mer, two subunits have been sequenced, both likely to be of the α-type according to current results (Solem, 2010; Groneberg, 2011).

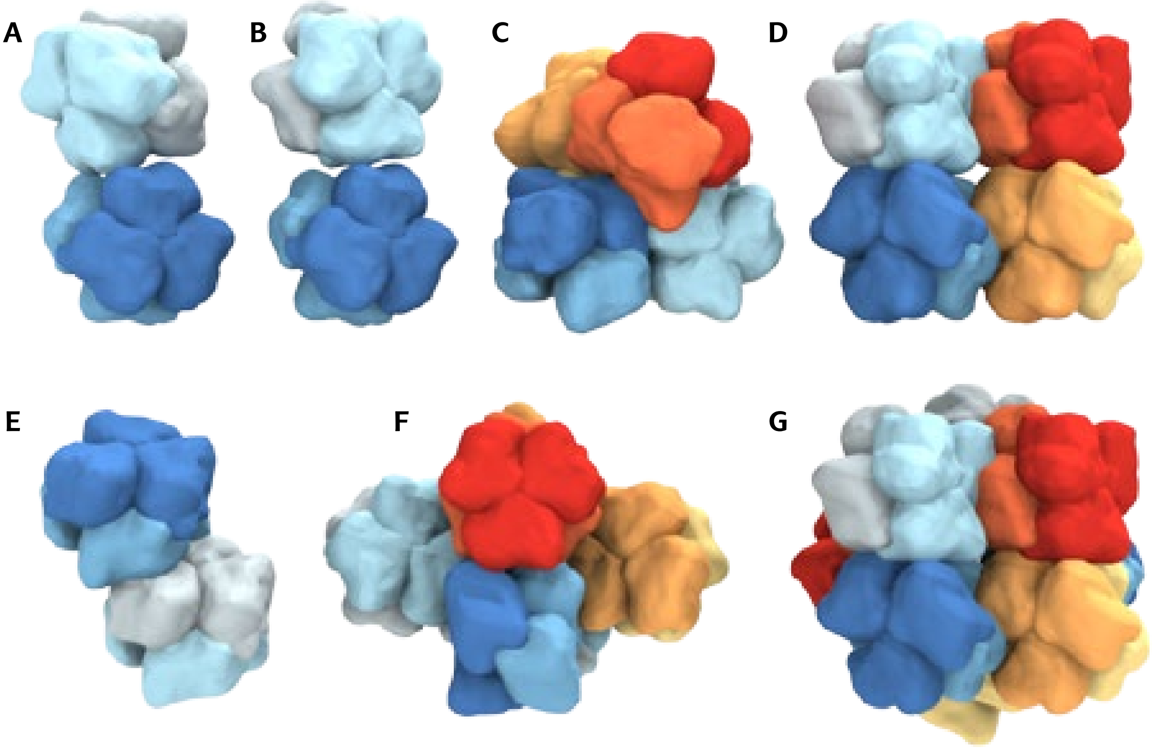


Fig. 3: Oligomer variants of arthropod hemocyanin. A: Typical crustacean 2x6mer B: Arachnidean 2x6mer C: Thalassinid 4x6mer D: 4x6mer of some arachnids E: 2x6mer of Stomadopoda (Crustacea) F: Myriapod 6x6mer G: 8x6mer of Xiphosura (modified from Markl & Decker, 1992).

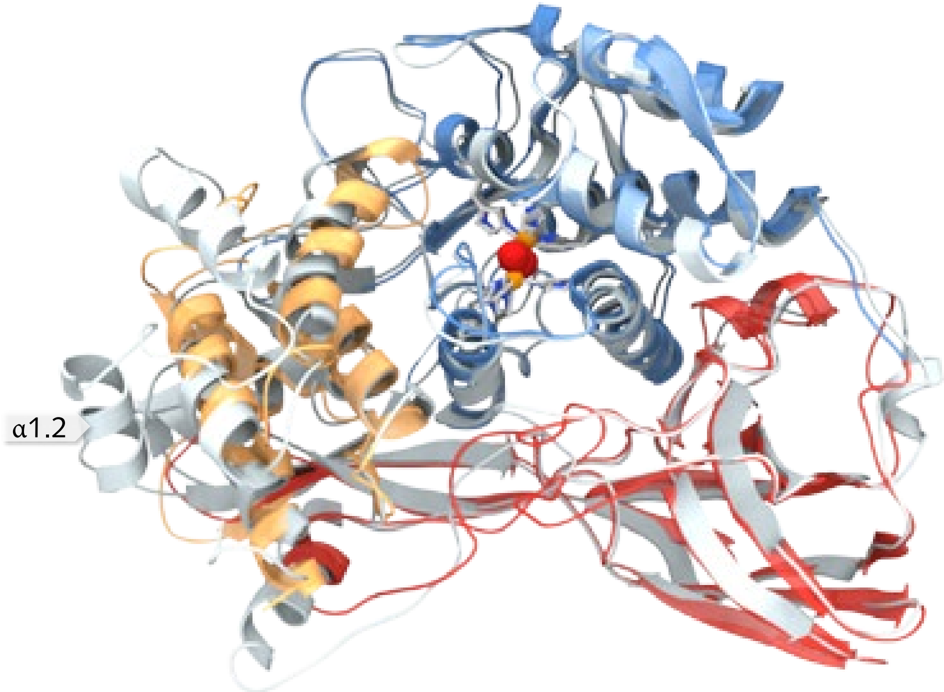


Fig. 4: Overlay of two hemocyanin x-ray structures. Hemocyanin of *Panulirus interruptus* (pdb entry 1HCY) depicted gray. The additional α1.2 helix as labeled is not present in the structure of *Limulus polyphemus* (pdb entry 1NOL), depicted orange (domain #1), blue (domain #2) and red (domain #3). The active site is rendered with its six histidine residues. Cu atoms colored light orange at 0,5 x atomic radius, oxygen red and nitrogen blue.

1.1.3 Crustacean 2x6mer hemocyanin

Recent studies have brought deeper insight into the arrangement of three different types of arthropod hemocyanins with 2x6mer architecture by using 3D-reconstructions from negative stain images (Arnold, 2012): The decapod type (3D-reconstructions of *Astacus leptodactylus, Carcinus maenas* and *Homarus americanus* hemocyanin), stomatopod type (*Odontodactylus scylarus*) and isopod type (*Ligia oceanica*).

The decapod type 2x6mer is characterized by the hexamers rotation angle of ~74° with respect to each other (Fig. 5D). In contrast the isopod type exhibits two stacked hexamers with no rotation or shift but tilted at an angle leaving a cleft and only two small contact areas (Fig. 5C). It has been proposed that both the isopod and decapod type stem from the same ancestral precursor (Fig. 5; Arnold, 2012). This precursor is still present today in the form of the thalassinid 4x6mer. Therefore the 4x6mer of *Callianassa truncata* can be split into 2x6mers two different ways, resulting in either the decapod or isopod configuration.

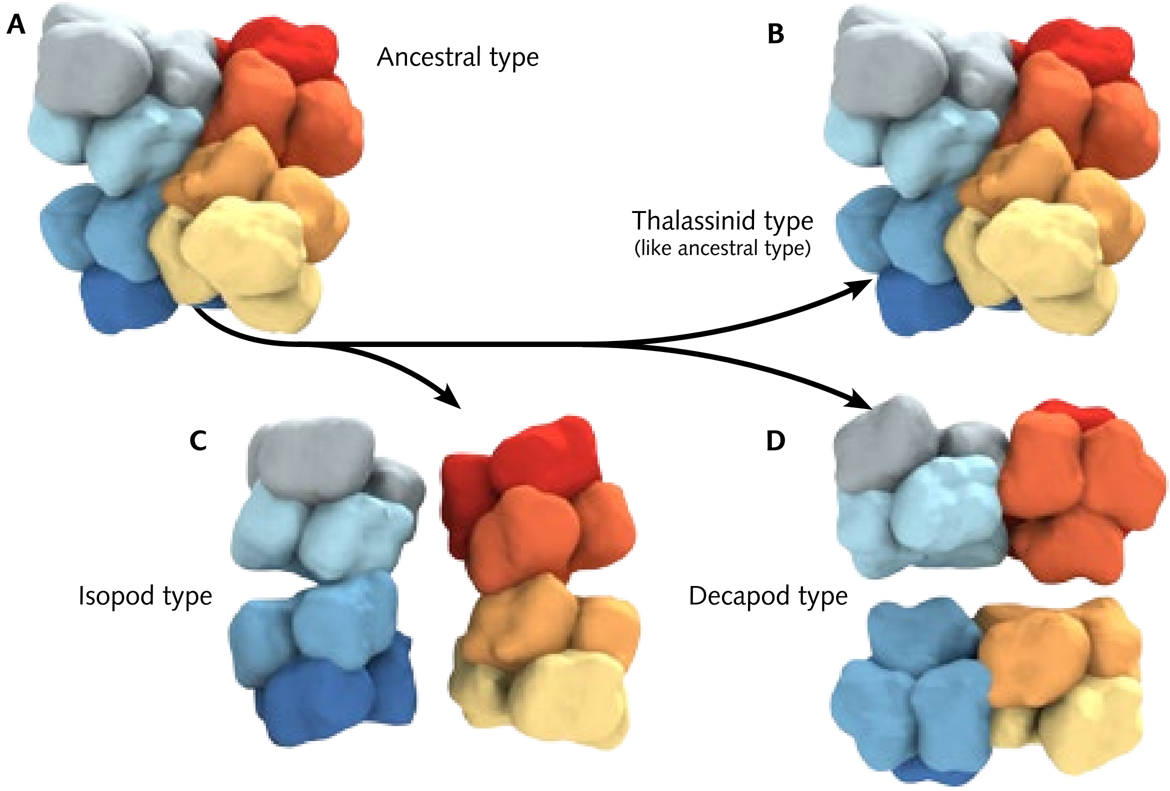


Fig. 5: Hypothetical evolution of isopod, decapod and thalassinid hemocyanin. When the isopod type branched off the ancestral 4x6mer split in a different way compared to when the isopod type developed later on. The ancestral from is still present in the form of thalassinid 4x6mers. Figure from Arnold 2012 (modified).

1.1.4 Hemocyanin in Xiphosura

The largest arthropod hemocyanin (3,5 MDa) occurs in horseshoe crabs such as *Limulus polyphemus*, with 48 active sites/subunits (8x6mer; Fig. 3G)(Magnus et al., 1994; Martin et al., 2007). A total of eight distinct subunit types form the 8x6mer (Table 1). Sequence data from *Limulus polyphemus* hemocyanin is still incomplete (Table 1) although many arthropod hemocyanin sequences are available (Burmester, 2001 & 2002).

Two notable 3D-reconstructions have been produced, beginning with 40 Å resolution by Taveau et al., 1997 and followed by reconstruction by Martin et al., 2007 that according to current standards has ~12 Å resolution.

Subunit type	Number of copies	Sequence
I	6	–
II	8	+
IIA	2	–
IIIA	8	+
IIIB	8	+
IV	8	+
V	4	–
VI	4	+

Table 1: Subunit type distribution in Limulus polyphemus hemocyanin. Authentic subunit sequence availability indicated by + = sequence available; – = no sequence available. References: Sullivan et al., 1974: Lamy et al., 1979; Brenowitz et al., 1981; Lamy et al., 1983

1.2 Functional interaction of hemocyanin subunits

Oxygen-binding behavior of hemocyanins is usually characterized by low to moderate oxygen affinity that can be modulated by a variety of factors, as an adaptation to the species-specific ecophysiology. Cooperativity of oxygen binding is comparatively low in molluscan hemocyanins (with hill coefficients around 2), but can be exceptionally high in arthropod hemocyanins (with Hill coefficients up to 9)(Loewe, 1978; Decker & Sterner, 1990; van Holde & Miller, 1995; review: Decker et al., 2007). It has been demonstrated that allosterism such as the cooperativity of oxygen binding is not restricted to the 1x6mer level, but includes the functional interaction of hexamers (Savel-Niemann et al., 1988).

2 Single Particle Analysis

3D-reconstruction from cryoEM images has become a *de facto* standard procedure in recent years (review: Thompson et al., 2016). This can in part be attributed to the ever increasing processing power of affordable CPUs. Additionally advances in software, sample preparation and instrumentation have lead to single particle reconstructions to routinely reach sub-nanometer resolutions. With the addition of structures solved by X-ray crystallography and homology modeling, an in-depth study of protein structure and thus understanding of function is possible.

The process of high resolution cryoEM starts with the vitrification of an unstained sample in an aqueous environment (Dubochet & McDowall, 1981; Dubochet et al., 1982). During vitrification, very quick (<20 ms) immersion of the sample into liquid ethane or propane leads to shock-freezing without the formation of ice crystals.

Some limitations must be kept in mind: vitrified biological samples offer an inherently low contrast due to similar scattering properties of the vitreous ice and the embedded protein itself. Also, biological specimens are highly susceptible to radiation damage. This necessitates dosage limitation (5-10 electrons/Å^2) which in turn leads to significant amounts of processing time spent to improve the low signal to noise ratio of the recorded data. An iterative cycle needs to be performed which makes single particle reconstructions prone to bias and artifacts if a point group symmetry is also imposed on the model.

3 Objectives

Some arthropod hemocyanins, notably the 8x6mer of horseshoe crabs, exhibit the highest cooperativity hitherto observed in proteins. Any future understanding of such subunit interaction phenomena in arthropod hemocyanins at the chemo-mechanical level requires a detailed knowledge of how the subunits and hexamers come together in the quaternary structure. The structure of the basic 1x6mer has been solved by X-ray crystallography, but revealing the exact assembly of the hexamers in the various 2x6mers, 4x6mers etc. is still a goal.

This thesis focuses on two well studied arthropod hemocyanin types: the 8x6mer from the horseshoe crab *Limulus polyphemus* and the 4x6mer of the thallasinid ghost shrimp *Callianassa truncata*. In the case of *L. polyphemus*, the goal was to build on the work of Martin et al. (2007): increase the resolution of the reconstruction with a new dataset acquired on a superior electron microscope. This opens up the possibility to further refine the existing molecular model by means of flexible fitting procedures.

For *C. truncata*, the precise architecture of the 4x6mer was unclear. Starting with a bias free 3D-reconstruction from negatively stained specimens this study aims to produce a high resolution 3D-reconstruction which, in conjunction with newly available sequencing data, should lead to a comprehensive molecular model of *C. truncata* hemocyanin.

B Material & Methods

1 Animals

1.1 *Limulus polyphemus* (horseshoe crab)

Limulus polyphemus, due to its remarkably shaped carapax commonly known as the "horseshoe crab", has little in common with crabs (*Crustacea*). Belonging to the phylum of *Arthropoda*, it is more closely related to modern-day scorpions and spiders (Smith et al., 2002). *Limulus polyphemus* is one of only three species in the *Limulidae* family.

Kingdom	*Animalia*
Phylum	*Arthropoda*
Subphylum	*Cheilcerata*
Class	*Merostomata*
Subclass	*Xiphosura*
Order	*Xiphosurida*
Family	*Limulidae*
Genus	*Limulus*
Species	*Limulus polyphemu*s

Fossils records of members of the *Limulidae* family dating as far back as 500 million years have been found (Shuster, 1982). Essentially unchanged compared to specimens living today, the animal is often referred to as a living fossil (Dunlap 1999). It is able to live in a wide range of water conditions (temperatures ranging from -5 to 35 °C and salinity between 11–36% have been reported)(Laughlin, 1983). In addition, *Limulus* is able to sustain short periods on shore.

A mature animal can measure up to 60 cm in length with a weight of approx. 5 kg. The typical lifespan amounts to 19 years. Fig. 6 provides an anatomical overview. Animals used for this study where imported from the Marine Biolab (Woodshole, USA). They were kept in the Institute of Zoology in seawater at 18 °C and fed with fish and squid.

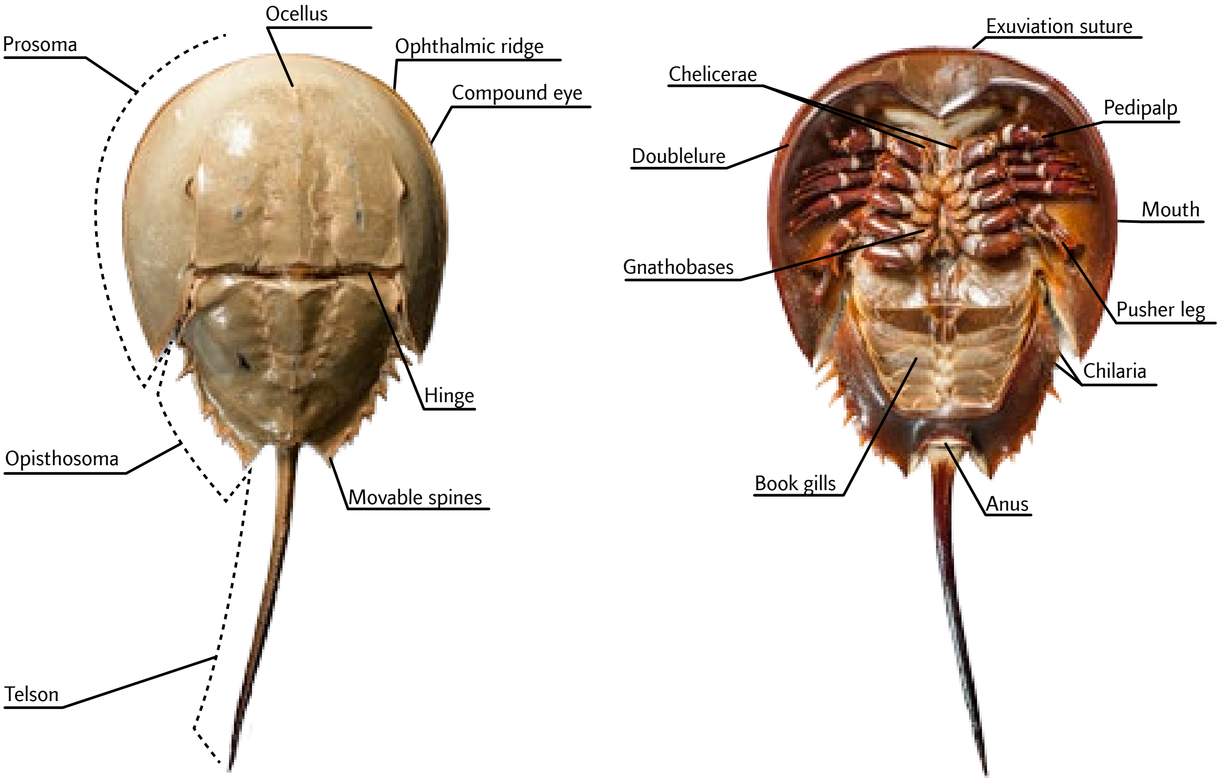


Fig. 6: Dorsal and ventral view of a typical Limulus polyphemus specimen. In this depiction a male animal is shown as indicated by the anatomy of the pedipalps which differ from adult females. The male carries a hook-like appendage which it uses to attach itself to the female during mating. Females in turn have pedipalps similar to the third to fifth appendages.

1.2 *Callianassa truncata* (ghost shrimp or mud shrimp)

Callianassa truncata which is also known as the ghost shrimp is a reference to its biotope. Animals usually grow to a size between 7 and 10 cm and live in extensive cave systems dug into soft sludge on the seafloor or on coral reefs. *C. truncata* can be found from the littoral zone down to the deep sea. Its habitat ranges from subtropical to arctic seas. The family consists of approximately 24 species.

Kingdom	*Animalia*
Phylum	*Arthropoda*
Subphylum	*Crustacea*
Class	*Malacostraca*
Order	*Decapoda*
Infraorder	*Thalassinidea*
Family	*Callianassidae*
Genus	*Callianassa*
Species	*Callianassa truncata*

Fig. 7: Lateral (left) and dorsal (right) view of Callianassa truncata. The major cheliped can clearly be identified.

The shrimps diet consists of depositions of silt (algae) and plankton. It is also worth mentioning the asymmetrically sized chelipeds which coined the nickname of one-armed shrimp. The animals used in this study were collected on the island of Giglio, Italy by Prof. Dr. Jürgen Markl and Dr. Julia Markl using a self-constructed sucker and scuba gear.

2 Biochemical Methods

2.1 Hemocyanin extraction and purification

A few ml of hemolymph were withdrawn from the dorsal sinus of a *Limulus polyphemus* with a syringe and subsequently centrifuged for 30 min at 3.300 g in an Eppendorf centrifuge. The cell free supernatant was then centrifuged for two hours at 130.000 g in an ultracentrifuge (Airfuge, Beckman, Munich). The hemocyanin pellets were resuspended in a low-salt buffer with 100 mM Tris HCl (pH 7,8), 10 mM $CaCl_2$, 10 mM $MgCl_2$. This procedure was performed by Dr. Wolfgang Gebauer.

2.2 Negative staining

Negatively stained specimens were prepared by the single droplet procedure (Harris and Agutter, 1970; Harris and Horne, 1991) using carbon coated grids that had been glow-discharged (Fig. 8). A single drop (5µl) of 0.1 mg/ml sample was picked up in a grid from a strip of parafilm. Most of the fluid was then drawn off by touching the side of the grid with a filter paper. Stabilizing buffer salts were removed by washing with three successive 10 µl droplets of water. The grid was then covered with a 10 µl droplet of stain (2% Uranyl acetate). The excess stain was then drawn off with a filter paper and the thin film of stain and sample material allowed to dry. The procedure was carried out at room temperature.

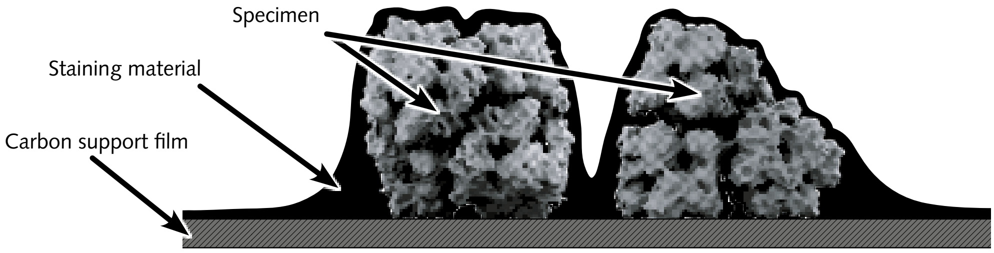


Fig. 8: Negative staining concept. Specimens rest on a carbon support film. They are engulfed in the dried staining material which increases contrast in the electron microscope.

3 Electron Microscopy

3.1 Cryo-plunging

For the process of vitrification of hemocyanin molecules, a modified CP3 cryo-plunger (Gatan Inc, USA) was used. The modifications – conceived and implemented by Dr. Frank Depoix – allow for controlled atmospheric conditions during the entire plunging process as opposed to commercially available off-the-shelf hardware.

Approximately 7-8 µl of purified protein (0.4 – 0.7 mg/ml) were applied to freshly glow-discharge treated holey carbon films, on copper EM grids (Quantifoil Micro Tools GmbH, Jena, Germany). After an incubation period, excess fluid was blotted and the grids plunged into liquid ethane.

3.2 Transmission-Electron-Microscopy (TEM)

3.2.1 TEM of *Limulus polyphemus* hemocyanin

Data was recorded by Dr. Frank Depoix in cooperation with Prof. Dr. Werner Kühlbrandt and Dr. Deryck Mills with a FEI Polara TEM (Max-Planck-Institute for Biophysics, Frankfurt am Main) at 300 kV acceleration voltage with 78k × instrumental magnification on Kodak SO-163 plate film. Cutout plate holders for low-dose cryo conditions where utilized. Nominal underfocus values lay between 1.1 and 3.1 µm.

3.2.2 TEM of *Callianassa truncata* hemocyanin

Negatively stained specimens were recorded with a Tecnai 12 transmission electron microscope at an accelerating voltage of 120 kV and 98k × instrumental magnification. CryoEM was performed with a FEG-equipped Tecnai F30 transmission electron microscope operating at an accelerating voltage of 300 kV and 49k × instrumental magnification on Kodak SO-163 plate film with a nominal underfocus between 1.8 and 3.3 µm. Cutout plate holders for low-dose cryo conditions where utilized.

3.2.3 Plate film development & digitization

All micrographs were developed for 12 min in full-strength Kodak D19 developer and digitized using a Heidelberg PRIMESCAN 7100 rotating drum scanner with a sampling size of 1 Å at specimen level to allow for a maximal resolution of 3 Å (a sample size of 1 Å is in theory capable to represent spatial frequencies up to 2 Å, however due to errors during image processing, a factor of 3 is suggested instead by LeBarron, 2008). 31 negatively stained micrographs and 82 cryoEM micrographs of *Callianassa truncata* hemocyanin and a total of 260 cryoEM micrographs of *Limulus polyphemus* were processed. Scanned micrographs were stored in a 16bit depth gray scale TIFF file format and subsequently pre-treated with the Silverfast 6 software in order to produce 8bit files.

4 Bioinformatic Methods

4.1 Single-particle 3D-reconstruction workflow overview

A major part of the iterative workflow seeks to improve the inherently low signal-to-noise ratio (SNR) of the projection images recorded within the TEM. With improved SNR imagery available, angular reconstitution is performed which then allows for the actual 3D-construction to be calculated. The resulting reconstruction is reprojected to produce projection images similar to those obtained in the microscope used as references for the following iteration. During this process, parameters are monitored to quantify the quality of the reconstruction. Fig. 9 presents an overview of the processes described in this chapter.

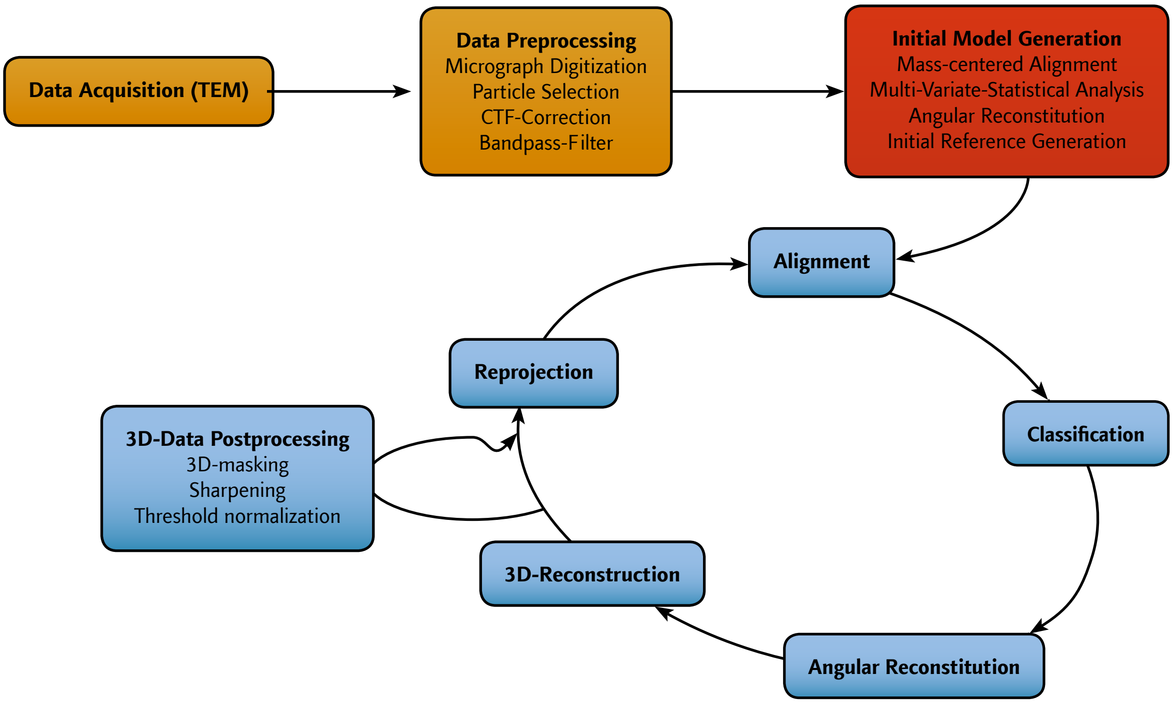


Fig. 9: Single-particle 3D-reconstruction work flow. See text and sub-chapters for detailed descriptions of each segment.

4.2 Particle selection and extraction

To better facilitate manual particle selection, image files were reduced to 1/4th size and treated with a low-pass block-convolution 2d-filter (IMAGIC-5 module BLOCK-2D). Manual particle selection was performed with the EMAN module BOXER which stores the coordinates of the selected particles. The coordinate values were then converted to match the full-size file version. Extraction into arrays of single particles was performed with a box size of 256 or 512 pixel square for the *C. truncata* and *L. polyphemus* dataset respectively.

4.3 Correction of the contrast-transfer-function (CTF)

Image formation in the TEM occurs due to elastic and inelastic scattering of electrons during interaction with the sample. Elastically scattered electrons experience a phase and amplitude shift. While in the case of unstained biological samples the amount of contrast derived from inelastic scattering and amplitude shift is negligible (Zhu et al., 1997), phase shift induced contrast carries most of the structural information. Thus, unstained biological samples can be approximated as weak phase objects which require high amounts of defocus in order to induce suitable phase contrast for imaging and subsequent image processing. The phase contrast itself is governed by the contrast-transfer-function (CTF) of the microscope (Fig. 10 panel A). As mentioned the CTF is a function of defocus but also of astigmatism, spherical aberration and electron wavelength (dependent on acceleration voltage). The function of the CTF is oscillating between +1 and -1 with each zero crossing producing a loss of information. In the microscope Thon rings (Thon, 1966) can be observed in Fourier transformations of the recorded image (power spectrum, Fig. 10 panel B).

Obviously the large amount of error (loss of information and inverted gray values) introduced by this effect has to be corrected in order to achieve high resolution reconstructions. For this purpose, the phases of negative parts within the CTF are flipped (=multiplied with -1, Fig. 10 panel C). This process requires precisely calculated zero crossing values and thus

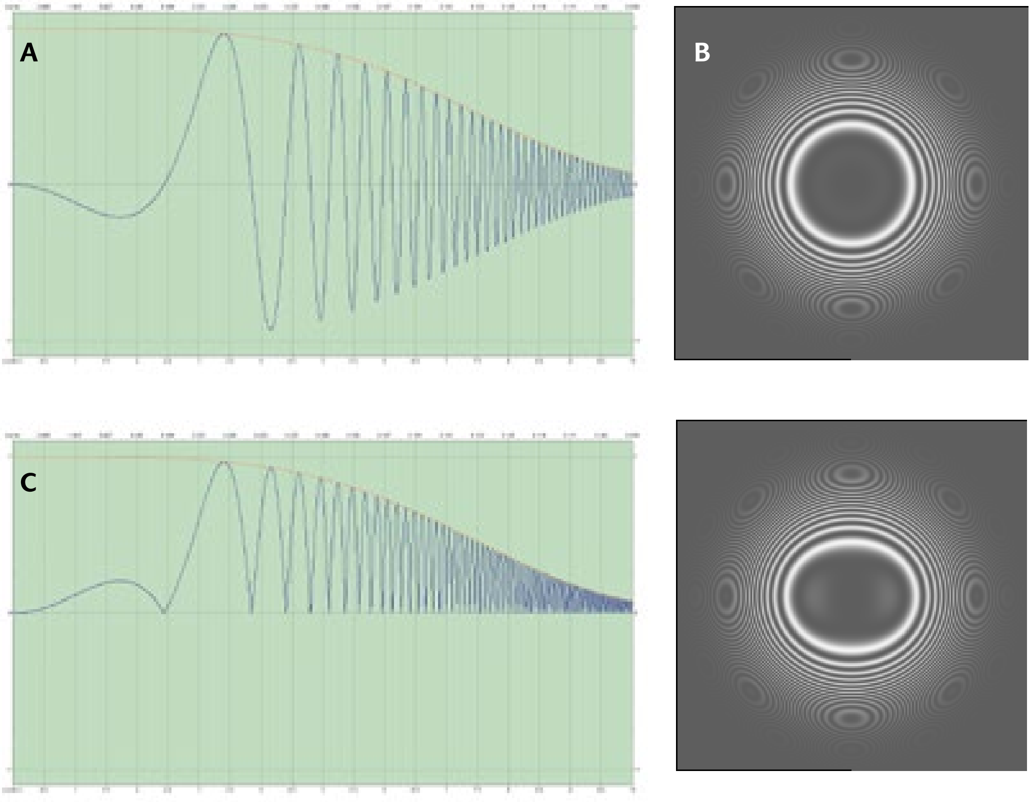


Fig. 10: Contrast-transfer-function (CTF) and power spectrum. Simulated CTF (A) with corresponding power spectrum (2D Fourier transformation) (B). The corrected or “flipped” CTF is shown in panel C. (all depictions were generated with the ctfexplorer software, www.maxsidorov.com/ctfexplorer/)

a determination of the actual defocus used. Both Ctffind3 and Findctf2d programs facilitate this process by simulating a CTF calculated from known TEM model specific parameters and fitting the Thon rings to the power spectrum of the recorded image thus determining the correct defocus. In addition, astigmatism and drift are easily detectable from the power spectrum and serve as an indicator of micrograph quality. In the case of Findctf2d, astigmatism is also corrected for each individual micrograph. Loss of information from zero crossings is compensated by recording data with a wide variety of defocus settings, equalizing the CTF for the whole of the dataset in later image processing without information gaps.

The CTF correction itself was performed by Findctf2d itself or in the case of defocus determination with Ctffind3 with the Imagic-5 module transfer.

4.4 Bandpass-filter

The CTF-corrected images are further treated to reduce the influence of noise and large scale gradients (e.g. due to uneven beam illumination or irregularities during film development). A high-pass filter removes low spatial frequencies representing structural information much larger than the particle of interest. In turn a low-pass filter removes high frequencies containing only noise but no structural information. The resulting filter curve is a bandpass-filter (Fig. 11).

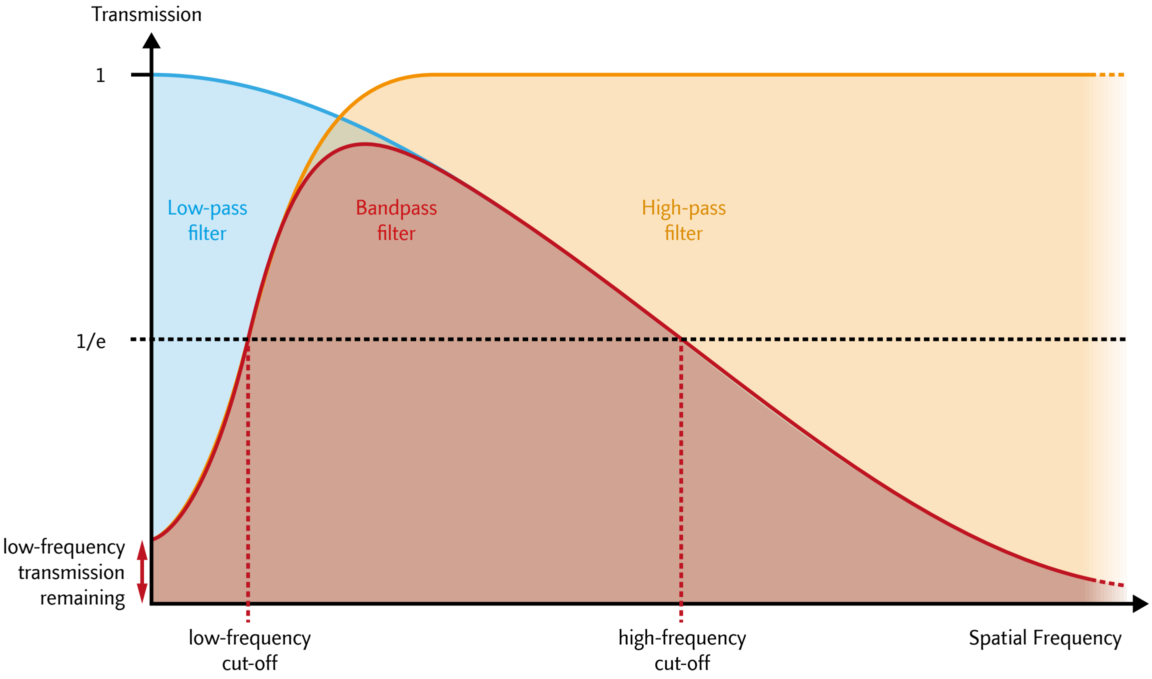


Fig. 11: Bandpass-filter. The bandpass-filter combines a low- and high-pass Gaussian filter to eliminate signals that might impede subsequent image processing.

The process of filtering the dataset before processing is important as to not disturb algorithms such as cross-correlation functions susceptible to strong signals. In addition a circular mask was applied to each particle to lessen the influence of adjacent particles or other unwanted signals. The radius of the mask was equal to 1/2 of the box size.

4.5 Multiple Reference Alignment

The major drawback of imaging vitrified biological samples is their radiation sensitivity and therefore the necessity to keep radiation exposure at a minimum. The resulting data is severely hampered by a low signal-to-noise ratio (SNR). In order to improve SNR levels, identical particle-projection images are averaged. The first step to facilitate averaging is to modify the translational and rotational positions of each particle with respect to several reference images by an algorithm termed multi-reference-alignment (MRA).

A reference reconstruction is used to compute projection images with defined angles from all directions (IMAGIC-5 module THREED-FORWARD). The angular step size is governed by the size and resolution of the dataset and phase in the reconstruction process.

First all images are aligned to reference one, and a correlation coefficient is assigned to each aligned image. Then the images are aligned to reference two and again correlation coefficients are calculated. Aligned images with the best alignment quality (highest correlation coefficient, with respect to reference one or two) are kept and the result is a merged data set with differently aligned images. This procedure is repeated until all references have been used (Fig. 12).

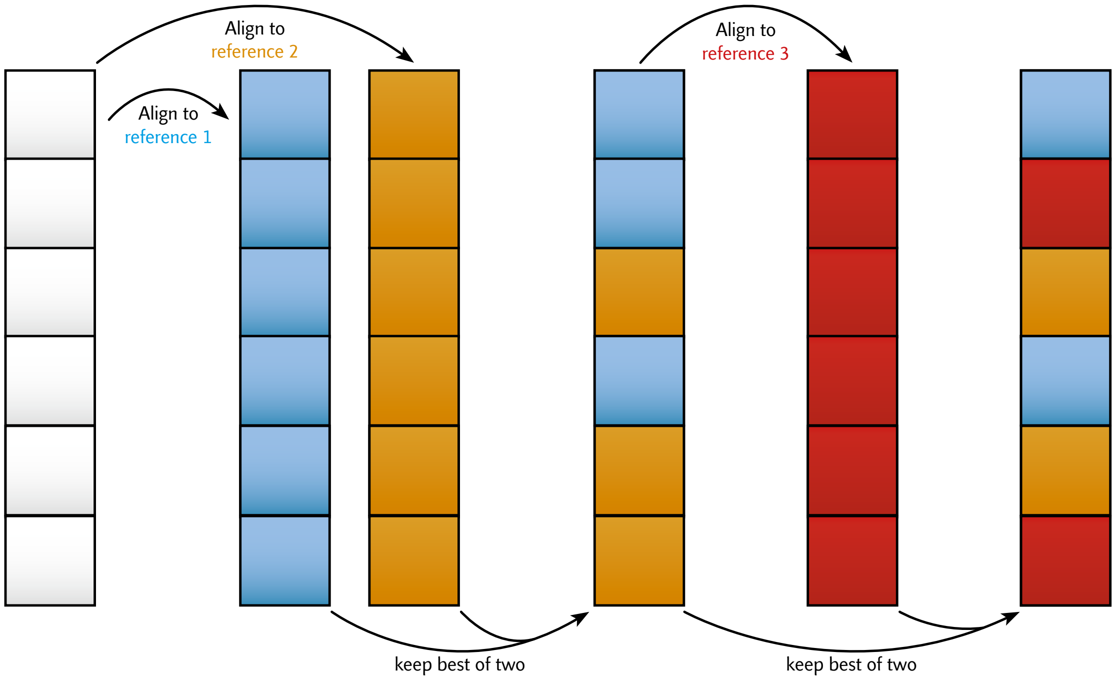


Fig. 12: Schematic MRA procedure overview. All particles are aligned to a reference after which a cross-correlation coefficient is determined. After alignment and correlation calculation to the following reference, the highest correlations are stored while lower reference correlations are discarded. The final output contains each particle with their highest correlation coefficient associated (according to van Heel et al., 2000. modified).

Each projections orientation parameters need to be found with respect to its corresponding reference: translation (X and Y plane) as well as rotation angle. In-plane Euler rotation is determined during rotational alignment (see 4.7.1). The translational alignment shifts the projection to align to the reference image. Both processes are run iteratively to optimize the alignment, each iteration starting with a rotational alignment followed by a translational one. Both algorithms are described in detail in the following section.

4.5.1 Translational alignment

Both TEM-projection and reference images are 2D-transformed into Fourier space (Fig. 13). Both Fourier transformations (FT) are multiplied resulting in a cross-correlation function (CCF, Saxton and Frank, 1977). When retransformed into real space, the CCF exhibits an off-center peak. Taking the angle and distance from the image center to the observed peak into account, a vector is calculated which equals the TEM-projections x/y displacement compared to the reference image.

4.5.2 Rotational alignment

TEM-projection and reference images are converted into a cylindrical coordinate system (also known as polar coordinates) followed by 1D-FTs, multiplication and retransformation to real space (Fig. 14). The result is a rotational correlation function (RCF). Depending on

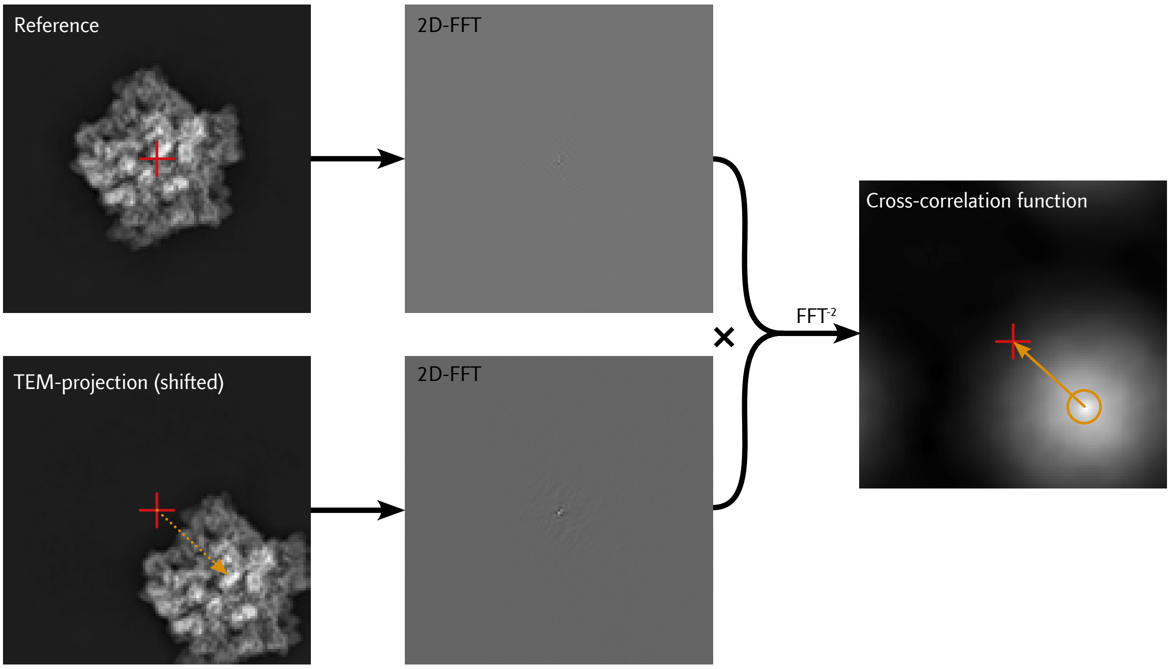


Fig. 13: Example of a translational alignment procedure. Both experimental TEM-projection image and reference are Fourier-transformed and multiplied in Fourier space. The TEM-projection was artificially shifted towards the bottom right for this example (dotted arrow). After retransformation into real space (=FFT^{-2}), the cross-correlation function shows a clear brightness peak (orange circle). The vector (orange arrow) drawn from the center of the peak to the image center corresponds to the vector the TEM-projection image is shifted in relation to the reference.

the point group symmetries inherent in the target molecule one or more brightness peaks within the RCF can be observed. These peaks, measured in degrees along the x-axis of the cylindrical coordinate system, correspond to the rotational shift between TEM-projection and reference image.

4.6 Classification of images

After successful alignment, the TEM-projection images are ready to be averaged. However, it has yet to be determined which projections are identical to each other and thus will result in an improved signal-to-noise-ratio when averaged. A multi-variate-statistics (MSA) approach is employed to classify and group identical projections into image-stacks (termed classes).

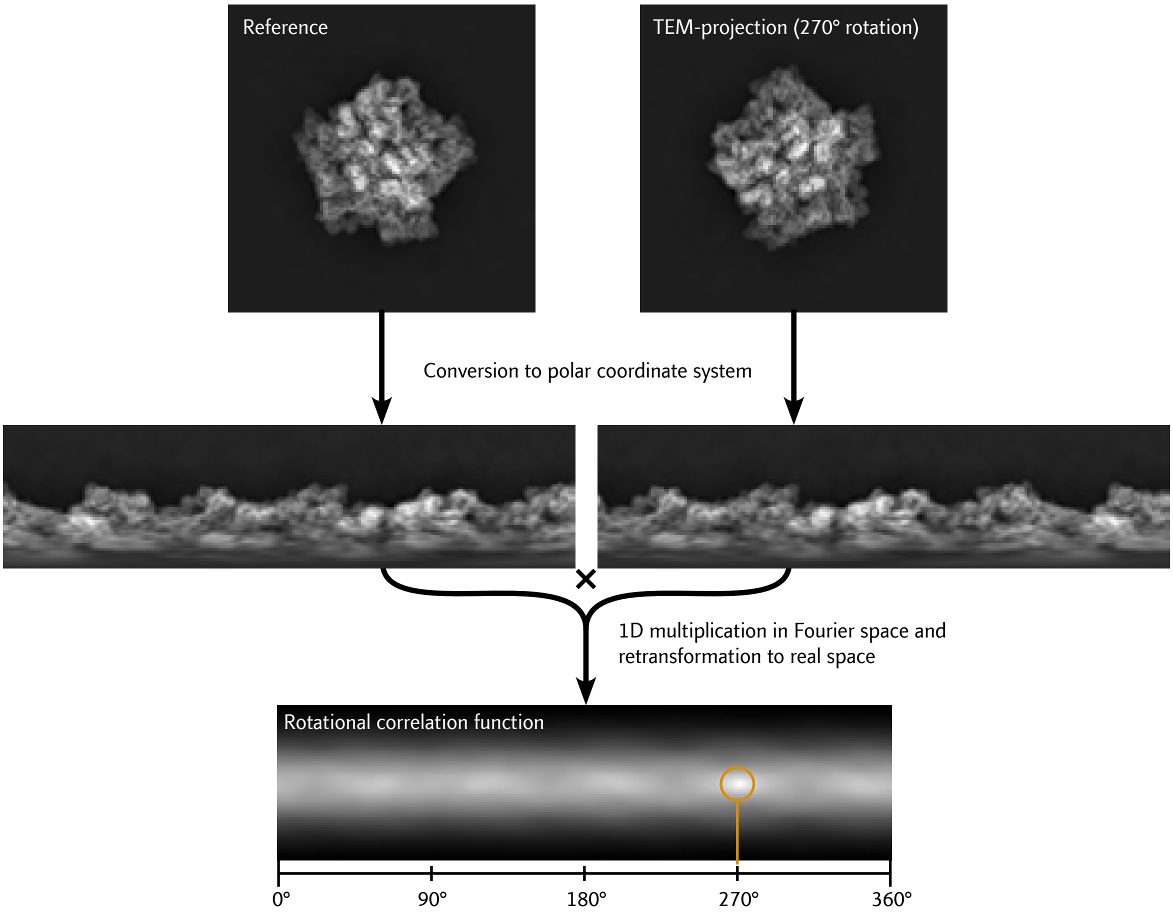


Fig. 14: Example of a rotational alignment procedure. Both experimental TEM-projection image and reference are converted into a cylindrical coordinate system representation. Both images are 1D-Fourier transformed, multiplied in Fourier space and retransformed to real space. The result is a rotational correlation function. A distinct brightness peak can be discerned (orange circle), whose position plotted on the x-axis in degrees coincides with angle of the projections rotation in relation to the reference (270° in this example).

4.6.1 Hyperspace construction

The MSA consist of a eigenvector-eigenvalue analysis (van Heel and Frank, 1981; van Heel, 1984). The process of this analysis starts with the creation of a multi-dimensional hyperspace. Each individual pixel of a projection-image is represented by a dimension within the hyperspace. As a result each individual image itself can be described by a single point within the hyperspace. Accordingly a set of images forms a point cloud where similar projections will be represented by points closer to each other and projections with little similarity will find their corresponding points being divided by greater distances. In addition, the point cloud may be separated into several sub-clouds representing various sub-classes of particles inherent to the sample.

4.6.2 Eigenvector analysis

As this approach produces a large amount of pixel-values, a data-reduction strategy is employed. For this purpose a new coordinate system is fitted to the point cloud in a way that its first axis points to the maximum inter-image variance within the cloud. The second axis, perpendicular to the first, then points to the second-largest variance and so forth.

A maximum of 69 unit vectors (eigenvectors) can be computed which are sufficient to represent all differences in the dataset which can attributed to structural features. The eigenvectors in the new coordinate system correspond to single points in hyperspace and therefore represent images, which are called eigenimages, whose significance is indicated by their eigenvalues (Schatz, 1992).

4.6.3 Classification

Optimal classification is achieved when the intra-class variance is kept at a minimum whereas at the same time inter-class variances are maximized. To that end a hierarchical ascendant classification algorithm (HAC) with moving elements refinement is utilized (van Heel, 1984; van Heel et al., 2000). At the beginning of the procedure, the number of classes equals the number of images in the dataset. The two classes with the least distance are merged producing a larger class and so on until all images belong to a single class. The algorithm tries to minimize intra-class variance during this procedure (Ward, 1982).

To produce the final amount of output classes as specified by the user a cut is made at the appropriate hierarchy level stemming from the merging process described above. The classification itself is then refined according to the optimum variance criterion allowing images to switch to other more suitable classes until the classes remain stable (Fig. 15). Lastly all projections that were assigned to the same class are averaged thereby improving the SNR level.

The MSA algorithm implemented in IMAGIC-5 is split into three modules: MSA-RUN constructs the hyperspace and eigenvector analysis, MSA-CLA contains the HAC classification algorithm and MSA-SUM performs the actual averaging process. To eliminate unwanted influences from the periphery of the boxed particles, a circular mask is applied during the MSA. It was built with the TEST-IMAGE module.

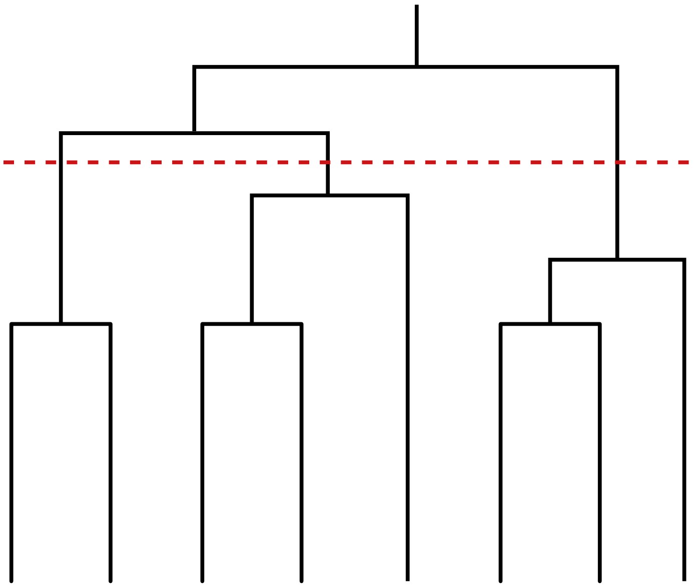


Fig. 15: Hierarchical ascendant classification tree. At the beginning of the procedure, the number of classes equals the number of images in the dataset. Two classes are merged producing a larger class until all images belong to a single class. A cut is made at the level that corresponds to number of classes specified by the user (dashed line, in this example resulting in a total of 3 classes). The projections are then averaged as assigned at the chosen hierarchy level (according to Schatz, 1992, modified)

4.7 Particle orientation determination

With the signal-to-noise ratio improved by the averaging process, the first step towards a 3D-reconstruction can be taken. As the molecules enclosed within the vitreous ice-layer have preferably random angular orientations, the precise relationship between those orientations must be determined. Several methods have been developed including random conical tilting (Radermacher et al., 1987; Radermacher, 1988), angular reconstitution (van Heel, 1987) and projection matching (Penczek et al., 1994).

In this study both angular reconstitution and projection matching were used. The input is derived solely from the TEM-projection images themselves. At the initial-model stage, angular reconstitution provides sufficient accuracy whereas in later refinement stages, projection matching is the method of choice.

4.7.1 Euler angles

Molecules enclosed within the vitreous ice-layer have all six degrees of freedom – translation through x, y, and z as well as corresponding rotation through α, β and γ (also known as Euler angles). The in-plane degrees of freedom are x, y and α whereas z, β and γ are out-of-plane degrees of freedom (Fig. 16).

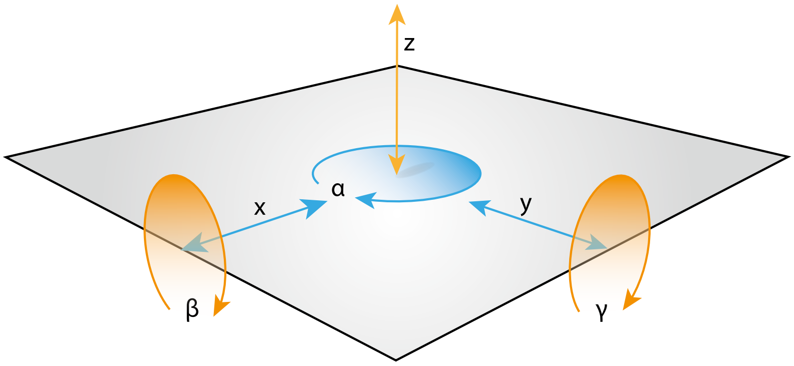


Fig. 16: Degrees of freedom for molecules embedded in vitreous ice. In-plane degrees of freedom are colored blue, out-of-plane degrees are colored orange.

As the data collected in the TEM is a projection along the z-axis, all TEM-projection images contain the same value for z. In turn this value can be neglected for the purpose of Euler angle determination. The degrees described by x and y equal 0 after a successful translational alignment. The α-angle describes the rotation of the object about the projection axis z. This rotation is being removed due to the rotational alignment. The only remaining unknown values which need to be determined are the out-of-plane rotations described by β and γ. The particle main axis is the molecules axis with the highest symmetry. In case of *L. polyphemus* and *C. truncata* hemocyanin, a D2 point-group-symmetry can be assumed. D2 stands for a two-fold dihedral symmetry. The tilt of the particle main axis relative to the projection axis z is described by β. The γ-angle describes the rotation around the particle main axis itself (Fig. 17).

4.7.2 Angular reconstitution

The common line projection theorem states that every 2D-projection of a 3D-object shares a 1D-common line with other 2D-projections of the same 3D-object (van Heel, 1987). A similar principle applies to Fourier-space (DeRosier and Klug, 1968).

By determining the common line, the relative orientations of the Euler-angles between averaged TEM-projection images can be found. A minimum of three different projections is always required to stabilize the spatial orientation of the projections in respect to each other as two projections are able to rotate freely around their common line (Fig. 18).

2D-class-average projections are converted into sinograms (all line projections, Fig. 19). A sinogram consists of 360 1D-projection lines through the 2D-class-average resulting from rotating the image in 1° steps and calculating a projection line at each step.

To find a common line, the sinograms of both projections are compared line by line to produce a cross-sinogram-correlation-function (CSCF). This function shows a peak (or several peaks depending on the symmetry of the sample) corresponding to the pair of matching common lines. As the iterative reconstruction process progresses, each addition of a projection into an existing euler-angle assigned image set leads to a search through all

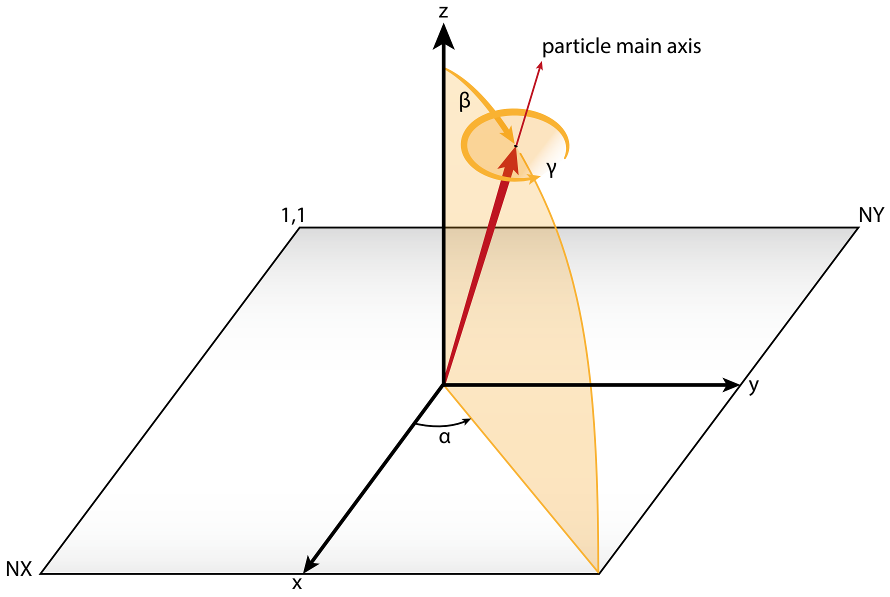


Fig. 17: Definition of Euler angles in the IMAGIC*-5 coordinate system.* The origin 1,1 is located top left, NX and NY specify the number of lines on their respective axis. After rotational and translational alignment, only the out-of-plane β and γ angles (orange) remain to be determined. The particle main axis is colored red. See text for further details (according to Stohr, 2007, modified).

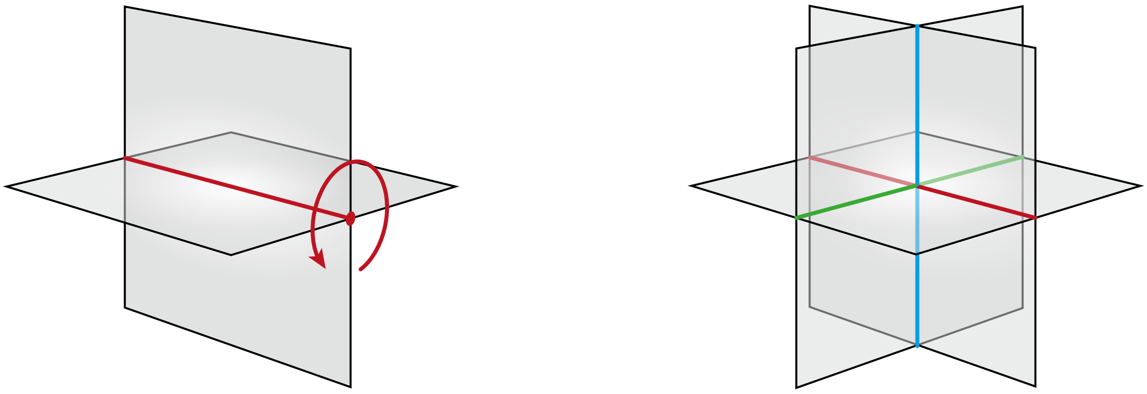


Fig. 18: At least three projections are needed to stabilize an angular reconstitution via common lines. Left: two projections share a common line both in real and Fourier space. However, they can freely be rotated along the common line axis (red). Right: a third projection introduces two additional common lines (green and blue) thus stabilizing the projections relative spatial orientations (according to Stohr, 2007, modified).

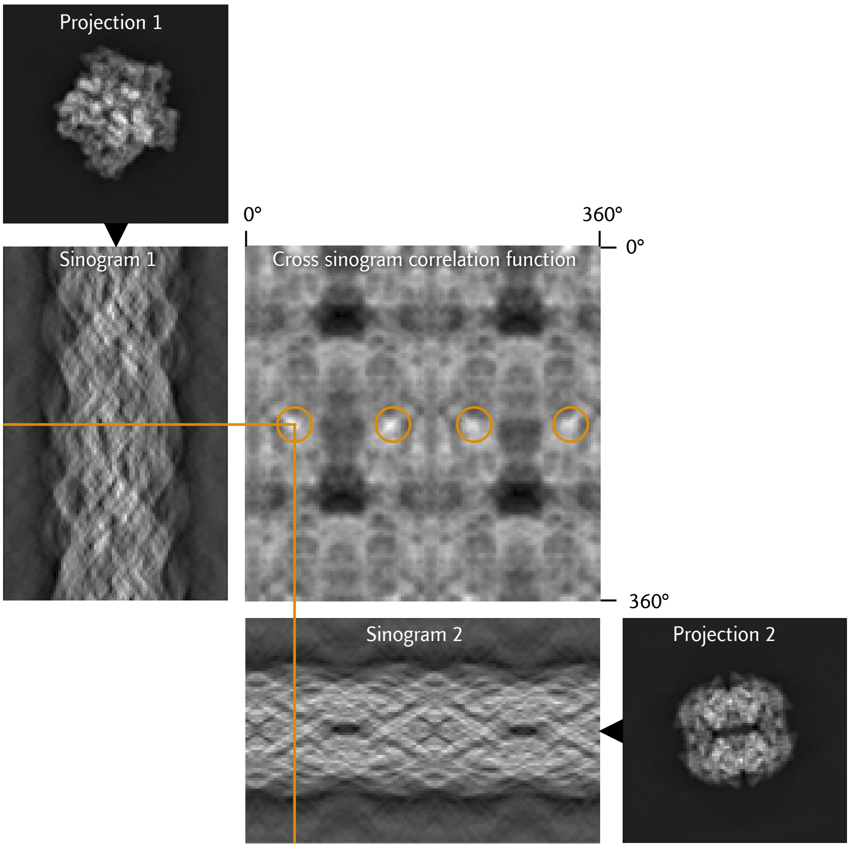


Fig. 19: Common line determination by cross sinogram correlation function. Two projection images from the *Limulus* dataset (projection 1 and 2, in this case reprojections have been used for improved SNR) are converted into their respective sinogram depictions (sinogram 1 and 2). A sinogram is a stack of 2D-line projections of an image rotated through 360°. The cross sinogram correlation function is the result of a multiplication of each line of sinogram 1 with each line of sinogram 2. Four brightness peaks are visible (orange circles) as expected because of the molecules D2 symmetry. An exemplary common line is also depicted (orange) (according to van Heel et al., 2000, modified).

orientations within the asymmetric triangle of the target molecules point-group-symmetry (Schatz et al., 1995). The asymmetric triangle is the unique building block which is multiplied during the application of the appropriate point-group-symmetry. It is described as part of a sphere spanning the area from one pole to the equator and covering, in the case of D2-symmetry, 180° of that hemisphere.

4.7.2.1 Anchor-set

After an initial model has been obtained (see 4.12) a set of reprojections can be produced. These lie within the particles asymmetric triangle (Orlova and van Heel, 1994, Schatz et al., 1995). In comparison to class-wise assigned Euler-angles, the anchor set provides known and evenly spaced Euler angles as well as better image quality. The determination of Euler angles was performed with IMAGIC-5s module EULER.

4.7.2.2 Projection matching

The projection matching technique relies solely on the cross-correlation values obtained during a multi-reference alignment procedure. As the accuracy of the alignment and reconstruction quality increases with each iterative cycle, the precision of the angular reconstitution is improved as well (Harauz and van Heel, 1986a).

It is assumed that all TEM-projection images that were aligned to the same reference during the MRA represent the same spatial orientation as the reference. The artificially projected reference has inherently known Euler angles and as such, all corresponding TEM-projection images are averaged and their Euler angles set equal to the references angles. Furthermore the user has the option to purge particles below a certain cross-correlation threshold in order to improve the quality of the class-averages.

While the projection matching method is able to produce high-resolution reconstructions, on the flip-side it is prone to bias by its strong dependency on cross-correlation values. For this study it was only used after a medium resolution reconstruction was achieved by MRA/MSA methods as described. Both Imagic-5 as well as Eman and Sparx offer projection-matching work flows (integrated into the major refinement loop refine).

4.8 3D-reconstruction

With the spatial orientation of each class-average determined, a 3D-reconstruction can be calculated. The algorithm employed for this study is that of a exact filter back projection (Harauz and van Heel, 1986b; Radermacher, 1988; Schatz et al., 1995).

The algorithm back-projects each class-average along their assigned Euler angles. A filter is needed to circumvent over-representation of low spatial frequencies (Fig. 20). This filter is calculated individually for each projection contributing to the 3D-reonstruction.

The final output consists of a stack of sections which conforms in size to the box-size and sampling step chosen during digitization and particle extraction (e.g. 512 sections at 1Å thickness per section and each section dimensioned at 512 x 512 pixels). The corresponding Imagic-5 module is called true-threed. In the case of Eman, the algorithm is contained within the major refinement loop refine. Visualization software such as Chimera or ePMV is then used to calculate surface renderings of the reconstruction.

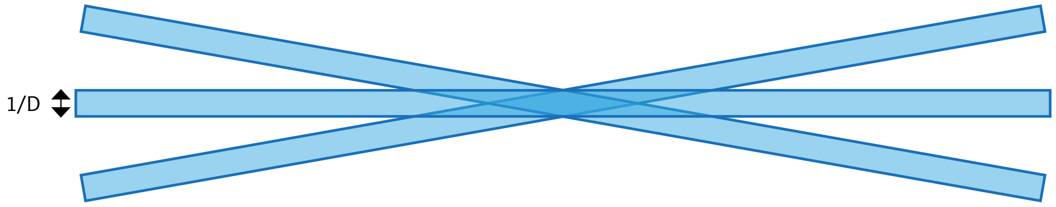


Fig. 20: The unfiltered backprojection algorithm introduces bias towards low spatial frequencies. A 2D-projection of a 3D object corresponds to a central section of the objects 3D-Fourier transformation (DeRosier and Klug, 1968). The central sections (each stemming from a 2D projection) have a thickness of 1/D, D being the largest linear dimension of the 3D-object. In Fourier space low frequencies are found at the center and thus overlapping areas (middle to deep blue) indicate low frequency overrepresentation (according to van Heel, 2000, modified).

4.9 3D-data post processing

4.9.1 3D-mask generation

A finished 3D-reconstruction may still contain a certain amount of background noise. This noise can be amplified by subsequent image processing tasks such as sharpening. A 3D-mask is built around the reconstruction to delete any such noise.

For this study, 3D-masks were generated with the option AUTOMASK2 from the PROC3D module of EMAN. It applies an automatically generated 'form fitting' soft (Gaussian) mask to the model. First the threshold for masking is determined by iso-surface generation in UCSF CHIMERA. The optimal threshold is found when no noise is connected to the molecule while at the same time the density representing the molecule itself is fully interconnected. Further parameters include: a radius in pixels (specifies the size of a sphere which barely comes into contact with the inside edge of the model, a value of 20 for the *Limulus* reconstructions) and the width in pixels of a soft mask shell (a width of five pixels was used for *L. polyphemus* reconstructions).

4.9.2 Sharpening

Loss of contrast as well as computational factors may lead to a decay of vital high frequency information carrying high resolution information. This may occur for example due to specimen movement, radiation damage, local charging or inaccurate angular reconstitution parameters. In x-ray crystallography the sum of these effects is modeled by a Gaussian decay of structure factors with a temperature factor (B-factor, Rosenthal and Henderson, 2003). A part of the strategy to regain lost high resolution features is to apply a high-pass filter to weaken the amplitude of spatial frequencies representing coarse structural information. IMAGIC-5 offers the module THREED-FILTER, EMAN contains PROC3D with appropriate filter options.

In addition inverse B-factor treatment is used to restore suppressed high-frequency information. Both processes are commonly known as "sharpening" as high frequency components are emphasized. In this study automatic B-factor determination and sharpening was performed with the EM-Bfactor software (Fernandez et al., 2008).

As high-resolution details are gradually superseded by noise, care must be taken not to introduce noise artifacts during the sharpening process. As a precaution the 3D-reconstruction is filtered to a nominal resolution after sharpening, thus removing any unsupported detail present.

4.9.3 Threshold normalization

In order to generate references which contain all areas of the molecule represented by comparable gray values, a normalization with EMANs PROC3D module, option NORMFILT was performed after 3D-masking and before sharpening. This step avoids heterogenous threshold-values throughout the 3D-reconstruction.

4.10 Reference image generation

The last step in preparation for a new reconstruction cycle is the generation of reference images for the multiple reference alignment. The 3D-reconstruction is used as a template to re-project projection images with known spatial orientation at a user-specified interval covering all possible orientations of the molecule. As the quality of the 3D-reconstruction improves, finer step-sizes are chosen. Reference image generation was performed with the Imagic-5 module threed-forward or as included in Emans refinement loop refine.

4.11 EMAN workflow

Single particle 3D-reconstruction in Eman is comparable to a work flow in Imagic-5 using the projection matching method (described in 4.7.2.2) if a sufficiently detailed initial model is available. For the *L. polyphemus* hemocyanin 3D-reconstruction presented in this study, the steps as described in 4.9 (3D-data post-processing) were performed after each iteration.

4.12 Initial model generation via C1-startup

The process of single-particle 3D-reconstruction is an iterative procedure usually run many times, feeding the result from the previous cycle into the next. It involves massive averaging procedures as well as the possibility to impose point-group symmetries. Consequently great care must be taken not to influence the outcome of the reconstruction from the very beginning with a false starting point. Nonetheless some sort of initial model must be used in order to jump-start the iterative workflow. Various approaches are feasible, in the case of this study the so called C1-startup method has been used.

To roughly center the particles without a specific reference, a reference-free alignment is performed first. This procedure calculates an overall mass center of all particles contained in the dataset and shifts the particles on the x and y axis only. Rotational alignment is not used at this time. Subsequently an initial MSA analysis and classification take place. The user then selects 6-9 classes with a strong SNR which are fed in random combinations of three class-averages into the Imagic-5 module true-threed. With the C1-startup option selected, common line angular reconstitution and 3D-reconstruction is performed as described in the previous sections. The user is then able to select the most promising initial model to commence iterative processing. Thus the initial model is effectively generated from the dataset itself without external references that might induce unwanted bias.

4.13 Quality control

The resolution of the 3D-reconstruction as determined by Fourier shell correlation (FSC) serves as an indicator of quality throughout the iterative process. The FSC quantifies the correlation between two 3D-reconstructions, each of is generated from either half of the available set of class-averages by measuring the normalized cross correlation as a function of spatial frequency (Harauz and van Heel, 1986; Orlova et al., 1997).

The widely accepted criterion for FSC resolution assessment is the 0.5-criterion which measures the resolution at the point where the FSC-curve intersects half its maximum value. This criterion is considered to be very conservative. Additionally several other criterion have been proposed of which only the 1/2-bit criterion is of interest for this study. Developed and published by van Heel and Schatz in 2005, a threshold curve is calibrated to roughly yield resolution values comparable to values obtained with x-ray crystallography methods. The curve represents the resolution level at which sufficient information has been gathered for interpretation.

FSC analysis in general is prone to computationally induced influences such as over-fitting of noise resulting in increased correlation of high spatial frequencies. An important aspect of quality control remains the appearance of the 3D-reconstruction itself. If a high resolution reconstruction is indicated by the FSC, correspondingly high resolution features such as distinguishable α-helices should be visible in the reconstruction (e.g. at resolutions below 9 Å).

4.14 Molecular modeling

The concept of homology protein modeling is that the conformation of a target-protein can be predicted by using experimentally determined protein structures as templates. Target and template have to share a similar amino acid sequence (Rost & Sander, 1996). This approach is feasible because a small change in protein sequence usually results in little to no change of the resulting 3D-structure (Hubbard & Blundell, 1987).

Homology modeling was performed using the software modeller 9v4 or the Swiss-model web server (Guex and Peitsch, 1997; Arnold et al., 2006). First, an alignment of the sequence to be modeled with related protein structures has to be calculated (ClustalW2 and Clustal Omega: Higgins and Sharp, 1988; Thompson et al., 1994; Thompson et al., 1997; Larkin et al., 2007). Then spatial restraints on the sequence using the alignment are extracted and the restraints to obtain the 3D-model are satisfied.

The spatial restraints include: Homology-derived restraints on the distances and dihedral angles in the target sequence, extracted from its alignment with the template structures. Stereochemical restraints such as bond length and bond angle preferences, obtained from the CHARMM-22 molecular mechanics forcefield. Statistical preferences for dihedral angles and non-bonded inter atomic distances, obtained from a representative set of known protein structures (Eswar et al., 2003).

Accuracy of the molecular models tends to increase with the target-template sequence identity. A model of a protein sequence with greater than 35% identity to a solved structure can expected to obtain an accuracy approaching that of a low resolution X-ray structure or a medium resolution NMR structure (Sanchez and Sali, 1997).

4.15 Flexible fitting

The flexible fitting workflow was developed utilizing the algorithm itself (DireX) followed by energy minimization and quality control steps.

4.15.1 DireX

DireX employs Deformable Elastic Network (DEN) restraints to reduce over fitting by decreasing the effective number of degrees of freedom used in the refinement. Missing or reduced density due to flexible parts of the protein can lead to artifacts in the structure refinement, which is addressed through the concept of restrained grouped occupancy refinement (Wang and Schroeder, 2012).

Each flexible fitting session cycled for 3000 iterations. The DEN parameters were set to move exclusively towards the reference structure and keep existing secondary structure elements intact.

4.15.2 Energy minimization with the YASARA server

Part of the comprehensive YASARA (Yet Another Scientific Artificial Reality Application, Krieger et al., 2009) package, the energy minimization algorithm applies a so called knowledge-based approach: A vast database of valid structures is used to statistically evaluate the model in question. A force-field incorporating the knowledge-based potentials then refines the structure. The energy minimization server is available on-line at http://www.yasara.org/minimizationserver.htm

4.15.3 What-If molecular model evaluation service

What If is a versatile molecular modeling package that is specialized on working with proteins and the molecules in their environment like water, ligands, and nucleic acids (Vriend 1990; Hekkelman et al., 2010). For this study only the protein model check component has been used to evaluate the quality of models modified by flexible fitting software.

4.15.4 Molprobity

The Molprobity web server (Lovell et al., 2003; Chen et al., 2010) was used to identify skewed rotamer angles of amino acid side chains after receiving the raw molecular model. Also, the Ramachandran plot calculation was used to gauge overall molecular model validity.

4.16 Computing Hardware

CPU-hour-intensive calculations were carried out on a Beowulf-type cluster with a total of 358 cores (Transtec AG, Tübingen). Each processing node contains two 2.3gHz AMD Opteron octacore with 16 GB RAM per CPU. Several types of OS were used: Debian 3.1 on the HPC-cluster and Ubuntu Linux as well as Microsoft Windows 7 64-bit on personal workstations.

4.17 Visualization

4.17.1 UCSF CHIMERA

CHIMERA is a highly extensible, interactive molecular graphics system. It was used for visualization and analysis of both cryoEM density maps and molecular models, rigid body fitting, rotamer modification (Pettersen et al., 2004; Goddard et al., 2007). It can be downloaded from http://www.cgl.ucsf.edu/chimera/download.html

4.17.2 Autodesk Maya 2012 with MentalRay

Maya is a comprehensive software package suitable for complete 3D-animation workflows (modeling, shading, animation, rendering). MentalRay was used for rendering output.

4.17.3 Extensible Python Molecular Viewer (ePMV)

ePMV is a framework for various 3D-animation software packages to allow for PDB and other structural data visualizations directly within the supported host software packages (Johnson & Ludin et al., 2011). In this study it has been used in conjunction with Autodesk Maya 2012.

4.17.4 Color schemes

Color schemes were adapted from http://www.ColorBrewer.org (Brewer, 2011)

4.17.5 Sequence alignment visualization

Sequence alignment graphs were created with ESPript (‘Easy Sequencing in PostScript’, Robert et al., 2014).

4.17.6 Quantification of c-alpha movement after flexible fitting

Histogram plots quantifying c-alpha movement were produced with the R project for statistical computing version 3.1.2 (R Core Team, 2016).

C Results & Discussion

1 3D-reconstruction of *Callianassa truncata* hemocyanin

1.1 Initial model

To facilitate an unbiased 3D-reconstruction process, an initial model was generated from negatively stained *Callianassa truncata* hemocyanin. While the resolution can be expected to stay limited, it is sufficient for the purpose of the initial model. Also, the high contrast of the micrographs might help to identify errors occurring in the 3D-reconstruction early on.

1.1.1 TEM with negative stain specimen preparation

Negatively stained specimen of isolated *C. truncata* hemocyanin were recorded with a Tecnai 12 transmission electron microscope at an accelerating voltage of 120 kV and 98k× instrumental magnification (Fig. 21).

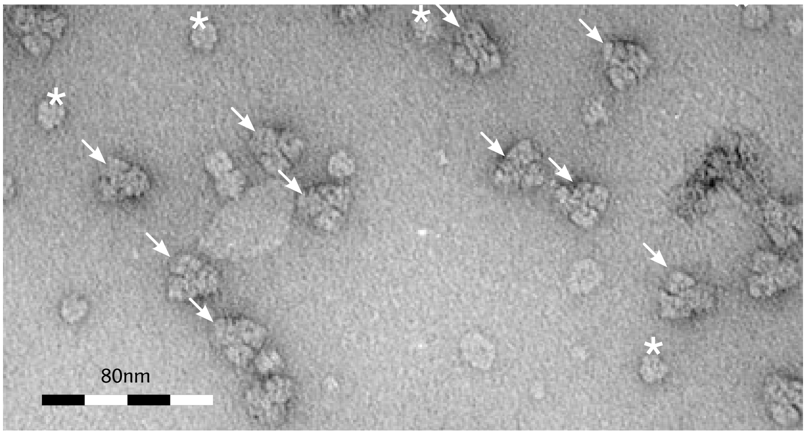


Fig. 21: Negative stain preparation of Callianassa truncata hemocyanin. Individual *C. truncata* hemocyanin molecules are visible (arrows). Dissociated hemocyanin (monomers, marked by asterisk) is present in the sample. A tendency to form molecule clusters is also apparent. Micrograph recorded at 98k x magnification. (scale bar total length = 80nm)

A total of 1091 particles were manually picked from the negatively stained micrographs and bandpass-filtered for further processing. A CTF-correction was not performed (Fig. 22).

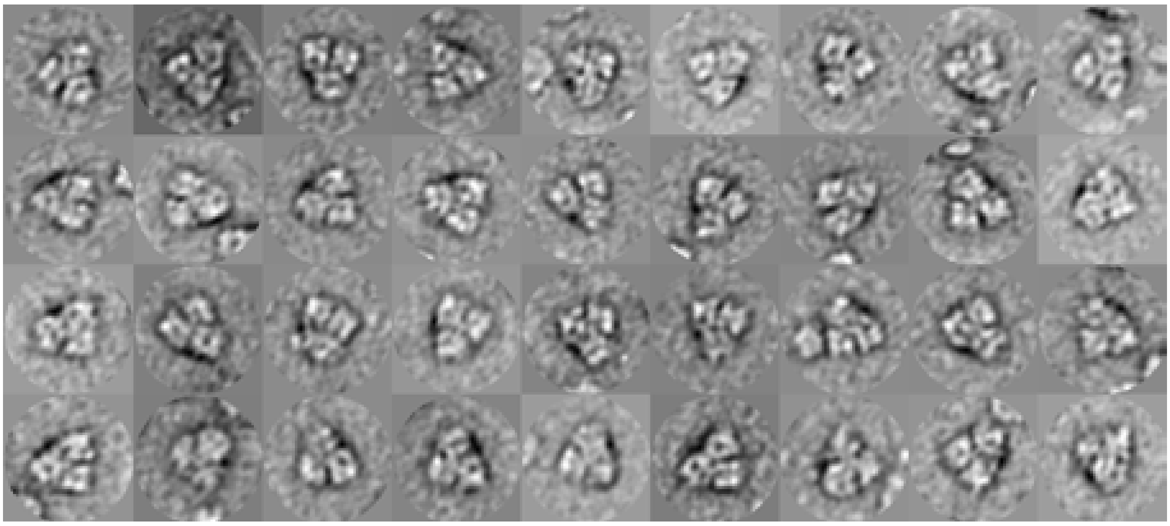


Fig. 22: Extracted and filtered particles from the negative stain preparation of Callianassa truncata hemocyanin. Many particles are in the preferred orientation of the molecule (a sample of 36 particles out of 1091 is shown).

1.1.2 Initial model

The initial model was obtained with the C1-startup method of IMAGIC-5, thus imposing no external references or other bias. After 16 iterations with MRA/MSA, a resolution of approx. 30-40 Å was achieved (Fig. 23). Due to preferred orientations of the molecule on the EM-grid carbon surface caused by the staining agent, only a very limited amount of tilt-

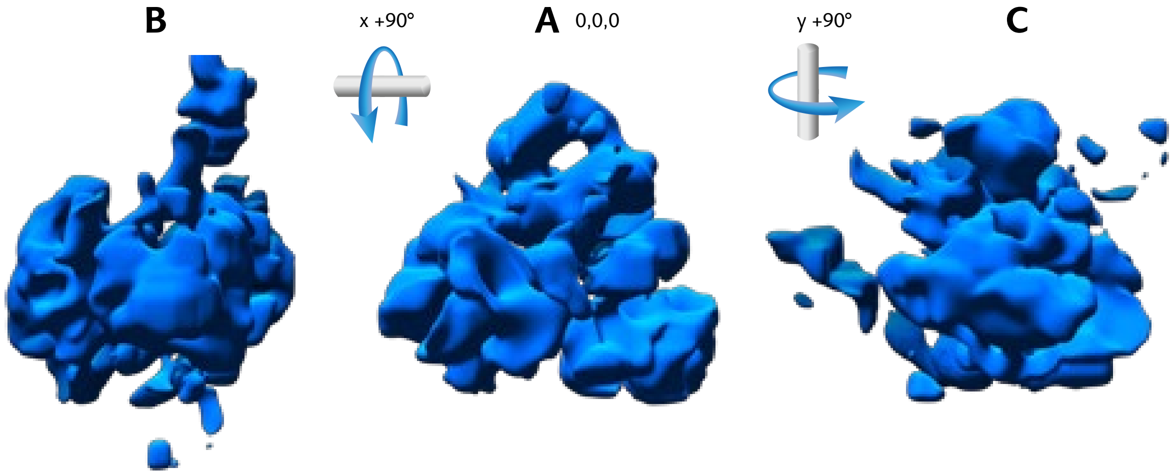


Fig. 23: Initial model of Callianassa truncata hemocyanin after 16 iterations. A: The preferred orientation of the molecule is represented by the particle main axis view. This roughly corresponds to a side view of the hemocyanin. B and C show tilted views.

ed projection images could be collected. This imposed a limit on the overall quality of the 3D-reconstruction.

The view along the particle main axis however represents a correct view of *C. truncata* hemocyanin with both 2x6mers discernible (Fig. 24).

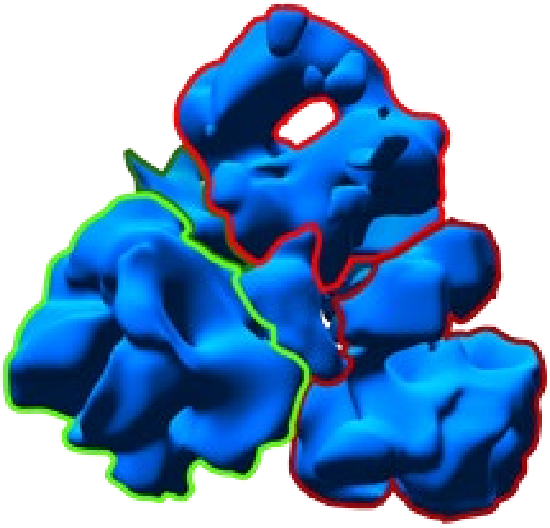


Fig. 24: Initial model main particle axis. All four hexamers are visible in this view of the 3D-reconstruction (arranged as dimer 1: red/orange and dimer 2: green/dark green).

1.2 3D-reconstruction from cryo-EM micrographs

1.2.1 Cryo-EM

Cryo-EM was performed with a FEG-equipped Tecnai F30 transmission electron microscope operating at an accelerating voltage of 300 kV and 49k × instrumental magnification. Cryo-EM micrographs (Kodak SO-163) were recorded with a nominal underfocus between 1.8 and 3.3 µm. After a visual quality inspection 82 cryoEM micrographs of *Callianassa truncata* hemocyanin were processed (Fig. 25).

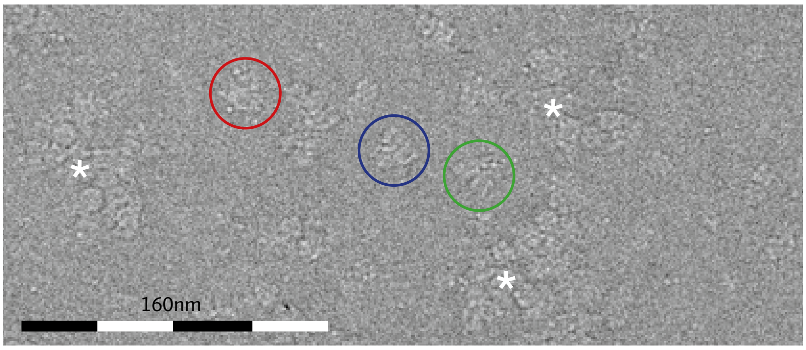


Fig. 25: Cryo-EM preparation of Callianassa truncata hemocyanin. Micrograph recorded at 49 kx magnification. Dissociated hemocyanin (dimers and monomers) is present in the sample. Molecule clusters can be observed (indicated by asterisks). Exemplary protein orientations: red circle – side view (see also Fig. 29 A); blue and green circles – tilted views (see also Fig. 29 B and C resp.;Total scale bar length: 160 nm)

From the cryo-EM micrographs, 7667 particles were manually selected with BOXER, and extracted into 468 × 468 pixel boxes. The contrast transfer function (CTF) for each micrograph as well as the correction for the phase reversals of the CTF was determined. The particles were 2 × 2 averaged to give a final sampling of 2 Å/pixel and subsequently band pass filtered and normalized.

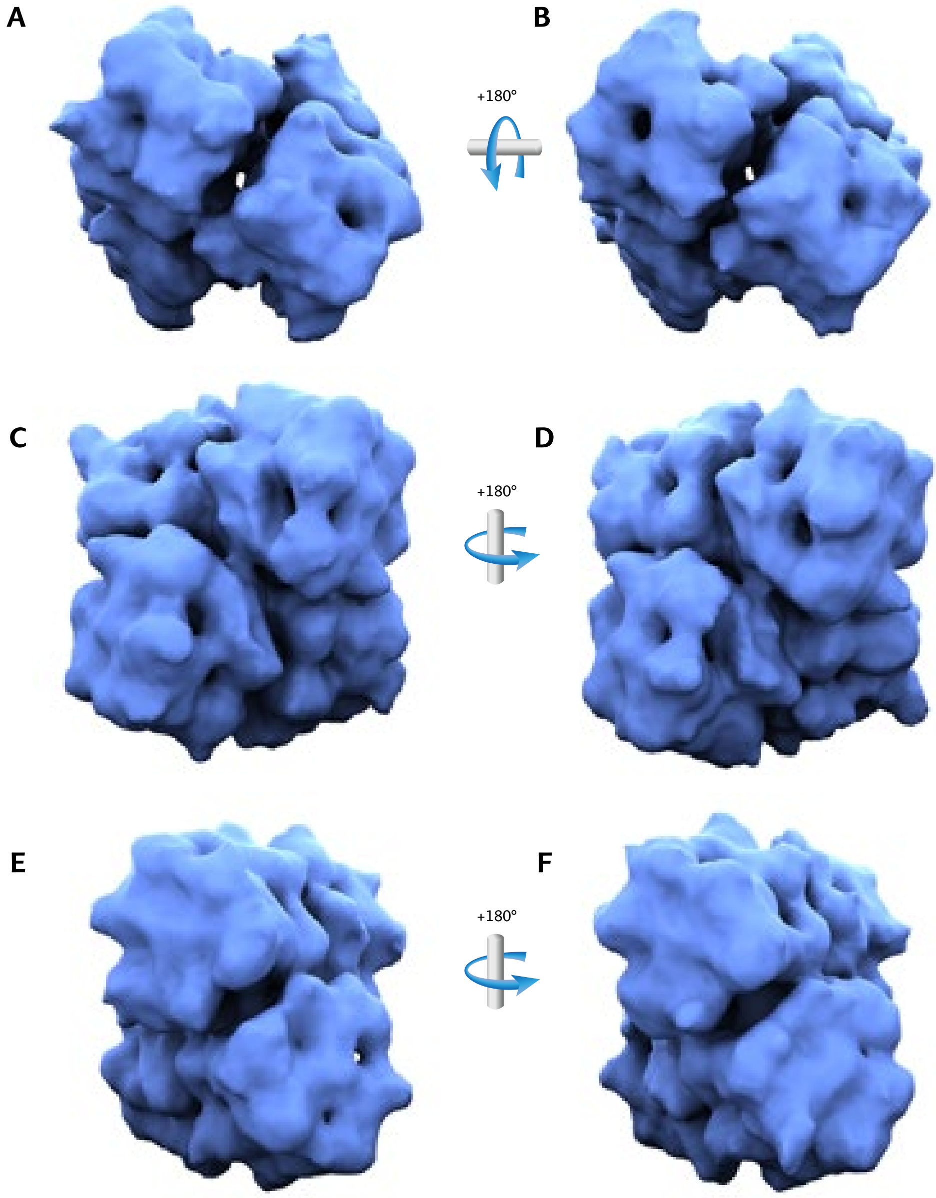


Fig. 26: Iso-surface comparison of the C1-3D-reconstruction. Panels A and B show a top-down view, representing the particle main axis and thus the main symmetry axis. Panels C/D and E/F show the second and third symmetry axis respectively. In each case, rotation by 180° shows an almost identical structure. As a consequence a D2 (two-fold dihedral symmetry) is clearly inherent to the molecule. As this reconstruction was calculated with no imposed symmetry, small differences are to be expected.

1.2.2 Reconstruction with imposed D2-symmetry

The initial map was calculated from class averages by applying the model obtained from negatively stained micrographs as a reference and then refined using projection matching with an initial angular increment of 10°.

Symmetric properties of *C. truncata* hemocyanin were unknown at the beginning of this project. After 20 iterative cycles performed with Eman, a D2 point-group symmetry was clearly inherent to the molecule. (Fig. 26). Differences between hexamers built of differing subunit types are very small and can be neglected, especially with a 3D-reconstruction at an intermediate resolution as presented in this study (Martin, 2006).

Subsequent refinement rounds were performed with a D2-symmetry applied and decreasing angular steps (smallest step: 2°) until no further improvement could be detected. The final map was obtained from 7330 particles. To facilitate overall reconstruction quality assessment, a comparison between reprojections and class-averages was performed (Fig. 27).

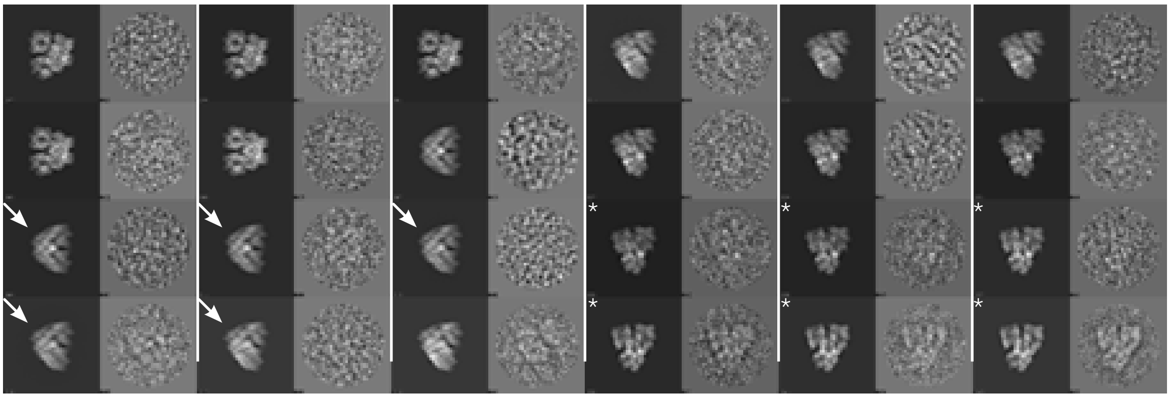


Fig. 27: Sample comparison of reprojections with their corresponding class-averages. Images with dark gray background are reprojections calculated from the final reconstruction. The right column with light gray background shows the corresponding class-average that was used. As both reprojection and class-average share the same Euler angles, ideally they should be identical. This figure shows a representative consecutive subset selected from the total of 2294 classes. Some tilted views are represented by classes with a lower signal-to-noise ratio (arrows) and thus occur less frequent in the dataset as opposed to an abundance of classes showing the preferred orientation of the particle in this preparation (marked by asterisks).

This analysis shows good overall correspondence but also indicates that the particle exhibits a preferred orientation even within the ice-layer of the cryo preparation. Many tilted views (roughly triangular shape) contribute classes with a lower signal-to-noise ratio when compared to the abundance of strong classes representing the preferred orientation.

1.2.3 Resolution assessment

The resolution of the final reconstruction is 9,8 Å according to the 0,5 criterion of the Fourier shell correlation (Fig. 28).

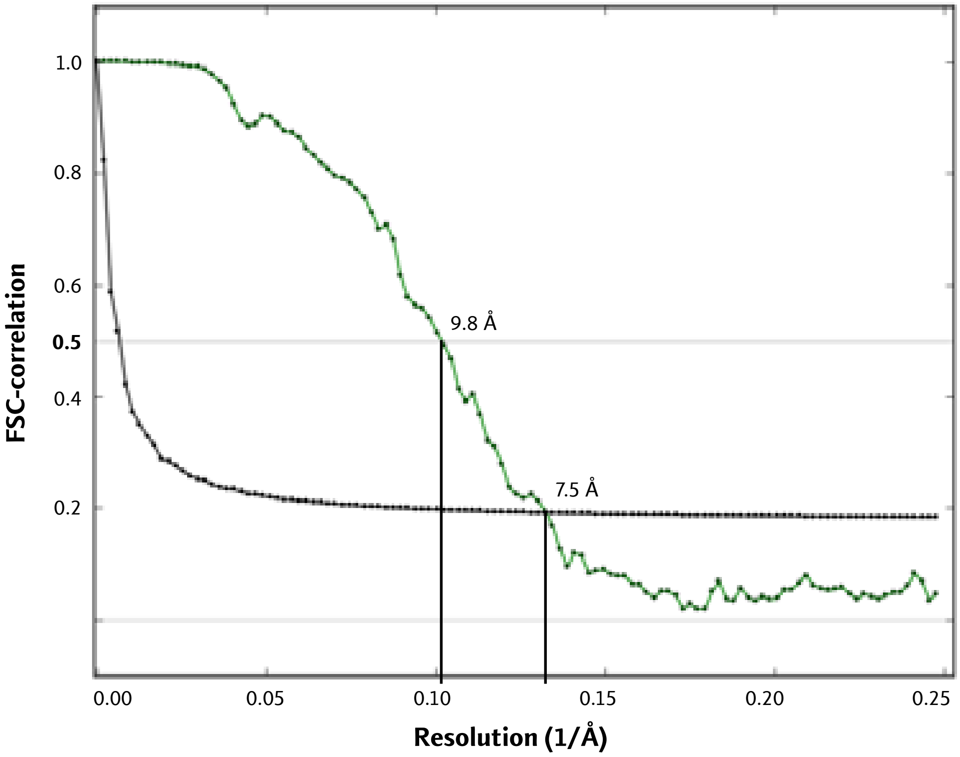


Fig. 28: Resolution of the final Callianassa truncata reconstruction as determined by Fourier shell correlation. The green curve represents the FSC of the final 3D-reconstruction, the black curve represents the 1/2bit-criterion threshold curve. The resolution is measured at 9,8 Å according to the 0,5-criterion and 7.5 Å according to the 1/2bit-criterion (plot created with plotfsc.py by Dr. Mario Dejung)

1.2.4 The 4x6mer

Fig. 29 shows the final reconstruction as an iso-surface rendering at a mass-correlated threshold. The 4x6mer can be broken down into two 2x6mers which form strong bridges at the center of the molecule. Overall dimensions are approx. 241 Å, 256 Å and 215 Å (width, height and depth as seen from the particle main axis, see Fig. 30). The 2x6mers are twisted around 62° with respect to their center axes (Fig. 30 D). The arrangement leads to an opening cleft between opposing hexamers when viewed side on (Fig. 30 C).

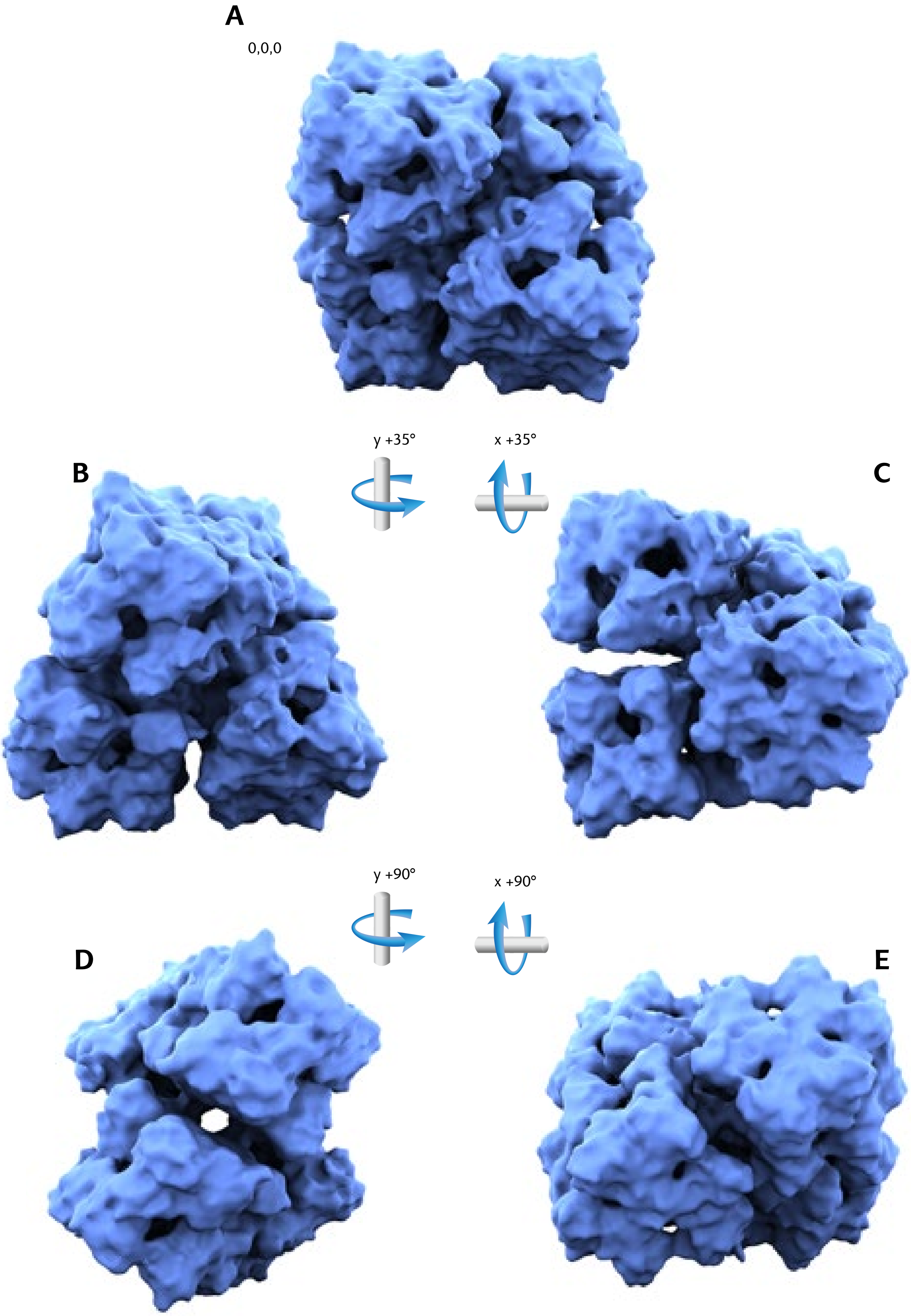


Fig. 29: Iso-surface representations of the final Callianassa truncata 3D-reconstruction. Shown at mass-correlated threshold with D2-symmetry imposed. The 35° tilted views where chosen to better convey the architecture of the 4x6mer. Each angular increment refers to total offset from the particle main axis (0,0,0) view.

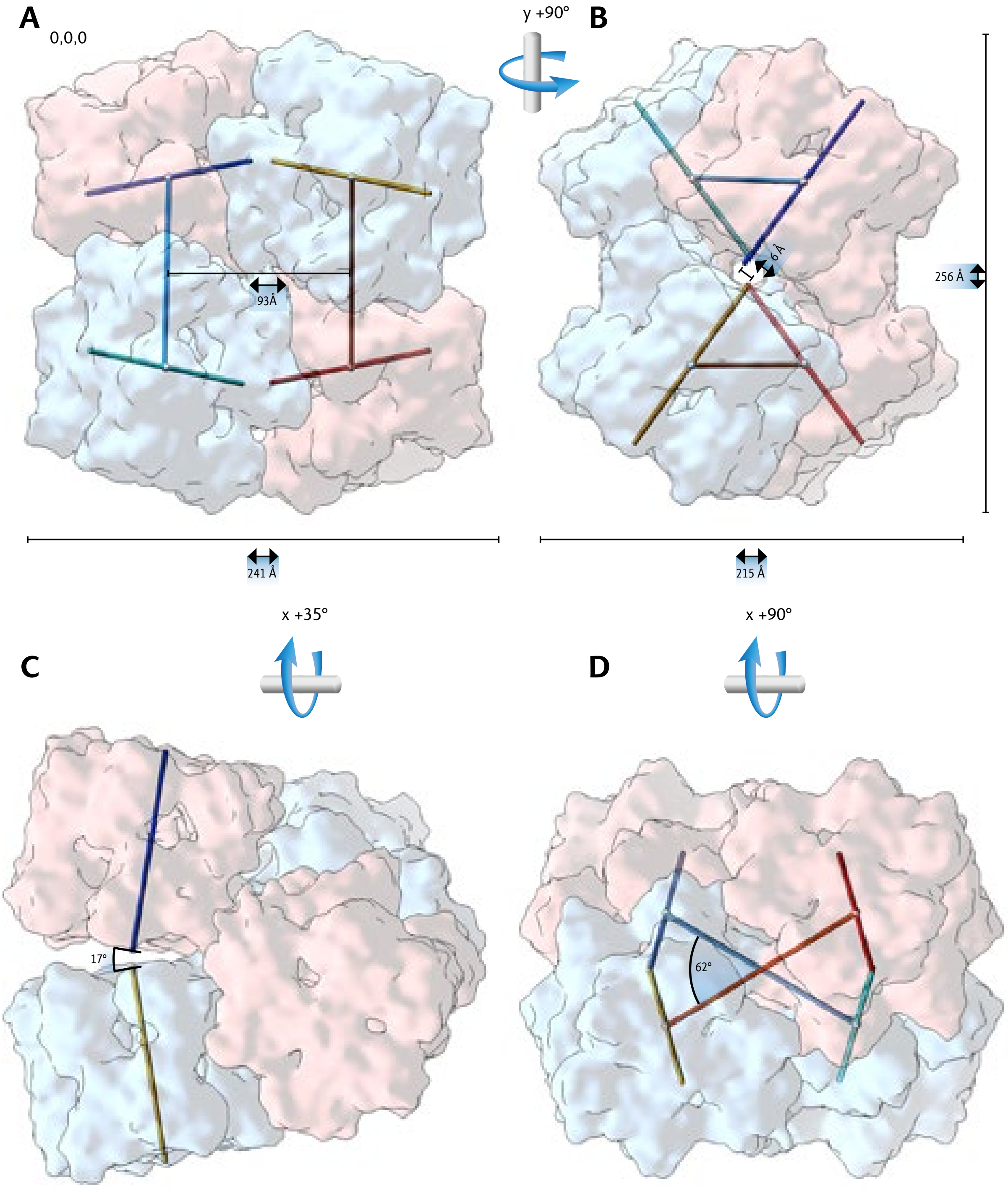


Fig. 30: The Callianassa truncata 4x6mer with hexamer axes and measurements. The main symmetry axis of each of the hexamers is indicated by a colored bar (2x6mer I: blue/green and 2x6mer II: yellow/red). A cyan/orange bar connects the hexamers of each 2x6mer by their center points. Hexamers within one 2x6mer unit are colored light red and blue respectively. Each angular increment refers to total offset from the 0,0,0 view. **A:** particle main axis with a distance between the two 2x6mers measured at 93 Å. **B:** side view showing the shift of two opposing hexamers at 6 Å. **C:** tilted view with the cleft between two opposing hexamers (as seen in panel B) opening at an angle of 19°. **D:** top-down view showing the 62° twist between the 2x6mers.

1.2.5 The two 2x6mer types contained within the 4x6mer

Several different types of crustacean hemocyanins with 2x6mer architecture are known (Markl & Decker 1992, Arnold 2012). The 4x6mer of *C. truncata* can be split two ways into distinct 2x6mers that are similar to other previously known 2x6mers (Fig. 31): The decapod type 2x6mer is characterized by the hexamers rotation angle of ~74° with respect to each other. In contrast the isopod type exhibits two stacked hexamers with no rotation or shift but tilted at an angle leaving a cleft and two small contact areas. It has been proposed by Arnold 2012 that both the isopod and decapod type stem from the same ancestral precursor.

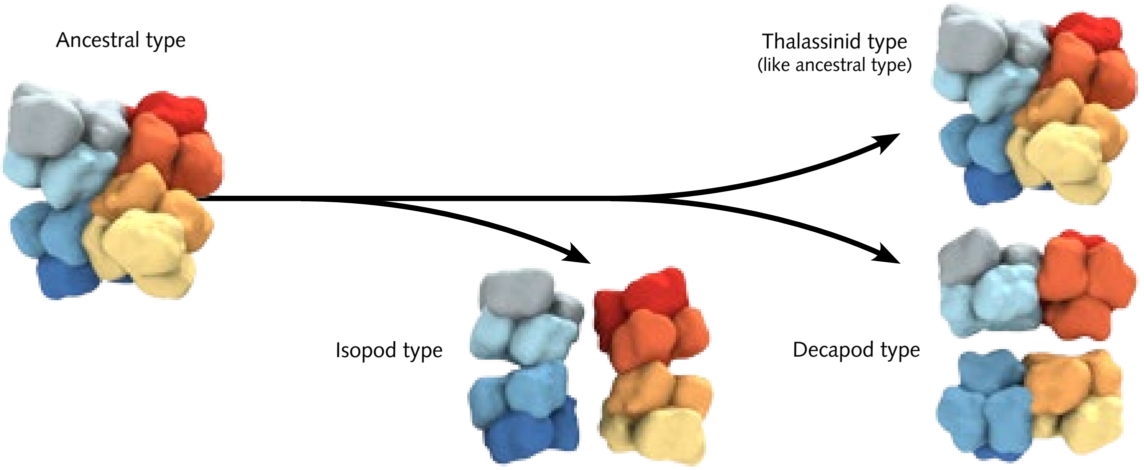


Fig. 31: Hypothetical evolution of isopod, decapod and thalassinid hemocyanin. When the isopod type branched off, the ancestral 4x6mer split in a different way compared to when the isopod type developed later on. The ancestral from is still present in the form of thalassinid 4x6mers. Figure from Arnold 2012 (modified).

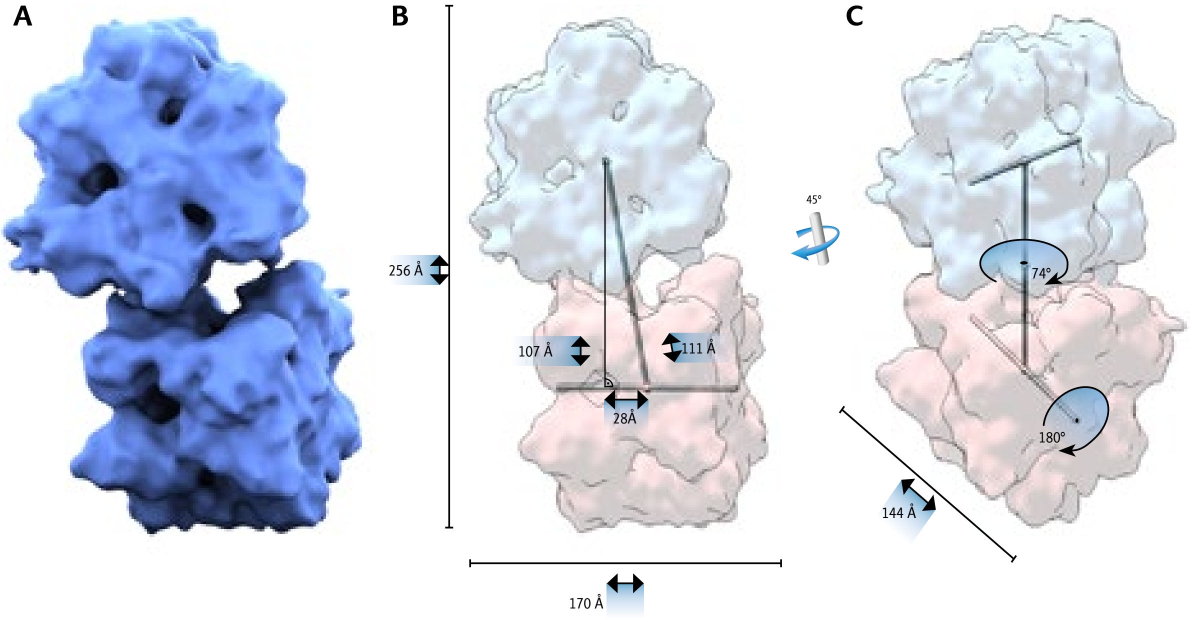


Fig. 32: The split Callianassa truncata 2x6mer – decapod type. **A:** surface rendering; **B and C:** overall dimensions of the C. truncata decapod type 2x6mer and measurements. The hexamer center axis is defined as the three-fold symmetry-axis around which the subunits are arranged in 120° steps. The view is centered on the main axis of the upper 1x6mer.

The decapod type 2x6mer reveals a twist between the hexamers of 74° (Fig. 32 C). The 1x6mers are also rotated 180° around their center axis with respect to each other. The distance of the center axes of the hexamers amounts to 107 Å while the center points are 111 Å apart, caused by an on-axis shift of 28 Å (Fig. 32 B). A total of three bridges connect the two 1x6mers.

Split differently the *C. truncata* 4x6mer reveals the second type of 2x6mer, the isopod type (Fig. 33). Two contacts connect the two 1x6mers. The cleft opens up from 15 to 32 Å with an angle of 17°. From visual inspection alone it might seem as if the hexamers are shifted slightly along the x-axis, this however cannot be confirmed when taking the mass centers as reference. The measured shift is minimal with 1,2 Å.

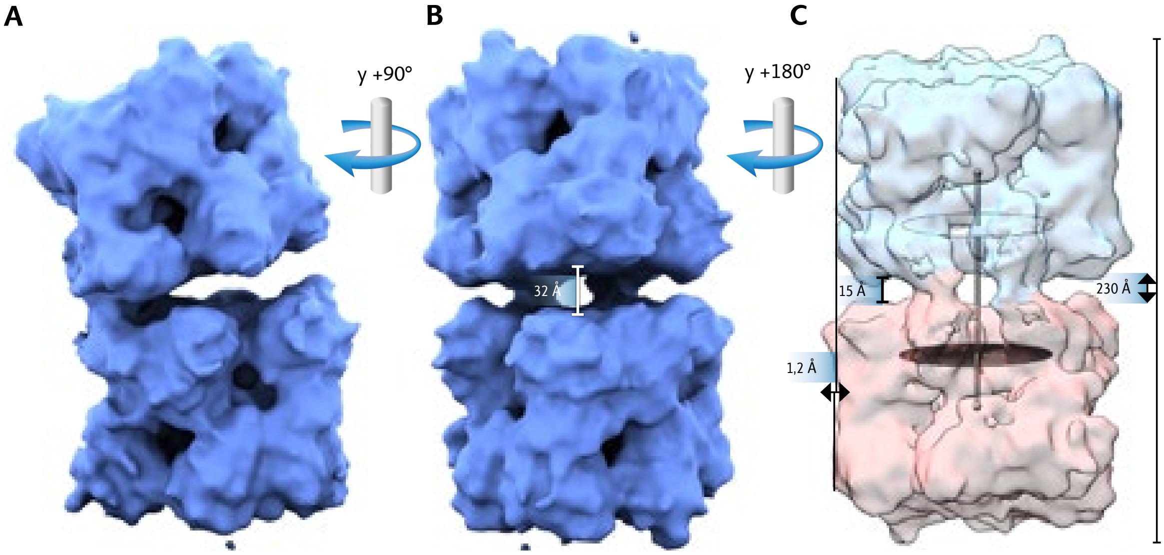


Fig. 33: The split Callianassa truncata 2x6mer – isopod type. A: Side view, the opening angle between the 1x6mers is 17° B: Side view rotated by 90° along the y-axis. C: Side view with the two connecting interfaces visible. The axis connects the mass centers of the two 1x6mers. They are 107 Å apart. The two disks span the mass centers of each inner half of the hexamers respectively to visualize the tilt angle between the 1x6mers. The 2x6mer measures 230 Å (largest expansion).

1.2.6 The 1x6mer

The basic subassembly of *C. truncata* is the 1x6mer. It exhibits a D3 point-group symmetry. Three roughly bean-shaped subunits are situated in a plane with the three remaining subunits underneath in a similar configuration, albeit rotated about 55° with respect to the top layer. All subunits are arranged around a central axis. This gives rise to the typical hexagonal appearance when viewed top-down and a roughly rectangular outline when seen from the side (Fig. 34 A and B).

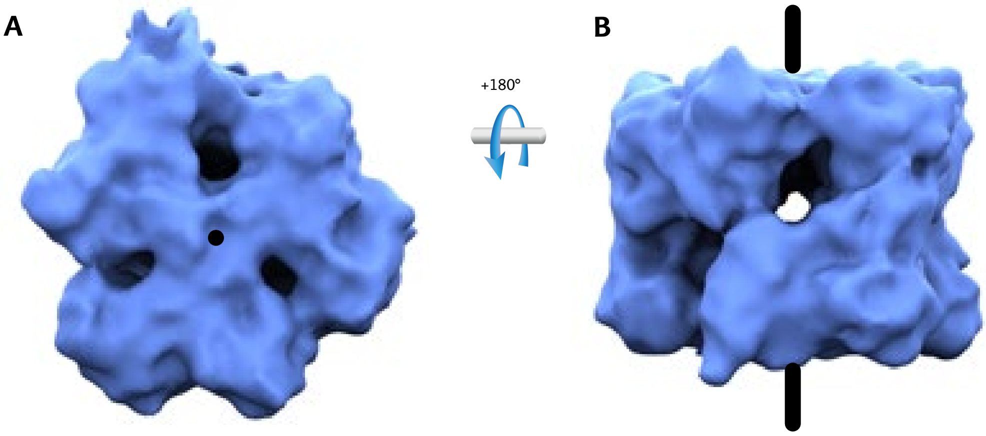


Fig. 34: Iso-surface rendering of the Callianassa truncata 1x6mer. Six roughly bean-shaped subunits are arranged in two stacked layers and twisted 55° around the central axis (black bars resp. dot). **A:** the hexagonal appearance of the 1x6mer as viewed top-down. The three subunits of the top layer can be distinguished clearly. **B:** side view of the hexamer giving a rectangular appearance. Also, four distinct subunits can be seen.

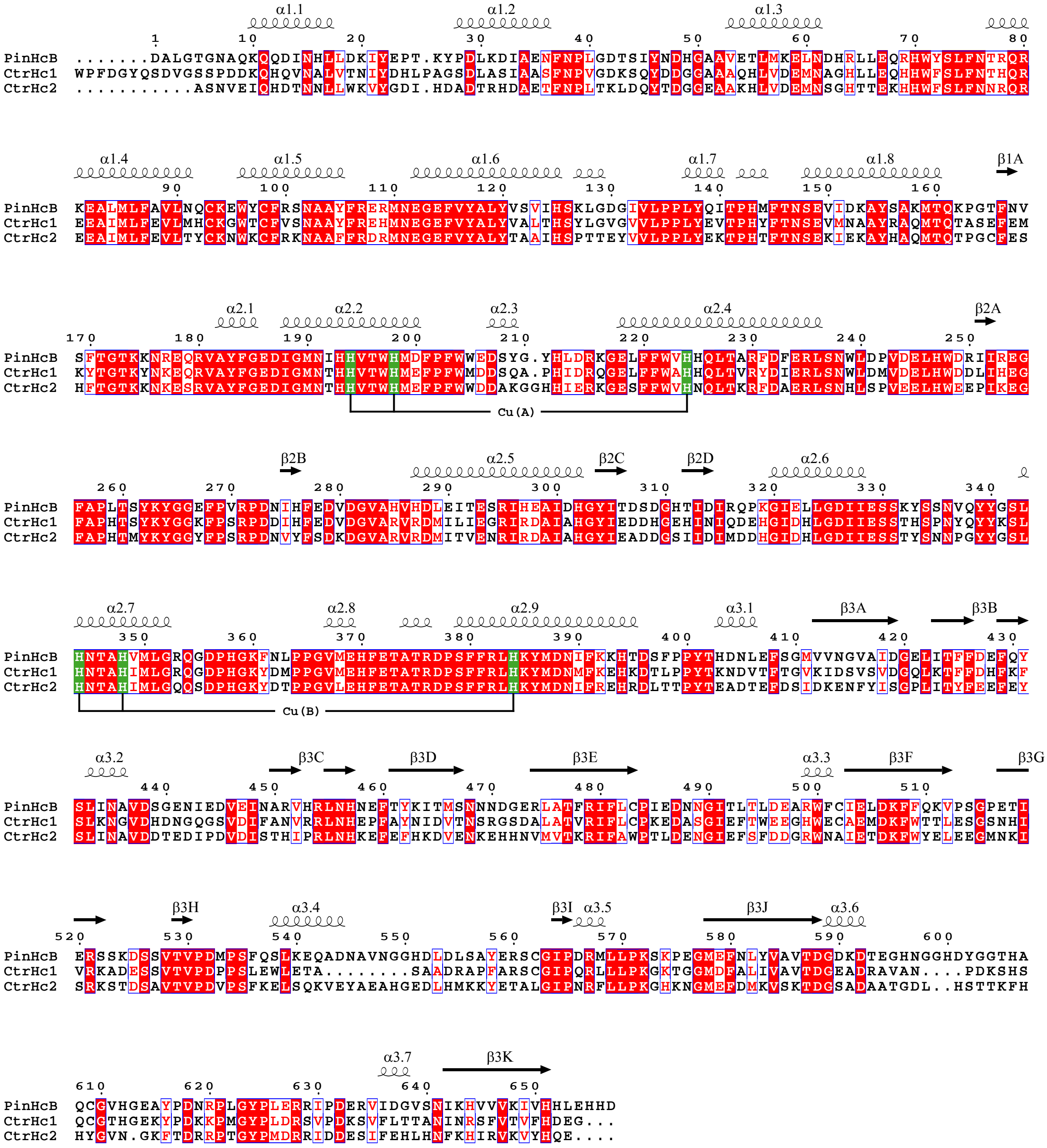
2 Molecular model of *C. truncata* hemocyanin

2.1 Homology-modeling strategy

In 2010, the molecular and biochemical characterization of *C. truncata* hemocyanin yielded the primary structure of the first subunit-type that builds the 4x6mer (*Callianassa truncata* type 1 hemocyanin (*Ctr*Hc1; Solem, 2010). This sequence was used to calculate a homology-model. A suitable template (similarity of 71%, BLOSUM 62) was found with the structure of a *Panulirus interruptus* hexamer as solved by x-ray crystallography (Volbeda & Hol, 1989).

Modeler 8v2 was used to build the initial model as well as for subsequent individual loop refinement iterations. The protein energy minimization server Chiron and UCSF Chimeras modification and refinement tools were used to further optimize and adapt the structure to the 3D-data presented in this study (Ramachandran et al., 2011).

In turn, for the sequence of a second subunit type with comparable sequence identity (*Ctr*Hc2; Groneberg, 2011), the same procedures were applied to obtain the molecular model. A sequence alignment with the *P. interruptus* reference and both *C. truncata* subunits is presented in Fig. 35



```
PinHcB .......DALGTGNAQKQQDINHLLDKIYEPT.KYPDLKDIAENFNPLGDTSIYNDHGAAVETLMKELNDHRLLEQRHWYSLFNTRQR
CtrHc1 WPFDGYQSDVGSSPDDKQHQVNALVTNIYDHLPAGSDLASIAASFNPVGDKSQYDDGGAAAQHLVDEMNAGHLLEQHHWFSLFNTHQR
CtrHc2 ...........ASNVEIQHDTNNLLWKVYGDI.HDADTRHDAETFNPLTKLDQYTDGGEAAKHLVDEMNSGHTTEKHHWFSLFNNRQR

PinHcB KEALMLFAVLNQCKEWYCFRSNAAYFRERMNEGEFVYALYVSVIHSKLGDGIVLPPLYQITPHMFTNSEVIDKAYSAKMTQKPGTFNV
CtrHc1 EEAIMLFEVLMHCKGWTCFVSNAAYFREHMNEGEFVYALYVALTHSYLGVGVVLPPLYEVTPHYFTNSEVMNAAYRAQMTQTASEFEM
CtrHc2 EEAIMLFEVLTYCKNWKCFRKNAAFFRDRMNEGEFVYALYTAAIHSPTTEYVVLPPLYEKTPHTFTNSEKIEKAYHAQMTQTPGCFES

PinHcB SFTGTKKNREQRVAYFGEDIGMNIHHVTWHMDFPFWWEDSYG.YHLDRKGELFFWVHHQLTARFDFERLSNWLDPVDELHWDRIIREG
CtrHc1 KYTGTKYNKEQRVAYFGEDIGMNTHHVTWHMEFPFWMDDSQA.PHIDRQGELFFWAHHQLTVRYDIERLSNWLDMVDELHWDDLIHEG
CtrHc2 HFTGTKKNKESRVAYFGEDIGMNTHHVTWHMEFPFWWDDAKGGHHIERKGESFFWVHNQLTKRFDAERLSNHLSPVEELHWEEPIKEG

PinHcB FAPLTSYKYGGEFPVRPDNIHFEDVDGVAHVHDLEITESRIHEAIDHGYITDSDGHTIDIRQPKGIELLGDIIESSKYSSNVQYYGSL
CtrHc1 FAPHTSYKYGGKFPSRPDDIHFEDVDGVARVRDMILIEGRIRDAIAHGYIEDDHGEHINIQDEHGIDHLGDIIESSTHSPNYQYYKSL
CtrHc2 FAPHTMYKYGGYFPSRPDNVYFSDKDGVARVRDMITVENRIRDAIAHGYIEADDGSIIDIMDDHGIDHLGDIIESSTYSNNPGYYGSL

PinHcB HNTAHVMLGRQGDPHGKFNLPPGVMEHFETATRDPSFFRLHKYMDNIFKKHTDSFPPYTHDNLEFSGMVVNGVAIDGELITFFDEFQY
CtrHc1 HNTAHIMLGRQGDPHGKYDMPPGVMEHFETATRDPSFFRLHKYMDNMFKEHKDTLPPYTKNDVTFTGVKIDSVSVDGQLKTFFDHFKF
CtrHc2 HNTAHIMLGQQSDPHGKYDTPPGVLEHFETATRDPSFFRLHKYMDNIFREHRDLTTPYTEADTEFDSIDKENFYISGPLITYFEEFEY

PinHcB SLINAVDSGENIEDVEINARVHRLNHNEFTYKITMSNNNDGERLATFRIFLCPIEDNNGITLTLDEARWFCIELDKFFQKVPSGPETI
CtrHc1 SLKNGVDHDNGQGSVDIFANVRRLNHEPFAYNIDVTNSRGSDALATVRIFLCPKEDASGIEFTWEEGHWECAEMDKFWTTLESGSNHI
CtrHc2 SLINAVDDTEDIPDVDISTHIPRLNHKEFEFHKDVENKEHHNVMVTKRIFAWPTLDENGIEFSFDDGRWNAIETDKFWYELEEGMNKI

PinHcB ERSSKDSSVTVPDMPSFQSLKEQADNAVNGGHDLDLSAYERSCGIPDRMLLPKSKPEGMEFNLYVAVTDGDKDTEGHNGGHDYGGTHA
CtrHc1 VRKADESSVTVPDPPSLEWLETA........SAADRAPFARSCGIPQRLLLPKGKTGGMDFALIVAVTDGEADRAVAN....PDKSHS
CtrHc2 SRKSTDSAVTVPDVPSFKELSQKVEYAEAHGEDLHMKKYETALGIPNRFLLPKGHKNGMEFDMKVSKTDGSADAATGDL..HSTTKFH

PinHcB QCGVHGEAYPDNRPLGYPLERRIPDERVIDGVSNIKHVVVKIVHHLEHHD
CtrHc1 QCGTHGEKYPDKKPMGYPLDRSVPDKSVFLTTANINRSFVTVFHDEG...
CtrHc2 HYGVN.GKFTDRRPTGYPMDRRIDDESIFEHLHNFKHIRVKVYHQE....
```

Fig. 35: Sequence alignment of Callianassa truncata subunits (CtrHc1 and CtrHc2) and the Panulirus interruptus modeling reference (PinHcB). Global similarity frames are printed with blue borders, strictly conserved amino acids are shaded red, similar amino acids in a similarity group are printed red. The Cu-binding histidine residues are shaded green. Secondary structure elements as deduced from X-ray structure *P. interruptus* are indicated at the top. The starting points of domains #1-3 are marked on the bottom.

2.1.1 *Callianassa truncata* type 1 hemocyanin subunit (*Ctr*Hc1)

The model for *Ctr*Hc1 is shown in Fig. 36. Quality control indicate good overall quality with some small problem areas scattered throughout the model (Fig. 37).

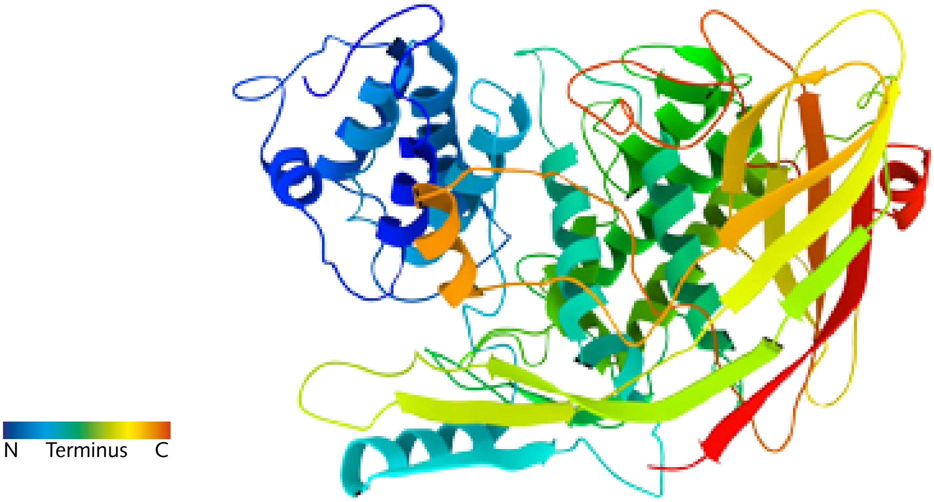


Fig. 36: Ribbon representation of CtrHc1. The structure is colored using a rainbow scheme starting at the N-terminus (violet) and ending at the C-terminus (red). The three major domains are colored as follows: domain #1 – blue/cyan; domain #2 – dark green, green; domain #3 – bright green, yellow, orange and red.

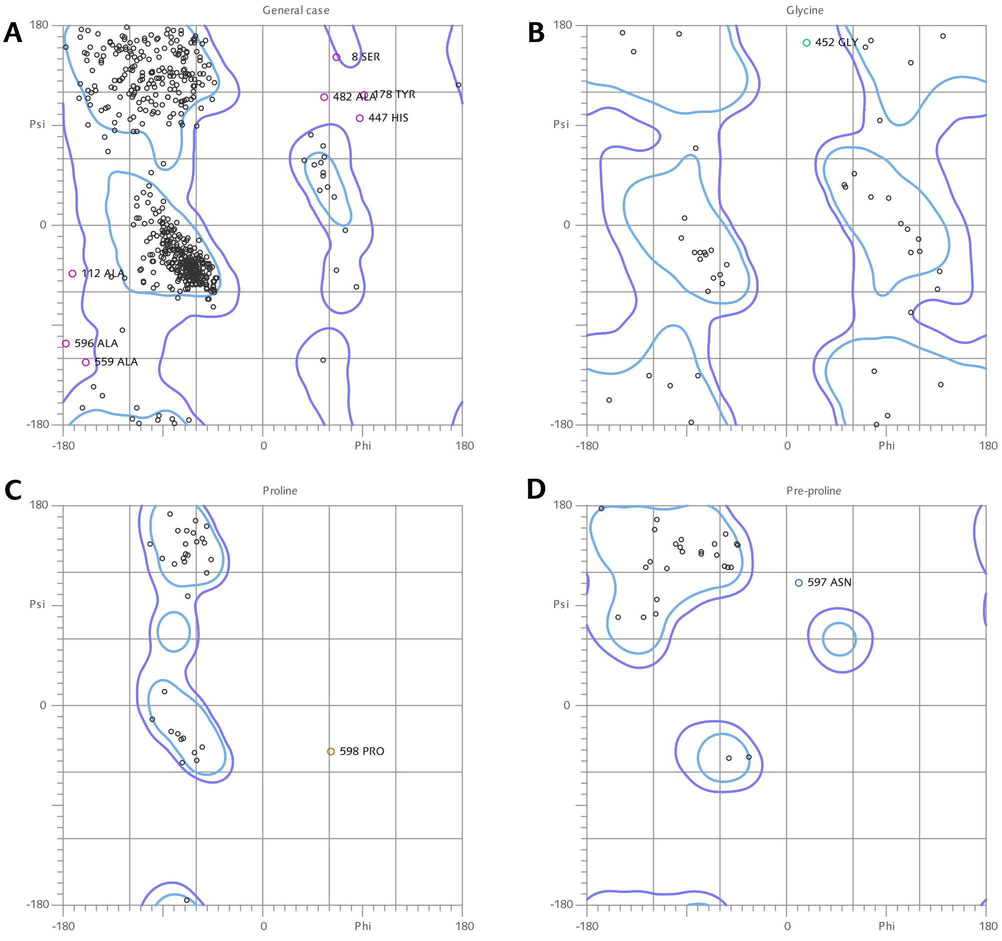


Fig. 37: Ramachandran plot of the CtrHc1 homology model. A total of 10 outliers were detected. **A:** general case. **B:** glycine outliers. **C:** proline outliers. **D:** pre-proline outliers.

2.1.2 *Callianassa truncata* type 2 hemocyanin subunit (*Ctr*Hc2)

The homology-model of *Ctr*Hc2 and the corresponding quality control are shown in Fig. 38 and Fig. 39 respectively.

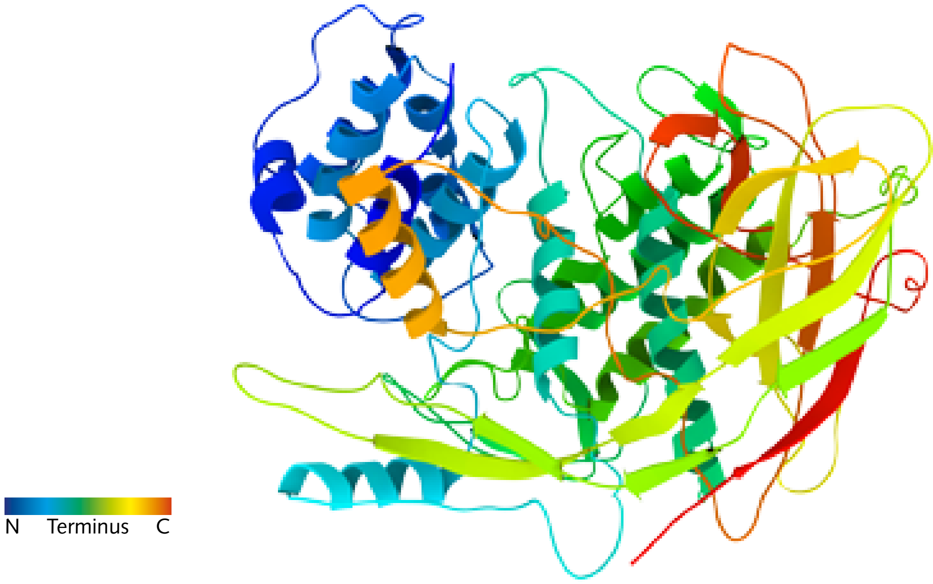


Fig. 38: Ribbon representation of CtrHc2. The structure is colored using a rainbow scheme starting at the N-terminus (violet) and ending at the C-terminus (red). The three major domains are colored as follows: domain #1 – blue/cyan; domain #2 – dark green, green; domain #3 – bright green, yellow, orange and red.

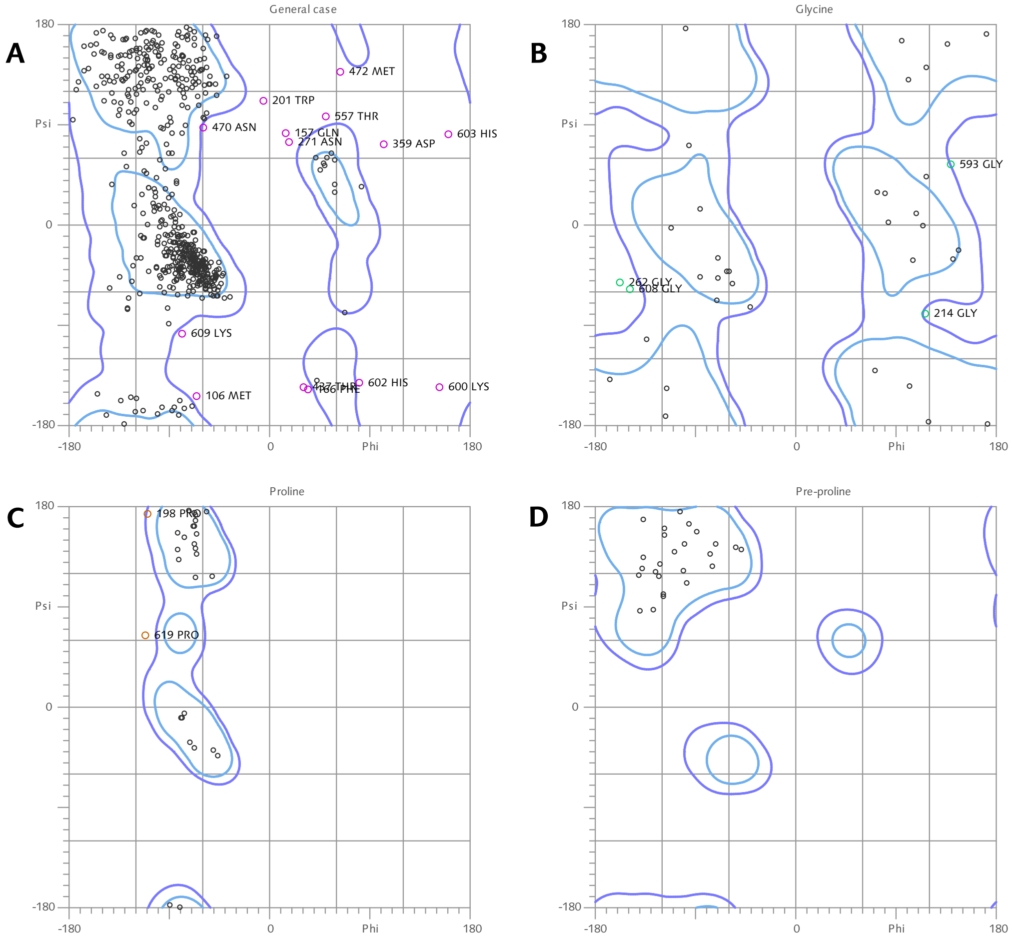


Fig. 39: Ramachandran plot of the CtrHc2 homology model. A total of 20 outliers were detected. **A:** general case. **B:** glycine outliers. **C:** proline outliers. **D:** pre-proline outliers.

2.1.3 Ramachandran outliers

In order to visualize potential problem areas, Ramachandran outliers were highlighted as described in the previous section (Fig. 40). The problems mainly concentrate on areas were loop refinement with the Modeller8v2 software has taken place.

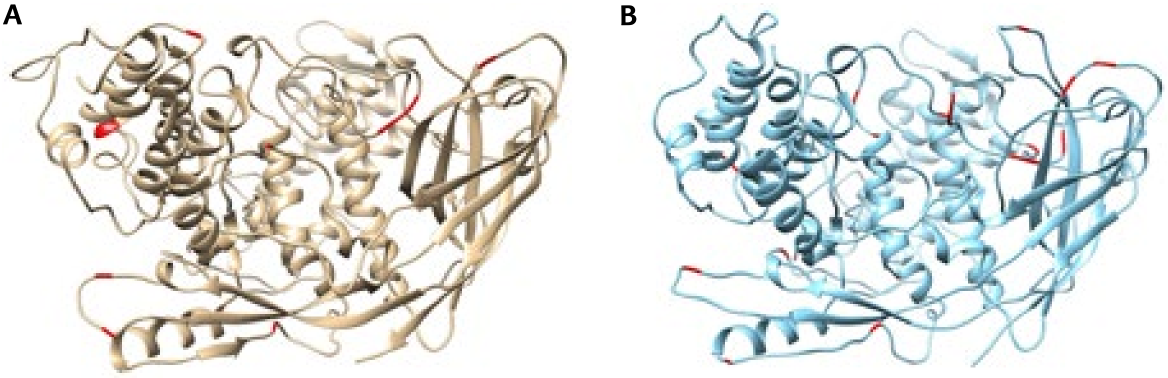


Fig. 40: Visualization of Ramachandran plot outliers. Outliers are highlighted as red sections **A:** The ten outliers of the *Ctr*Hc1 homology model. **B:** the *Ctr*Hc2 homology model contains a total of 20 outliers.

2.1.4 Molecular model of *Callianassa truncata* hemocyanin

For the purpose of this study a molecular model of C. truncata was built with UCSF Chimeras matchmaker function using the two previously described subunits. Based on earlier findings, a *Ctr*Hc1:*Ctr*Hc2 ratio of 5:1 per hexamer was assumed (Solem, 2010; Groneberg, 2011). Individual subunits were docked into their corresponding places using the *Panulirus interruptus* crystal structure as a scaffold (Volbeda & Hol, 1989; Fig. 40). Based on the chain designations used in the structure of a *Panulirus* hexamer, subunits were labeled with designators a1-a5 and b. *Ctr*Hc2 is always at position b. Positions a1-a5 are copies of *Ctr*Hc1.

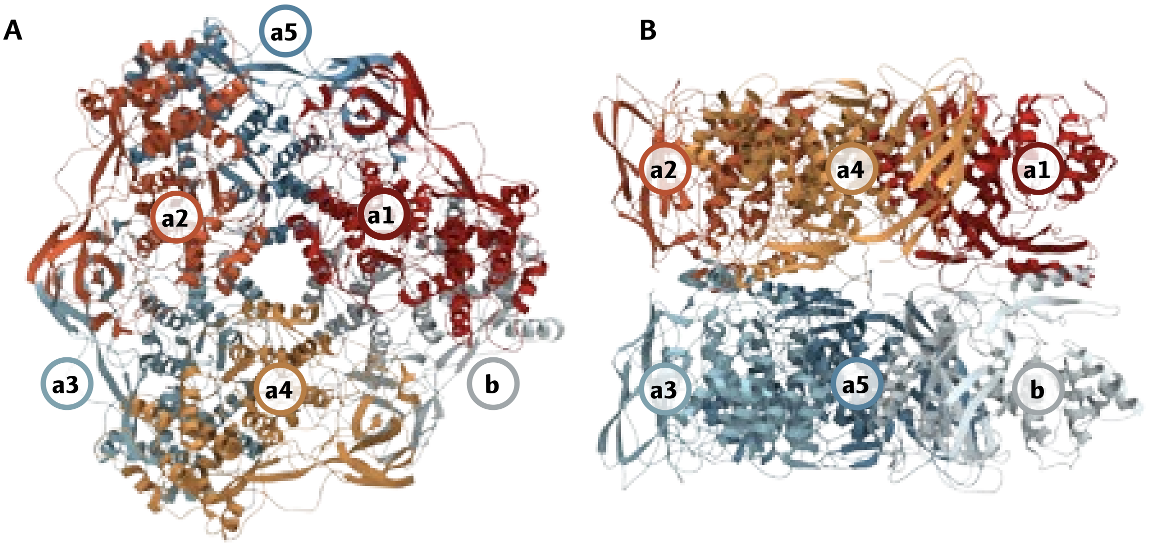


Fig. 41: An assembled hexamer consisting of five CtrHc1 subunits and one CtrHc2 subunit. Hexamer designations labeled a1-a5 and b. **A:** top-down view along the central axis of the hexamer. Subunits a1, a2 and a4 are located on the upper half of the 1x6mer, subunits b, a3 and a5 on the lower half. **B:** Side view of the hexamer.

Thus a complete molecular model of the 4x6mer was obtained (Fig. 42). The hexamers were placed so that the four *Ctr*Hc2-type subunits are oriented towards the center of the 4x6mer.

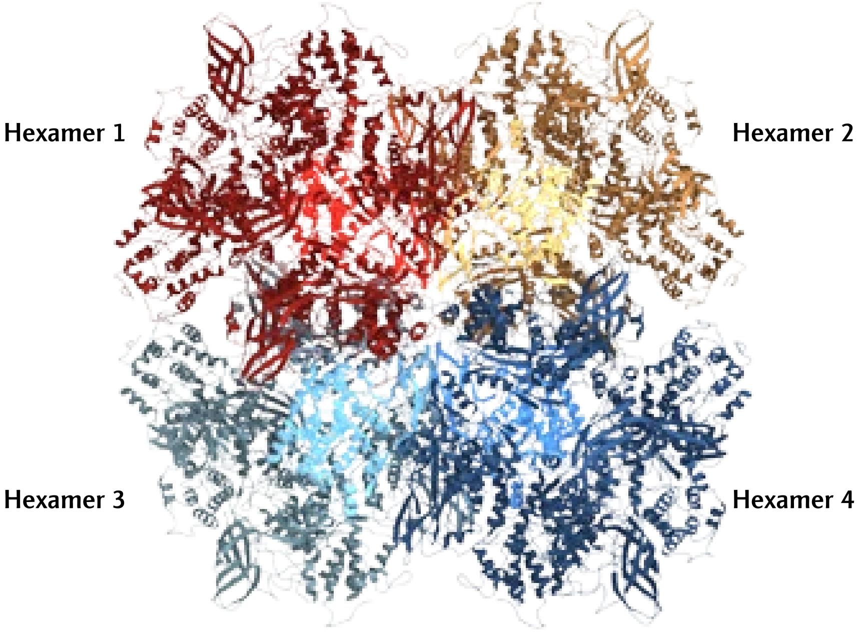


Fig. 42: Complete molecular model of the Callianassa truncata 4x6mer. Individual hexamers are colored red and yellow (hexamer 1 and 2) as well as dark cyan and blue (hexamer 3 and 4). The four *Ctr*Hc2 subunits at the center core are colored red (hexamer 1), yellow (hexamer 2), cyan (hexamer 3) and blue (hexamer 4). The 4x6mer is oriented as seen from the particle main axis.

2.1.5 Rigid-body fitting

The molecular model was docked into the 3D-reconstruction by UCSF CHIMERAS FIT-INTO-MAP rigid-body fitting algorithm. After docking the hexamers as a whole, each monomer was allowed to reiterate its position individually to improve overall fitting quality. Only a slight shift occurred during this procedure. When viewed at the mass-correlated threshold, 12% of all atoms are on the outside of the contour level of the reconstruction. This can be attributed to the hexamers hollow interior (Fig. 43 C and D).

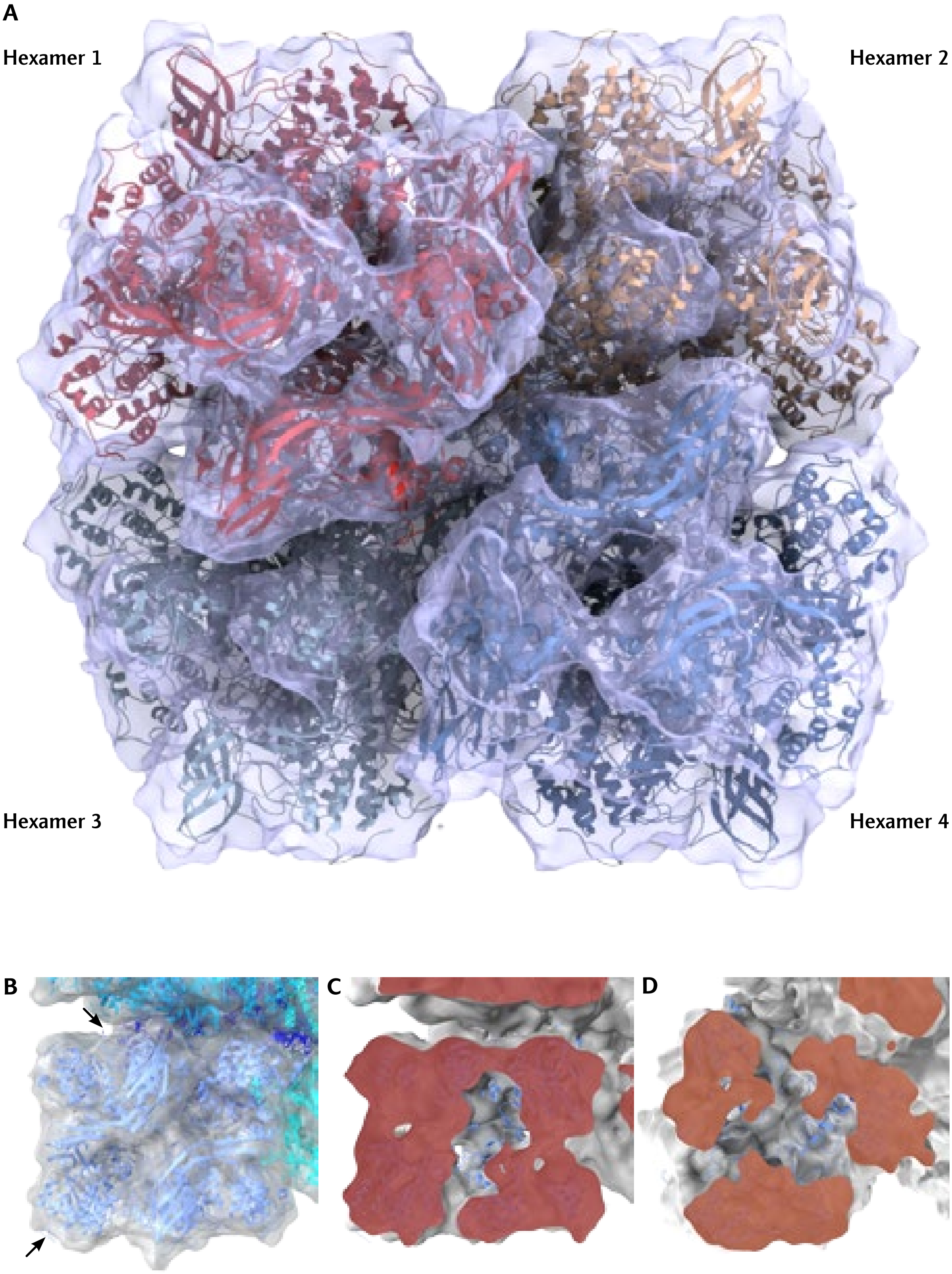


Fig. 43: Total and cutaway views of the 3D-reconstruction with the molecular model in place. **A:** the molecular model docked into the 3D-reconstruction, seen along the particle main axis. **B:** from the outside, a few protrusions can be seen (marked by arrows), most notably the N-terminus. **C:** the side view cutaway shows part of the molecular model protruding on the inside of the hexamer. **D:** cutaway as seen from a top-down point of view. Both C and D show cuts through the center of the 1x6mer.

2.2 Interface analysis

A total of twelve potential contacts facilitated by four distinct types of interface between the hexamers can be identified (Table 2 and Fig. 44/Fig. 45). All interfaces are also clearly visible as bridges at nominal threshold levels in the reconstructed density map.

Overall there are three subunits involved in contacts (designated b, a3 and a4) with subunit b type *Ctr*Hc2 participating in every contact but one, which indicates its importance to the assembly of the 2x6 and 4x6mer. Most contacts with the exception of the b–a4 bridge are densely packed in the center of the 4x6mer. This stems from the small contact areas caused by the large cleft of the isopod type 2x6mer.

Interface	Residues in first subunit	Residues in second subunit
b–a4	542**YA**543, **H**546, **E**548, **H**551	535**SVTVPD**540
b–b	Both subunits: 411**ENFYISGP**418, 458**EFHKD**462, 636**FKHIRVK**642	
b–a3	469**HN**470, 589**DAATGD**594	53**QYDDGGA**59, 100**HCK**102,
a3–a3	Both subunits: 50**DK**51	

Table 2: Amino acid residues involved in inter-hexamer interfaces.

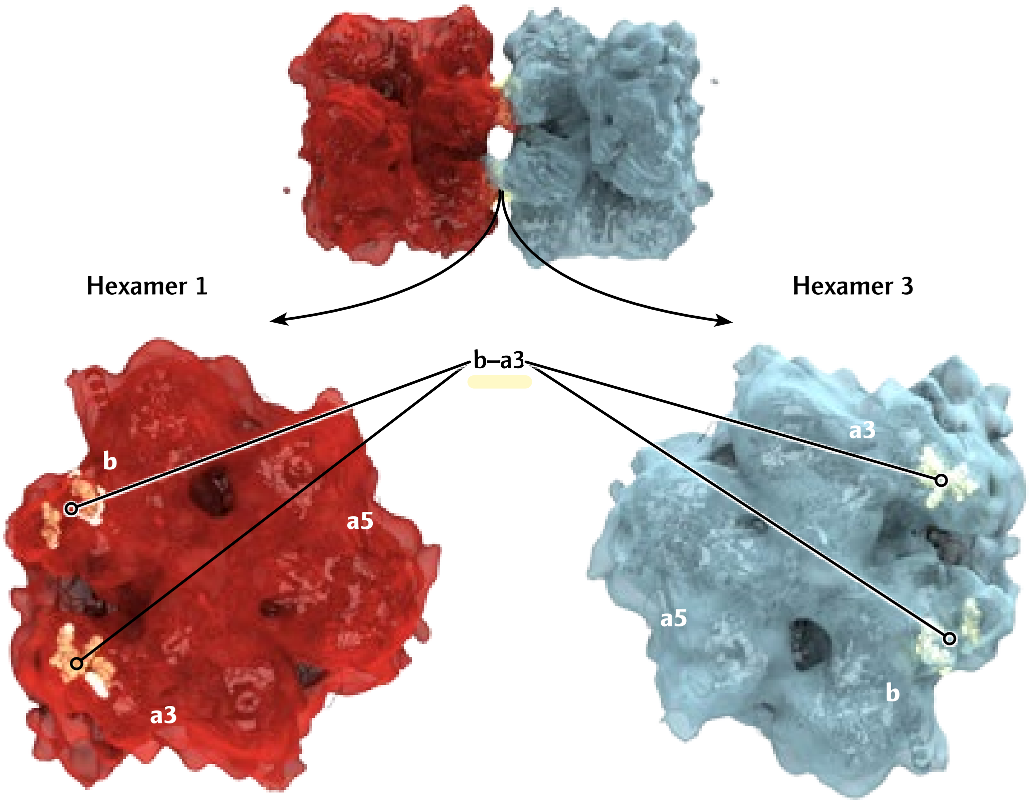


Fig. 44: Overview of interface locations in the isopod type 2x6mer. Subunit designations are labeled a1-a5 and b. The 2x6mer of the isopod type is connected by the b–a3 interface. The 2x6mer is split at the interface plane. The hexamers are turned outwards by 90° so that the interface plane is facing the viewer.

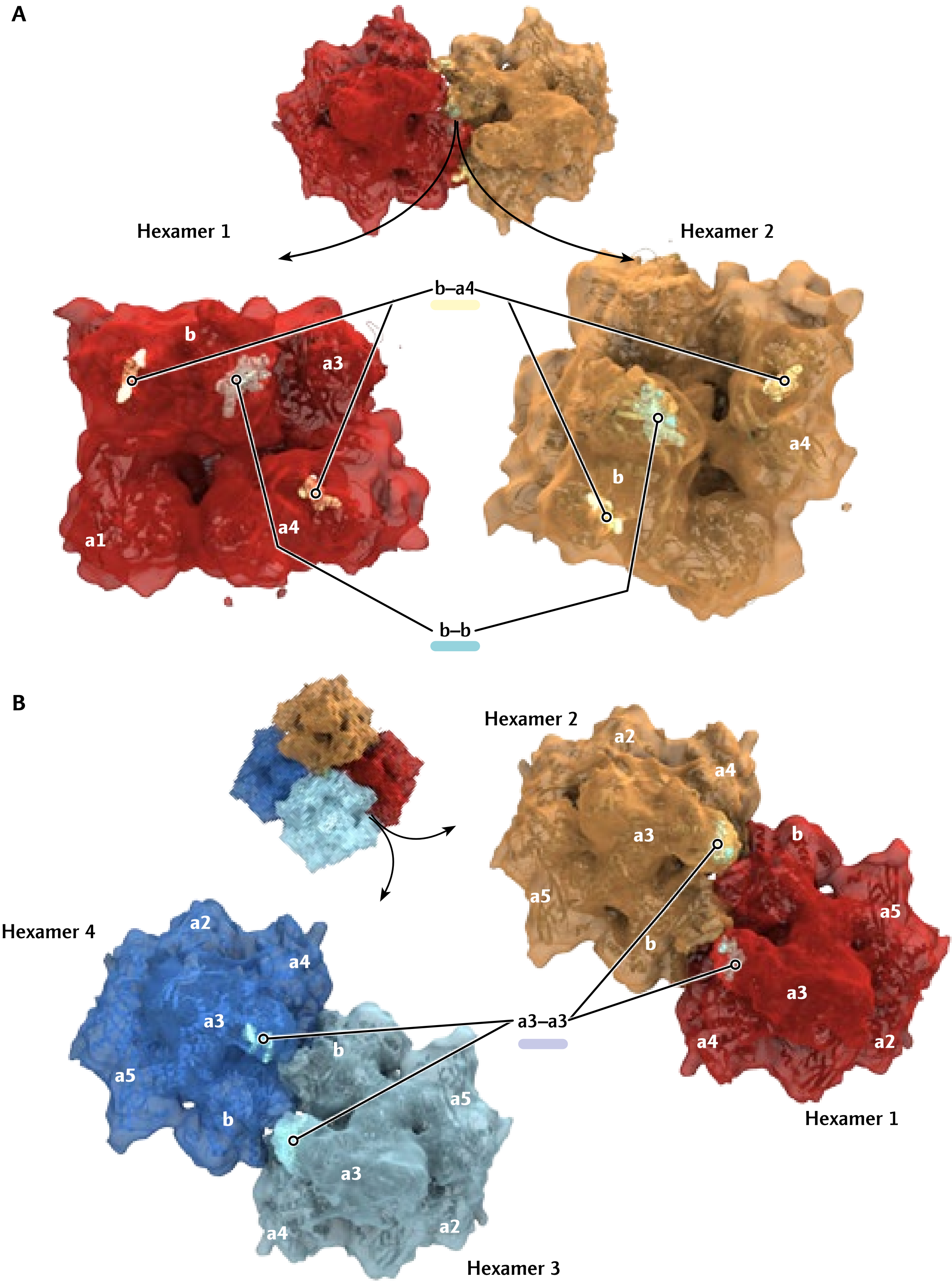


Fig. 45: Overview of interface locations in the decapod type 2x6mer and the 4x6mer. Subunit designations are labeled a1-a5 and b. Panel A: Location of the b–a4 and b–b interfaces in the decapod variant. The 2x6mer is split at the interface plane. The hexamers are turned outwards by 90° so that the interface plane is facing the viewer. Panel B: Location of the a3–a3 interface of the 4x6mer. The 4x6mer is split at the interface plane. The subunits are turned outwards to expose the contact zones.

All depictions of amino acids in the following section employ a color code in addition to text labels (Table 3). To help the viewers depth perception, the depictions are rendered using realistic soft lighting with shadows and depth of field. As a result colors might appear in slightly different shades depending on the lighting situation.

🟩	Isoleucine	ILE	I	🟩	Valine	VAL	V
🟦	Glutamine	GLN	Q	🟧	Threonine	THR	T
⬜	Glycine	GLY	G	🟪	Histidine	HIS	H
🟥	Glutamic Acid	GLU	E	🟪	Tryptophan	TRP	W
🟨	Cysteine	CYS	C	🟦	Phenyalanine	PHE	F
🟥	Aspartic Acid	ASP	D	⬜	Alanine	ALA	A
🟧	Serine	SER	S	🟨	Methionine	MET	M
🟦	Lysine	LYS	K	🟩	Leucine	LEU	L
🟫	Proline	PRO	P	⬛	Arginine	ARG	R
🟦	Asparagine	ASN	N	🟦	Tyrosine	TYR	Y

Table 3: Residue abbreviations and corresponding color code used throughout the figures.

The subunit chains are colored from N to C-terminus using a rainbow color scheme. The approximate starting points of the three domains are indicated below:

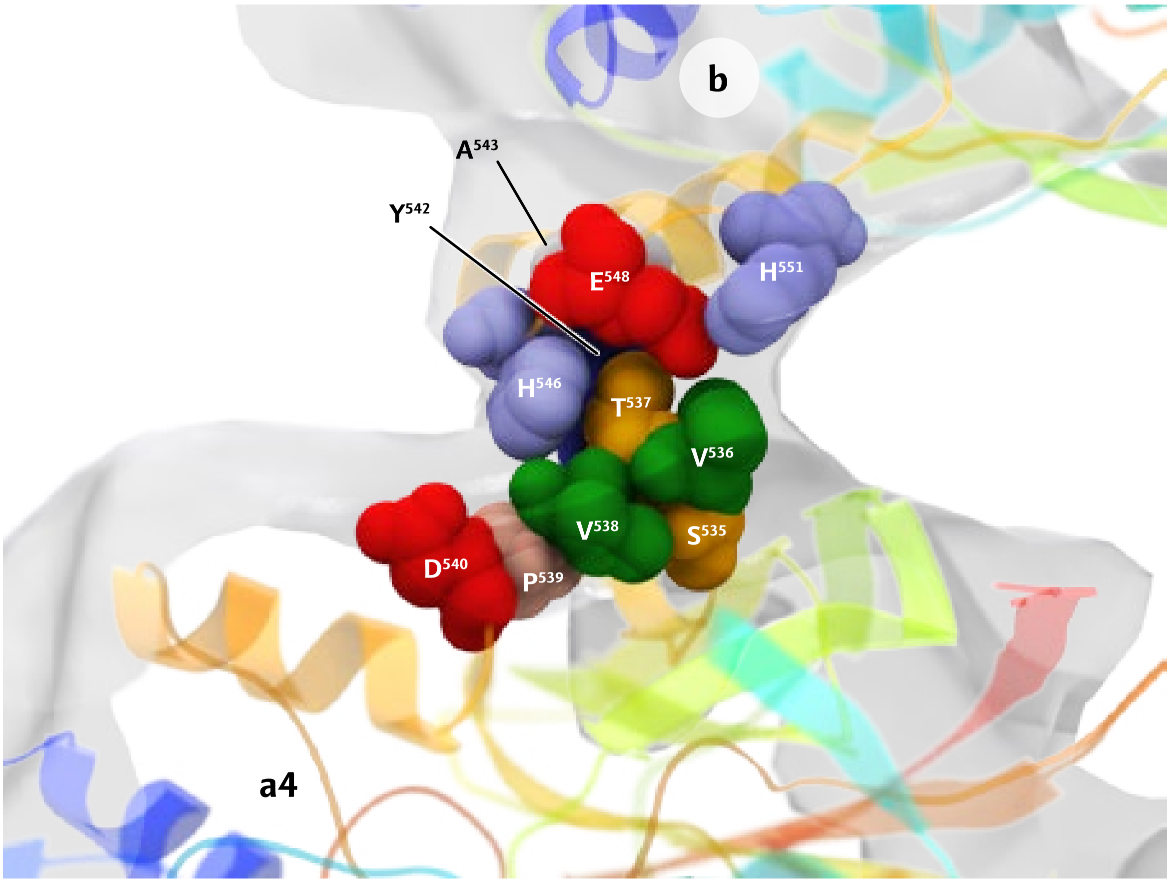


Fig. 46: The b–a4 interface between 1x6mers of the 2x6mer. The relevant slice of the reconstructed density map is displayed for reference.

2.2.1 Interfaces between the 1x6mers of the decapod type 2x6mer

2.2.1.1 The b–a4 interface

This small interface yields a distinct bridge between the 1x6mers forming the decapod type 2x6mer. It involves a contact between domain #3 of both subunits (Fig. 46). Due to the imposed D2 symmetry, a total of four bridges can be observed in the reconstruction at mass correlated threshold levels.

The α3.3 helix of subunit b protrudes towards the β3H→α3.3 loop of subunit a4. The molecular model was modified by loop modeling to adapt to the structure as observed in the reconstruction. The core of the interface has the potential to from hydrogen bonds (H546, T537, Y542, S535). The charged residues are too far apart to form salt bridges (with the exception of H546 and E548, both from subunit b). This arrangement could serve to stabilize the interface. The interface is flanked by two histidine residues from subunit b, H546 and H551. Cooperative behavior dictates that allosteric interaction must somehow be transmitted between the components of the hemocyanin. In this case the arrangement of the b–a4 interface might act similar to a ball joint.

2.2.1.2 The b–b interface

The second connection within the decapod type 2x6mer is the parallel association of domain #3, namely strands β3D, β3A and β3K from both subunits (Fig. 48A and B). It forms a chamber-like structure. On the inside the chamber has a width of ~6-9 Å. The top is dominated by ten charged residues (E411,461KD462,K515,K637) which surround two asparagine residues (N412; Fig. 47A). It should be noted that there are clashes between the strands in the area of N412 which cannot be remedied without moving the backbone. Asparagine is frequently involved in protein active or binding sites. Its polar side-chain can interact with other polar or charged atoms (Betts et al., 2003). The walls of the chamber consist of 14 residues, namely 413FYIS416, I639 and V641. The walls are flanked by four histidine residues (H460 and H638) as well as further charged residues (E458, K515, R640, K642: Fig. 47B). The bottom forms an opening with residues 418PLI420 and 643VY644 (Fig. 47C). Some additional residues are located on the outside of the described structure: E458, F459

Though not identical the interface bears some resemblance to the II–II interface of *L. polyphemus* (Martin et al., 2007), which also features the complementary association of domain #3 forming a chamber-like structure. It is reasonable to assume the b–b interface described here might be able to bind an allosteric ligand in a similar fashion.

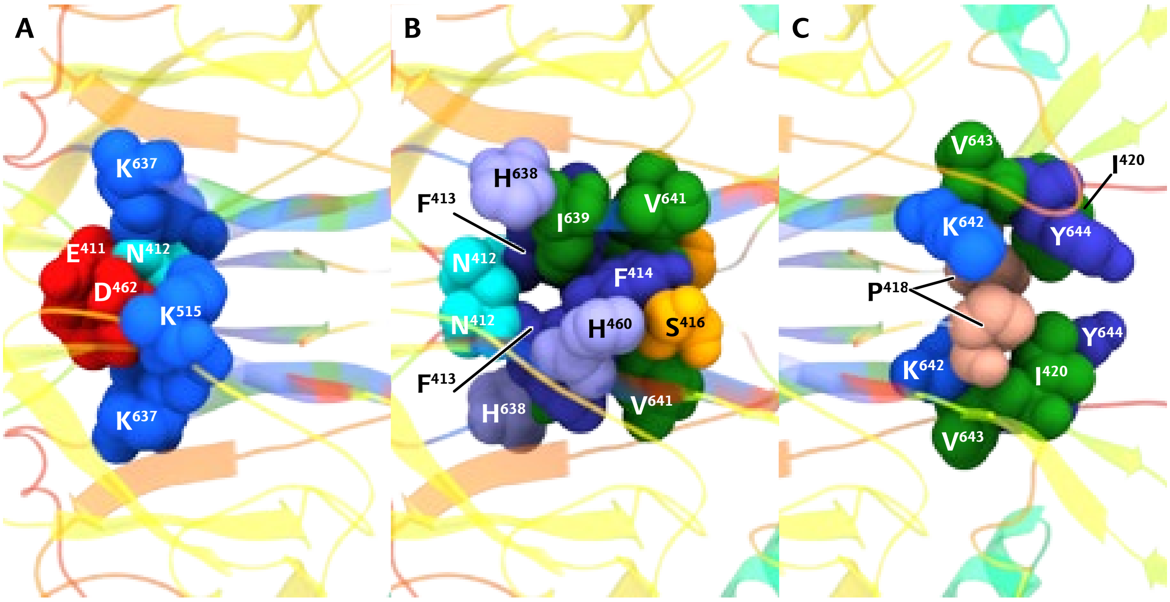


Fig. 47: The b–b interface between 1x6mers of the 2x6mer – breakdown into cap, middle and lower parts. The viewpoint is rotated by 90° around the x-axis, Panel A: Cap Panel B: Middle part with N412 displayed to help viewer orientation Panel C: lower section of the b–b interface.

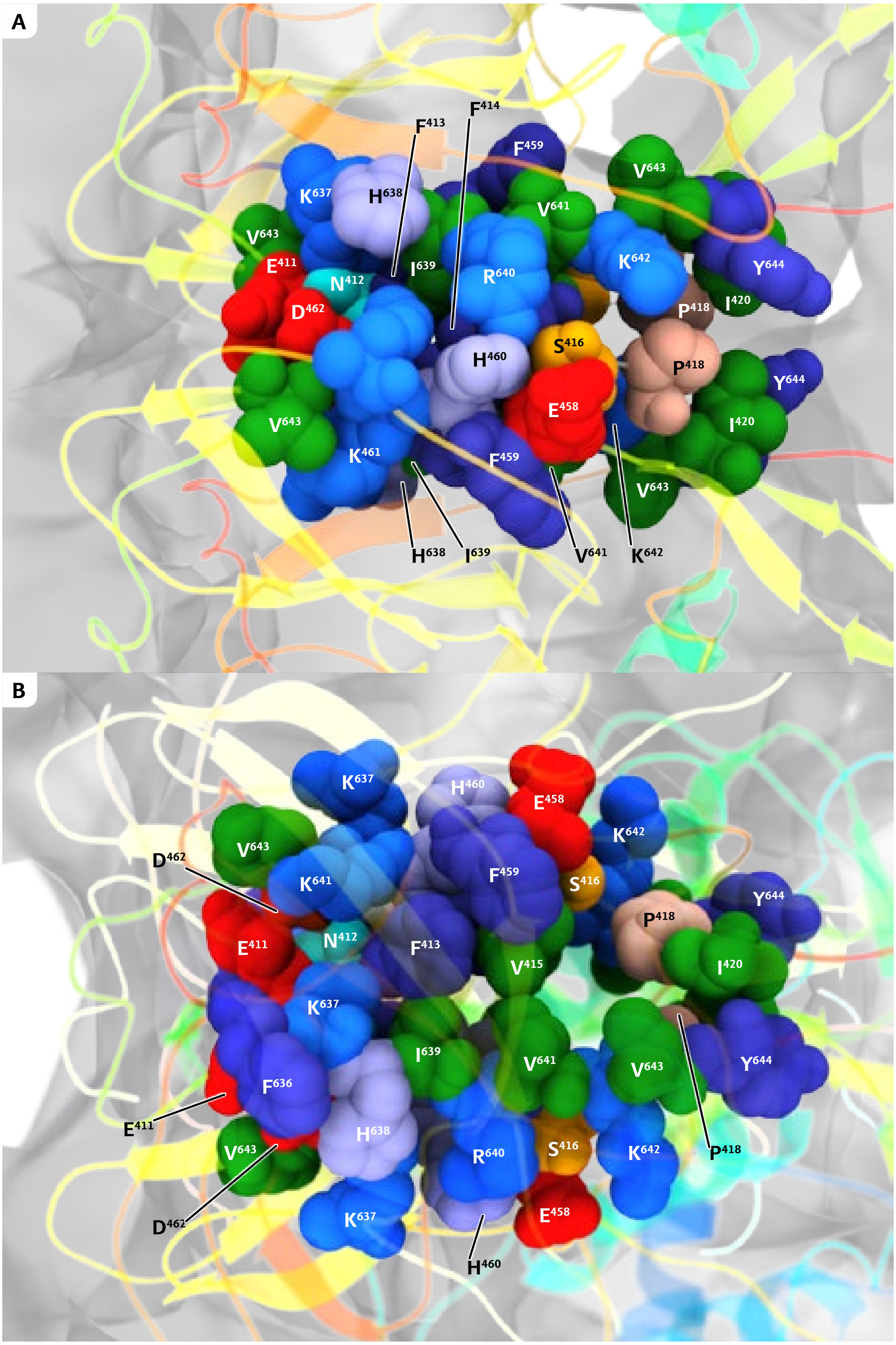


Fig. 48: The b–b interface between 1x6mers of the 2x6mer. Panel A: Front of the interface. Panel B: Back of the interface, viewpoint rotated by 90° around the x-axis. The secondary structure elements between the interface and the viewer are rendered more transparent as to not obstruct the depiction. The reconstructed density map is displayed for reference.

2.2.1.3 The b–a3 interface

Two copies of the b-a3 interface connect the 1x6mers to form the isopod type 2x6mer. The interface itself is subdivided into two adjacent contact points (Fig. 49). The contact is brought about by the long β3J→α3.7 loop of subunit a3 coming in contact with helices α1.3 and α1.5 of subunit b. Both components can be distinguished in the reconstructed density map at mass correlated threshold levels. Despite the prominent bridges visible in the reconstruction, a high frequency of glycine and alanine residues along with a relatively small contact area leads to relatively few opportunities for interaction.

The first contact K102-D589 has the potential to form a salt bridge (distance ~5 Å). K102 is flanked by a cysteine and histidine residue (C101 and H100). D589 is accompanied by a threonine T592. The second part of the contact centers on another potential salt bridge: S52-E511. In addition N470-Q53 can form a hydrogen bond. This site is flanked by a histidine residue (H469).

The b–a3 interface might be able to perform a rocking motion by modifying either one or both of the main salt bridges.

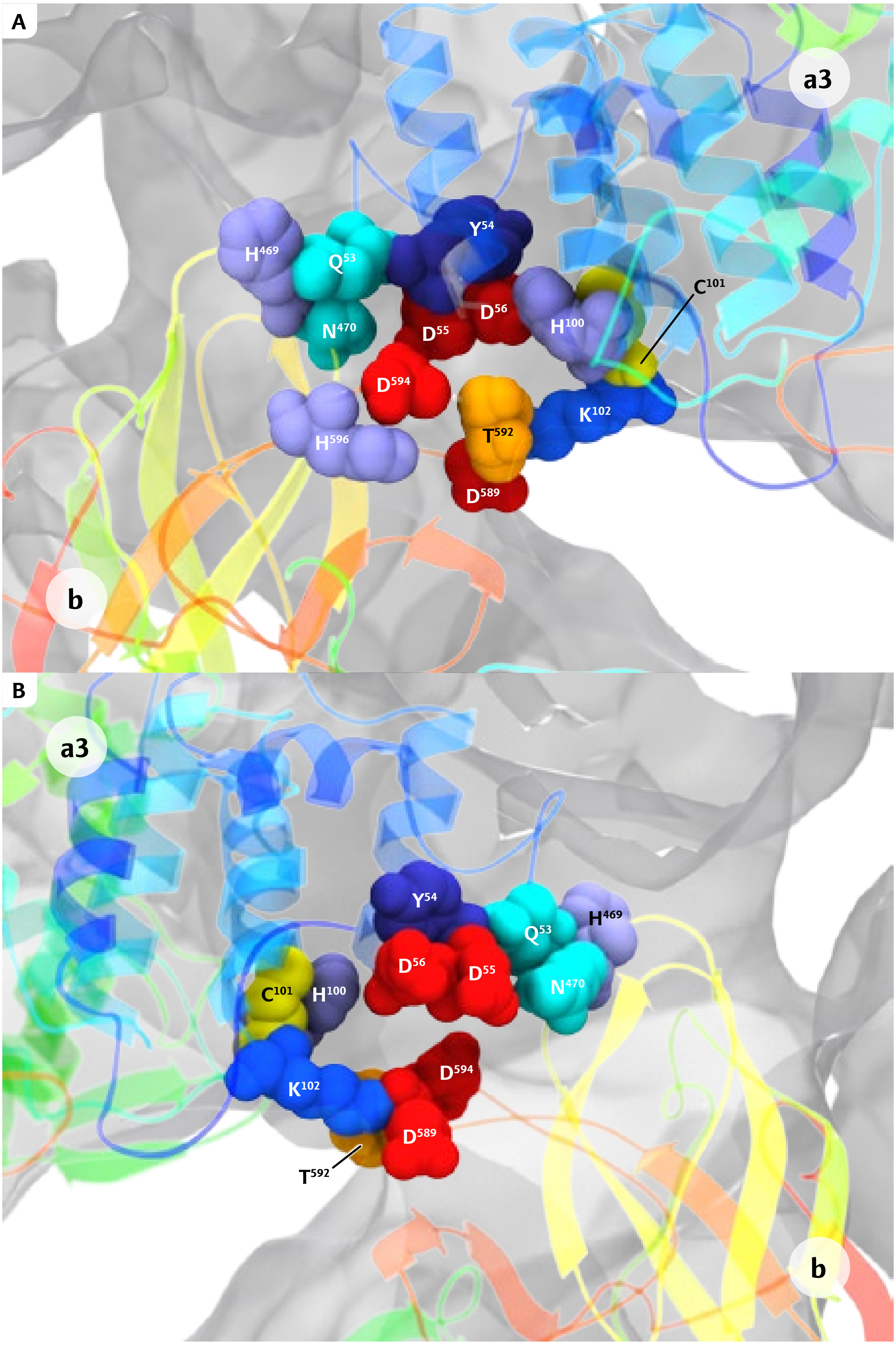


Fig. 49: The b–a3 interface. Two copies of the b-a3 interface connect the 1x6mers to form the isopod type 2x6mer. Panel A: Front of the interface. Panel B: Back of the interface, viewpoint rotated by 90° around the x-axis. The reconstructed density map is displayed for reference.

2.2.1.4 The a3–a3 interface

In the 4x6mer the a3–a3 interfaces are located adjacent to the b–a3 interface. The interface spans across the two 2x6mers. Based on the present data, despite an elongated area of possible interaction the actual interface contact is fairly small (Fig. 50 A and B). Indeed in the reconstructed volume a small bridge is visible followed by a distinct cleft between both subunits.

There is some overlapping which could not be resolved by pair-wise modeling because the backbone distance is only 2,5 Å in the contact area. The interface is formed by both subunits α1.2→α1.3 loop bringing both an aspartic acid and lysine residue together (DK(50-51)). This indicates the possibility of two strong salt bridges. In the current model the loop might be incorrectly modeled since both aspartic residues clash. There are no other residues present in the surrounding area which might facilitate ligand binding or allosteric interaction between the subunits.

2.2.2 Summary: cooperativity and allosteric mechanisms in *Ctr*Hc

Compared to the structure of highly cooperative hemocyanins like those found in *E. californicum* (Hartmann & Decker, 2002) or *L. polyphemus* (Martin et al., 2007) the range for movement appears to be more limited in the present molecular model of *C. truncata* hemocyanin. Due to the dense concentration of interfaces at the core of the 4x6mer with the involvement of only two subunit types and the resulting tight interconnectivity, even a slight rotation of one 1x6mer would affect numerous neighboring interfaces.

A higher resolution reconstruction will be required to model long loops correctly and verify the interface architecture. Nevertheless it has been proven that subunit *Ctr*Hc type 2 is essential to the assembly of both the 2x6mers and the 4x6mer of *C. truncata*.

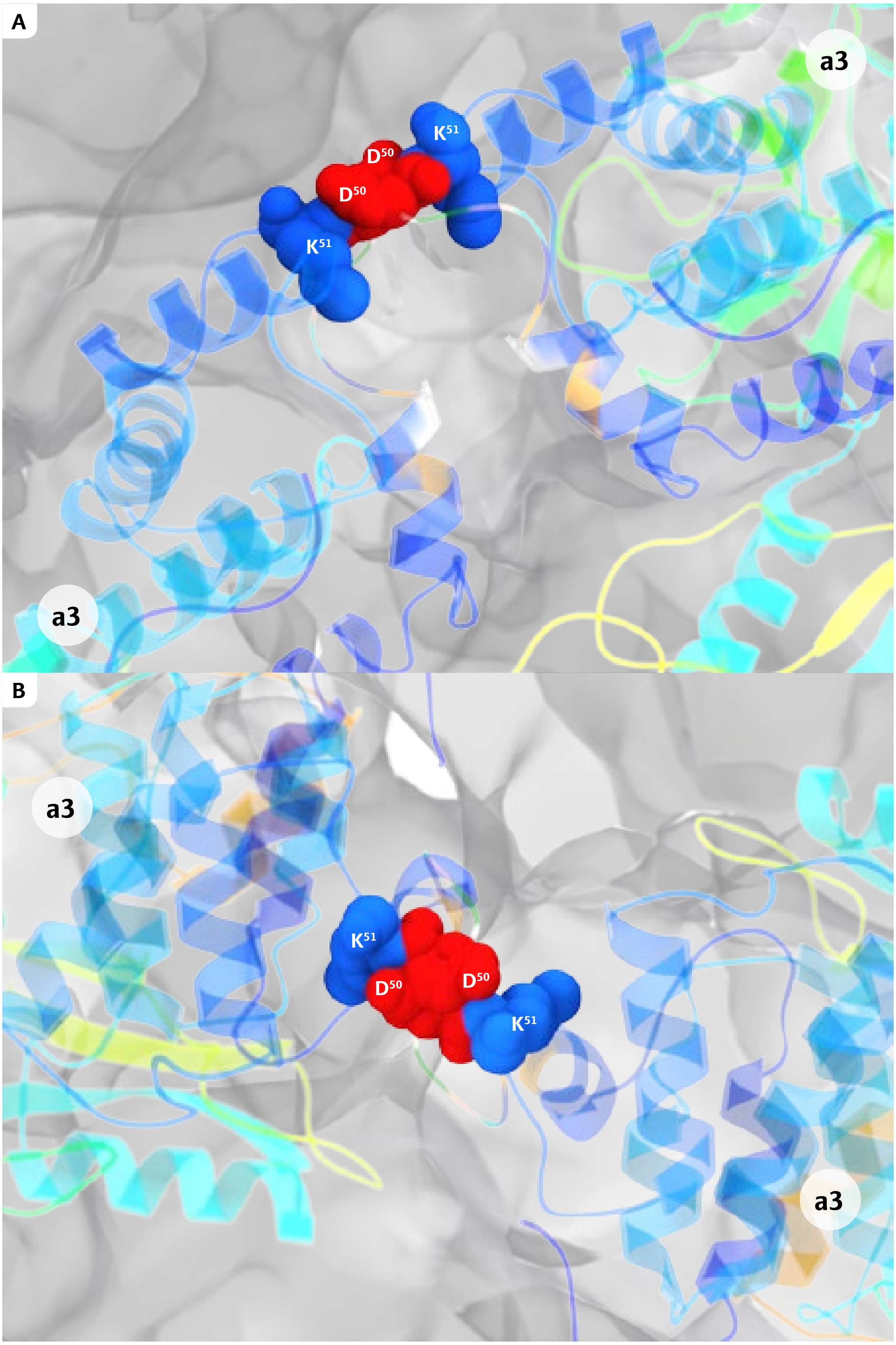


Fig. 50: The a3–a3 interface. Panel A: Front of the interface. Panel B: Back of the interface, viewpoint rotated by 90° around the x-axis. The reconstructed density map is displayed for reference.

3 3D-Reconstruction of *Lpo*Hc

In an established procedure I used the Eman 1.7 software package to produce 3D-reconstructions from the acquired dataset. Subsequently the data was processed *de novo* by Dr. Mario Dejung with the Sparx software package (Hohn et al., 2007; Tang et al., 2007; Baldwin & Penczek, 2007). The 1.7 version of Eman can be characterized as being a generation behind Sparx in terms of advanced 3D-reconstruction algorithms. The result obtained from the Sparx processing run with 38032 particles yielded a higher resolution (Eman 8,7 Å vs Sparx 9,0 Å; 0,5 criterion – see 3.1 on page 64). Thus it was used for further analysis in this study. The reconstruction was kindly provided by Dr. Mario Dejung, who performed all Sparx related work.

Overall the Sparx reconstruction appears to feature a more evenly distributed resolution, especially apparent in certain α-helical elements, whereas for the Eman result some masses have been cut of by the surface rendering threshold at similar levels (Fig. 51 and Fig. 52).

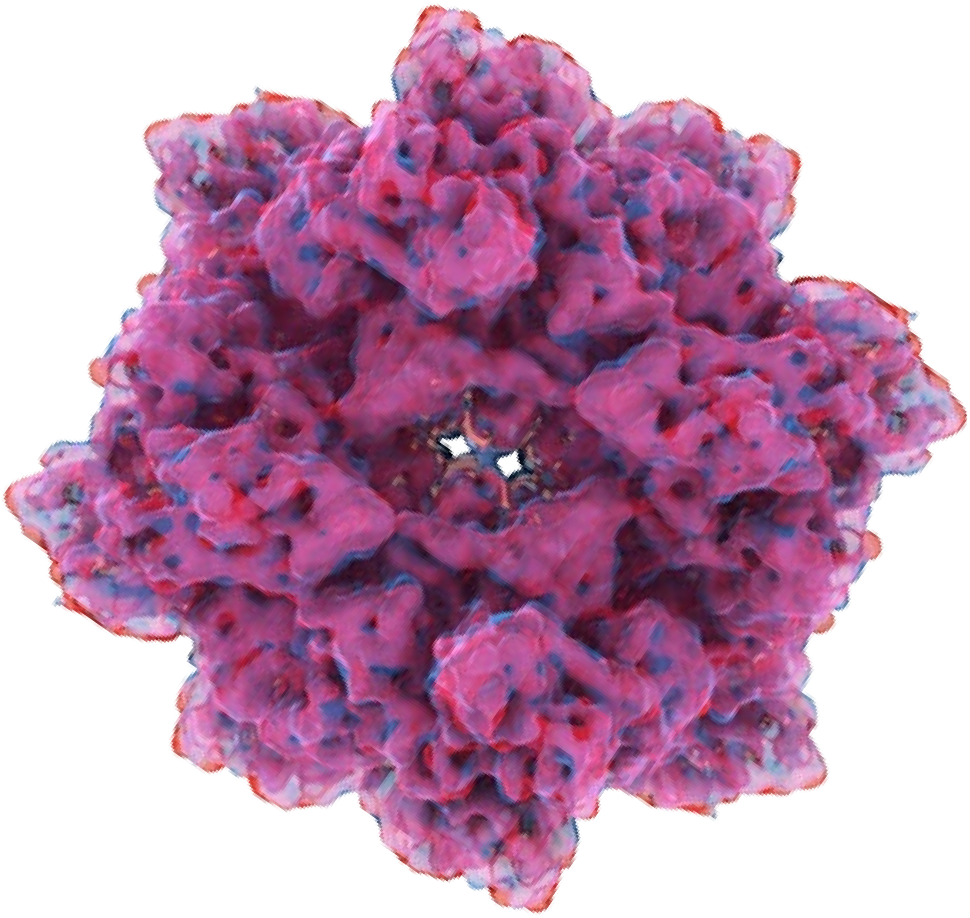


Fig. 51: Composite surface rendering comparison. View along the Z axis. The contour level threshold of both maps has been adjusted to show equal amounts of mass, albeit at less than mass correlated threshold levels to accentuate secondary structure elements. The Eman reconstruction is tinted red, the result produced by Sparx depicted blue. In areas with overlapping mass the colors combine to a show a magenta tint.

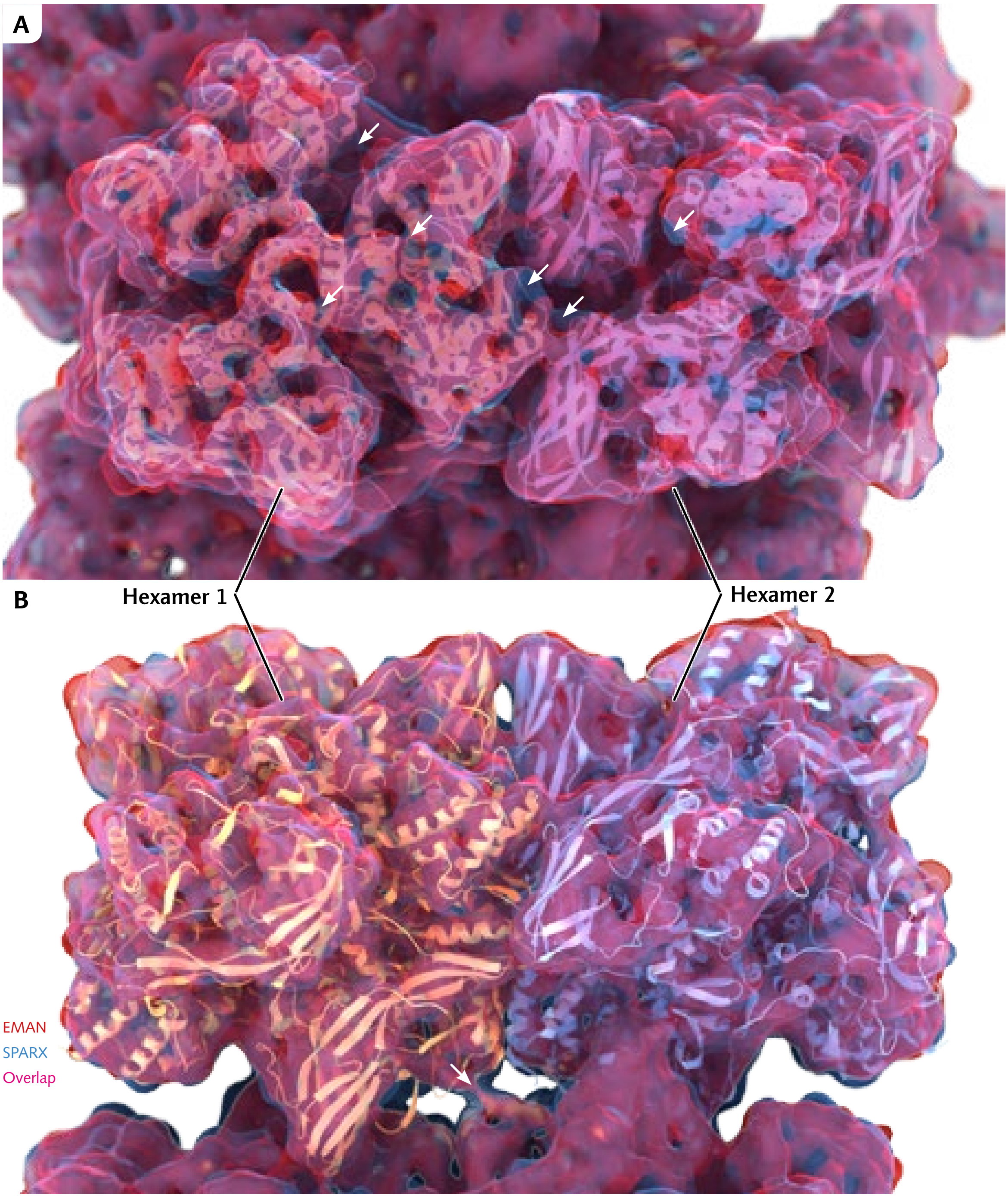


Fig. 52: Composite surface rendering comparison of Hexamers 1 and 2. Panel A: Top down view Panel B: Slanted view at 45° The contour level threshold of both maps has been adjusted to show equal amounts of mass, albeit at less than mass correlated threshold levels to accentuate secondary structure elements. The Eman reconstruction is tinted red, the result produced by Sparx depicted blue. In areas with overlapping mass the colors combine to a show a magenta tint. The molecular model is superimposed to facilitate structure validation. Notable features (marked by arrows): The Sparx reconstruction shows a more evenly distributed resolution in certain α-helical elements, whereas the Eman result exhibits missing areas (cut off by the high surface rendering threshold value).

3.1 Reconstruction Overview

3.1.1 Fourier Shell Correlation

According to the 0,143 criterion, which is the relevant criterion for the methods used to produce the reconstruction (Rosenthal & Henderson, 2003), the resolution is measured at 7,1 Å (Fig. 53). For comparison purposes other criteria for resolution assessment result in a resolution of 8,7 Å (conservative 0,5 criterion) and 7,8 Å (1/2bit criterion). The measurements were provided by Mario Dejung.

3.1.2 Volumetric Surface Representation

Fig. 54 presents the final reconstruction in the form of a volumetric surface rendering at mass correlated threshold levels. No sharpening artifacts are visible within a large range of threshold levels. Only at very low threshold values do some noise and few sharpening artifacts appear, but stay at reasonable amounts (data not shown). The impressions from the surface renderings are supported by the Fourier Shell Correlation measurements and appearance.

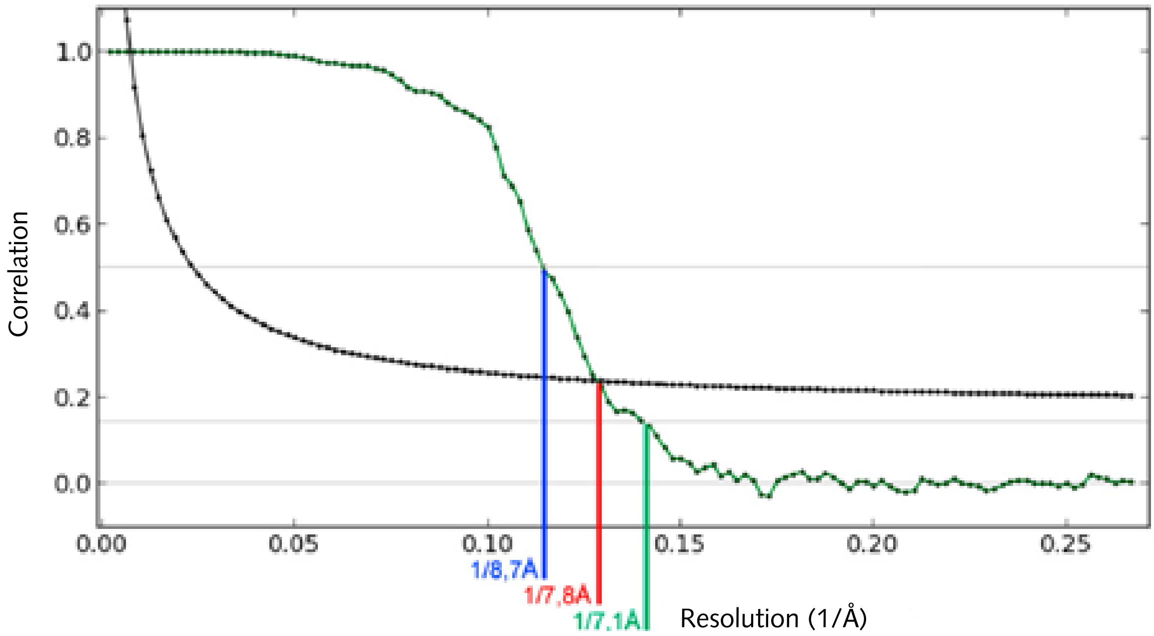


Fig. 53: Fourier Shell Correlation (FSC) of the Limulus polyphemus reconstruction. The correlation curve is depicted dark green. The 0,143 criterion with 7,1 Å resolution (intersection marked by green line) is applicable in this case according to the methods used to produce the reconstruction (Rosenthal & Henderson, 2003). Other criteria for resolution assessment are presented for comparison purposes: the conservative 0,5 criterion results in 8,7 Å resolution (intersection marked by blue line). The 1/2bit criterion curve (black) intersects at 7,8 Å (marked by red line). Graph provided by Dr. Mario Dejung.

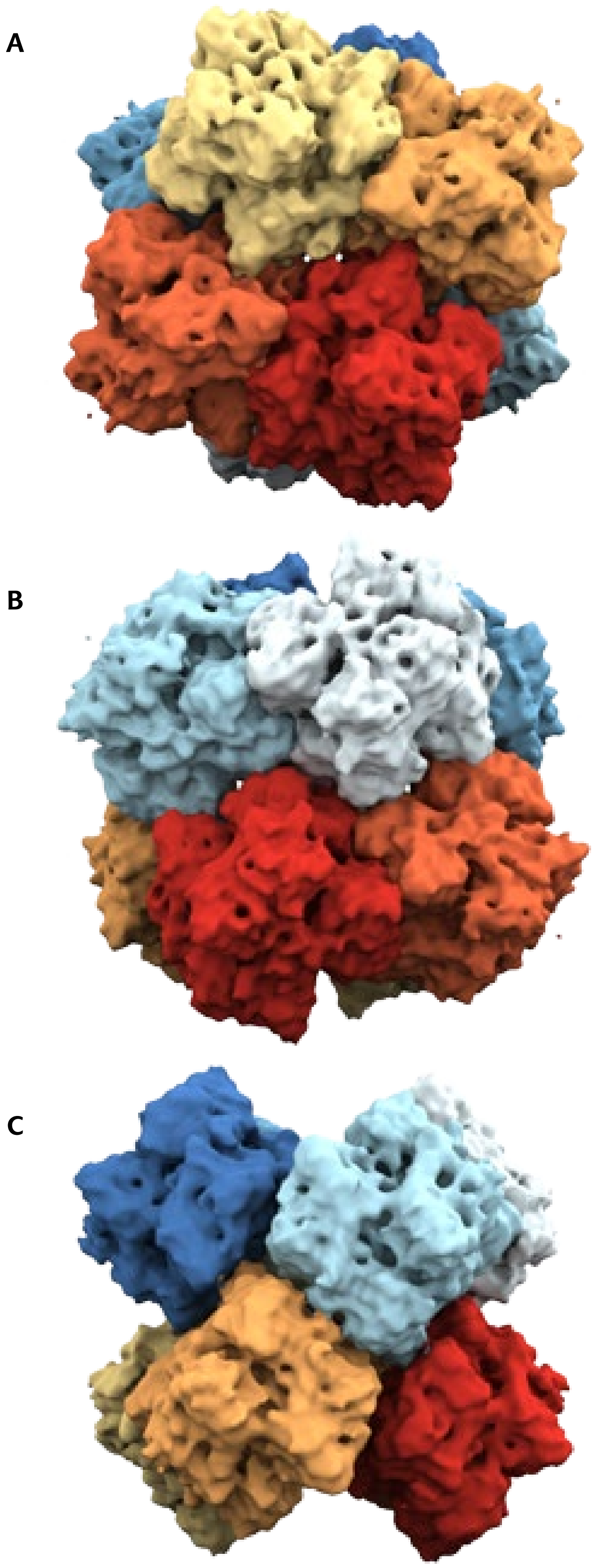


*Fig. 54: Volumetric surface rendering of the 3D-reconstruction of Lpo*Hc. Hexamers 1 through 4 are colored dark blue to light blue. Hexamers 5 though 8 are colored red to yellow. Panel A: view along the Z axis. Panel B: view along the X axis. Panel C: view along the Y axis.

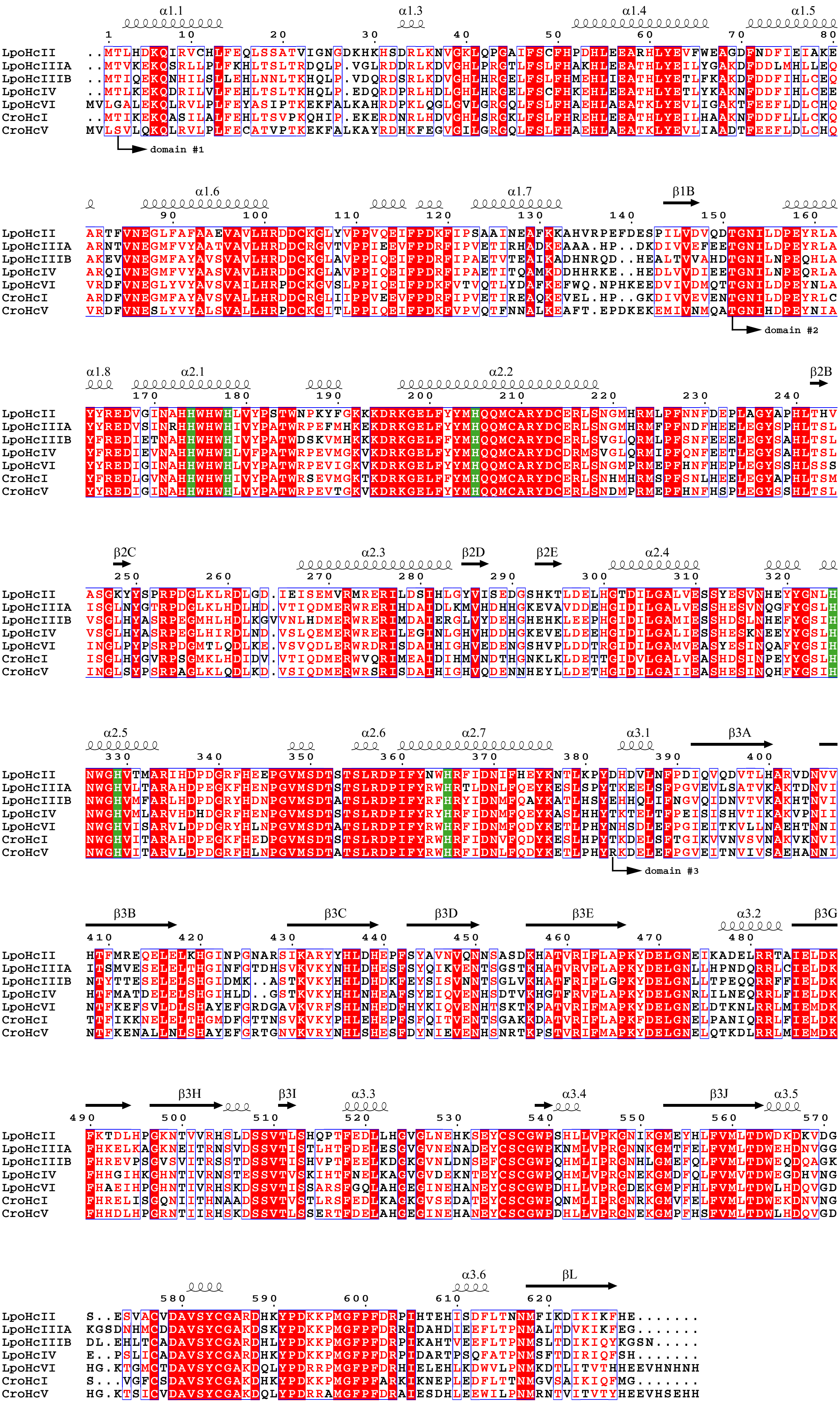
4 Flexible Fitting of *Lpo*Hc

4.1 Molecular model

The molecular model used in this study differs from previously published works (Martin et al., 2007) in that sequences from *Eurypelma californicum* are no longer used. As the sequence from *Carcinoscorpius rotundicauda*, the Southeast Asian horseshoe crab became available (Jiang et al., 2005a and b), subunits I and V were rebuilt with the new data using homology modeling. In addition the *L. polyphemus* subunit IIIB (Rehm et al., 2012) was incorporated into the model.

Fig. 55 summarizes the sequences used in an alignment. For orientation purposes the secondary structure elements as deduced from the X-ray structure of subunit II are also indicated (Hazes et al., 1993).

Fig. 55: Figure on facing page: Alignment of Limulus polyphemus and Carcinoscorpius rotundicauda subunits. Alignment calculated with Clustal Omega. Global similarity frames are printed with blue borders, strictly conserved amino acids are shaded red, similar amino acids in a similarity group are printed red. The Cu-binding histidine residues are shaded green. Secondary structure elements as deduced from X-ray structure of subunit II are indicated at the top. The starting points of domain #1-3 are marked at the bottom.



```
LpoHcII   ..MTLHDKQIRVCHLFEQLSSATVIGNGDKHKHSDRLKNVGKLQPGAIFSCFHPDHLEEARHLYEVFWEAGDFNDFIEIAKE
LpoHcIIIA ..MTVKEKQSRLLPLFKHLTSLTRDQLP.VGLRDDRLKDVGHLPRGTLFSLFHAKHLEEATHLYEILYGAKDFDDLMHLLEQ
LpoHcIIIB ..MTIQEKQNHILSLLEHLNNLTKHQLP.VDQRDSRLKDVGHLHRGELFSLFHMEHLIEATHLYETLFKAKDFDDFIHLCEQ
LpoHcIV   ..MTLKEKQDRILVLFEHLTSLTKHQLP.EDQRDPRLHDLGHLHRGELFSCFHKEHLEEATHLYETLYKAKNFDDFIHLCEE
LpoHcVI   MVLGALEKQLRVLPLFEYASIPTKEKFALKAHRDPKLQGLGVLGRGQLFSLFHAEHLAEATKLYEVLIGAKTFEEFLDLCHQ
CroHcI    ..MTIKEKQASILALFEHLTSVPKQHIP.EKERDNRLHDVGHLSRGKLFSLFHREHLEEATHLYEILHAAKNFDDFLLLCKQ
CroHcV    MVLSVLQKQLRVLPLFECATVPTKEKFALKAYRDHKFEGVGILGRGQLFSLFHAEHLAEATKLYEVLIAADTFEEFLDLCHQ

LpoHcII   ARTFVNEGLFAFAAEVAVLHRDDCKGLYVPPVQEIFPDKFIPSAAINEAFKKAHVRPEFDESPILVDVQDTGNILDPEYRLA
LpoHcIIIA ARNTVNEGMFVYAATVAVLHRDDCRGVTVPPIEEVFPDRFIPVETIRHADKEAAA.HP..DKDIVVEFEETGNILDPEYRLA
LpoHcIIIB AKEVVNEGMFAYAVSVAVLHRDDCKGLAVPPIQEIFPDRFIPAETVTEAIKADHNRQD..HEALTVVAHDTGNILNPEQHLA
LpoHcIV   ARQIVNEGMFVYAASVAVLHRDDCKGLAVPPIQEIFPDRFIPAETITQAMKDDHHRKE..HEDLVVDIEETGNILNPEQRLA
LpoHcVI   VRDFVNEGLYVYAVSVAILHRPDCKGVSLPPIQEIFPDKFVTVQTLYDAFKEFWQ.NPHKEEDVIVDMQTTGNILDPEYNLA
CroHcI    ARDFVNEGMFAYAVSVALLHRDDCRGLIIPPVEEVFPDRFIPVETIREAQKEVEL.HP..GKDIVVEVENTGNILDPEYRLC
CroHcV    VRDFVNESLYVYALSVALLHRPDCKGITLPPIQEIFPDKFVPVQTFNNALKEAFT.EPDKEKEMIVNMQATGNIHDPEYNIA

LpoHcII   YYREDVGINAHHWHWHLVYPSTWNPKYFGKKKDRKGELFYYMHQQMCARYDCERLSNGMHRMLPFNNFDEPLAGYAPHLTHV
LpoHcIIIA YYREDVSINRHHWHWHIVYPATWRPEFMHKEKDRKGELFYYMHQQMCARYDCERLSNGMHRMFPFNDFHEELEGYSPHLTSL
LpoHcIIIB YFREDIETNAHHWHWHIVYPATWDSKVMHKKKDRKGELFYYMHQQMCARYDCERLSVGLQRMLPFSNFEEELEGYSAHLTSL
LpoHcIV   YFREDIEVNAHHWHWHLVFPATWRPEVMGKVKDRKGELFYYMHQQMCARYDCDRMSVGLQRMIPFQNFEETLEGYSAHLTSL
LpoHcVI   YYREDIGINAHHWHWHLVYPATWRPEVIGKVKDRKGELFYYMHQQMCARYDCERLSNGMPRMEPFHNFHEPLEGYSSHLSSS
CroHcI    YFREDLGVNAHHWHWHIVYPATWRSEVMGKTKDRKGELFYYMHQQMCARYDCERLSNHMHRMSPFSNLHEELEGYAPHLTSM
CroHcV    YYREDIGINAHHWHWHLVYPATWRPEVTGKVKDRKGELFYYMHQQMCARYDCERLSNDMPRMEPFHNFHSPLEGYSSHLTSL

LpoHcII   ASGKYYSPRPDGLKLRDLGD.IEISEMVRMRERILDSIHLGYVISEDGSHKTLDELHGTDILGALVESSYESVNHEYYGNLH
LpoHcIIIA ISGLNYGTRPDGLKLHDLHD.VTIQDMERWRERIHDAIDLKMVHDHHGKEVAVDDEHGIDILGALVESSHESVNQGFYGSLH
LpoHcIIIB VSGLHYASRPEGMHLHDLKGVVNLHDMERWRERIMDAIERGLVYDEHGHEHKLEEPHGIDILGAMIESSHDSLNHEFYGSIH
LpoHcIV   VSGLHYASRPEGLHIRDLND.VSLQEMERWRERILEGINLGHVHDDHGKEVELDEEHGIDILGALIESSHESKNEEYYGSLH
LpoHcVI   INGLPYPSRPDGMTLQDLKE.VSVQDLERWRDRISDAIHIGHVEDENGSHVPLDDTRGIDILGAMVEASYESINQAFYGSIH
CroHcI    ISGLHYGVRPSGMKLHDIDV.VTIQDMERWVQRIMEAIDIHMVNDTHGNKLKLDETTGIDVLGALVEASHDSINPEYYGSLH
CroHcV    INGLSYPSRPAGLKLQDLKD.VSIQDMERWRSRISDAIHIGHVQDENNHEYLLDETHGIDILGAIIEASHESINQHFYGSIH

LpoHcII   NWGHVTMARIHDPDGRFHEEPGVMSDTSTSLRDPIFYNWHRFIDNIFHEYKNTLKPYDHDVLNFPDIQVQDVTLHARVDNVV
LpoHcIIIA NWGHVLTARAHDPEGKFHENPGVMSDTSTSLRDPIFYRWHRTLDNLFQEYKESLSPYTKEELSFPGVEVLSATVKAKTDNVI
LpoHcIIIB NWGHVMFARLHDPDGRYHDNPGVMSDTATSLRDPIFYRFHRYIDNMFQAYKATLHSYEHQLIFNGVQIDNVTVKAKHTNVI
LpoHcIV   NWGHVMLARVHDHDGRFHENPGVMSDTSTSLRDPIFYRYHRFIDNMFQEYKASLHHYTKTELTFPEISISHVTIKAKVPNII
LpoHcVI   NWGHVISARVLDPDGRYHLNPGVMSDTATSLRDPIFYRWHRFIDNMFQDYKETLPHYNHSDLEFPGIEITKVLLNAEHTNNI
CroHcI    NWGHVITARAHDPEGKFHEDPGVMSDTSTSLRDPIFYRWHRFIDNVFQDYKESLHPYTKDELSFTGIKVVNVSVNAKVKNVI
CroHcV    NWGHVITARVLDPDGRFHLNPGVMSDTATSLRDPIFYRWHRFIDNLFQDYKETLPHYRKDELEFPGVEITNVIVSAEHANNI

LpoHcII   HTFMREQELELKHGINPGNARSIKARYYHLDHEPFSYAVNVQNNSASDKHATVRIFLAPKYDELGNEIKADELRRTAIELDK
LpoHcIIIA ITSMVESELELTHGINFGTDHSVKVKYNHLDHESFSYQIKVENTSGSTKHATVRVFLAPKYDELGNLLHPNDQRRLCIELDK
LpoHcIIIB NTYTTESELELSHGIDMK..ASTKVKYHHLDHDKFEYSISVNNTSGLVKHATFRIFLGPKYDELGNLLTPEQQRRFFIELDK
LpoHcIV   HTFMATDELELSHGIHLD..GSTKVKYHHLNHEAFSYEIQVENHSDTVKHGTFRVFLAPKYDELGNRLILNEQRRLFIELDK
LpoHcVI   NTFKEFSVLDLSHAYEFGRDGAVKVRFSHLNHEDFHYKIQVENHTSKTKPATVRIFMAPKYDELGNELDTKNLRRLMIEMDK
CroHcI    TTFIKKNELELTHGMDFGTTNSVKVKYPHLEHEPFSFQITVENTSGAKKDATVRIFLAPKFDELGNELPANIQRRLFIELDK
CroHcV    NTFKENALLNLSHAYEFGRTGNVKVRYNHLSHESFDYNIEVENHSNRTKPSTVRIFMAPKYDELGNELQTKDLRRLMIEMDK

LpoHcII   FKTDLHPGKNTVVRHSLDSSVTLSHQPTFEDLLHGVGLNEHKSEYCSCGWPSHLLVPKGNIKGMEYHLFVMLTDWDKDKVDG
LpoHcIIIA FHKELKAGKNEITRNSVDSSVTISLHTFDELESGVGVNENADEYCSCGWPKNMLVPRGNNKGMTFELFVMLTDWEHDNVGG
LpoHcIIIB FHREVPSGCVSVITRSSTDSSVTISHVPTFEELKDGKGVNLDNSEFCSCGWPQHMLIPRGNHLGMEFQLFVMLTDWEQDQAGK
LpoHcIV   FHHGIHKGHNTIVRNSTESSVTVSKIHTFNELKAGVGVDEKNTEYCSCGWPQHMLVPRGNEKGMDFQLFVMVTDWEGDHVNG
LpoHcVI   FHAEIHPGHNTIVRHSKDSSVTISSARSFGQLAHGEGINEHANEYCSCGWPDHLLVPRGDEKGMPFHLFVMLTDWLHDQVGD
CroHcI    FHRELISGQNIITHNAADSSVTVSTLRSFEDLKAGKGVSEDATEYCSCGWPQNMLIPRGNRKGMVFELFVMLTDWEKDNVNG
CroHcV    FHHDLHPGRNTIIRHSKDSSVTLSSERTFDELAHGEGINEHANEYCSCGWPDHLLVPRGNEKGMPFHSFVMLTDWLHDQVGD

LpoHcII   S..ESVACVDAVSYCGARDHKYPDKKPMGFPFDRPIHTEHISDFLTNNMFIKDIKIKFHE.......
LpoHcIIIA KGSDNHMCDDAVSYCGAKDSKYPDKKPMGFPFDRRIDAHDIEEFLTPNMALTDVKIKFEG.......
LpoHcIIIB DL.EHLTCADAVSYCGARDHLYPDKKPMGFPFDRPIKAHTVEEFLTPNMSLTDIKIQYKGSN.....
LpoHcIV   E..PSLICADAVSYCGARDHKYPDKKPMGFPFDRPIDARTPSQFATPNMSFTDIRIQFSH.......
LpoHcVI   HG.KTGMCTDAVSYCGAKDQLYPDRRPMGFPFDRHIELEHLKDWVLPNMKDTLITVTHHEEVHNHNH
CroHcI    S..VGFCSDAVSYCGAKDHKYPDKKPMGFPFARKIKNEPLEDFLTTNMGVSAIKIQFMG.......
CroHcV    HG.KTSICVDAVSYCGAKDQLYPDRRAMGFPFDRAIESDHLEEWILPNMRNTVITVTYHEEVHSEHH
```

4.2 Preparation for the fitting process

The Sparx-3D-reconstruction obtained from Mario Dejung was split into 12 segments using the Chimera functions color-zone and split map. Due to the D2-symmetry imposed on the reconstruction, the flexible fitting was performed on an isolated 2x6mer (Fig. 56). The resulting structures were subsequently copied and applied to the remaining symmetrically identical positions within the density map to complete the final model.

The radius value of the color-zone-tool was chosen so that masses between subunits would be cut as close as possible and at the same time resulting in balanced and symmetric cuts. This is to avoid large empty masses (otherwise occupied by adjacent subunits) that might present a strong influence for the fitting algorithm. After segmentation the box-size of each subunits density map was reduced to 50 voxels to aid stable and fast processing.

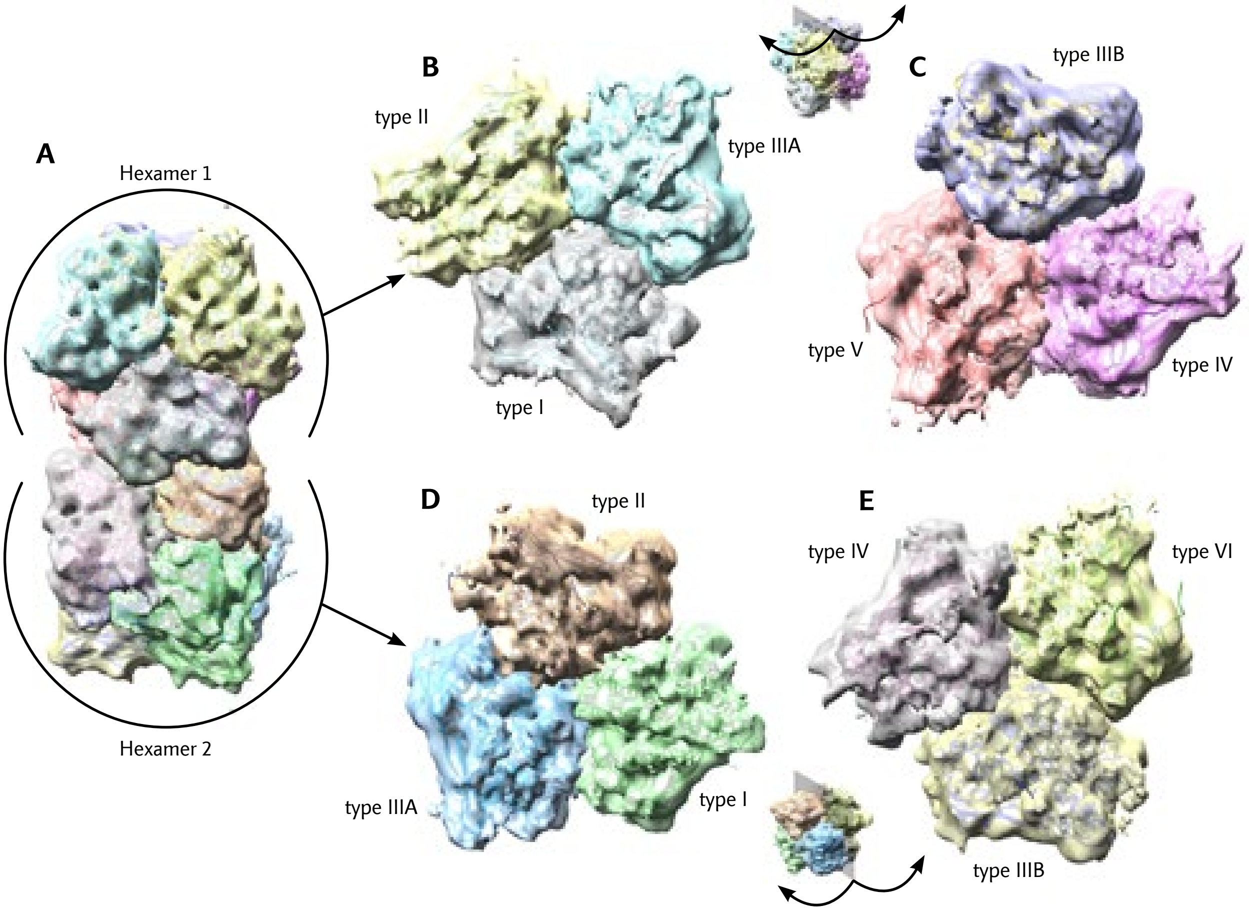


Fig. 56: The 2x6mer as used for the flexible fitting process. **A:** The 2x6mer looking onto the symmetry axis of hexamer 1. **B:** Subunits of hexamer 1 – type I (light gray), II (yellow) and IIIA (cyan), as seen from the inside – the interfaces connecting to the inner half the hexamer are exposed in this view **C:** Second half of hexamer 1, the subunits are of type IIIB (dark blue), IV (magenta) and V (light red). **D and E:** A cutaway showing the interfaces connecting the subunits of hexamer 2, subunit types I (green), type II (brown) and IIIA (blue) in panel D, subunit types IIIB (ocher), IV (dark gray) and VI (yellow) in panel E.

4.3 Flexible Fitting

As described in 4.15.2 the flexible fitting was carried out with the DireX software. The output was optimized with the YASARA web-server and finally any remaining rotamer errors manually corrected in CHIMERA (as indicated by MolProbity).

4.3.1 Quality Assessment

Evaluations were carried out at each step of the process, with both the MolProbity web server and the What-If molecular model evaluation service (see chapter 4.15.3).

In order to gain an overall impression of the quality of the flexible fitting as well as its effects on the integrity of the molecular model, Z-scores calculated with the What-If package are shown in Fig. 57 through Fig. 62. This is an overall summary of the quality of the structure as compared with current reliable structures. To indicate the normality of a score, the score is expressed as a Z-value or Z-score. This is the number of standard deviations that the score deviates from the expected value. A property of Z-values is that the root-mean-square of a group of Z-values (the RMS Z-value) is expected to be 1.0. Z-values above 4.0 and below -4.0 are very uncommon (modified from the What-if documentation; Vriend 1990). The RMS Z-values give an impression of how well the model conforms to common refinement restraint values.

The following RMS Z-value parameters are presented:

- Bond lengths and angles, Omega angle restraints: values below 1,0 can be described as ‘tight’ angles and lengths whereas values above 1.0 indicate ‘loose’ angles and lengths (Vriend 1990).
- Side chain planarity and dihedral distribution: Values close to 1.0 indicate a valid structure (Vriend 1990)

The following Z-value parameters are presented:

- Packing quality: Low-resolution structures and structures solved partially or wholly by computational methods tend to form clusters of atoms in some places and holes or voids in others. This is referred to as bad packing of the atoms. Guidelines for first generation packing quality: Every molecule with a global quality below -2.7 is guaranteed wrong. A molecule with a quality below -2.0 might be mis-folded or poorly refined. Every molecule with a global quality below -1.2 does not belong in a database of reliable structures (Vriend 1990).
- Second generation packing quality: This test is based on the same principle as the first generation packing quality test, however it uses a newer, more refined scale. Every molecule with a quality Z-score for all atoms below -5.0 is guaranteed wrong. A molecule with a quality below -3.0 might be mis-folded or poorly refined. Every molecule with a global quality below -2.0 does not belong in a database of reliable structures (Vriend 1990).
- Inside – outside distribution normality: Returns an inside – outside normality score that is calibrated against PDB files of water-soluble molecules (Vriend 1990).

- Ramachandran plot appearance, chi-1/chi-2 rotamer normality and backbone conformation: parameters are compared with current reliable structures.

Although first selected to minimize the flexible fitting results, Chirions unresolvable failure to produce reliable output with certain subunits (especially type IIIB and IV subunits) along with an overall improved performance led to the decision to use the YASARA service instead. Numerous other energy optimization protocols were also evaluated (data not shown).

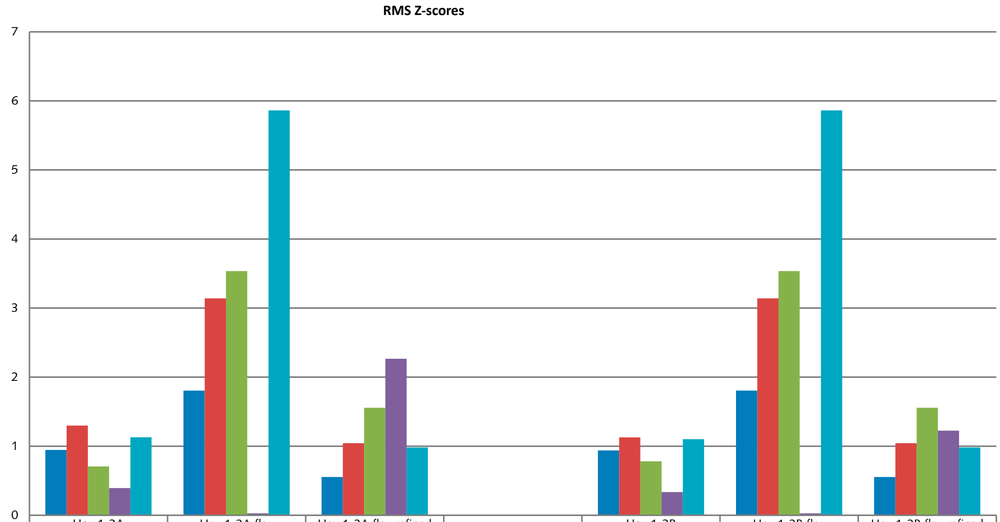
RMS Z-scores

	Hex 1-1	Hex 1-1 flex	Hex 1-1 flex refined		Hex 1-2	Hex 1-2 flex	Hex 1-2 flex refined
■ Bond lengths	0.950	1.783	0.522		0.948	1.824	0.529
■ Bond angles	1.274	3.029	0.992		1.269	3.121	0.994
■ Omega angle restraints	0.730	3.137	1.519		0.692	3.632	1.589
■ Side chain planarity	0.298	0.025	1.164		0.337	0.027	1.321
■ Improper dihedral distribution	0.996	5.397	1.000		0.984	6.092	1.049

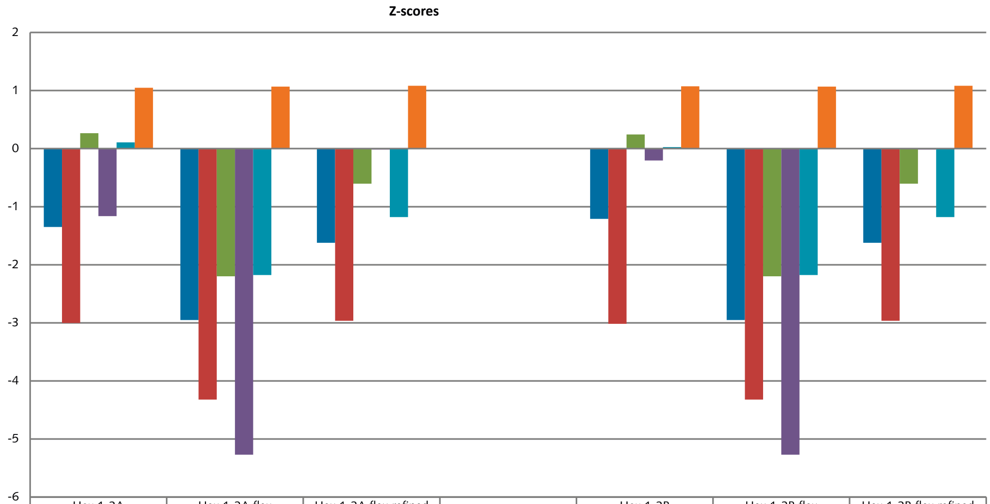
Z-scores

	Hex 1-1	Hex 1-1 flex	Hex 1-1 flex refined		Hex 1-2	Hex 1-2 flex	Hex 1-2 flex refined
■ 1st generation packing quality	-1.128	-2.646	-1.408		-1.070	-2.917	-1.769
■ 2nd generation packing quality	-3.181	-4.016	-2.548		-2.826	-4.236	-2.938
■ Ramachandran plot appearance	0.401	-2.071	-0.516		0.532	-1.946	-0.666
■ chi-1/chi-2 rotamer normality	-0.598	-5.182	0.229		0.600	-4.765	0.316
■ Backbone conformation	0.125	-1.740	-0.741		0.387	-1.911	-0.798
■ Inside/Outside distribution	1.045	1.046	1.0511		.0521	.060	1.062

Fig. 57: What-if results for subunits type I and II, segments from hexamer 1. For RMS Z-scores, closer to 1.0 is better, for structure Z-scores, positive is better than average.

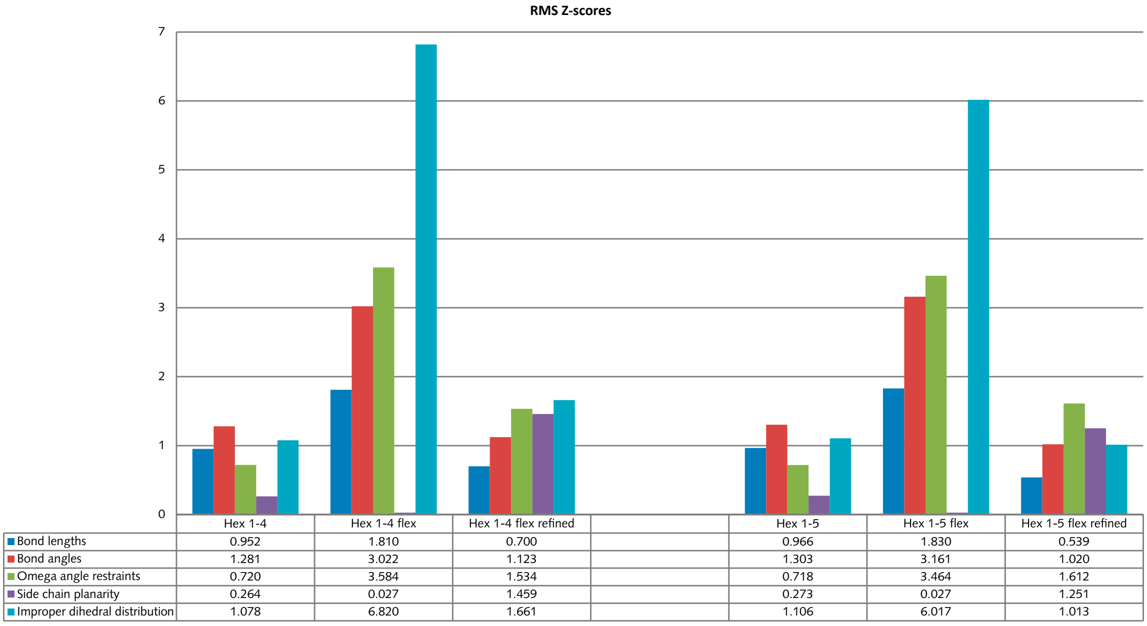


	Hex 1-3A	Hex 1-3A flex	Hex 1-3A flex refined		Hex 1-3B	Hex 1-3B flex	Hex 1-3B flex refined
🟦 Bond lengths	0.946	1.805	0.554		0.939	1.805	0.554
🟥 Bond angles	1.298	3.140	1.044		1.127	3.140	1.044
🟩 Omega angle restraints	0.706	3.536	1.556		0.782	3.536	1.556
🟪 Side chain planarity	0.393	0.027	2.266		0.335	0.027	1.226
🟦 Improper dihedral distribution	1.129	5.863	0.981		1.101	5.863	0.981

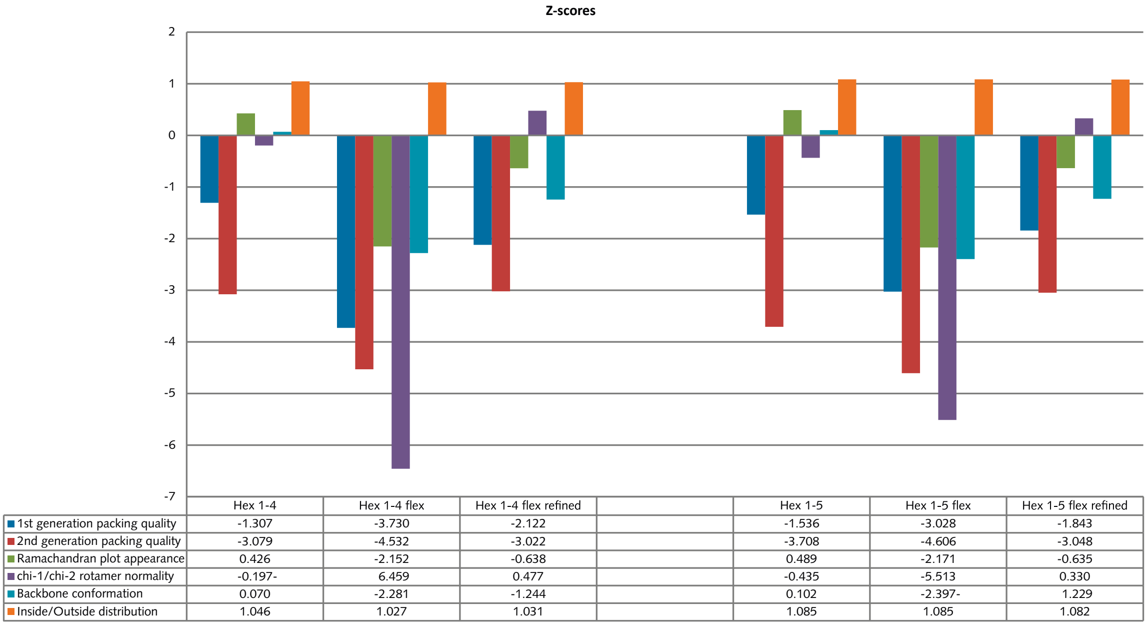


	Hex 1-3A	Hex 1-3A flex	Hex 1-3A flex refined		Hex 1-3B	Hex 1-3B flex	Hex 1-3B flex refined
🟦 1st generation packing quality	-1.350	-2.952	-1.622		-1.210	-2.952	-1.622
🟥 2nd generation packing quality	-3.002	-4.321	-2.964		-3.017	-4.321	-2.964
🟩 Ramachandran plot appearance	0.265	-2.199	-0.603		0.243-	2.199-	0.603
🟪 chi-1/chi-2 rotamer normality	-1.163	-5.269	-0.016		-0.204	-5.269	-0.016
🟦 Backbone conformation	0.108	-2.177	-1.179		0.026	-2.177	-1.179
🟧 Inside/Outside distribution	1.0471	.067	1.081		1.073	1.067	1.081

Fig. 58: What-if results for subunits type IIIA and IIIB, segments from hexamer 1. For RMS Z-scores, closer to 1.0 is better, for structure Z-scores, positive is better than average.

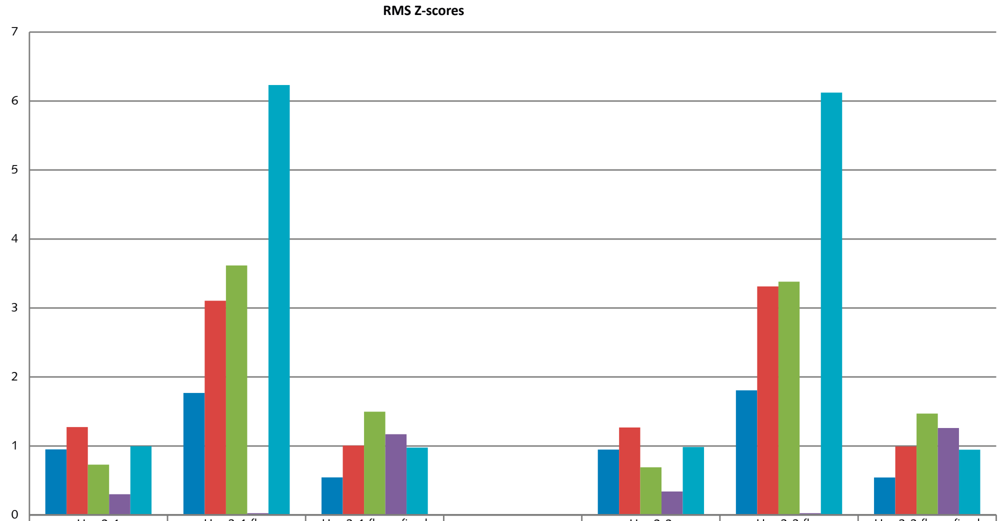


	Hex 1-4	Hex 1-4 flex	Hex 1-4 flex refined		Hex 1-5	Hex 1-5 flex	Hex 1-5 flex refined
🟦 Bond lengths	0.952	1.810	0.700		0.966	1.830	0.539
🟥 Bond angles	1.281	3.022	1.123		1.303	3.161	1.020
🟩 Omega angle restraints	0.720	3.584	1.534		0.718	3.464	1.612
🟪 Side chain planarity	0.264	0.027	1.459		0.273	0.027	1.251
🟦 Improper dihedral distribution	1.078	6.820	1.661		1.106	6.017	1.013

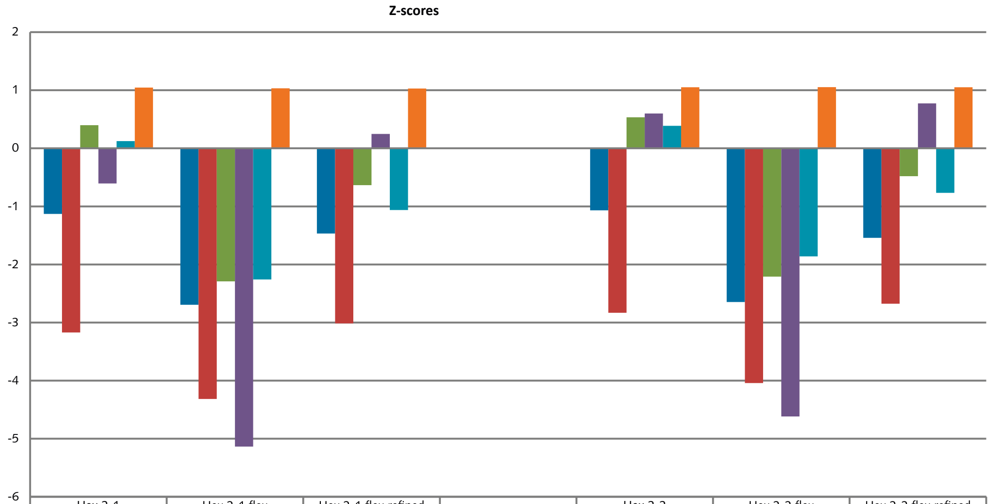


	Hex 1-4	Hex 1-4 flex	Hex 1-4 flex refined		Hex 1-5	Hex 1-5 flex	Hex 1-5 flex refined
🟦 1st generation packing quality	-1.307	-3.730	-2.122		-1.536	-3.028	-1.843
🟥 2nd generation packing quality	-3.079	-4.532	-3.022		-3.708	-4.606	-3.048
🟩 Ramachandran plot appearance	0.426	-2.152	-0.638		0.489	-2.171	-0.635
🟪 chi-1/chi-2 rotamer normality	-0.197-	6.459	0.477		-0.435	-5.513	0.330
🟦 Backbone conformation	0.070	-2.281	-1.244		0.102	-2.397-	1.229
🟧 Inside/Outside distribution	1.046	1.027	1.031		1.085	1.085	1.082

Fig. 59: What-if results for subunits type IV and V, segments from hexamer 1. For RMS Z-scores, closer to 1.0 is better, for structure Z-scores, positive is better than average.

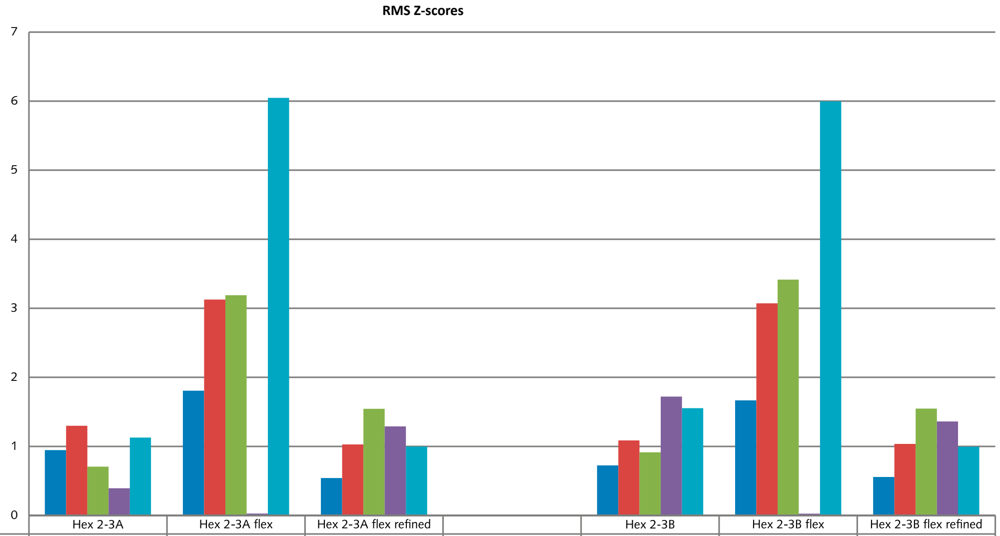


	Hex 2-1	Hex 2-1 flex	Hex 2-1 flex refined		Hex 2-2	Hex 2-2 flex	Hex 2-2 flex refined
🟦 Bond lengths	0.950	1.769	0.544		0.947	1.806	0.543
🟥 Bond angles	1.274	3.105	1.006		1.268	3.313	0.994
🟩 Omega angle restraints	0.730	3.616	1.497		0.691	3.382	1.470
🟪 Side chain planarity	0.301	0.026	1.170		0.340	0.026	1.261
🟦 Improper dihedral distribution	0.997	6.232	0.976		0.984	6.122	0.946

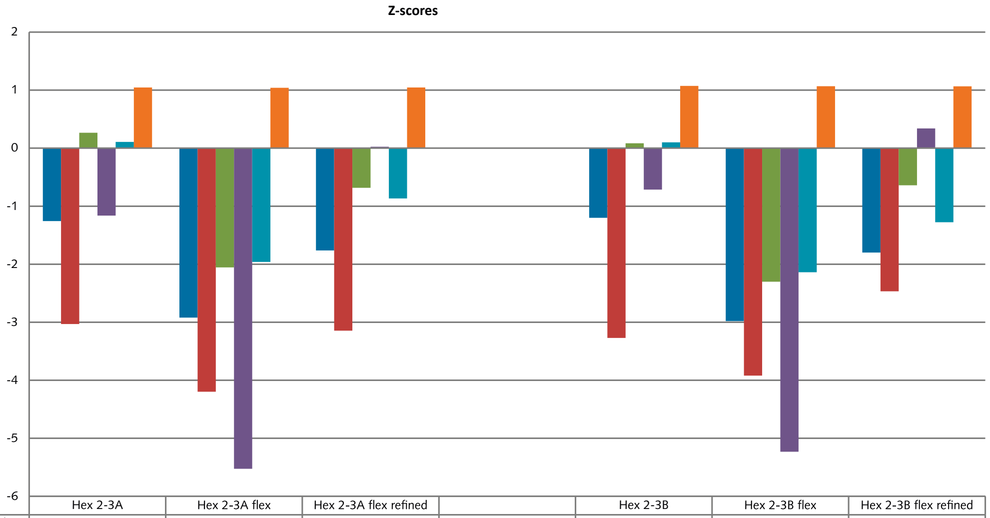


	Hex 2-1	Hex 2-1 flex	Hex 2-1 flex refined		Hex 2-2	Hex 2-2 flex	Hex 2-2 flex refined
🟦 1st generation packing quality	-1.129	-2.693	-1.468		-1.068	-2.647	-1.541
🟥 2nd generation packing quality	-3.171	-4.315	-3.015		-2.832	-4.042	-2.675
🟩 Ramachandran plot appearance	0.397	-2.290	-0.635		0.534-	2.210-	0.480
🟪 chi-1/chi-2 rotamer normality	-0.606	-5.136	0.247		0.599-	4.6170	.773
🟦 Backbone conformation	0.124	-2.258	-1.062		0.387-	1.861-	0.766
🟧 Inside/Outside distribution	1.046	1.033	1.030		1.0511	.053	1.051

Fig. 60: What-if results for subunits type I and II, segments from hexamer 2. For RMS Z-scores, closer to 1.0 is better, for structure Z-scores, positive is better than average.

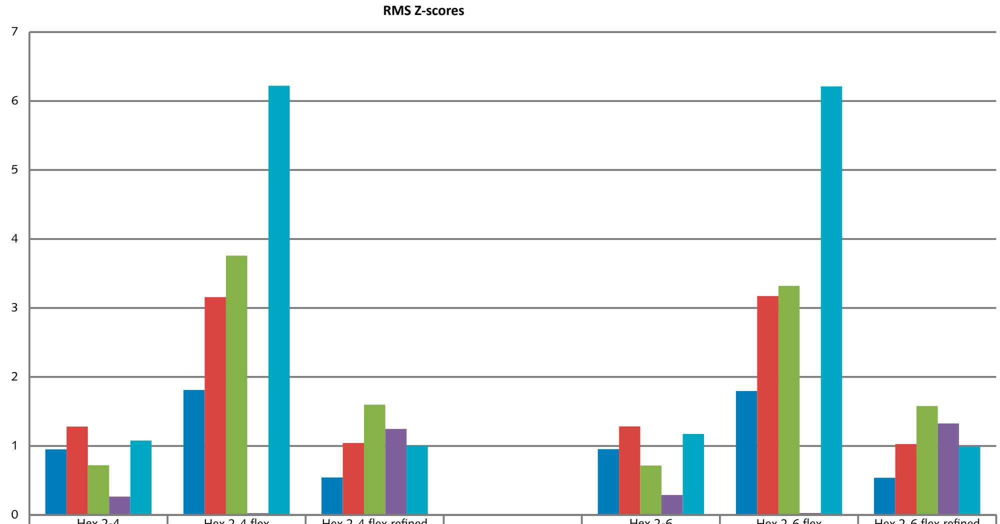


	Hex 2-3A	Hex 2-3A flex	Hex 2-3A flex refined		Hex 2-3B	Hex 2-3B flex	Hex 2-3B flex refined
🟦 Bond lengths	0.946	1.806	0.5410		.725	1.667	0.557
🟥 Bond angles	1.299	3.127	1.0291		.086	3.072	1.036
🟩 Omega angle restraints	0.706	3.190	1.5450		.914	3.415	1.547
🟪 Side chain planarity	0.394	0.028	1.2891		.722	0.026	1.361
🟦 Improper dihedral distribution	1.127	6.047	0.9981		.554	5.998	0.995

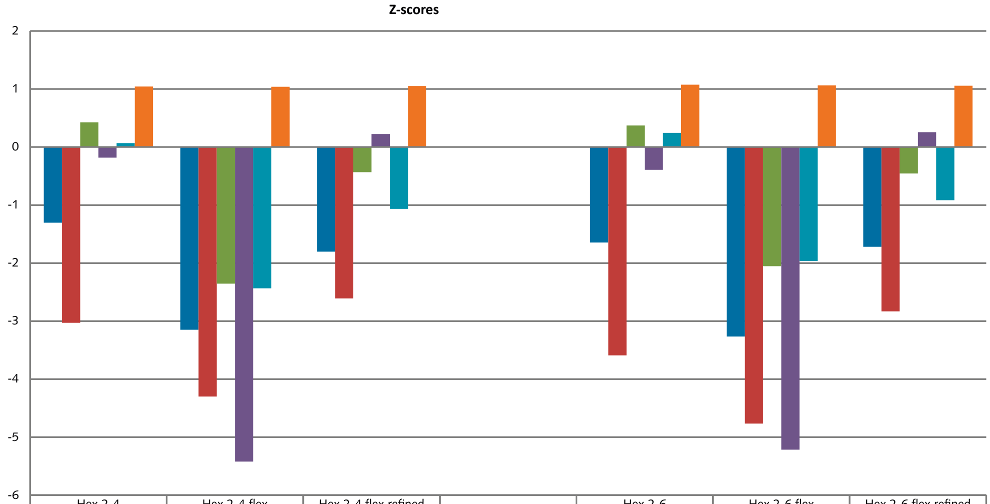


	Hex 2-3A	Hex 2-3A flex	Hex 2-3A flex refined		Hex 2-3B	Hex 2-3B flex	Hex 2-3B flex refined
🟦 1st generation packing quality	-1.257	-2.920	-1.762		-1.200	-2.980-	1.799
🟥 2nd generation packing quality	-3.031	-4.197	-3.146		-3.270	-3.920	-2.467
🟩 Ramachandran plot appearance	0.265	-2.055	-0.6840		.084-	2.302-	0.639
🟪 chi-1/chi-2 rotamer normality	-1.162	-5.526	0.025		-0.715	-5.231	0.339
🟦 Backbone conformation	0.108	-1.960	-0.866		0.101	-2.138	-1.277
🟧 Inside/Outside distribution	1.046	1.041	1.045		1.073	1.0671	.066

Fig. 61: What-if results for subunits type IIIA and IIIB, segments from hexamer 2. For RMS Z-scores, closer to 1.0 is better, for structure Z-scores, positive is better than average.



	Hex 2-4	Hex 2-4 flex	Hex 2-4 flex refined		Hex 2-6	Hex 2-6 flex	Hex 2-6 flex refined
🟦 Bond lengths	0.951	1.811	0.543		0.954	1.796	0.539
🟥 Bond angles	1.281	3.158	1.042		1.283	3.172	1.027
🟩 Omega angle restraints	0.720	3.759	1.597		0.715	3.321	1.578
🟪 Side chain planarity	0.266	0.026	1.246		0.288	0.028	1.326
🟦 Improper dihedral distribution	1.078	6.221	1.001		1.174	6.213	0.989



	Hex 2-4	Hex 2-4 flex	Hex 2-4 flex refined		Hex 2-6	Hex 2-6 flex	Hex 2-6 flex refined
🟦 1st generation packing quality	-1.303	-3.149	-1.803		-1.645	-3.265	-1.720
🟥 2nd generation packing quality	-3.029	-4.300	-2.608		-3.590	-4.766	-2.832
🟩 Ramachandran plot appearance	0.427	-2.355	-0.433		0.373	-2.054	-0.455
🟪 chi-1/chi-2 rotamer normality	-0.183	-5.420	0.224		-0.394	-5.215	0.256
🟦 Backbone conformation	0.068	-2.435	-1.067		0.243	-1.965	-0.915
🟧 Inside/Outside distribution	1.043	1.039	1.051		1.075	1.064	1.057

Fig. 62: What-if results for subunits type IV and VI, segments from hexamer 2. For RMS Z-scores, closer to 1.0 is better, for structure Z-scores, positive is better than average.

4.3.2 YASARA energy calculations and Z-scores

In addition to the What-if evaluations the YASARA software calculates energy levels and its own global molecule Z-score (Table 4). While absolute energy levels, whether positive or negative are not important, a successful energy minimization step is often characterized by a negative value (Krieger et al., 2009). The Z-scores, while improved after the processing, only border on an acceptable value around -2,0 or better (Fig. 63).

	Start	End
Hexamer 1 – type I		
Energy (jK/mol)	374758,5	-321044,1
Score	-7,73	-1,73
Hexamer 1 – type II		
Energy (jK/mol)	6172910716	-321889,3
Score	-7,6	-1,78
Hexamer 1 – type IIIA		
Energy (jK/mol)	310356,3	-328544
Score	-7,99	-2,09
Hexamer 1 – type IIIB		
Energy (jK/mol)	12765266,9	-314420,1
Score	-8,04	-2,26
Hexamer 1 – type IV		
Energy (jK/mol)	917220961,3	-320110,1
Score	-8,69	-2,13
Hexamer 1 – type V		
Energy (jK/mol)	40662330,8	-330600,9
Score	-8,32	-2,08
Hexamer 2 – type I		
Energy (jK/mol)	5309386,7	-316730,4
Score	-8,13	-1,98
Hexamer 2 – type II		
Energy (jK/mol)	434795659,5	-322599
Score	-7,66	-1,56
Hexamer 2 – type IIIA		
Energy (jK/mol)	130837400,4	-329983,7
Score	-7,9	-2,17

	Start	End
Hexamer 2 – type IIIB		
Energy (jK/mol)	39774518,6	-320033,8
Score	-8,13	-2,06
Hexamer 2 – type IV		
Energy (jK/mol)	11578198,8	-330588,3
Score	-8,07	-2,07
Hexamer 2 – type VI		
Energy (jK/mol)	621542,1	-330852,1
Score	-7,8	-1,82

Table 4: YASARA Energy calculations and Z-scores.

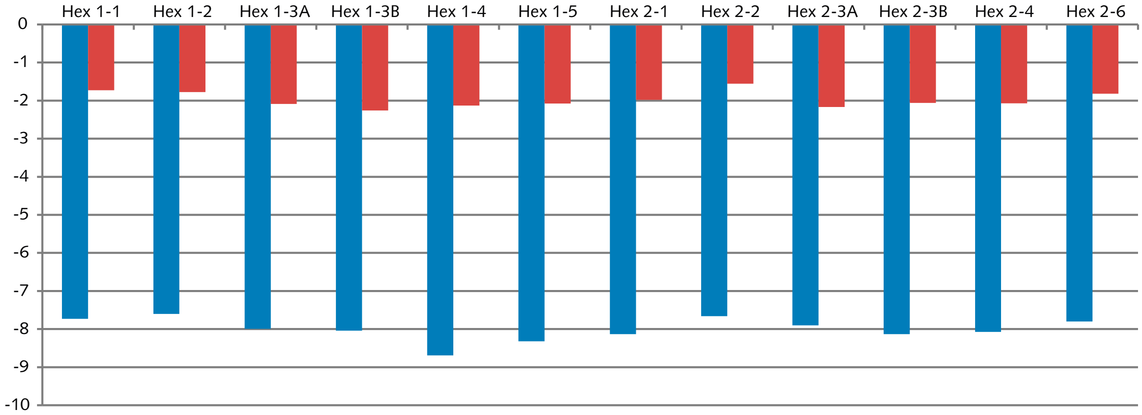


Fig. 63: YASARA Z-score results. Z-score values before energy minimization (unmodified flexible fitting result) are depicted blue, values after optimization red.

4.3.3 C-alpha atom shift after flexible fitting

Fig. 64 through Fig. 69 show the shift in Å for each C-alpha backbone atom (top) and corresponding histogram plots of shift frequency (bottom). C-alpha shift graph remarks: Bars on the top and bottom indicate secondary structure elements (helix: black; sheet: gray; loop: white space). Histogram plots of C-alpha backbone shift frequency. The plots are binned to 0.1 Å steps. In addition a normal curve as calculated from the histogram data is projected on top of the bins (blue). A marker on the x-axis indicates the maximum of the normal curve.

In general, the secondary structure protection provided by the DireX flexible fitting algorithm performs reasonably well, at least for longer elements. However very short helical elements are prone to being shifted. The long α1.1→α1.3 loop in domain #1 exhibits larger scale changes in some cases (Fig. 64B; Fig. 66 A), while in other monomers shifts are fairly even throughout the molecule. The C-terminus is also prone to be shifted by the fitting algorithm.

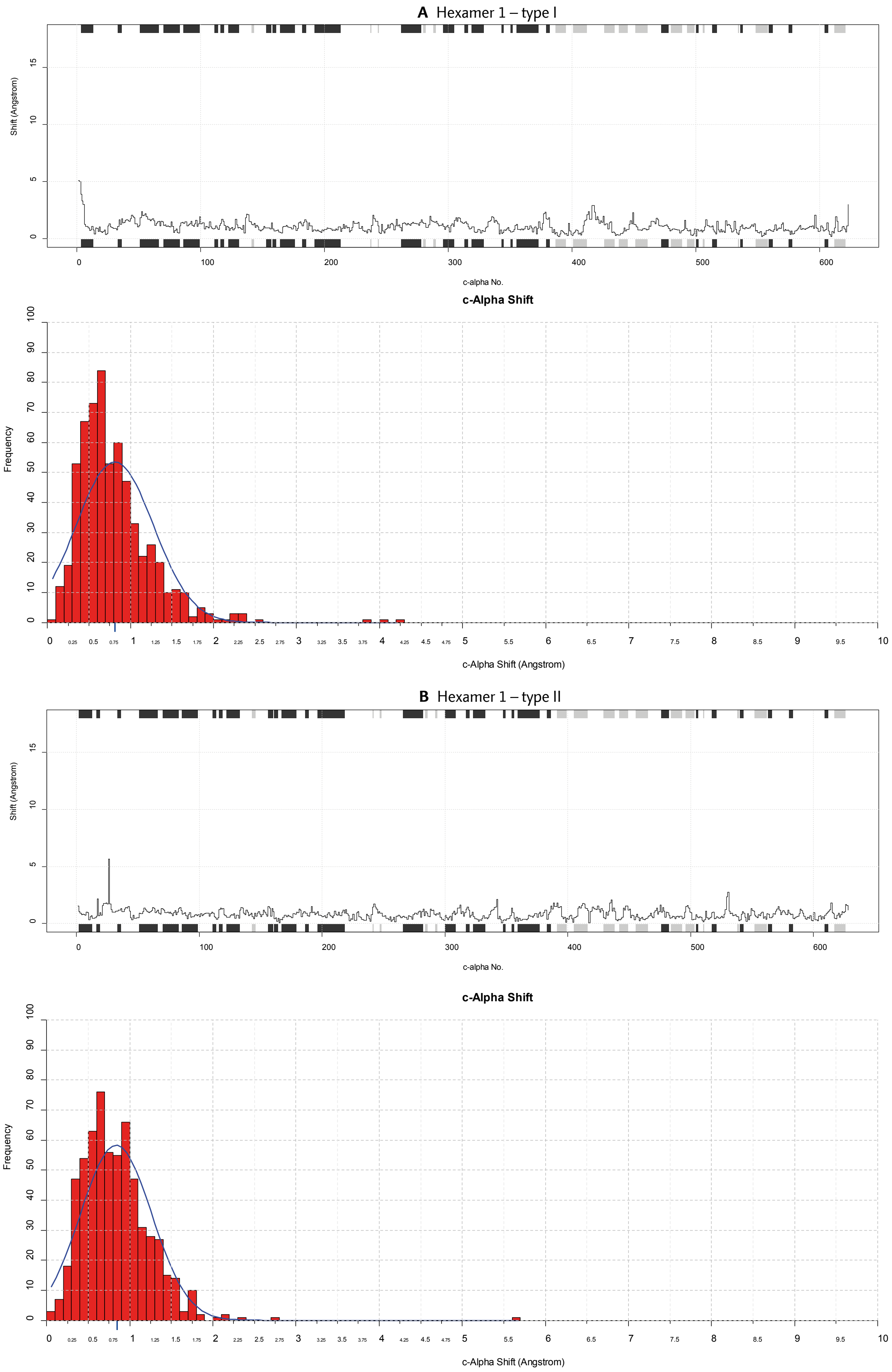


Fig. 64: Flexible Fitting c-alpha shift and histogram distribution. Panel A: Hexamer 1 – type I. Panel B: Hexamer 1 type II.

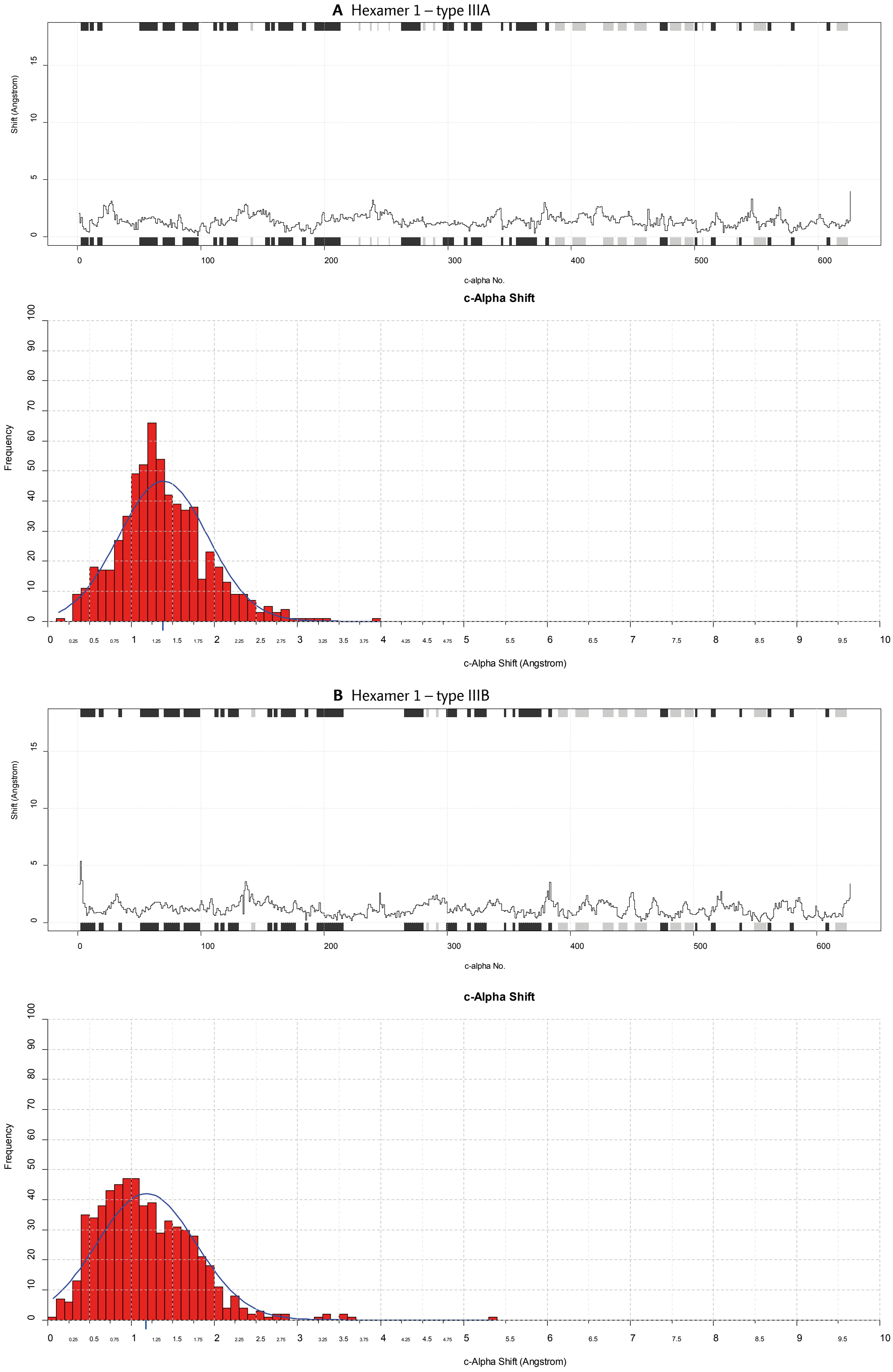
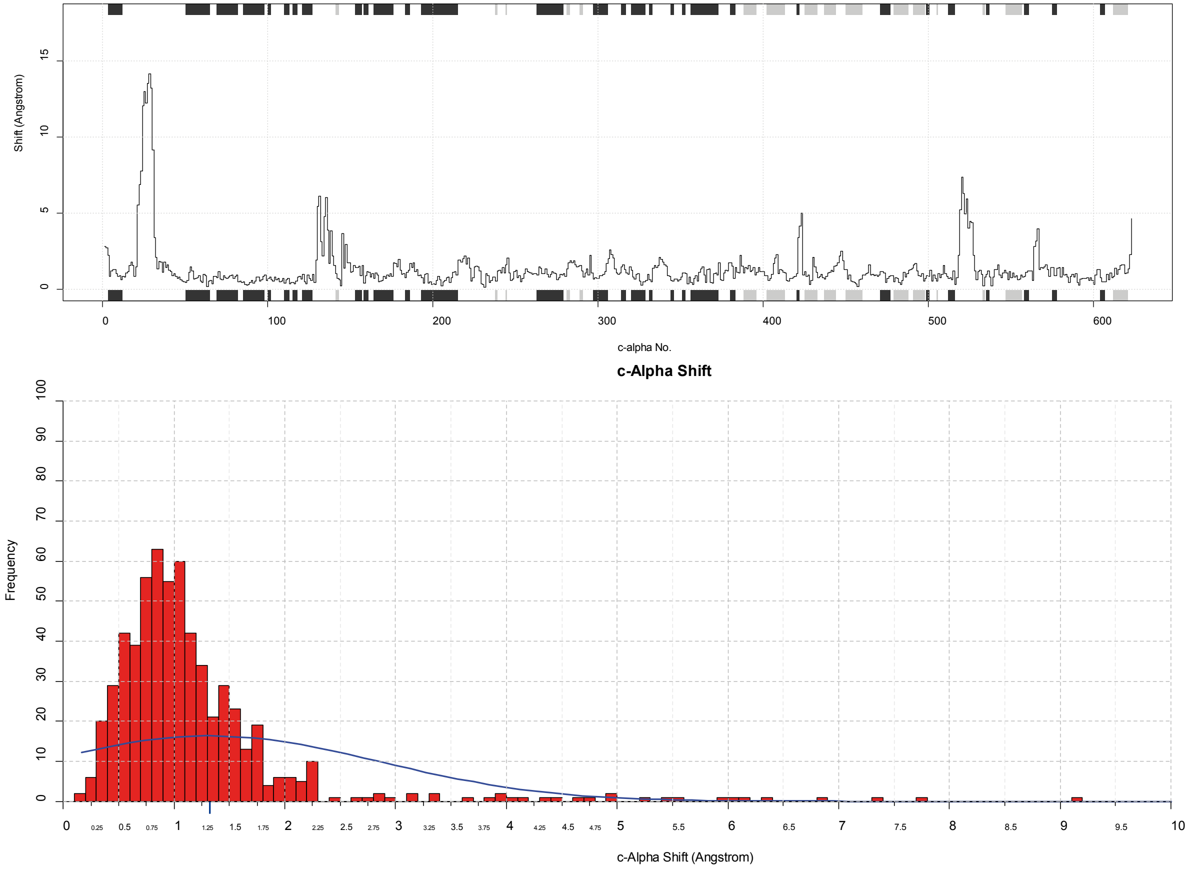


Fig. 65: Flexible Fitting c-alpha shift and histogram distribution. Panel A: Hexamer 1 – type IIIA. Panel B: Hexamer 1 type IIIB.

A Hexamer 1 – type IV



B Hexamer 1 – type V

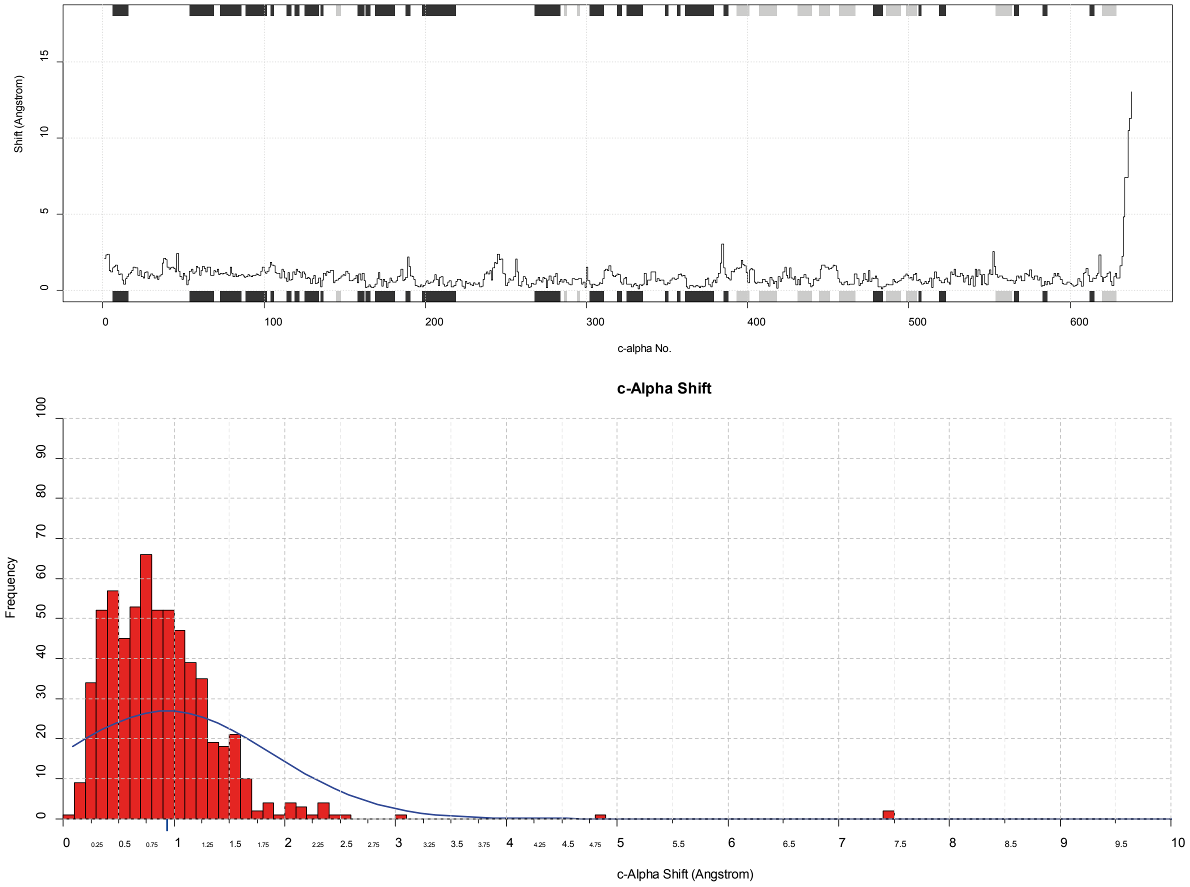
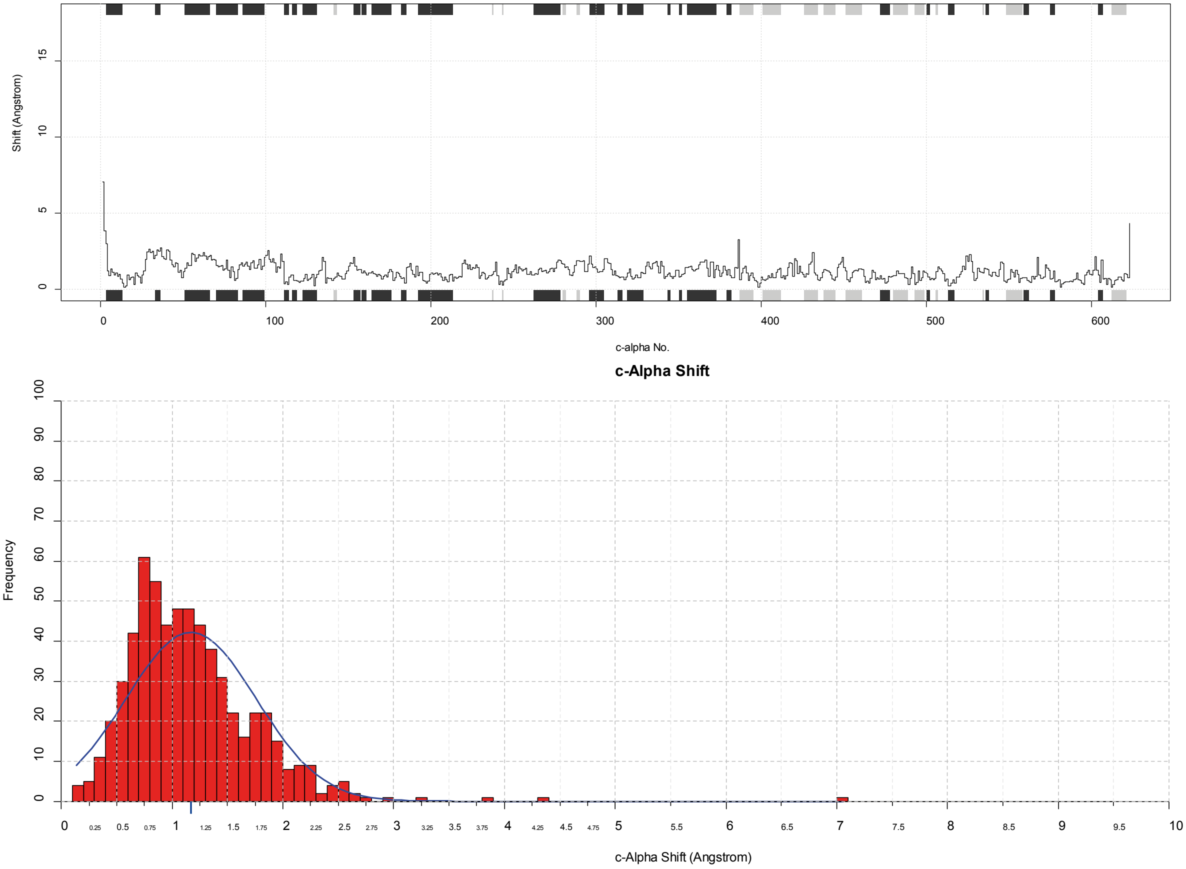


Fig. 66: Flexible Fitting c-alpha shift and histogram distribution. Panel A: Hexamer 1 – type IV. Panel B: Hexamer 1 type V.

A Hexamer 2 – type I



B Hexamer 2 – type II

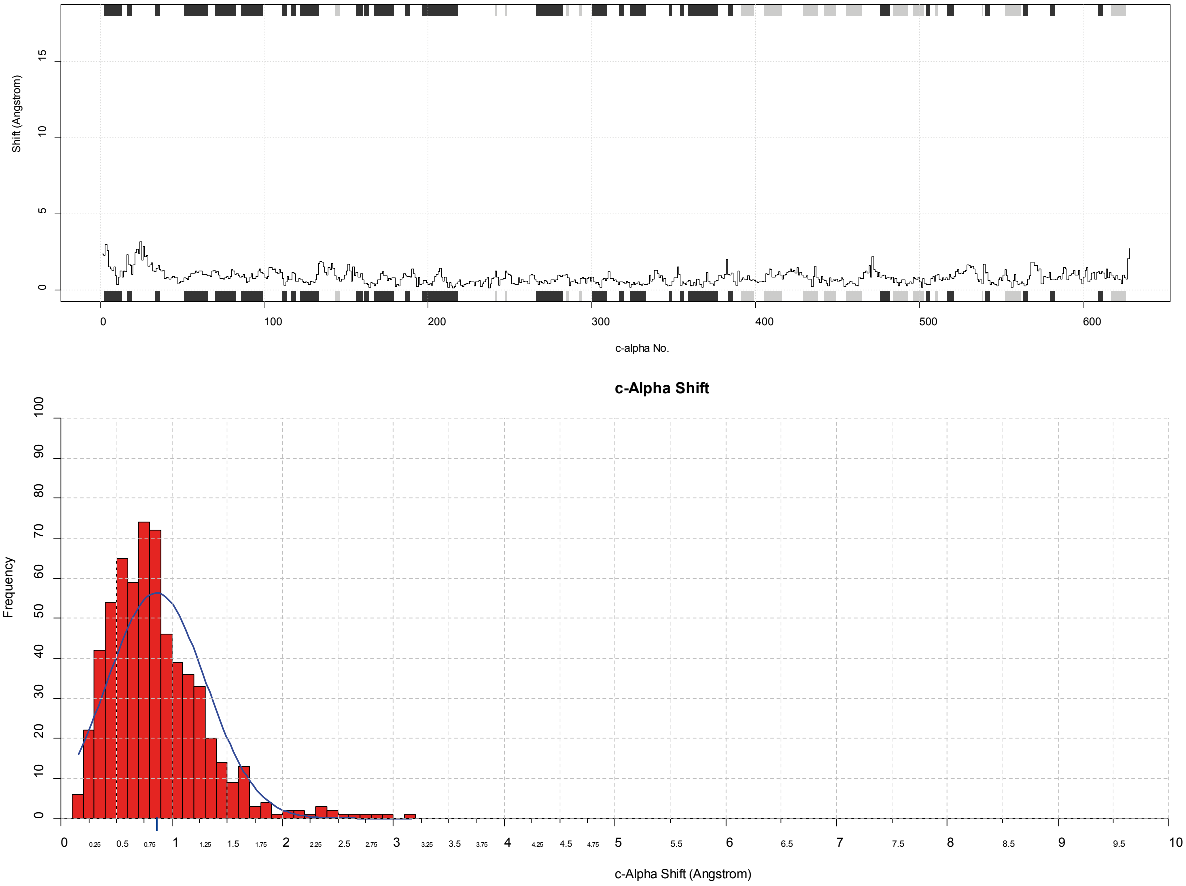
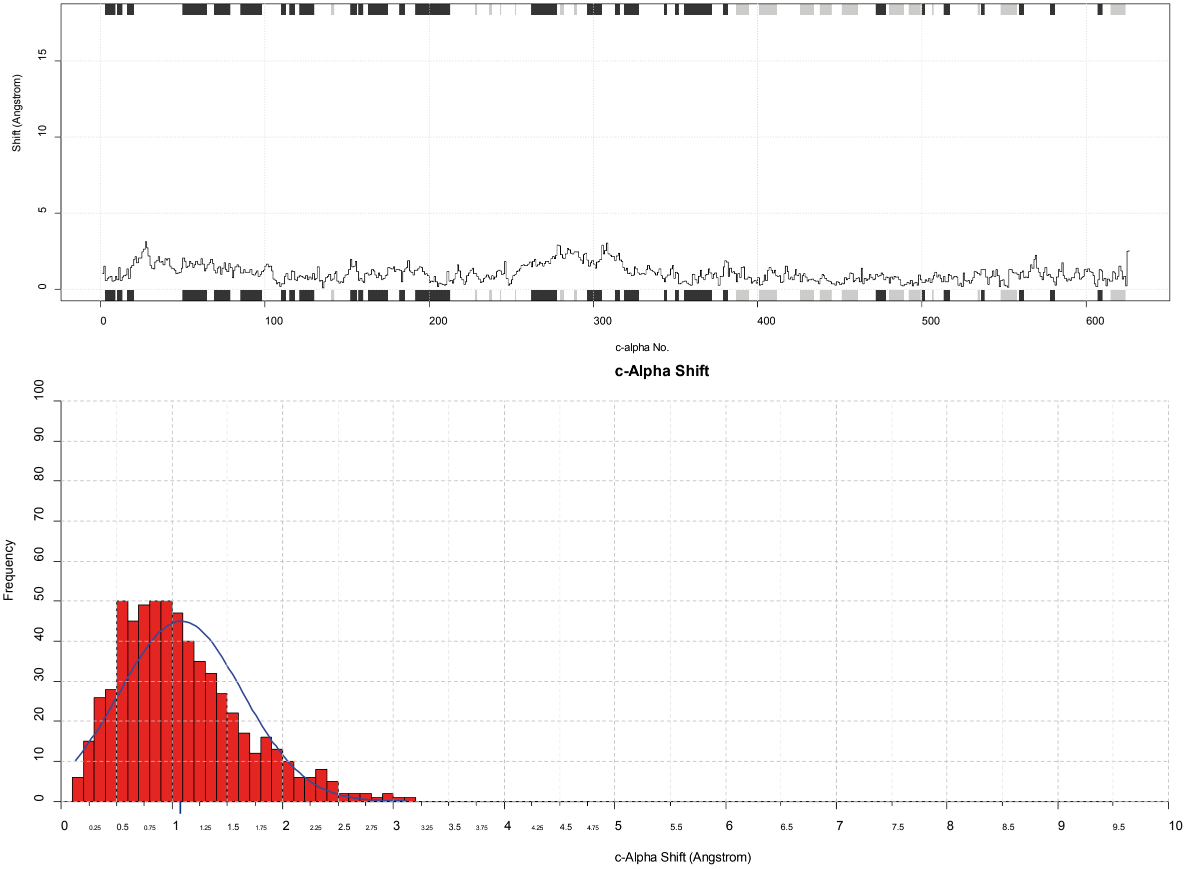


Fig. 67: Flexible Fitting c-alpha shift and histogram distribution. Panel A: Hexamer 2 – type I. Panel B: Hexamer 2 type II.

A Hexamer 2 – type IIIA



B Hexamer 2 – type IIIB

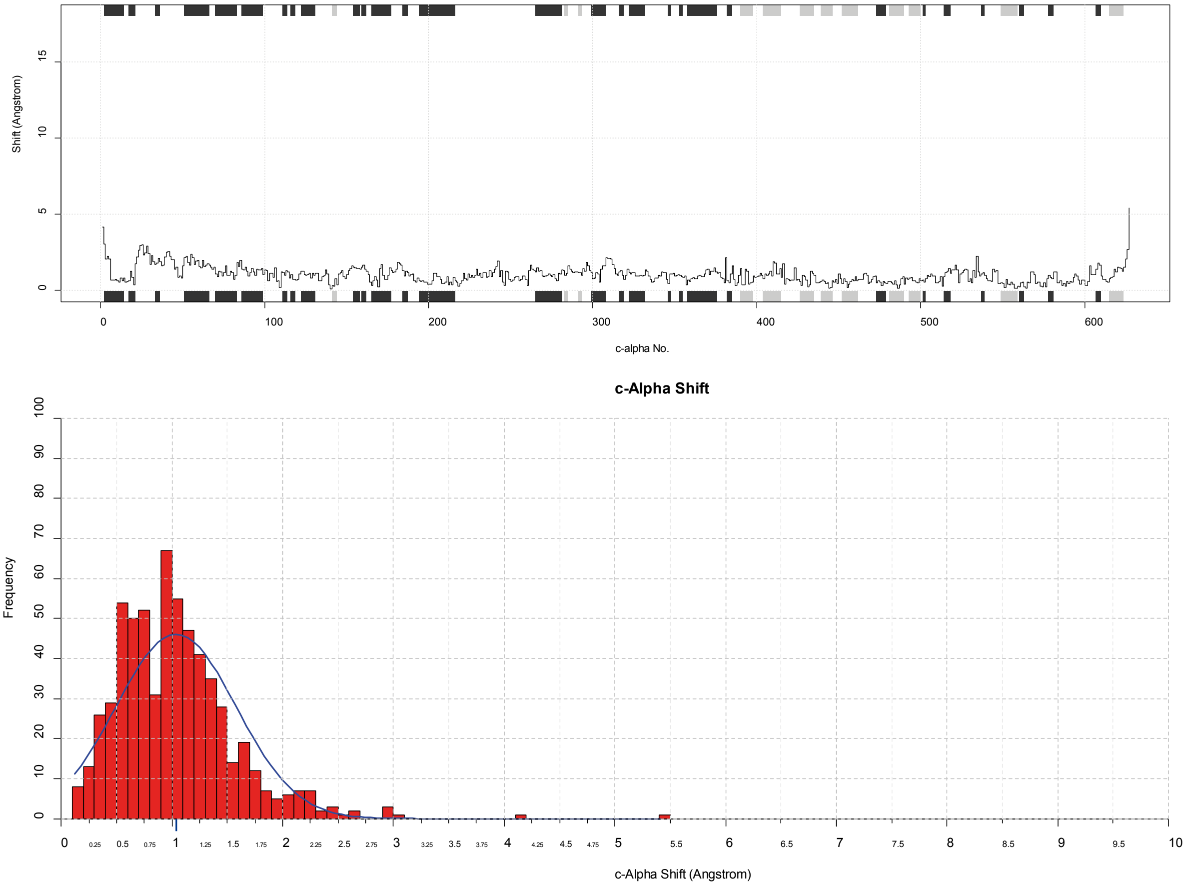
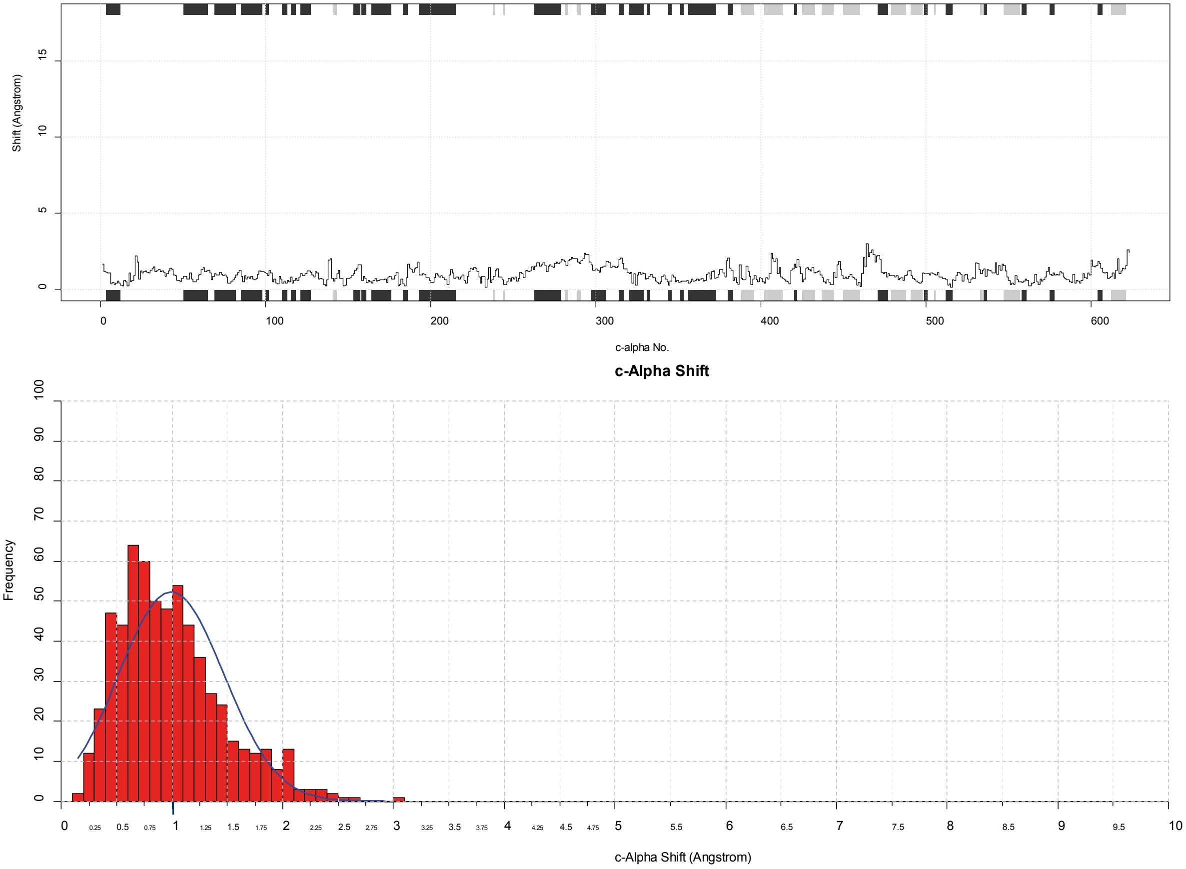


Fig. 68: Flexible Fitting c-alpha shift and histogram distribution. Panel A: Hexamer 2 – type IIIA. Panel B: Hexamer 2 type IIIB.

K Hexamer 2 – type IV



L Hexamer 2 – type VI

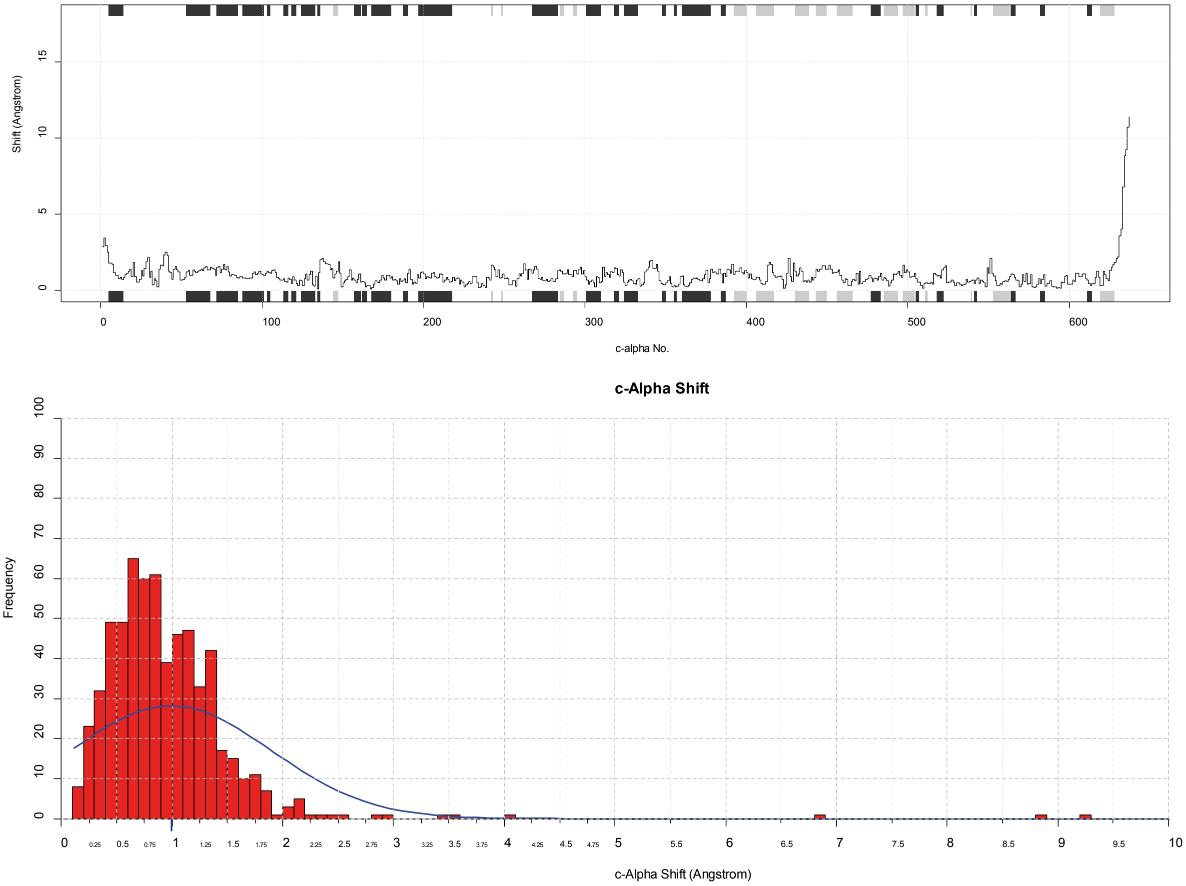


Fig. 69: Flexible Fitting c-alpha shift and histogram distribution. Panel A: Hexamer 2 – type IV. Panel B: Hexamer 2 type VI.

4.4 Molecular Model of *Lpo*Hc

After quality assessment the flexibly fitted molecular model of *Lpo*Hc was completed by using the symmetry functions of the Chimera software, thus reproducing the exact locations of each subunit throughout the entire model. Fig. 70 and Fig. 71 provide an overview and detailed renderings of the completed model.

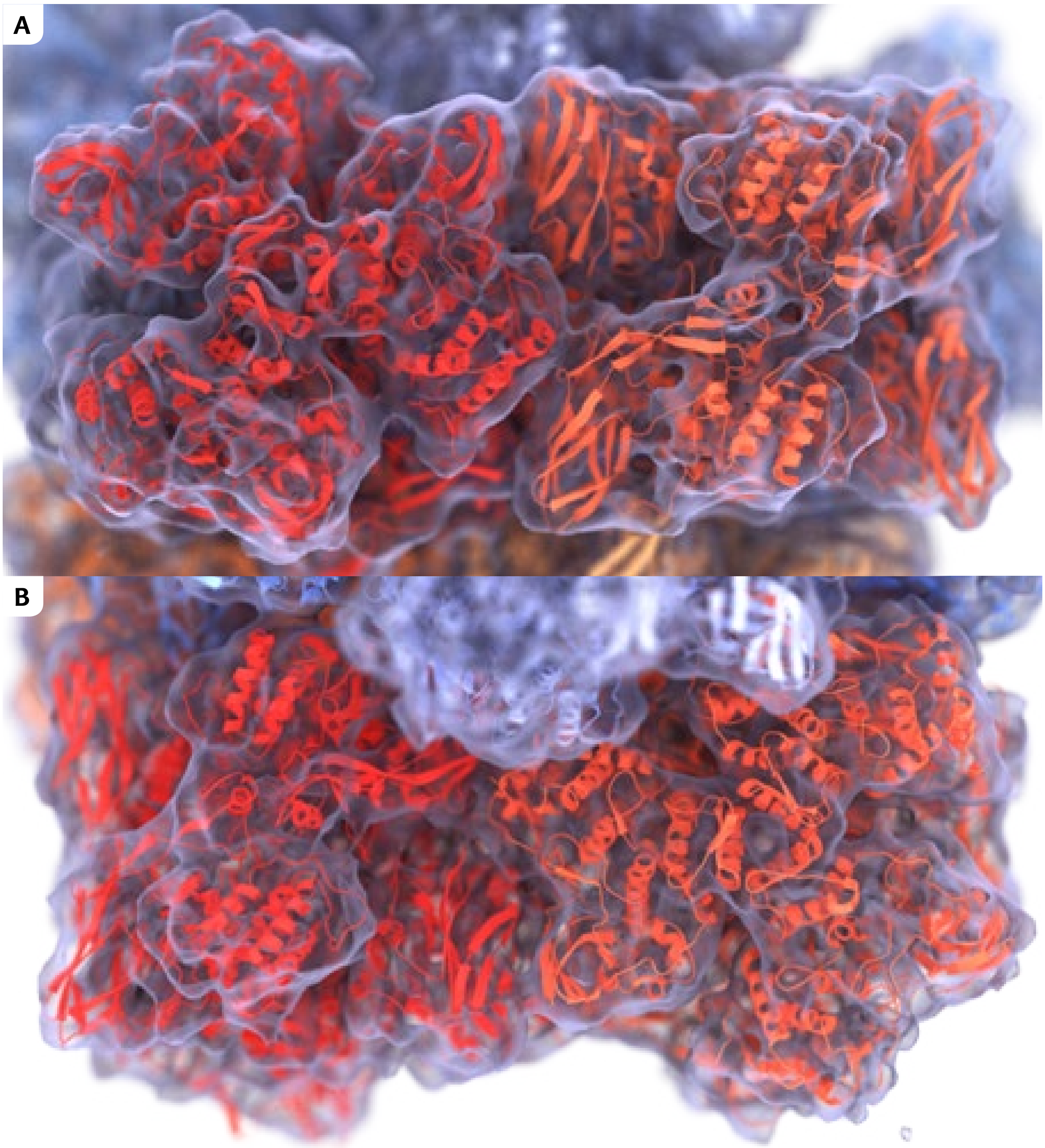


Fig. 70: Volumetric surface rendering of the 3D-reconstruction of Lpo*Hc with molecular model.* Panel A: On axis view of hexamer 1 (red) and side view of hexamer 2 (orange). Panel B: side view of hexamer 1 (red) and on axis view of hexamer 2 (orange).

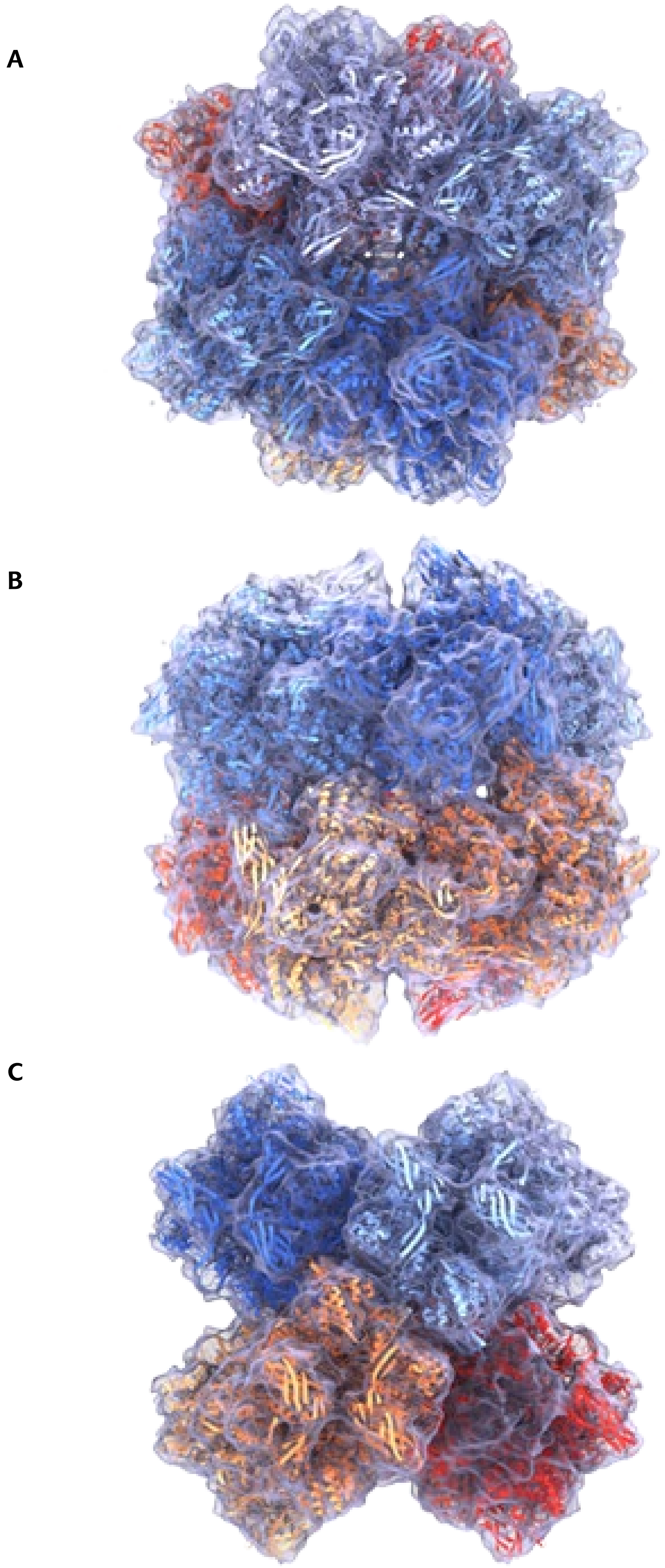


*Fig. 71: Volumetric surface rendering of the 3D-reconstruction of Lpo*Hc *with molecular model.* Hexamer 1 through 4 are colored dark blue to light blue. Hexamer 5 though 8 are colored red to yellow. Panel A: view along the Z axis. Panel B: view along the X axis. Panel C: view along the Y axis.

4.5 Flexible Fitting – subunit analysis

4.5.1 Hexamer 1 – subunit type I

Fig. 72: Flexible Fitting result for subunit type I of hexamer 1. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 72 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,937 (rigid body) vs 0,942 (flexible). To calculate the correlation coefficient, a simulated map at 8 Å resolution was used.

The largest changes can be observed in domain 1. Helix α1.1 is shifted outside the map-contour when viewed at high threshold levels (Fig. 73A). At mass-correlated threshold, an unfilled mass is visible towards which α1.1 was pushed by the algorithm by a maximum of 5 Å at the N-terminus. Moving along the backbone the shift decreases.

The next notable shift occurs for α1.4. Observable at high threshold levels, the fitting of the helix is improved by a constant shift of 1,6 Å. Even so this movement is not adequate to fully push the helix into the reconstructed mass (Fig. 73B).In case of α1.5 the helix is moved by 1,5 Å and pushed slightly outside of the map contour (Fig 49C).

The reconstruction exhibits an empty mass between α1.1 and α1.5 (Fig. 73C, gray circle, map shown at lower threshold level to show empty mass) which might present a pull for the fitting algorithm that results in the shifting of α1.5. Helix α1.6 is shifted along the helical axis by approx. 1,3 Å without changing fitting quality. The helices in the α1.7→α1.8 loop are shifted by 0,8 Å improving fit especially for the second loop (Fig. 73D). The secondary structure is slightly distorted in this case: DSSP does not recognize helical structures in this region (Wouter et al., 2015, Kabsch & Sander 1983).

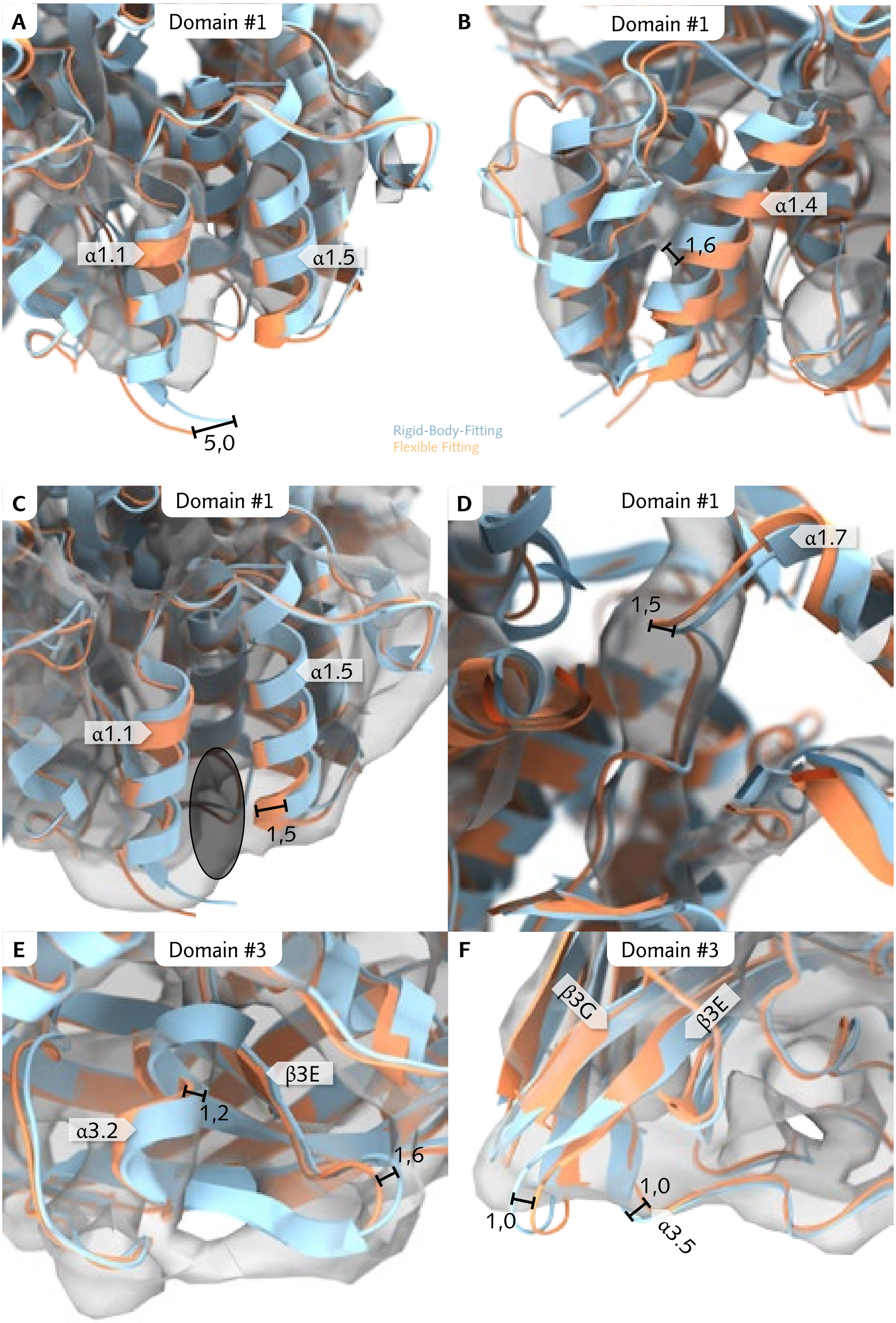


Fig. 73: Flexible Fitting result for subunit type I of hexamer 1 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

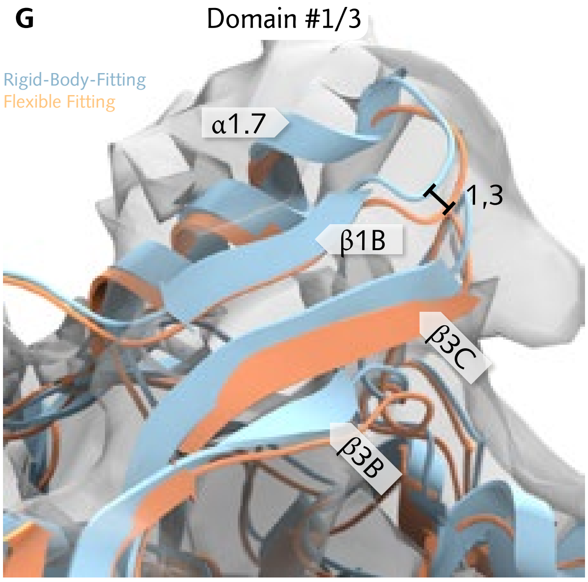


Fig. 74: Flexible Fitting result for subunit type I of hexamer 1 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

α1.7 exhibits a shift by 0,8 Å along the helical axis with no immediate impact on fitting quality. The β3E→α3.2 loop undergoes a 180° turn reversing direction (Fig. 73E). In this region a shift by 1,3 Å matches the reconstructed volume and reduces the number of atoms outside of the map-contour. The α1.7→β1B loop undergoes a 1,3 Å shift, again to pull the loop further inside the reconstructed volume (Fig. 74).

The anti-parallel β-barrel of domain #3 is pulled into the reconstruction, at the outer edges by as much as 1,0 Å (Fig. 73F). A closer look at the active site (Fig. 75) reveals the largest shift to occur for H324 with a mean and maximum shift of 0,5 Å and 1,2 Å respectively (reference for this measurement is the plane of the imidazole functional group). Other components exhibit minor shifting and/or tilting.

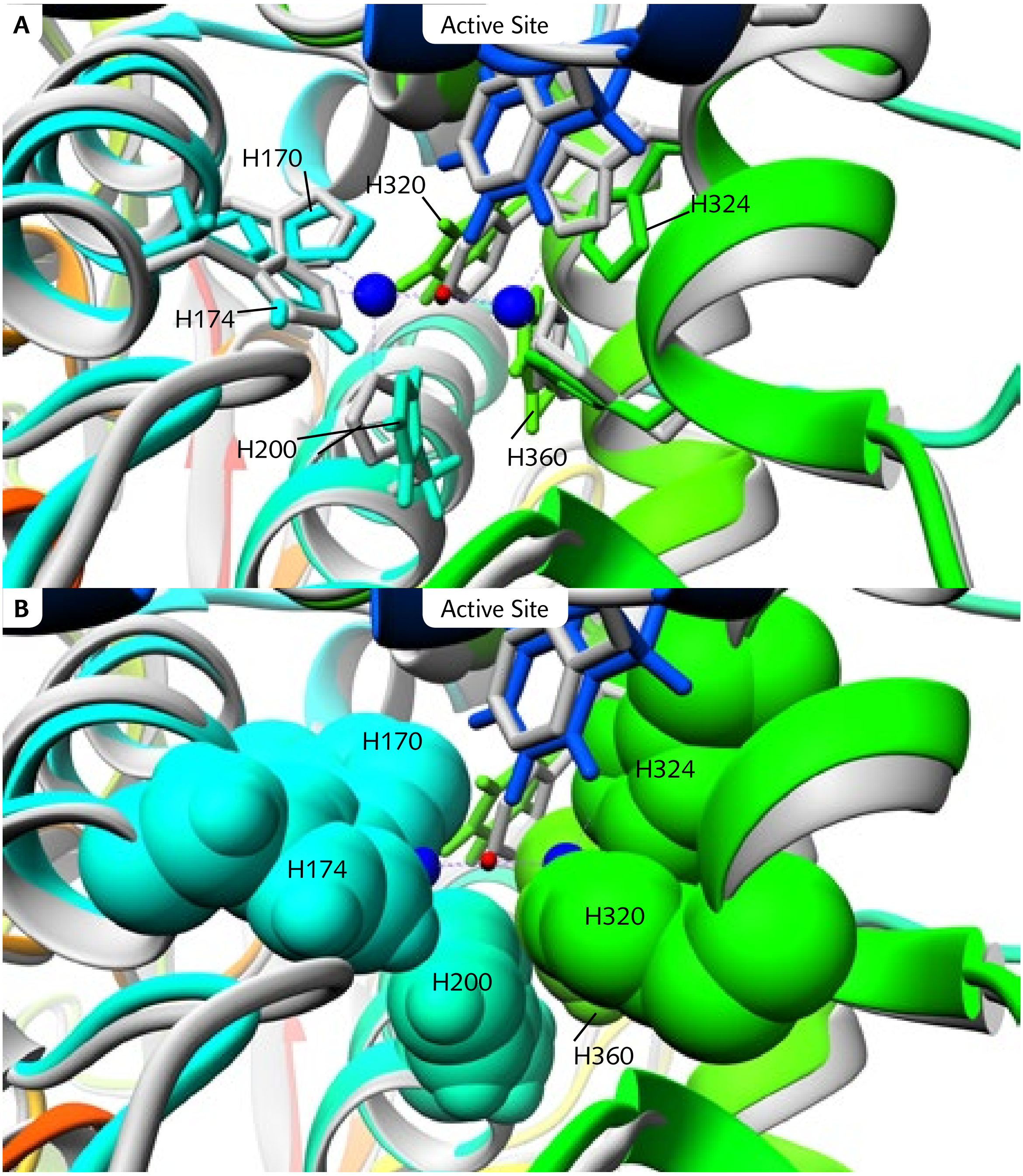


Fig. 75: Active Site of hexamer 1 subunit type I. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O2 (red), flexible fitting shown with rainbow color code from N to C terminus. Panel B: Flexibly fitted histidines are depicted with their atoms VDW radius.

4.5.2 Hexamer 1 – subunit type II

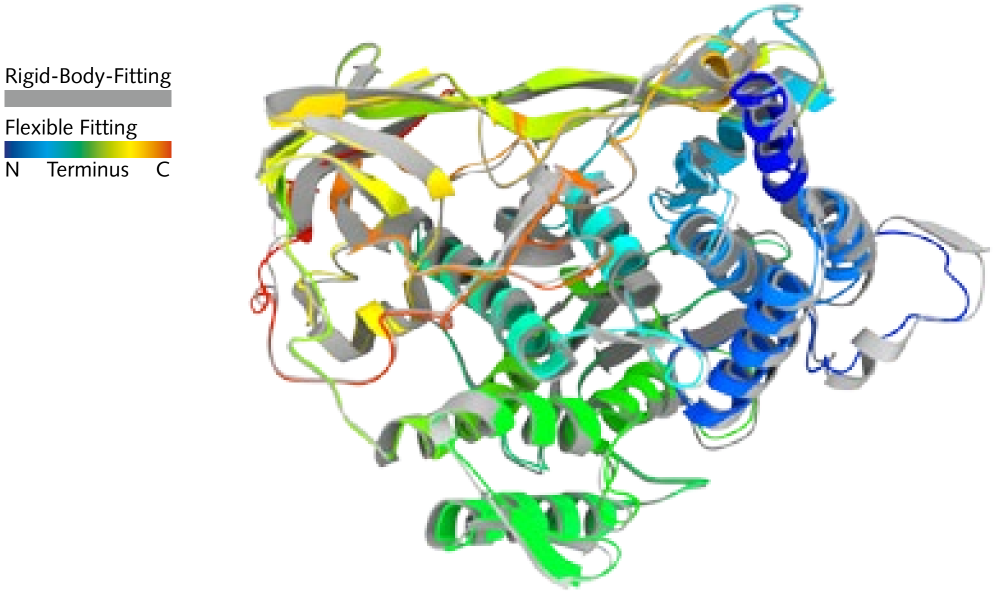


Fig. 76: Flexible Fitting result for subunit type II of hexamer 1. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 76 provides an overall comparison between the original structure and the flexible fitting result. To calculate the correlation of the molecular models of subunit type II of hexamer 1, simulated maps at 8 Å resolution were used. The resulting correlation coefficient is 0,9397 (rigid body) vs 0,9405 (flexible).

α1.1, α1.4 and α1.5 are shifted by 1,0 Å towards the center of the reconstruction, slightly worsening the fit in some areas (Fig. 77A). The α1.1→α1.3 strand is pushed towards the empty mass present in the reconstruction (Fig. 77A). Helix α1.3 is also pulled into the reconstruction by 1,0 Å at the expense of added distortion (Fig. 77B).

Domain #2 remains without notable changes (Fig. 77C). The anti-parallel β-barrel in domain #3 is moved by 1,4 – 1,5 Å, improving overall fit in the region (Fig. 77D).

Fig. 78 provides a detailed look at the active site. 1NOL is used as reference. The largest shift occurs for H177 with a mean and maximum shift of 0,7Å and 1,5 Å respectively (reference for this measurement is the plane of the imidazole functional group). H173 also exhibits a notable mean and maximum shift of 0,5Å and 1,1 Å respectively. Other components exhibit minor shifting and/or tilts. In the flexible fitting configuration both copper atoms can be dislocated by approx. 1,5 Å.

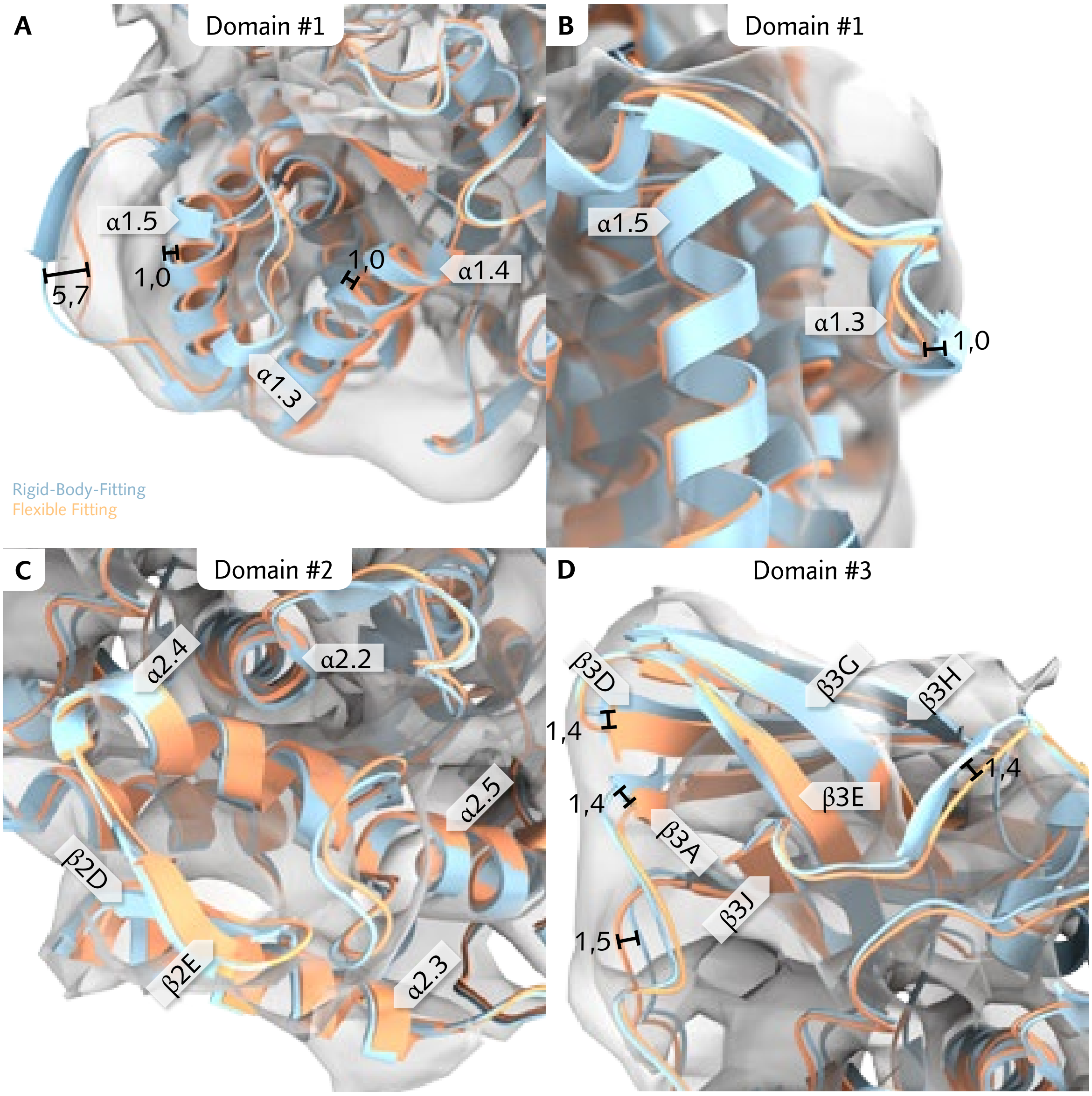


Fig. 77: Flexible Fitting result for subunit type II of hexamer 1 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for details. All distances measured in Å.

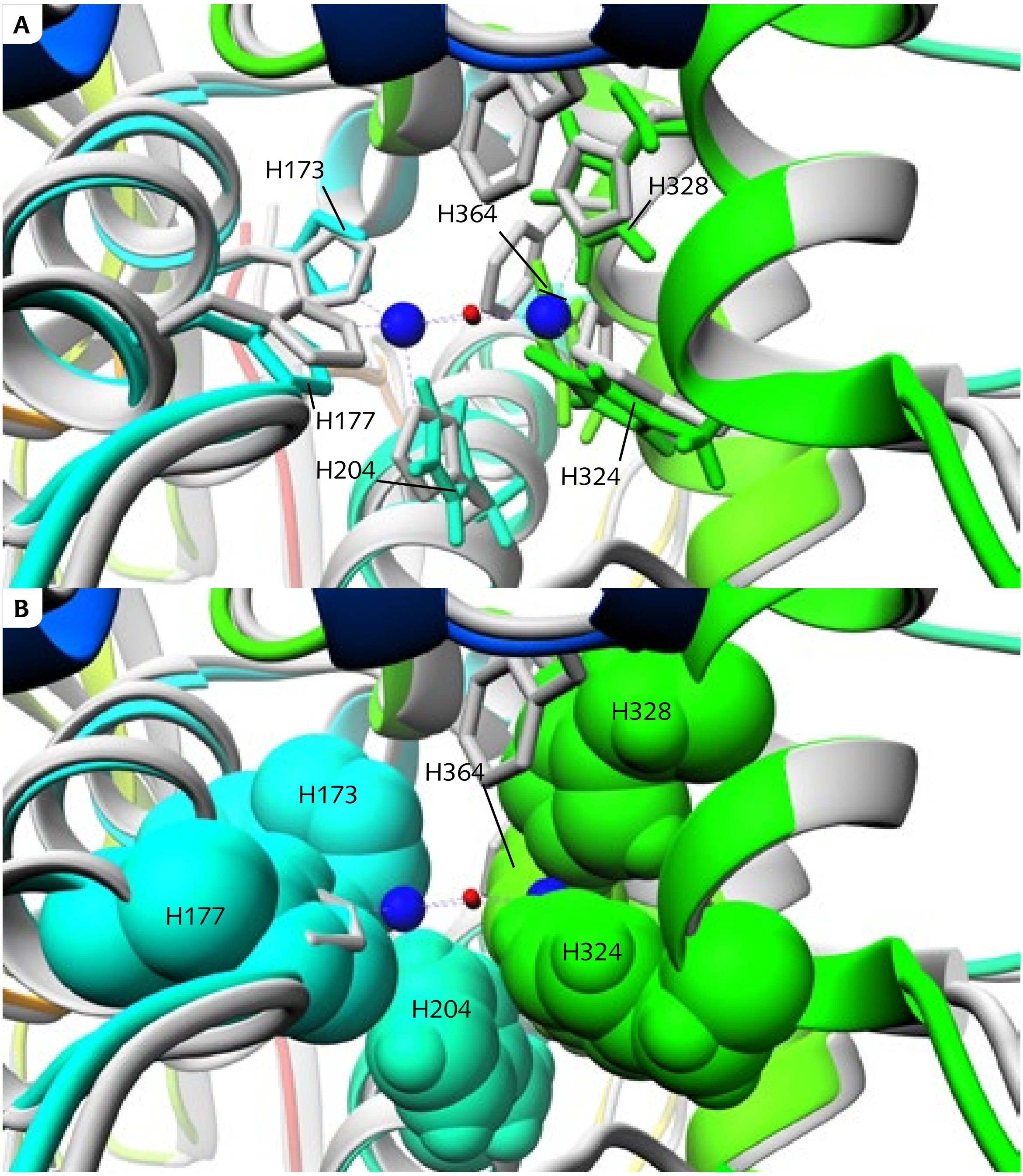


Fig. 78: Active Site of hexamer 1 subunit type II. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O2 (red), flexible fitting shown with rainbow color code from N to C terminus. Panel B: Flexibly fitted histidines are depicted with their atoms VDW radius.

4.5.3 Hexamer 1 – subunit type IIIA

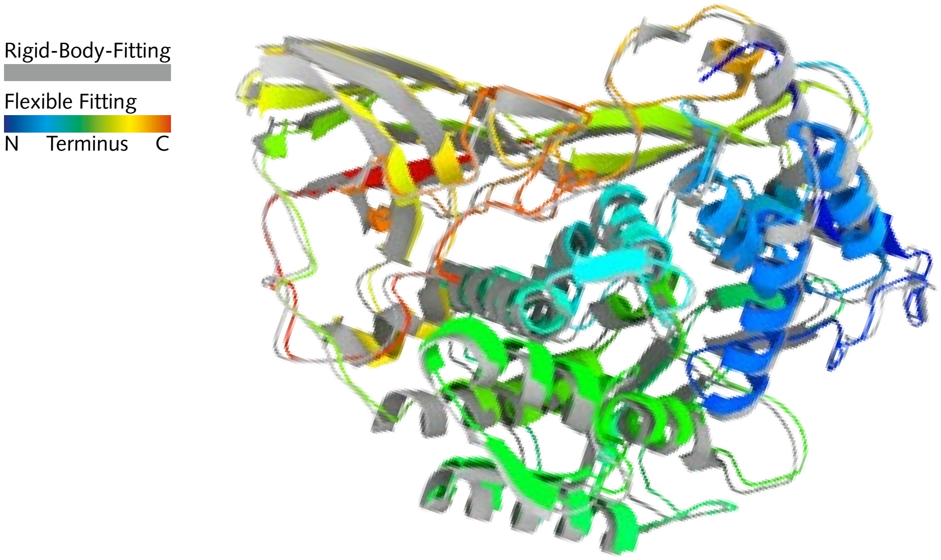


Fig. 79: Flexible Fitting result for subunit type IIIA of hexamer 1. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 79 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9335 (rigid body) vs 0,9409 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

Compared to other fitting results in this study, larger shifts throughout the subunit can be observed (see section 4.3.3). Starting with domain #2 (Fig. 80A) various helices are shifted with values ranging from 1,5 Å to 2,3 Å. Helix α2.3 is turned into the reconstruction, moving the C-terminal end of the helix by 1,5 Å (Fig. 80B). The short anti-parallel β-sheets β2D→β2E are shifted accordingly. The β-hairpin loop β3B→β3C that reaches from domain #3 to domain #1 is shifted with values ranging from 1,8 to 2,2 Å (Fig. 80C). The β-barrel dominating domain #3 is largely left unchanged, with the exposed β3D→β3E hairpin being pulled further into the reconstruction (1,5 Å, Fig. 80D).

Within the active site (using 1NOL as reference), the largest shift occurs for H169, tilted by 43,1° with a mean and maximum shift of 1,7Å and 2.3 Å respectively (reference for this measurement is the plane of the imidazole functional group). H200 also exhibits a notable mean and maximum shift of 0,8Å and 2,0 Å respectively, with a tilt of 27,7°. Other components exhibit minor shifting and/or tilts. All rotational values are plane/plane angles of the imidazole functional group.

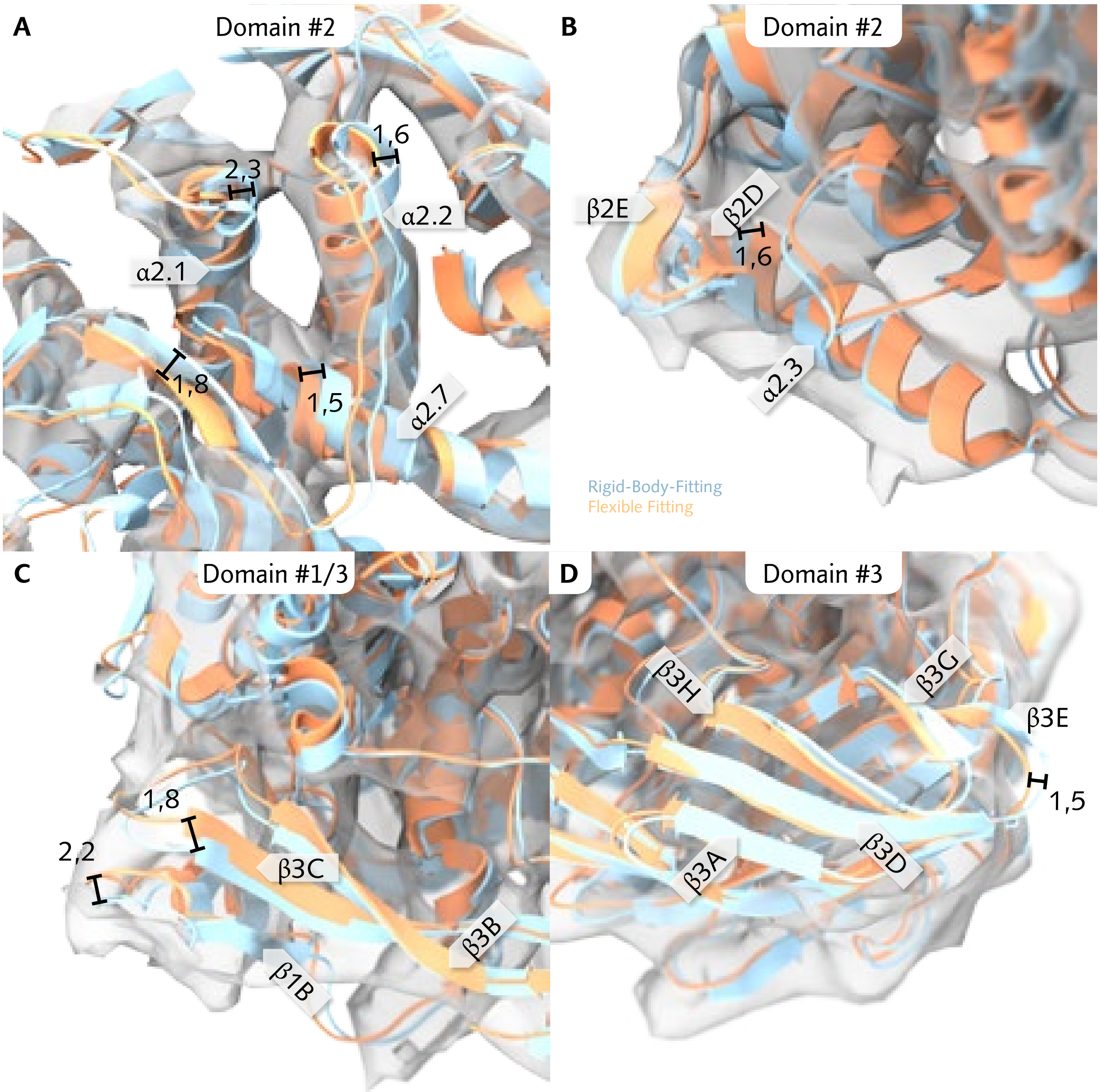


Fig. 80: Flexible Fitting result for subunit type IIIA of hexamer 1 subunit type – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

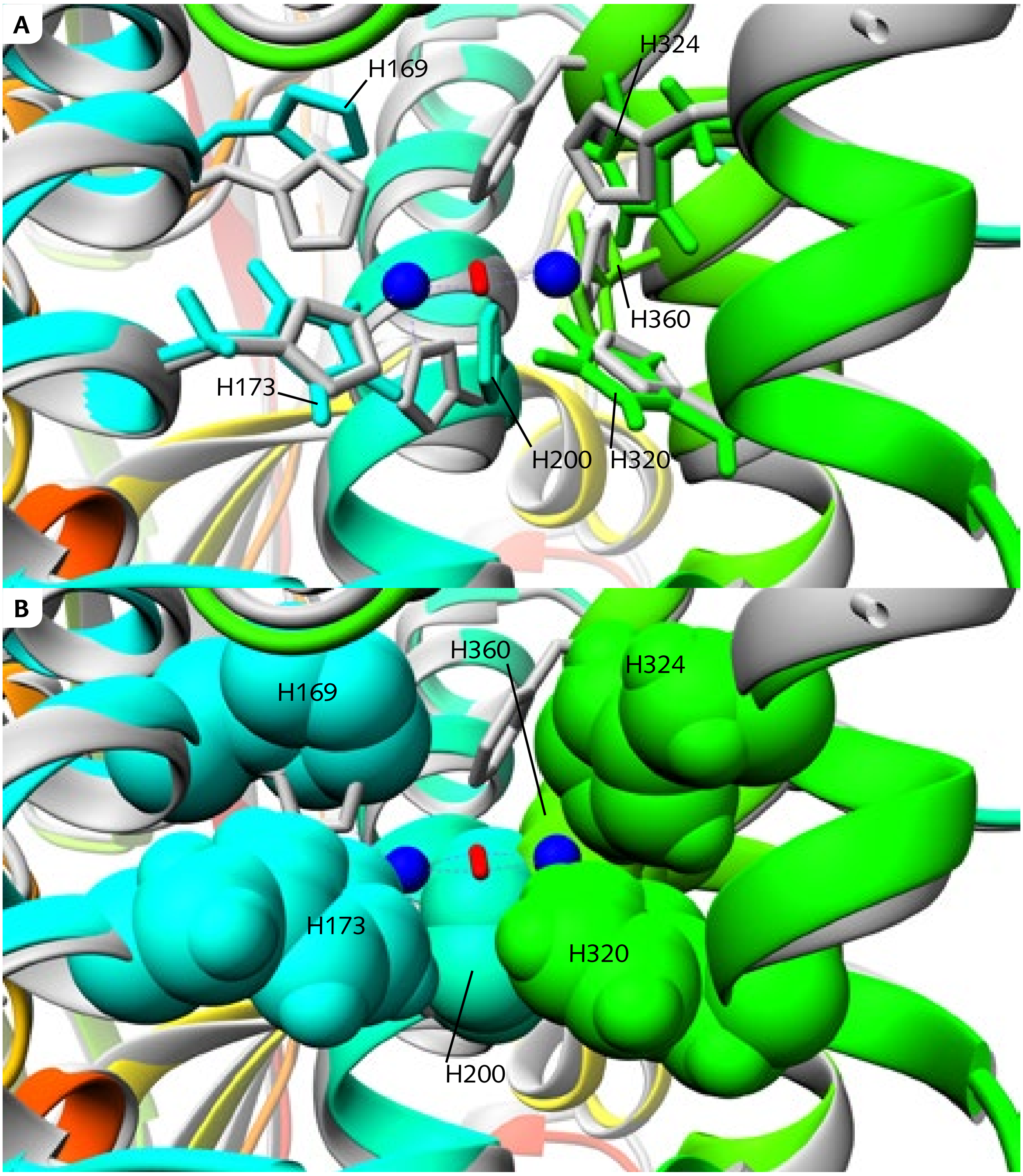


Fig. 81: Active Site of hexamer 1 subunit type IIIA. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus. Panel B: Flexible fitted histidines are depicted with their atoms VDW radius.

4.5.4 Hexamer 1 – subunit type IIIB

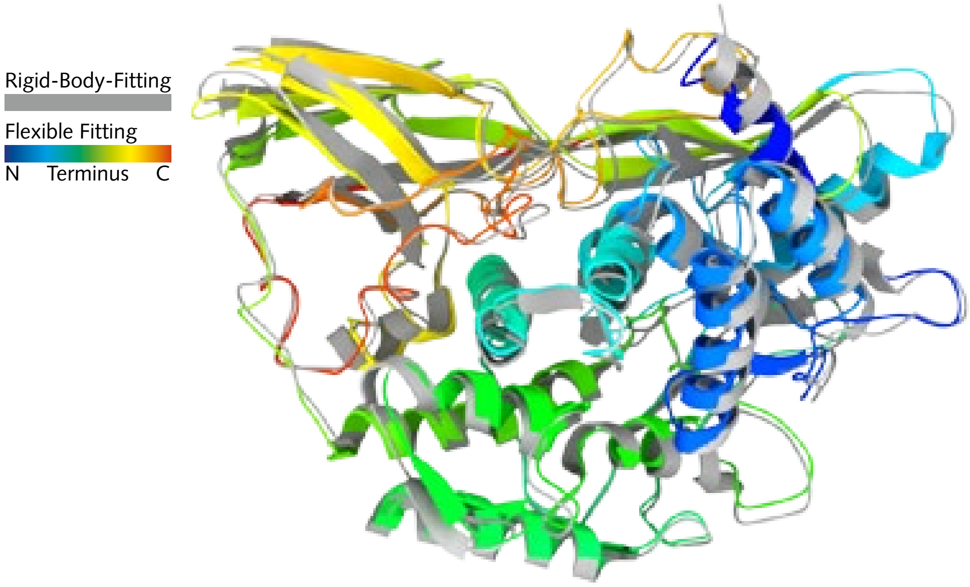


Fig. 82: Flexible Fitting result for subunit type IIIB of hexamer 1. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 82 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9391 (rigid body) vs 0,9488 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

Compared to other fitting results in this study, larger shifts throughout the subunit can be observed (see section 4.3.3). Starting with domain #1 (Fig. 83A and B), α1.1 is shifted by 1,4 Å, improving fit. The fit for α1.4 is worsened by a shift of 1,8 Å (Fig. 83B, alternate viewpoint on domain #1). Helix α1.5 is moved by 1,2 Å, improving the fitting result. In domain #2 (Fig. 83C) shifts between 0,9 Å (α2.7) and 1,6 to 2,0 Å (α2.3 and β2E respectively) can be observed. These changes improve overall the overall fit, especially in the region of β2E.

Components of domain #3 are shifted by as much as 1,9 Å (Fig. 83D) without discernible impact on fitting quality.

Within the active site (Fig. 84, using 1NOL as reference), H171 exhibits the largest shift by the flexible fitting algorithm: 0,7 Å mean and 2,0 Å maximum with a rotation of 43°. This is followed by H202 (0,5 Å mean and 1,3 Å max with 20° rotation) and H175 (0,3 Å mean and 0,7 Å max, 8,4° rotation). H320 shows 0,1 Å mean and 0,5 Å max (12,1° rotation) while H327 is shifted by 0,1 Å mean and 0,4 Å max, rotating 15,4°. The shift in position of H363 appears to be negligible. All rotational values are plane/plane angles of the imidazole functional group.

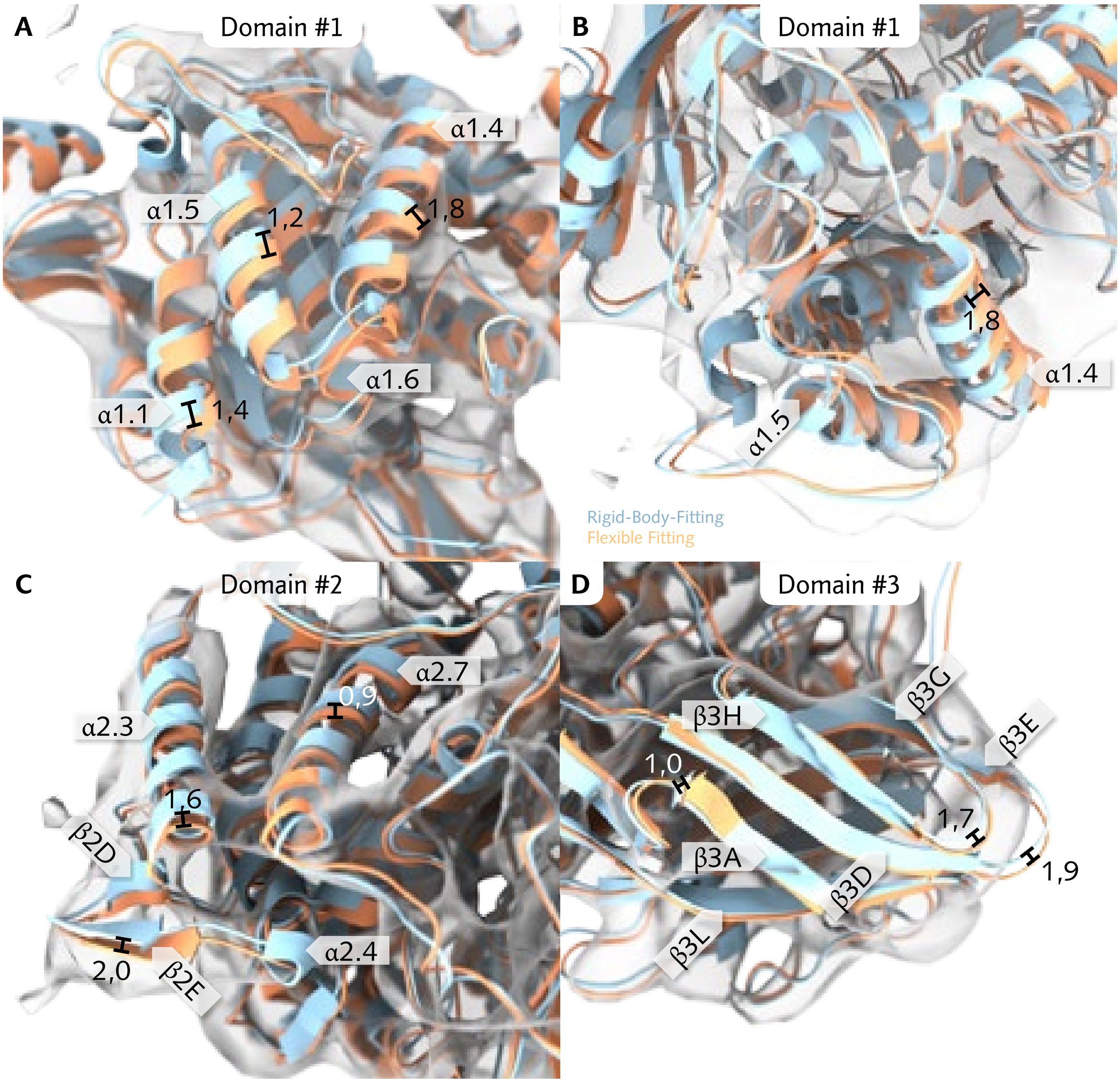


Fig. 83: Flexible Fitting result for subunit type IIIB of hexamer 1 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

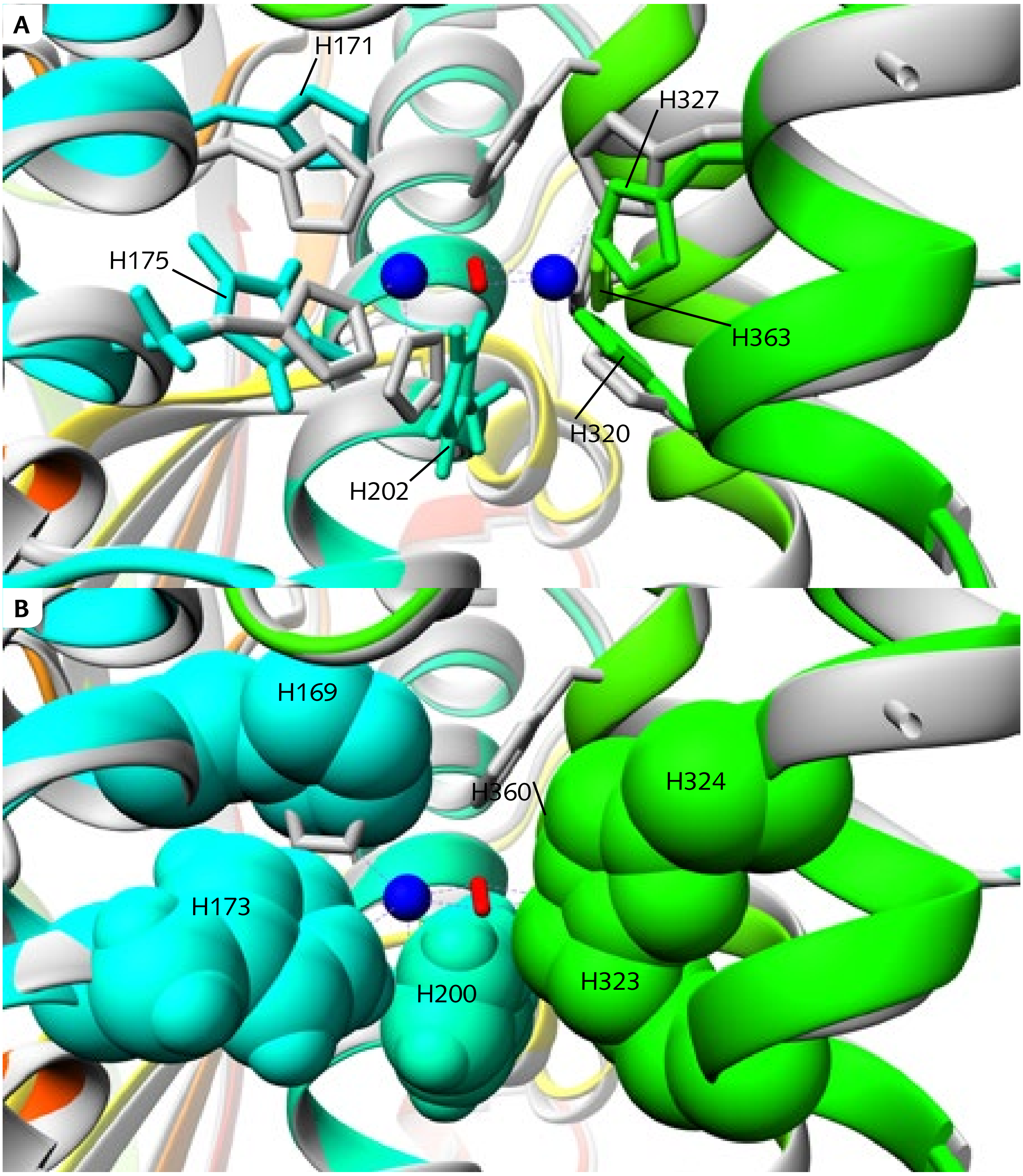


Fig. 84: Active Site of hexamer 1 subunit type IIIB. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.5 Hexamer 1 – subunit type IV

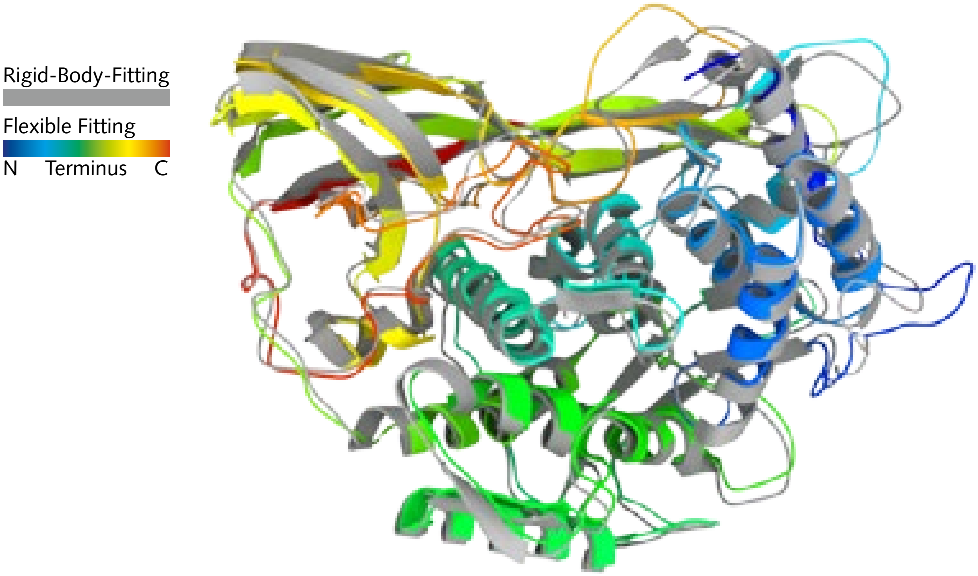


Fig. 85: Flexible Fitting result for subunit type IV of hexamer 1. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 85 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9417 (rigid body) vs 0,9489 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

A large shift occurs in domain #1. The α1.1→α1.3→α1.4 loop undergoes a large change in configuration (Fig. 86A). It should be noted that α1.3 is detected neither in the original nor the flexible fitting molecular model. Fig. 86B displays the reconstruction at a lower than standard threshold level to better show the mass the loop was fitted into. Following the backbone to the β1B→α1.7 loop, a shift of 4,8 Å at its maximum reshapes the loop to better fit the contour of the reconstruction (Fig. 86C).

The helices of domain #2 containing the active site are shifted by ~1,0 Å (Fig. 86D). In this region the β2D→β2E region is shifted by 1,9 Å, improving fit. Domain #3 sees its β-barell components shifted with values ranging from 1,3 to 1,7 Å (Fig. 86E). A notable fitting improvement is made by pulling the β3D→β3E loop into the reconstructed volume by 2,5 Å.

The active site shows structural damage from the flexible fitting process (Fig. 87) A large shift of 1,0 Å mean and 3,1 Å max has occurred to H361, with a rotation of 74,6°. In this configuration H361 cannot take part in forming the binding site. Less notable shifts can be observed for H170 (0,5 Å mean, 1,7 Å max, 35° rotation) and H325 (0,6 Å mean, 1,6 Å max, 20° rotation). H201 and H174 exhibit minor shifting, 0,1 Å mean, 0,2 Å max, 5,7° rotation and 0,3 Å mean, 1,0 Å max, 19,3° rotation respectively. All rotational values are plane/plane angles of the imidazole functional group.

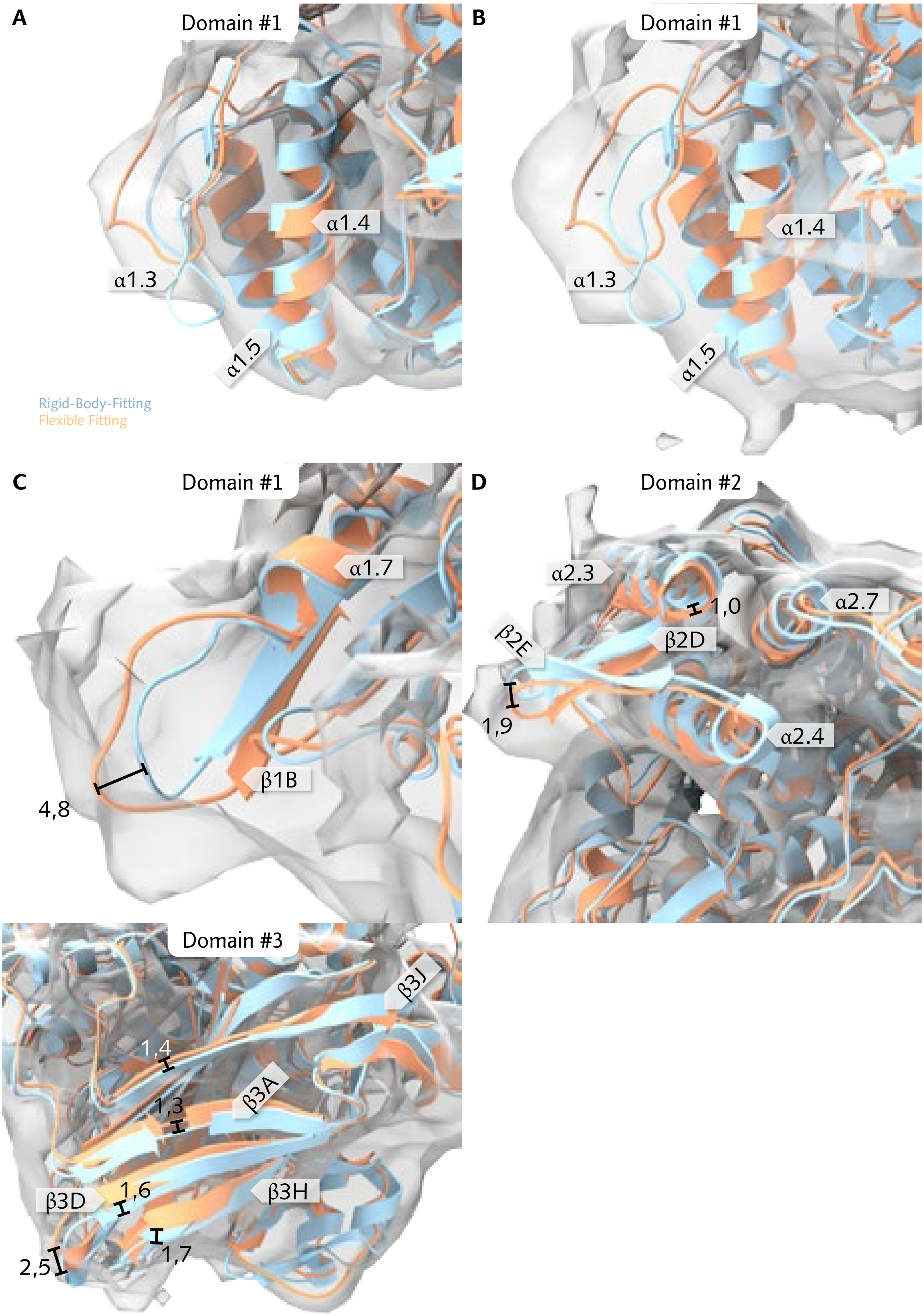


Fig. 86: Flexible Fitting result for subunit type IV of hexamer 1 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

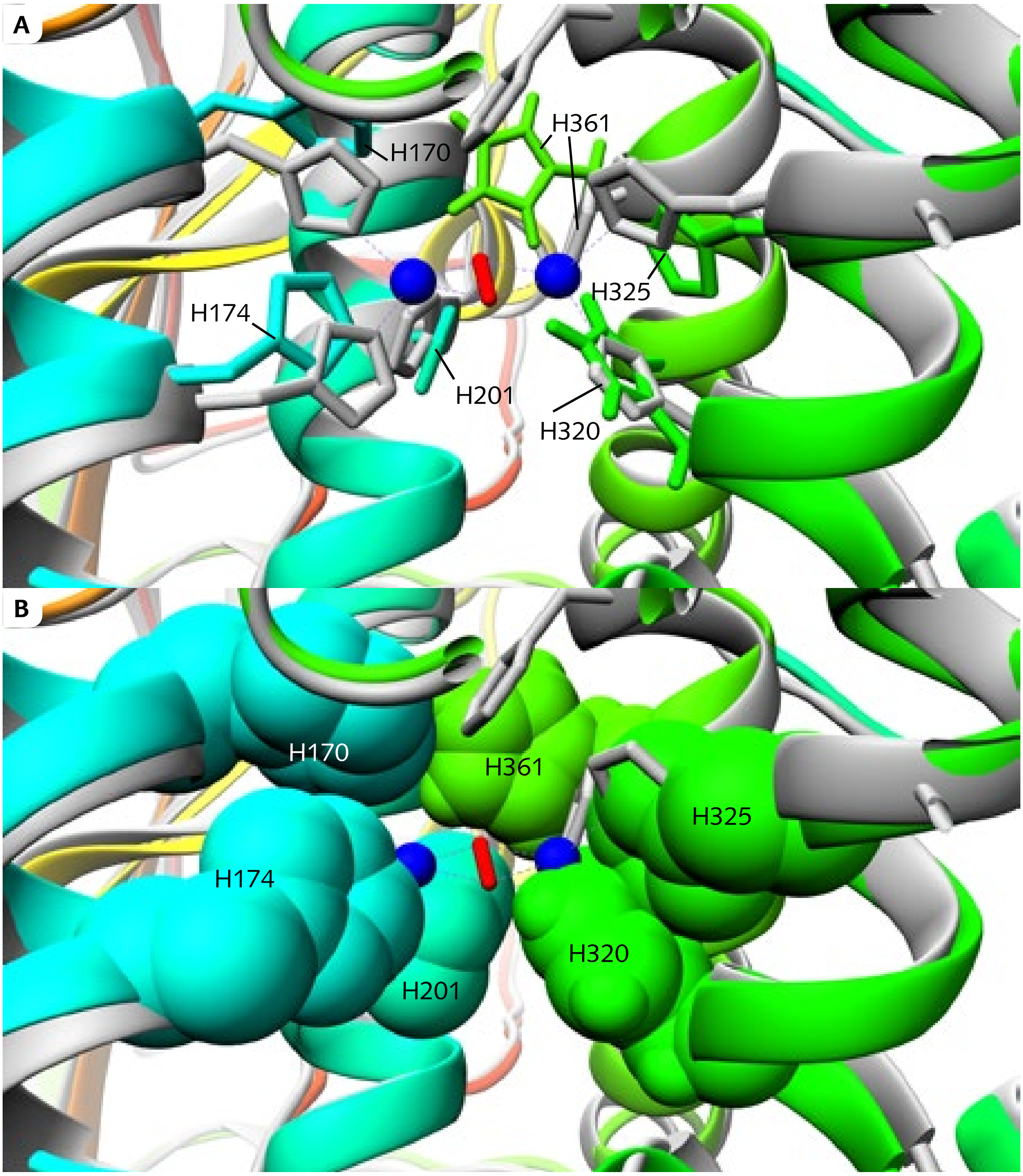


Fig. 87: Active Site of hexamer 1 subunit type IV. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.6 Hexamer 1 – subunit type V

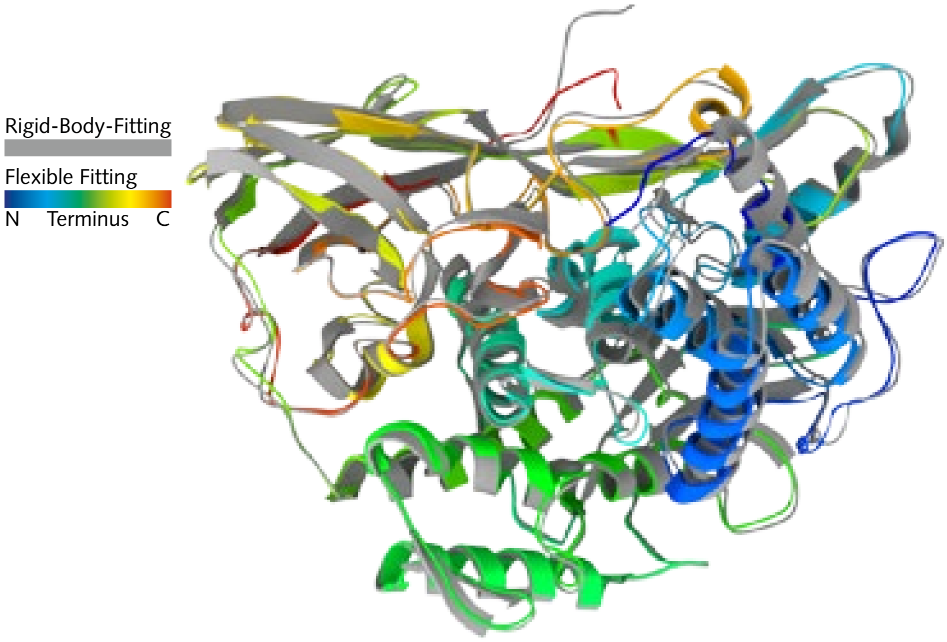


Fig. 88: Flexible Fitting result for subunit type V of hexamer 1. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 88 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9331 (rigid body) vs 0,9432 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

The components of domain #1 are shifted by 0,9 to 1,2 Å (Fig. 89A), improving the fitting quality, especially of the α-helices.

Domain #2 remains without large scale changes (Fig. 89C) with the exception of the α2.1→α2.2 loop which is pulled into the reconstructed mass (1,6 Å shift, Fig. 89B). Domain #3 is modified slightly by the fitting algorithm (Fig. 89D). Most notably the elongated C-terminus is pulled into the reconstructed volume (Fig. 89E).

Fig. 90 provides a detailed comparison of the active site with the 1NOL PDB as reference. H179 shows the largest shift of 0,6 Å mean and 1,5 Å max with 17,6° rotation. It is followed by H366 with 0,5 Å mean and 1,4 Å max shift and a rotation of 17,7°. Less drastic shifts can be observed for H175 (0,3 Å mean, 1,0 Å max, 24,9° rotation), H330 (0,4 Å mean, 0,9 Å max, 10,1° rotation) and H206 (0,3 Å mean, 0,8 Å max, 10° rotation). H326 has a negligible movement of 0,1 Å mean, 0,2 Å max and 7,8° rotation. All rotational values are plane/plane angles of the imidazole functional group.

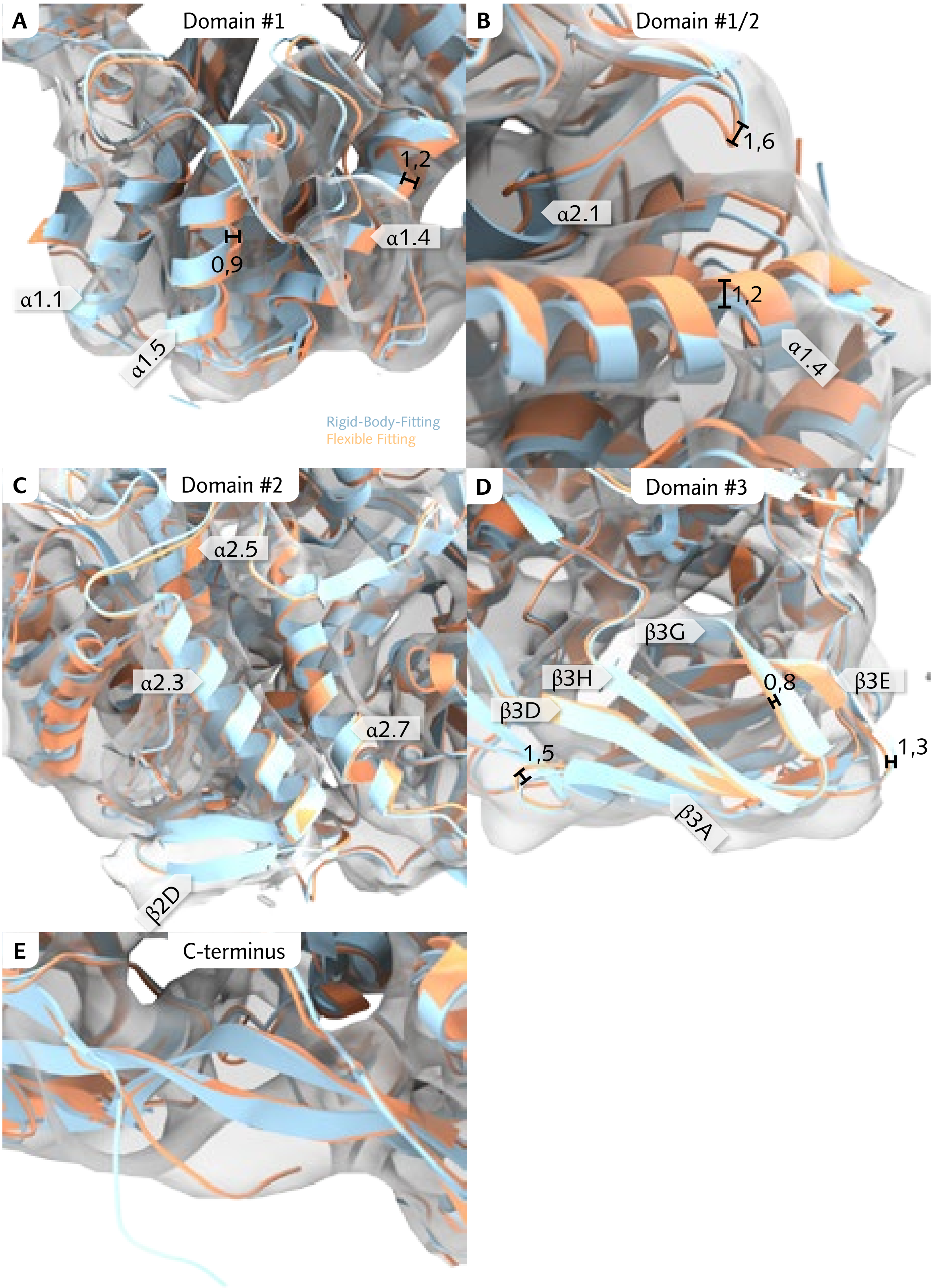


Fig. 89: Flexible Fitting result for subunit type V of hexamer 1 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

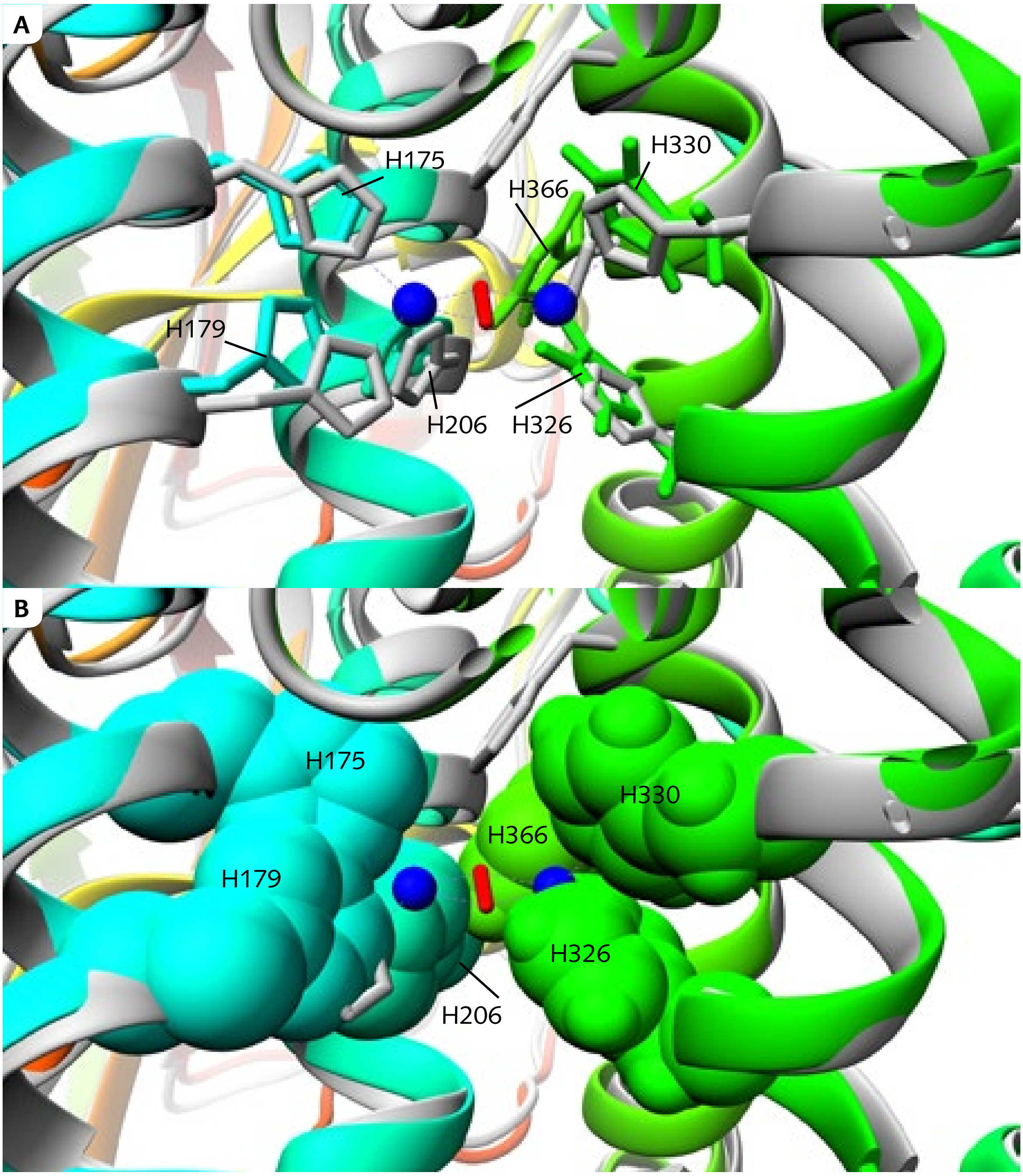


Fig. 90: Active Site of hexamer 1 subunit type V. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.7 Hexamer 2 – subunit type I

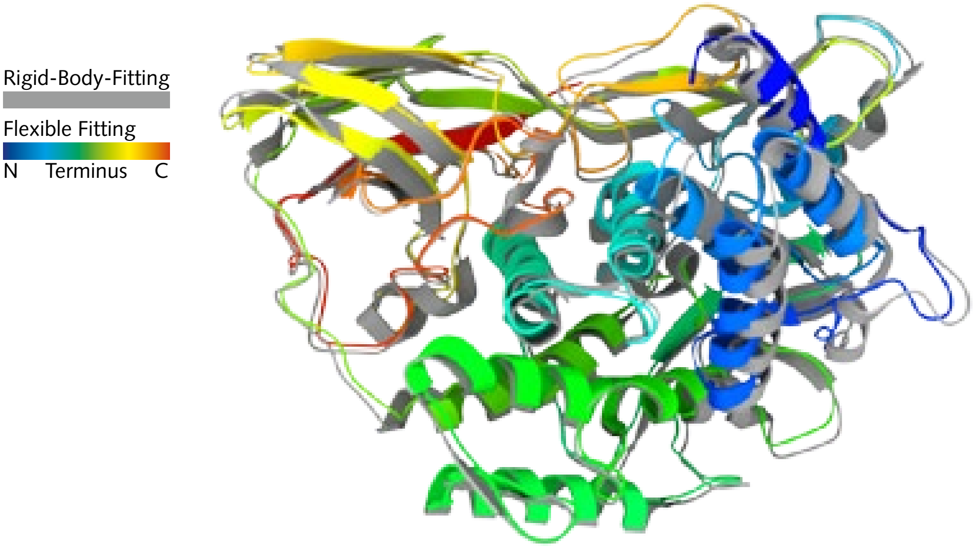


Fig. 91: Flexible Fitting result for subunit type I of hexamer 2. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 91 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9368 (rigid body) vs 0,945 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

The N-terminal end of α1.1 is shifted by 2,7 Å (Fig. 92A) using the C-terminal end of the helix as a pivot point. Helices α1.5 and α1.6 shift both shift by 1,5 Å (Fig. 92A). Another larger shift occurs to α1.4: a movement of 2,1 Å brings the helix into the reconstructed volume (Fig. 92B).

Domain #2 is affected with multiple shifts (Fig. 92C) ranging from values of 1,7 to 1,9 Å for secondary structure elements (Fig. 92C and D). The β2D→β2E region shifted by 1,1 to 1,4 Å, closer matching the reconstructed volume.

The dominating β-barell of domain #3 remains largely unchanged with the exception of interconnecting loops (Fig. 92E). The β3D→β3E loop is pulled into the volume by 1,4 Å; α3.6 is shifted by as much as 2,1 Å on exposed sections, although it is not recognized by the DSSP algorithm (Fig. 92E).

A closer look at the active site compared to the 1NOL PDB shows that H200 exhibits the largest shift with 0,4 Å mean and 1,2 Å max and a rotation of 16,4°. H360 is shifted by 0,4 Å mean and 1,1 Å max, 15° rotation. The remaining parts show small shifts (H320 0,3 Å mean and 0,5 Å max, rotation 10,7°; H324 0,2 Å mean and 0.6 Å max, 10,9°; H170 0,1 Å mean, 0,5 Å max, rotation 16,7°). All rotational values are plane/plane angles of the imidazole functional group.

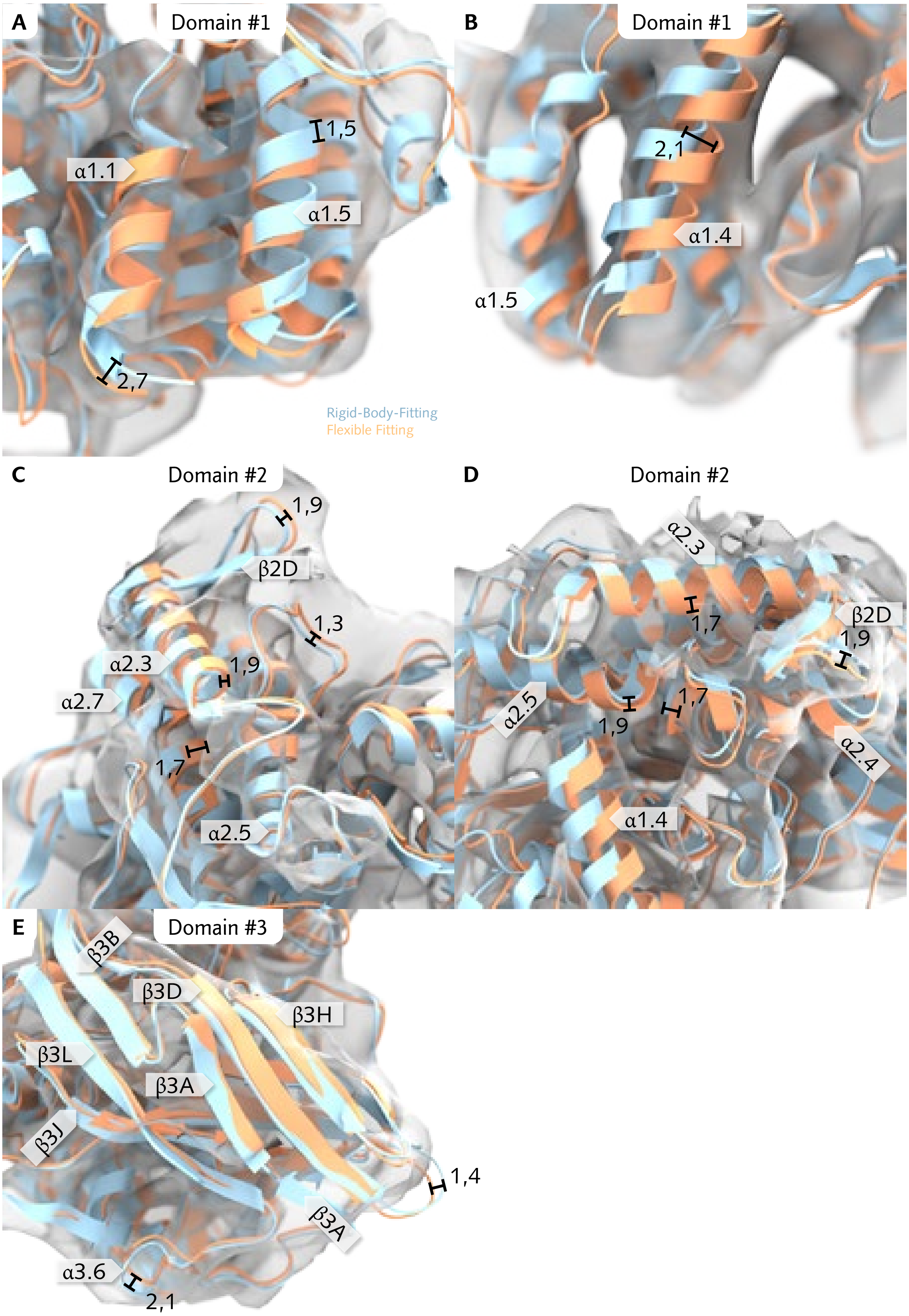


Fig. 92: Flexible Fitting result for subunit type I of hexamer 2 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

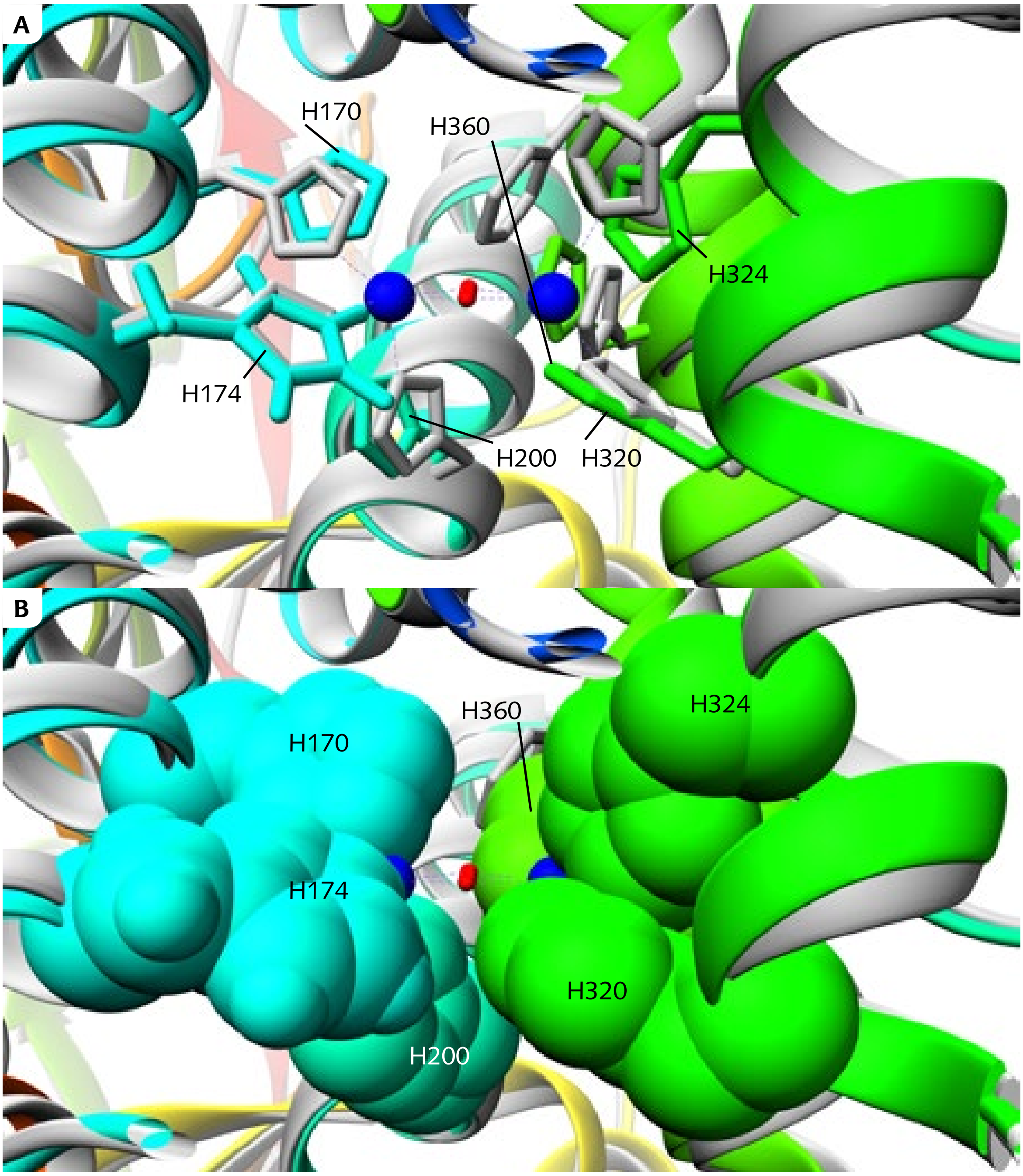


Fig. 93: Active Site of hexamer 2 subunit type I. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.8 Hexamer 2 – subunit type II

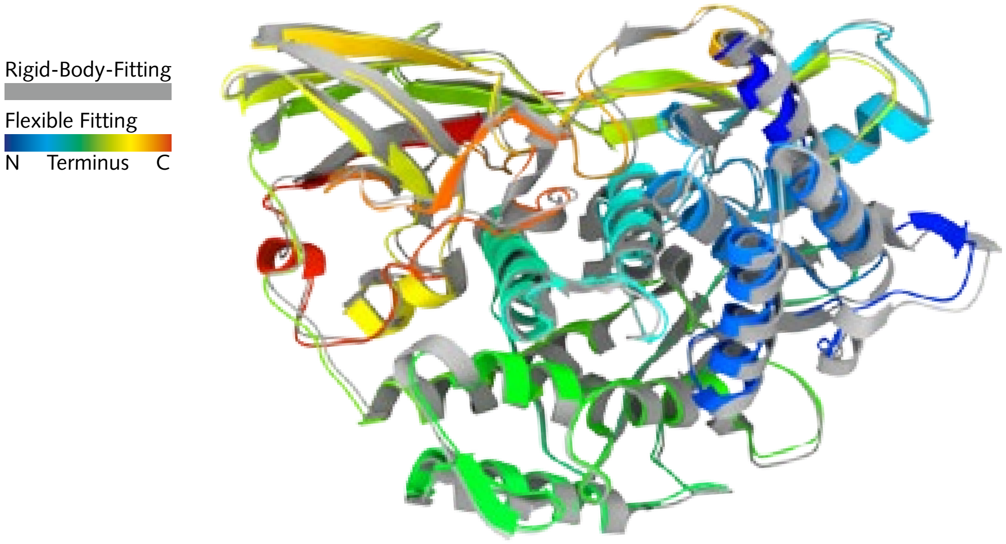


Fig. 94: Flexible Fitting result for subunit type II of hexamer 2. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 94 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9454 (rigid body) vs 0,976 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

Some smaller scale changes can be observed in domain #1 (Fig. 95A). Helix α1.1 is distorted by the fitting at the C-terminal end elongating the α1.1→α1.3 loop. Helix α1.3 and α1.5 are shifted towards the center of the subunit by 1,3 and 0,8 Å respectively (Fig. 95A). Helix α1.7 is moved by 1,1 Å (Fig. 95B), slightly improving fit at the C-terminal end of the helix.

Domains #2 and #3 are devoid of large structural modifications (Fig. 95C/D), for example α2.7 is shifted by 0,7 Å.

A comparison of the active site with the 1NOL PDB (Fig. 96) reveals three histidines which exhibit notable shifting: H364 with 0,7 Å mean, 1,8 Å max, 21,6° rotation; H177 with 0,5 Å mean, 1,4 Å max, 16,6° rotation and H328 with 0,5 Å mean, 1,3 Å max and 14,1° rotation. Less drastic shifting occurs for H204 (0,2 Å mean, 0,7 Å max, 23,9° rotation), H173 (0,1 Å mean, 0,4 Å max, 8,1° rotation) and H324 (0,1 Å mean, 0,3 Å max, 8,3° rotation). All rotational values are plane/plane angles of the imidazole functional group.

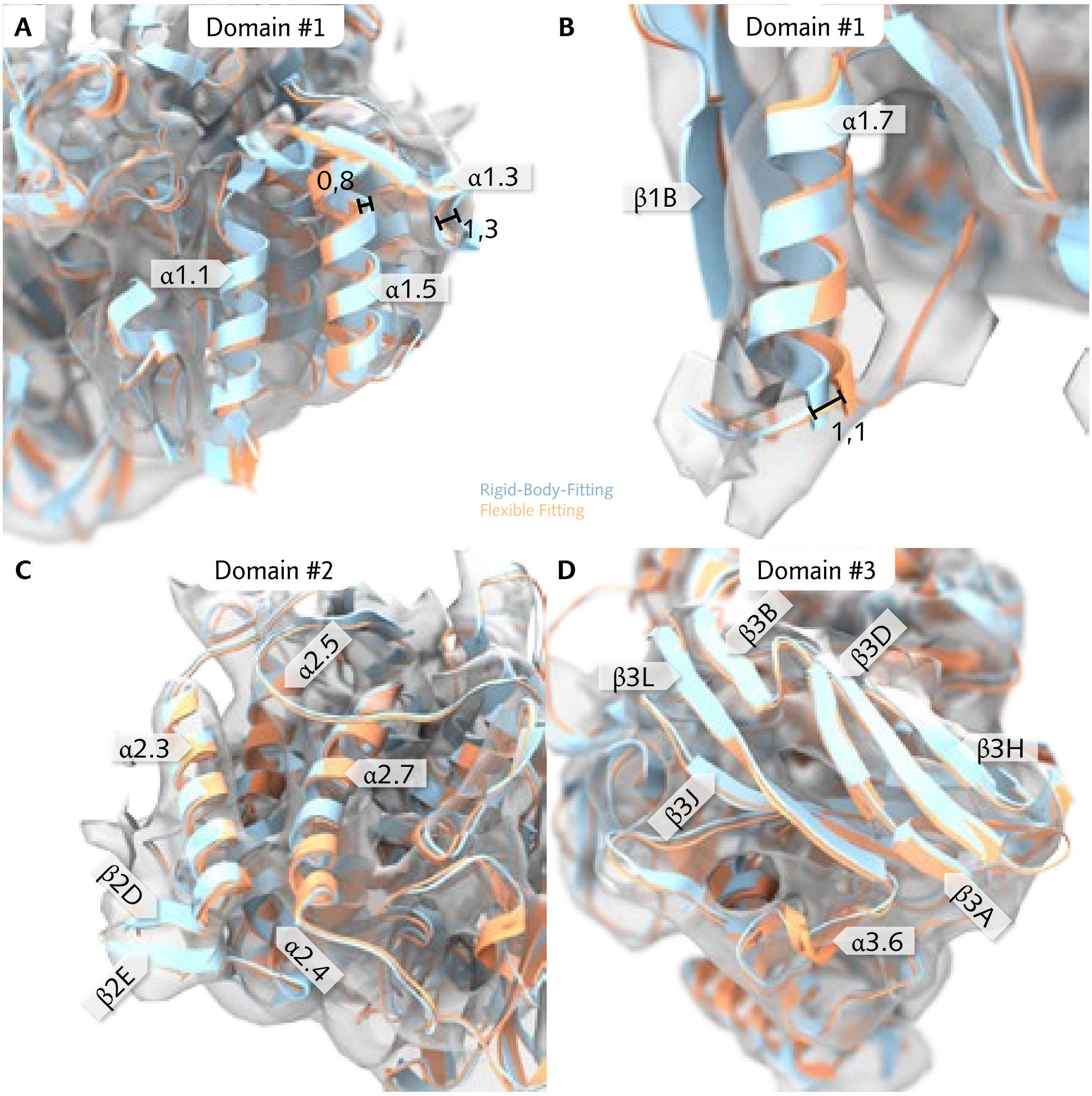


Fig. 95: Flexible Fitting result for subunit type II of hexamer 2 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

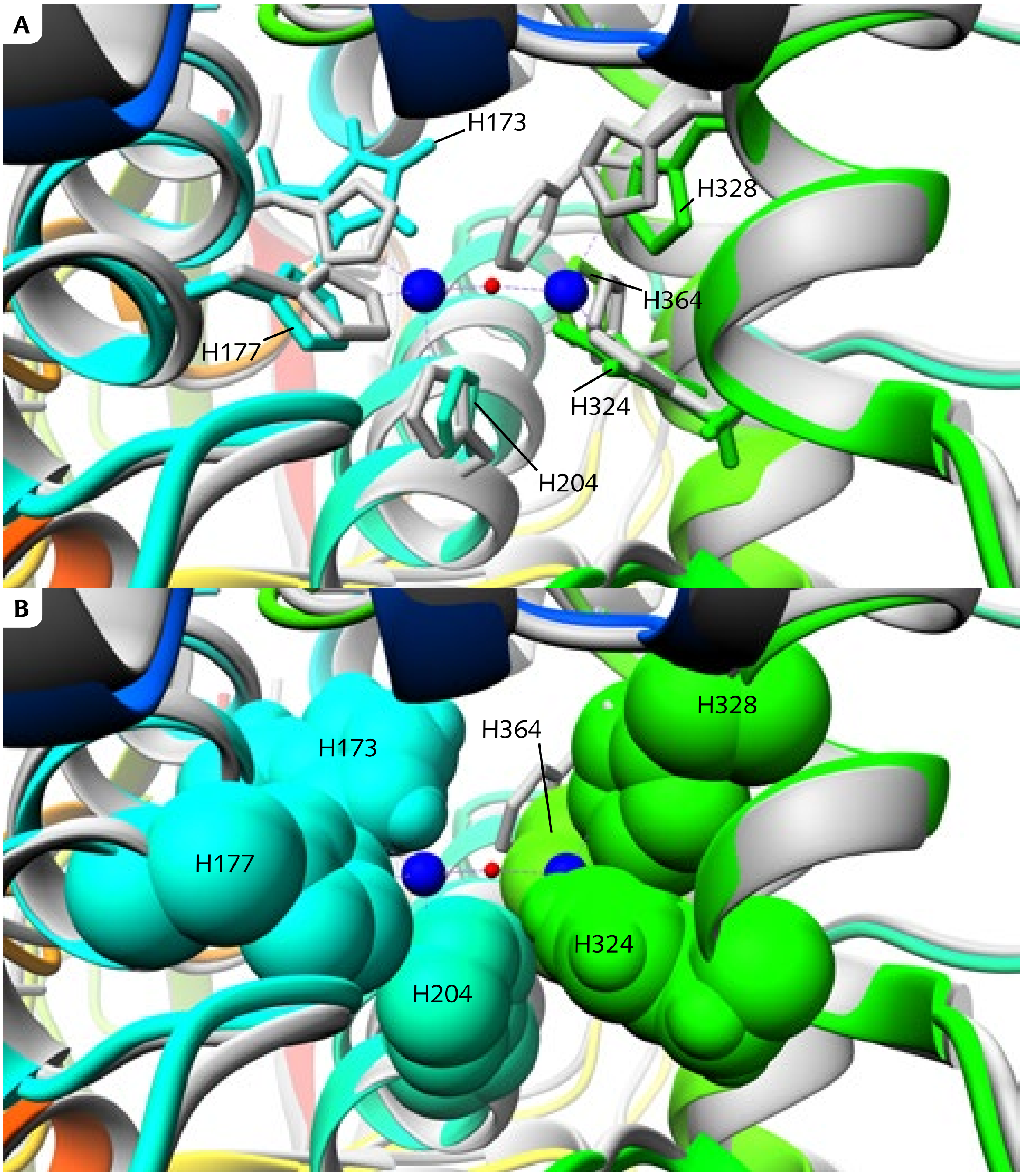


Fig. 96: Active Site of hexamer 2 subunit type II. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.9 Hexamer 2 – subunit type IIIA

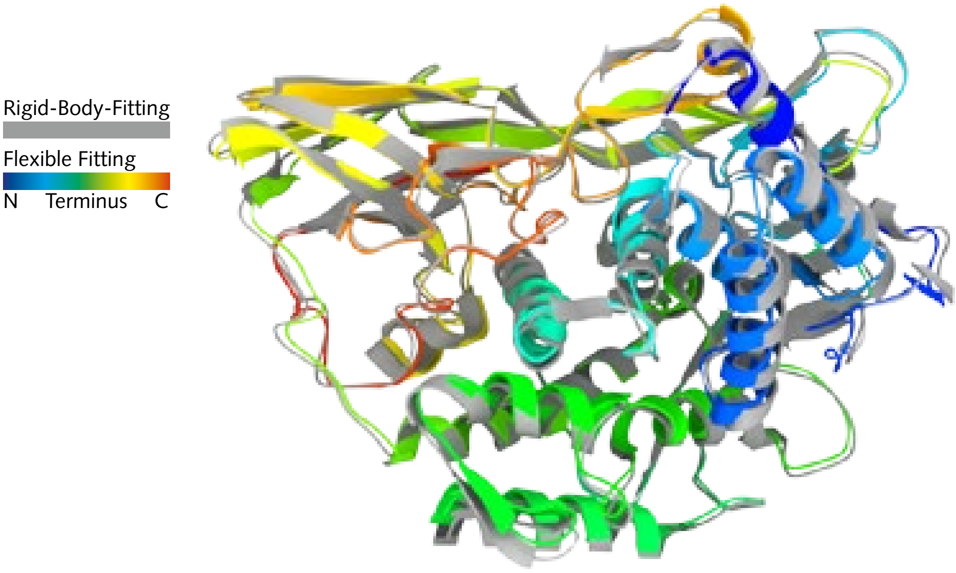


Fig. 97: Flexible Fitting result for subunit type IIIA of hexamer 2. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 97 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9483 (rigid body) vs 0,9539 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

Notable in domain #1 is the shift of α1.4 and α1.5 towards the center of the reconstruction by 0,9 and 1,4 Å respectively. Loops are also modified by as much as 1,9 Å leading to an improved fit in this region (Fig. 97A). In domain #2 a larger movement shifts α2.3 (1,6 Å), α2.4 (1,9 Å) and α2.5 (0,8 Å) away from the periphery of the reconstructed volume. In contrast α2.7 is only shifted by 0,5 Å (Fig. 97B and C). For the β2D→β2D region a shift of 2,1 Å can be observed. This movement pushes β2E outside of the reconstructed volume (Fig. 97B and C).

Smaller movements can be observed in domain #3, however some secondary structure elements are slightly distorted (Fig. 97D).

Fig. 99 provides a comparison of the active site with 1NOL as reference. In contrast to the other active sites, several larger shifts occur: H169 with 0,8 Å mean and 2,3 Å max (42.1° rotation), H200 with 0,7 Å mean and 2,0 Å max (35,4° rotation) and H324 with 0,6 Å mean, 1.7 Å max (28,1° rotation). The remaining parts show minor shifting: H173 with 0,3 Å mean and 0,9 Å max (18,5° rotation), H360 with 0,4 Å mean and 1,0 Å max (16,8° rotation) as well as H320 with 0,1 Å mean and 0,4 Å max (10,1° rotation). All rotational values are plane/plane angles of the imidazole functional group.

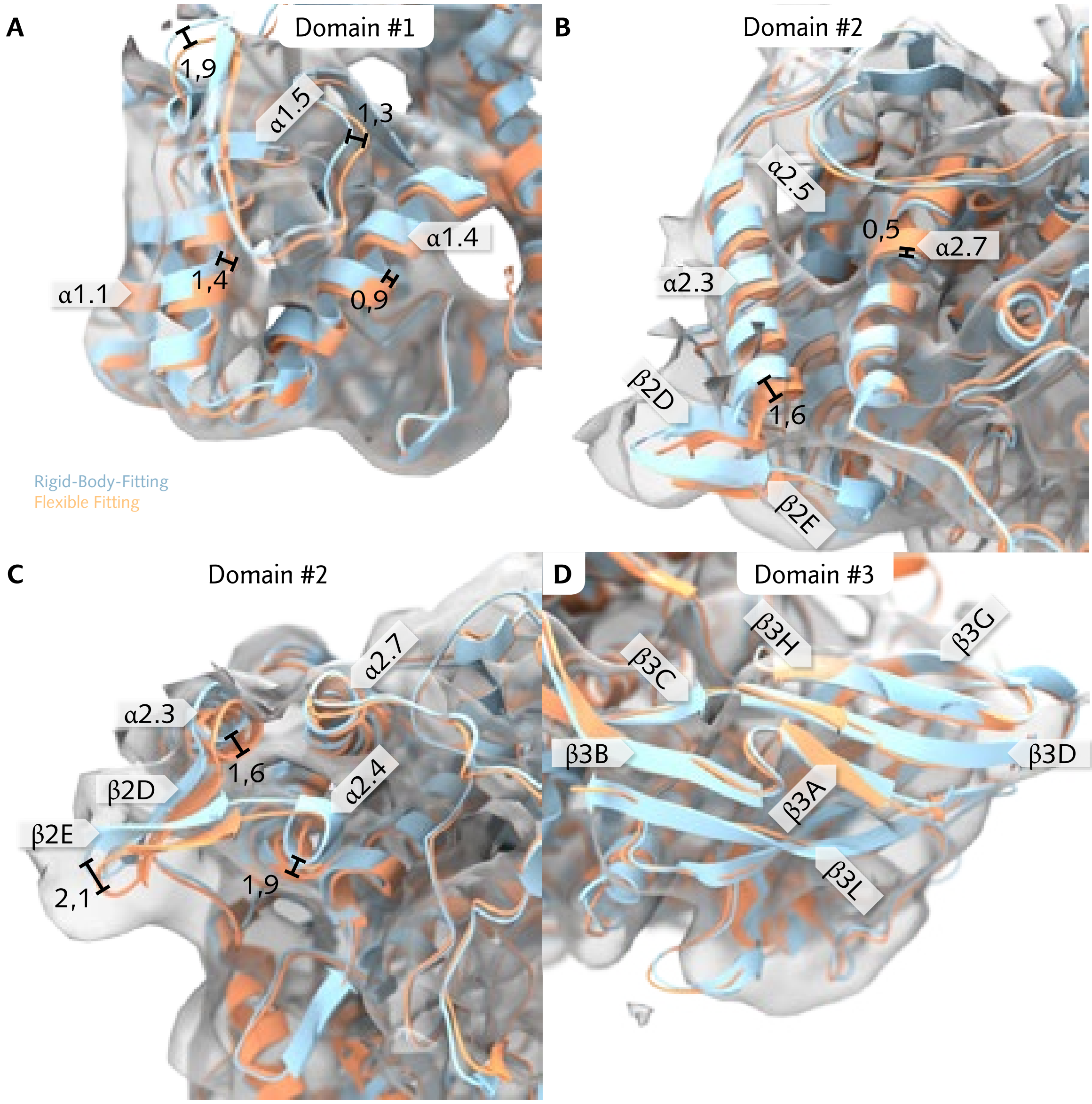


Fig. 98: Flexible Fitting result for subunit type IIIA of hexamer 2 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

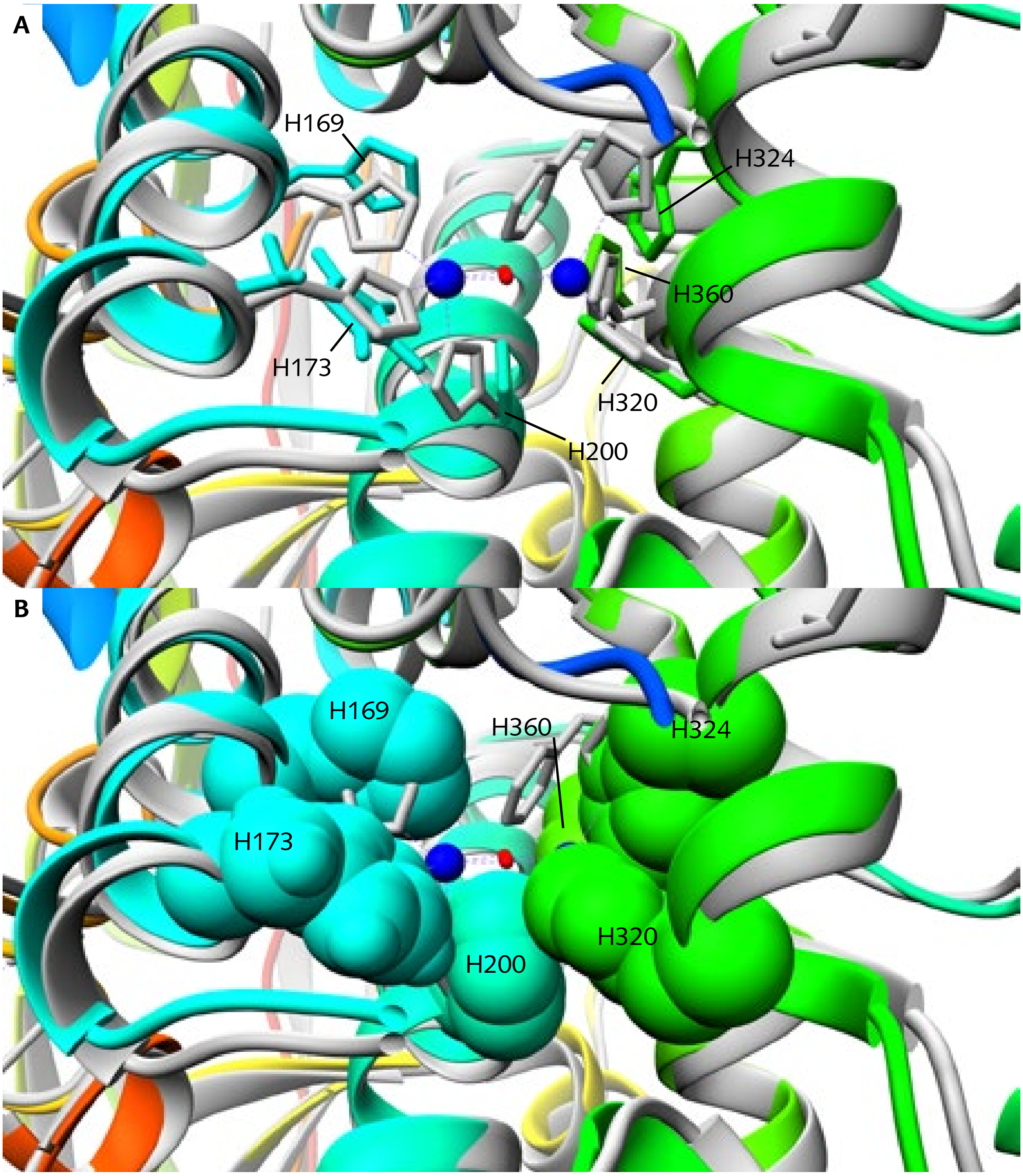


Fig. 99: Active Site of hexamer 2 subunit type IIIA. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.10 Hexamer 2 – subunit type IIIB

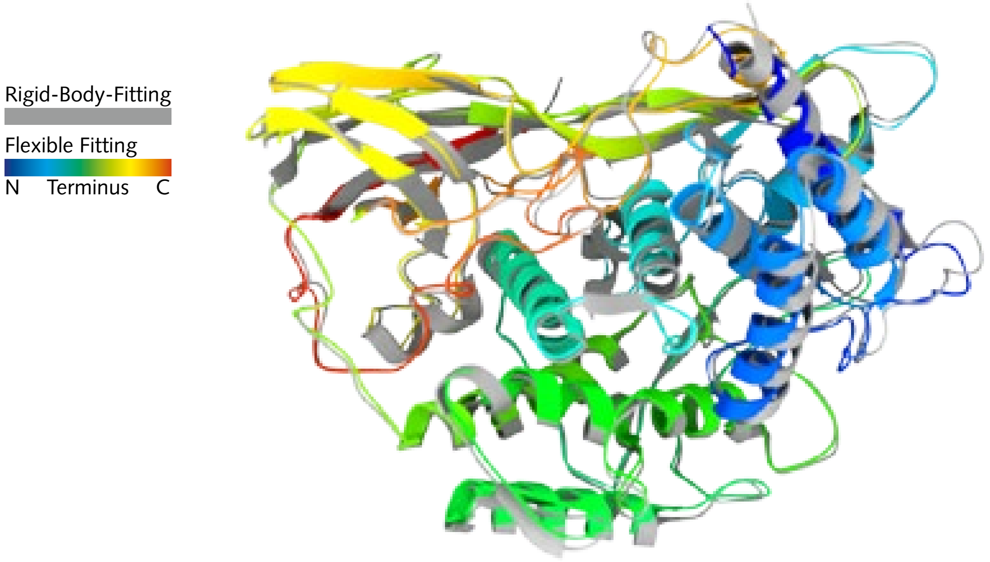


Fig. 100: Flexible Fitting result for subunit type IIIB of hexamer 2. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 100 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9428 (rigid body) vs 0,9519 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

Many components of domain #1 are shifted away from the periphery of the reconstructed volume (Fig. 101A and B). Helix α1.4 and α1.5 shift by 1,9 and 1,6 Å respectively. The α1.1→α1.3 loop is moved by 2,3 to 3,0 Å (α1.3 is slightly distorted in the process) to better fit into the reconstructed volume. Domain #2 is affected by multiple shifts ranging from values of 0,7 to 1,6 Å for secondary structure elements (Fig. 101C). The β2D→β2E region is shifted by 1,2 Å. Domain #3 remains unchanged without larger notable modifications (Fig. 101D)

Fig. 102 provides a comparison of the active site with 1NOL as reference. The largest deviation can be observed on H202 with 0,6 Å mean and 1,7 Å max (27,2° rotation). Smaller shifts are exhibited by H171 (0,3 Å mean and 1,2 Å max, 34,4° rotation), H175 (0,2 Å mean and 0,8 Å max, 18,9° rotation) and H323 (0,3 Å mean and 0,7 Å max, 10,5° rotation). H363 and H327 show very minor shifting, approx. 0,1 Å. All rotational values are plane/plane angles of the imidazole functional group.

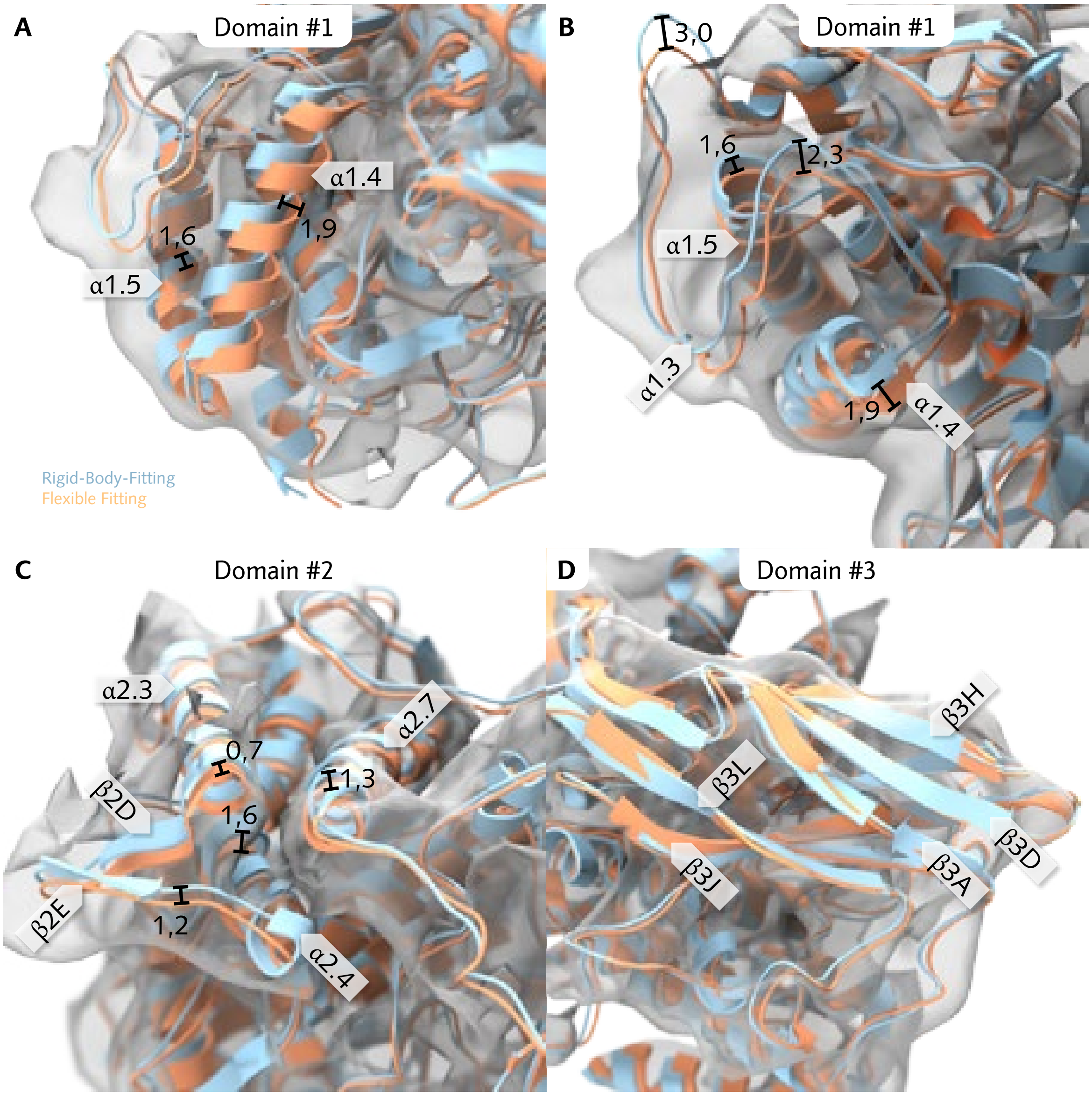


Fig. 101: Flexible Fitting result for subunit type IIIB of hexamer 2 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

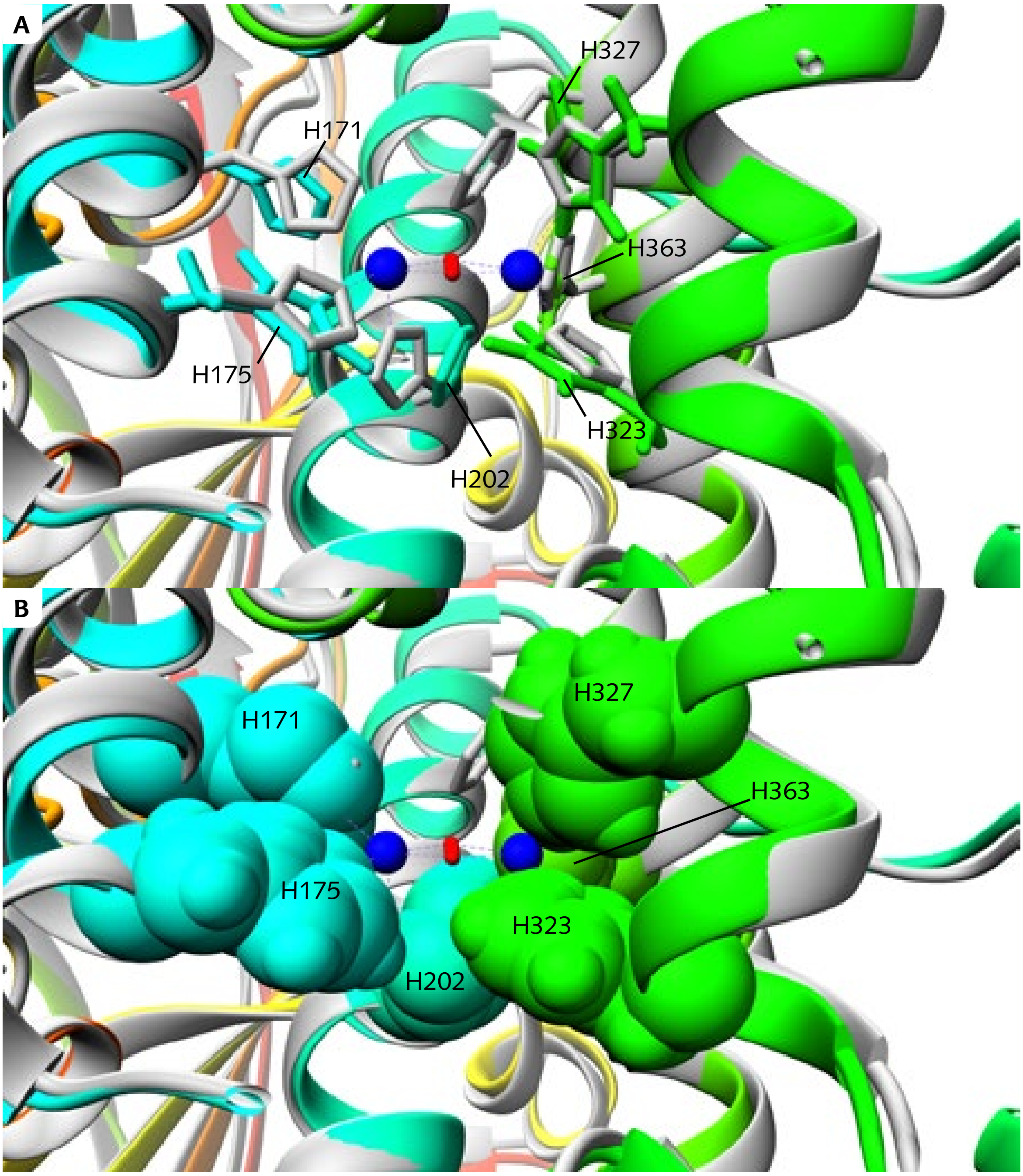


Fig. 102: Active Site of hexamer 2 subunit type IIIB. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.11 Hexamer 2 – subunit type IV

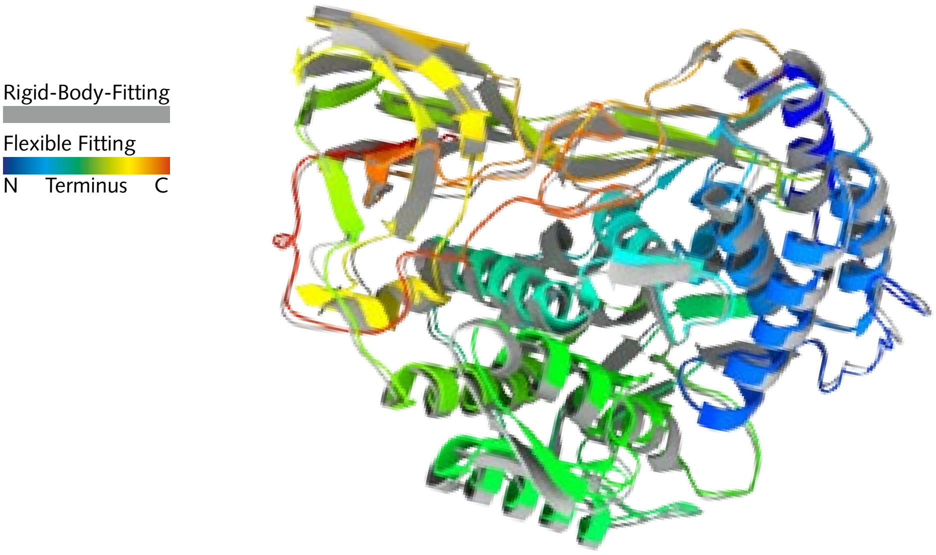


Fig. 103: Flexible Fitting result for subunit type IV of hexamer 2. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 103 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9413 (rigid body) vs 0,948 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

Domain #1 exhibits shifts between 0,6 and 1,2 Å (α1.5 and α1.4, Fig. 104A). The shifted α-helices produce a better fit, especially when viewed at high threshold levels. The α1.1→α1.3 loop is shifted by 1,1 Å with little impact on overall fitting quality (Fig. 104A).

Larger shifts can be observed for α-helical components of domain 2# (Fig. 104A and C). Values range from 0,7 Å for α2.7 to 1,8 Å for α2.3. The β2D→β2E region shifts by as much as 2,0 Å. These modifications improve the overall fitting quality of the domain. Domain #3 remains unchanged without larger notable shifts (Fig. 104D).

Fig. 105 provides a comparison of the active site with 1NOL as reference. Notable shifts occur for H201 (0,4 Å mean and 1,0 Å max, 18,9° rotation) and H325 (0,4 Å mean and 0,9 Å max, 9,2° rotation). The other four components remain close to their original orientations: H361 (0,2 Å mean and 0,6 Å max, 6,9° rotation), H174 (0,1 Å mean and 0,2 Å max, 9,1° rotation), H170 (0,1 Å mean and 0,3 Å max, 17,8° rotation) and H321 (0,1 Å mean and 0,4 Å max, 13,4° rotation). All rotational values are plane/plane angles of the imidazole functional group.

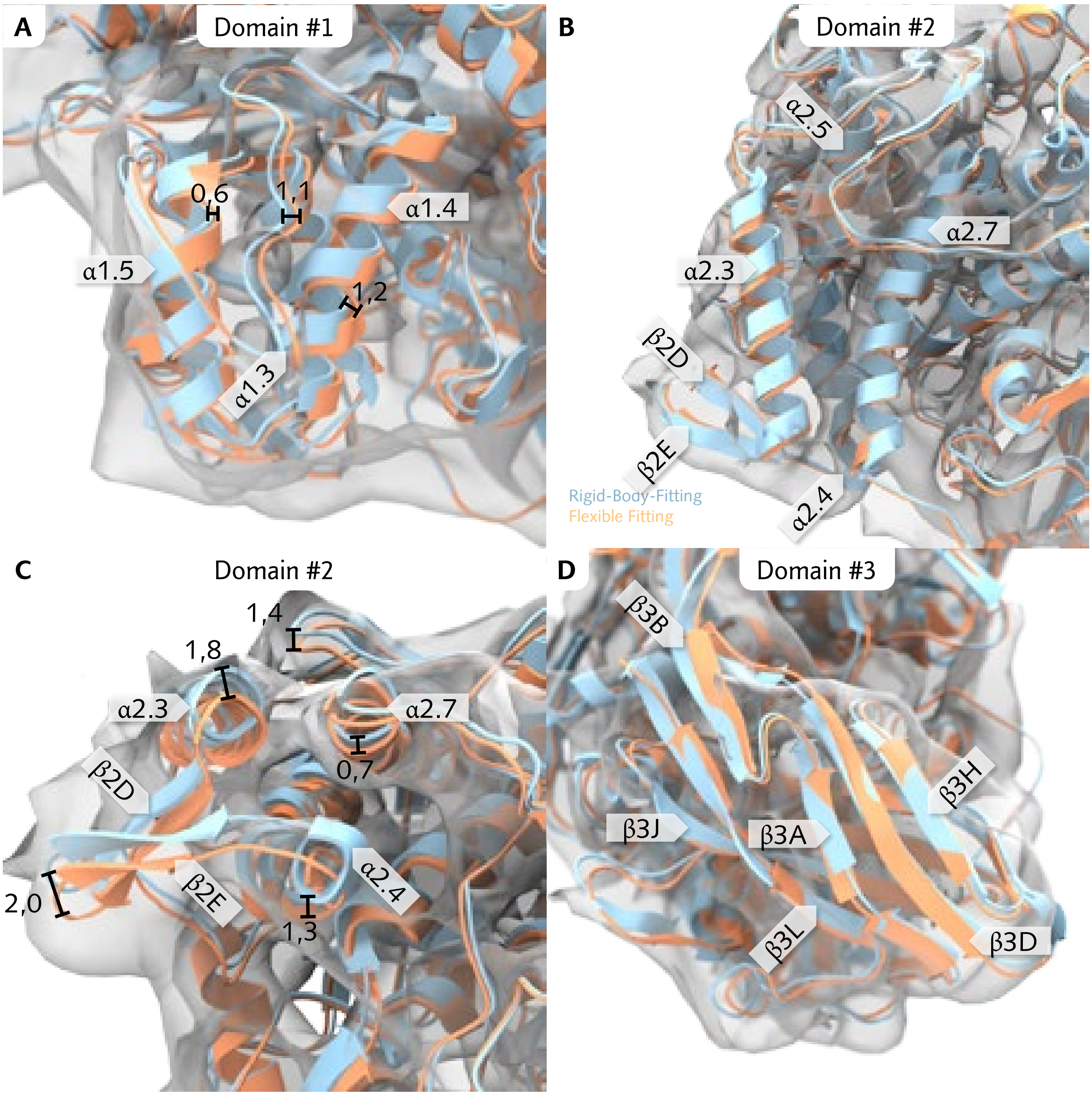


Fig. 104: Flexible Fitting result for subunit type IV of hexamer 2 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

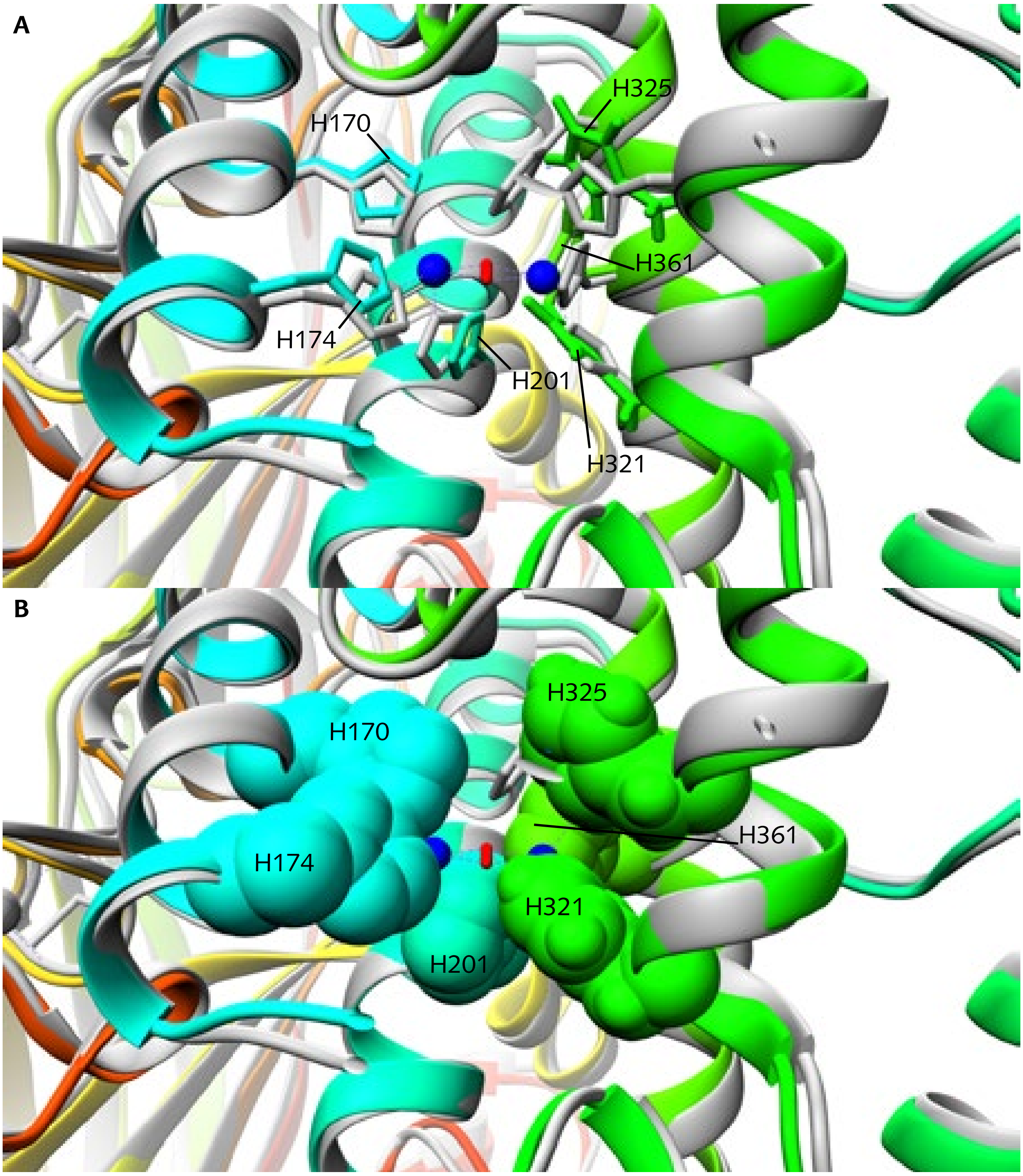


Fig. 105: Active Site of hexamer 2 subunit type IV. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.5.12 Hexamer 2 – subunit type VI

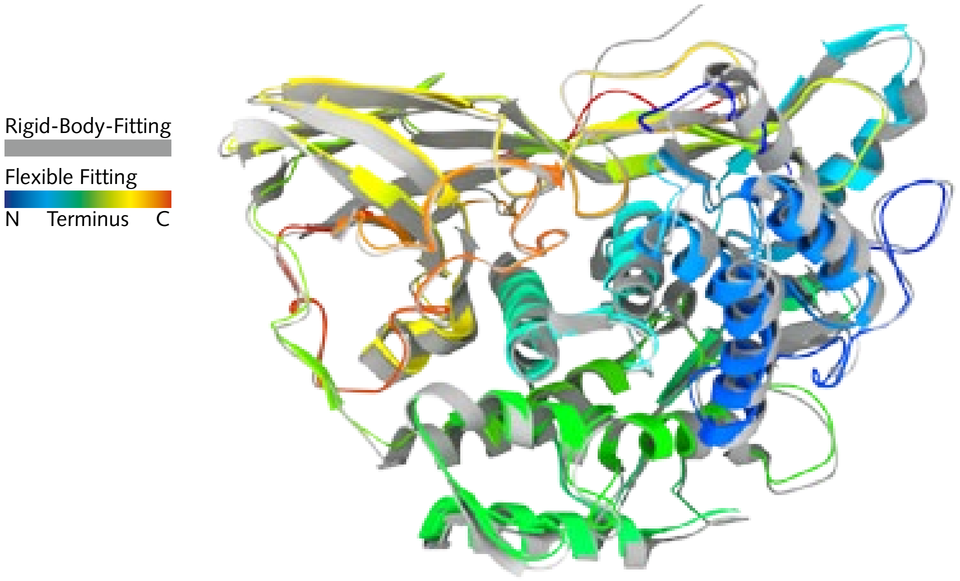


Fig. 106: Flexible Fitting result for subunit type VI of hexamer 2. Rigid-body-fitting result colored gray, flexible fitting result with rainbow color code (see legend).

Fig. 106 provides an overall comparison between the original structure and the flexible fitting result. The correlation coefficient is 0,9369 (rigid body) vs 0,9473 (flexible). To calculate the correlation coefficient, simulated maps at 8 Å resolution were used.

The N-terminus is moved by 2,9 Å. Secondary structure elements of domain #1 exhibit shifts between 1,0 and 1,2 Å (Fig. 107A, α1.5 and α1.4 respectively). Shifts throughout domain #2 remain below 1,0 Å (Fig. 107A). The β2D→β2E region is of note, shifting by 0,9 Å to improve the fitting quality (Fig. 107B). The α2.5→α2.6 loop is shifted by 2,0 Å towards the center of the reconstructed volume (Fig. 107C). Domain #3 remains unchanged without larger notable shifts (Fig. 107D). It should be noted however that the secondary structure is slightly distorted in this area.

Fig. 108 provides a comparison of the active site with 1NOL as reference. Two histidine residues exhibit larger shifts: H178 with 0,7 Å mean and 1,9 Å max (26,1° rotation) as well as H205 with 0,4 Å mean and 1.4 Å max (33,2° rotation). Smaller shifts occur for H329 (0,4 Å mean and 1,1 Å max, 16,4° rotation) and H174 (0,4 Å mean and 1,0 Å max, 18,1° rotation). H365 (0,3 Å mean and 0,7 Å max, 8,8° rotation) and H325 (0,2 Å mean and 0,4 Å max, 1,3° rotation) are less affected. All rotational values are plane/plane angles of the imidazole functional group.

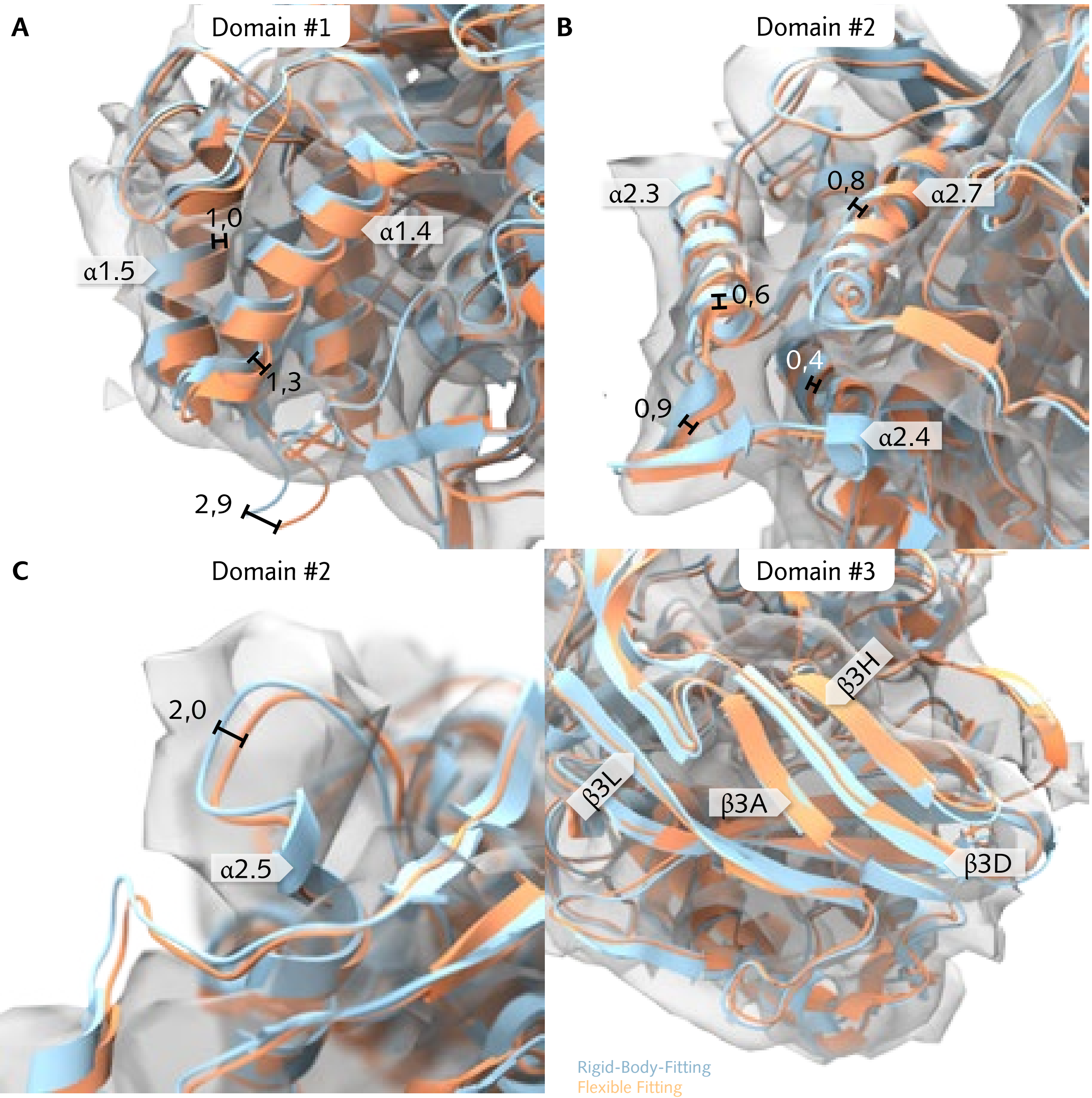


Fig. 107: Flexible Fitting result for subunit type VI of hexamer 2 – regions of interest. Rigid-Body-Fitting depicted light blue, flexible fitting result orange. See text for detailed description. All distances measured in Å.

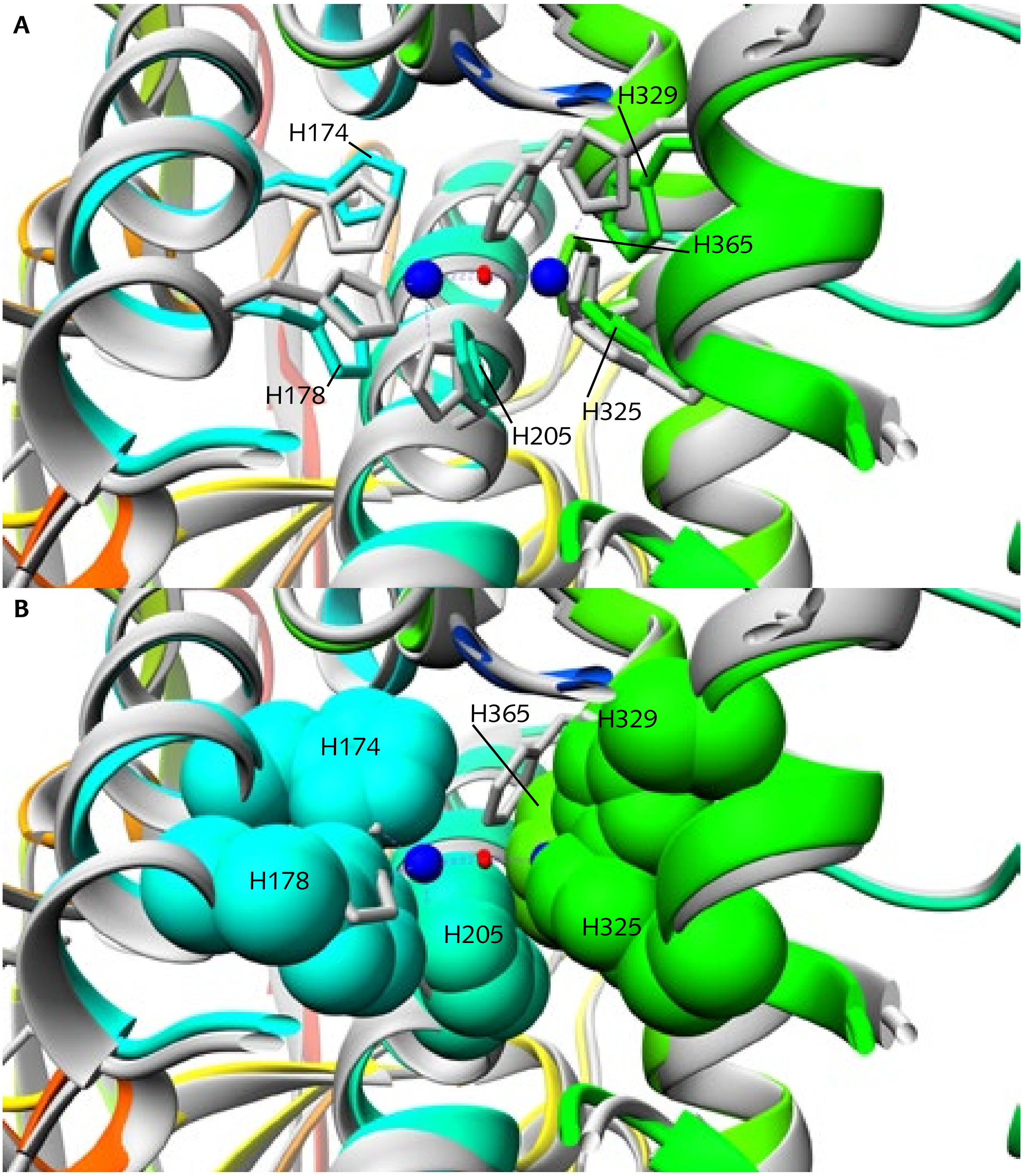


Fig. 108: Active Site of hexamer 2 subunit type VI. Panel A: Reference PDB 1NOL depicted in gray with copper atoms (blue) and O_2 (red), flexible fitting shown with rainbow color code from N to C terminus.

4.6 Flexible Fitting – Interface analysis

This section compares the interfaces as previously described by Martin et al., 2007. Again it should be noted that not all subunits of *Lpo*Hc have been successfully sequenced. Subunit type II, IIIA, IIIB, IV and VI have authentic sequences available. The authentic sequence of subunit IIIB is newly available. For subunits I and V a homology model based on the sequences of *Carcinoscorpius rotundicauda* hemocyanin was used instead. Subunits I and IIA are not participating in any inter-hexamer contacts, but subunit V is both involved in interfaces forming the 2x6mer as well as the 4x6mer. Furthermore the locations of subunits V and VI are not known, the designations are interchangeable.

4.6.1 Interfaces forming the 2x6mer

Four different types of interface with six distinct bridges are visible in the reconstructed density map (Fig. 109) as previously described (Martin et al., 2007). A II–II interface with two subregions, a V–VI interface, two equivalent II–IV interfaces with two subregions as well as a V–VI interface with the equivalent IV–VI interface (flop and flip face of the 4x6mer respectively)

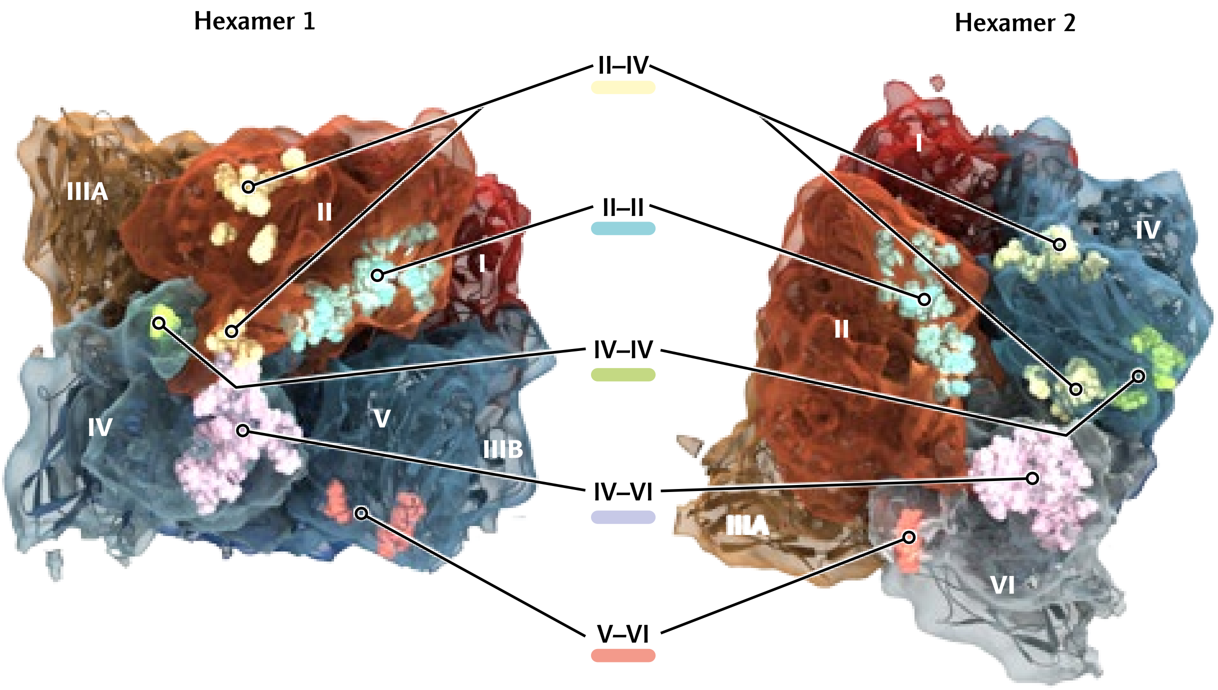


Fig. 109: Interface location between the hexamers of the basic 2x6mer. The hexamers are split at their plane of contact and rotated by 90° towards the viewer, thus exposing the contact surfaces. The subunits participating in the interface are type II (dark orange), type IV (pale blue) in both hexamers, type V (dark blue) in hexamer 1 and type VI (light gray) in hexamer 2. There are six interfaces comprised of four types, depicted by an atom-sphere visualization: Two II–IV bridges (yellow), the II–II bridge (cyan), the IV–IV (green) and IV–VI (violet) bridges, which are structurally equivalent and the V-VI bridge (red).

4.6.1.1 The II–II interface

The II–II interface is formed by complementary association of domain #3 (Fig. 110). Both subunits contribute the β3A→β3B loop and β3C→β3D loop The overall structure of the interface remains unchanged, although several shifts of amino acid position occur compared to the reference (Table 5; Martin et al., 2007). There is a slight tendency for both subunits to shift inward towards the interface, tightening the contact zone by ~0,75 Å. This still leaves the contact apparatus intact. Clashes remain in the H503–S442 connection as well as P440–P440 and D437–R411–F409. An overlap of H503 and D403 as observed in the reference study is not present in both rigid- and flexible fit models.

Subunit	Residue	RMSD (Å)
II	H399	0,515
	A400	0,432
	R401	0,867
	V402	0,754
	D403	0,437
	H407	1,461
	F409	1,536
	R411	1,964
	Y433	1,286
	Y434	1,735
	H435	3,061
II cont'd	L436	1,343
	D437	1,692
	H438	1,778
	E439	1,886
	P440	0,615
	S441	0,296
	F442	1,060
	Y443	1,609
	H503	0,389
	K625	1,532
	H627	2,121

Table 5: RMSD values of all participating residues of interface II–II. The values were calculated comparing all atoms of each residue individually.

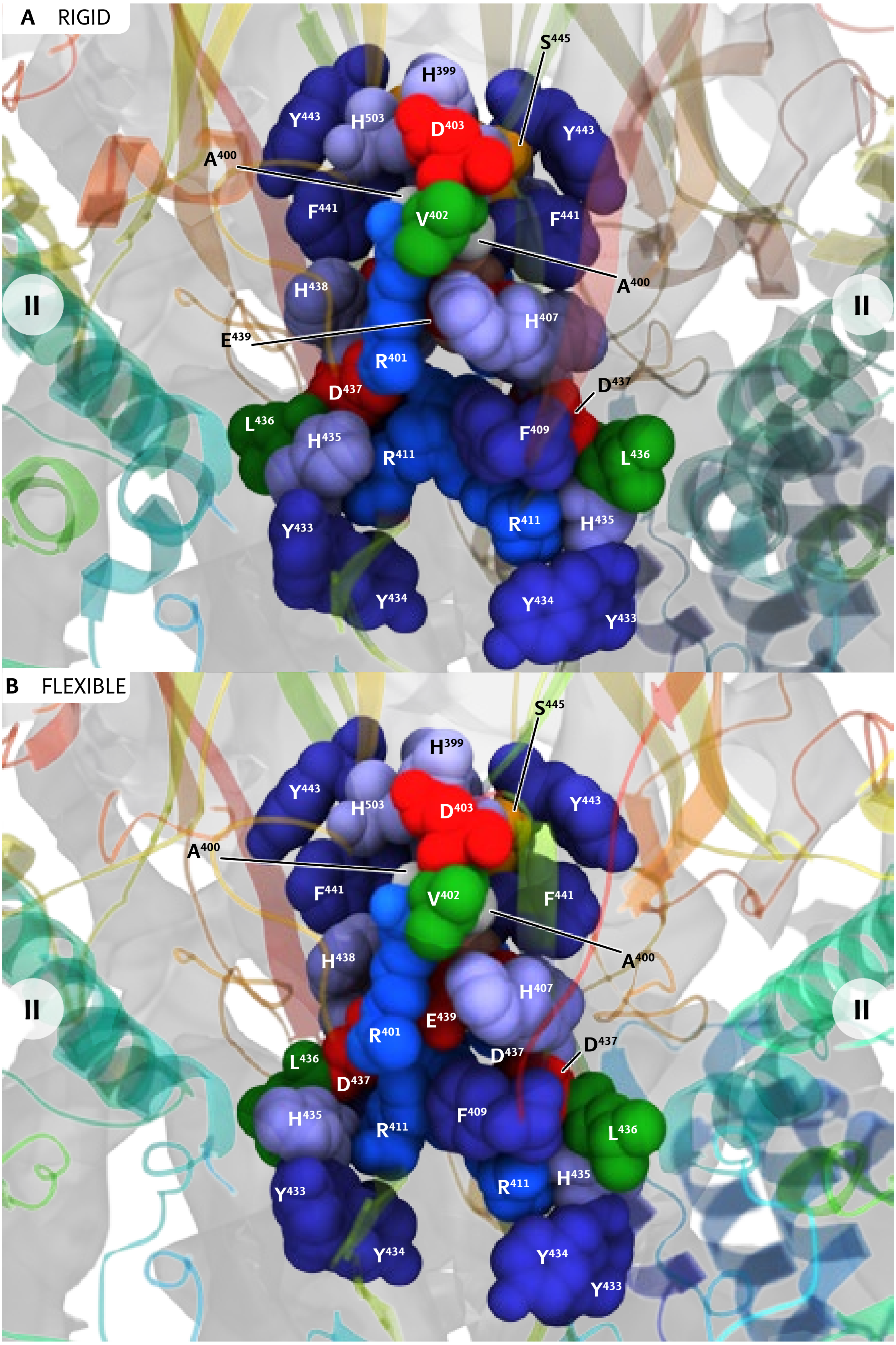


Fig. 110: The II–II interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay.

The top part of the interface might be formed by direct backbone interaction between eight residues (A399, V401, E438 and P439, Fig. 111 A and B).

The upper part (Fig. 111 C and D) consists of four histidine residues (H399 and H503), two serine residues (S442), two aspartic acid residues (D403) and four aromatic residues (F441 and Y443) flanking S442. Shifts in this region are minimal with the exception of S442 and Y443 (RMSD 1,060 Å and 1,609 Å respectively).

The middle section (Fig. 111 E and F) is formed by six histidine residues (H407, H338 and H629) as well as two phenylalanine residues and four acidic residues (D337 and E339). This part has been deemed important for the binding of an allosteric ligand.

The most notable shift after flexible fitting occurs for H435, which is part of the lower section of the described chamber (Fig. 111 G and H). It can be assumed that this is not of functional significance.

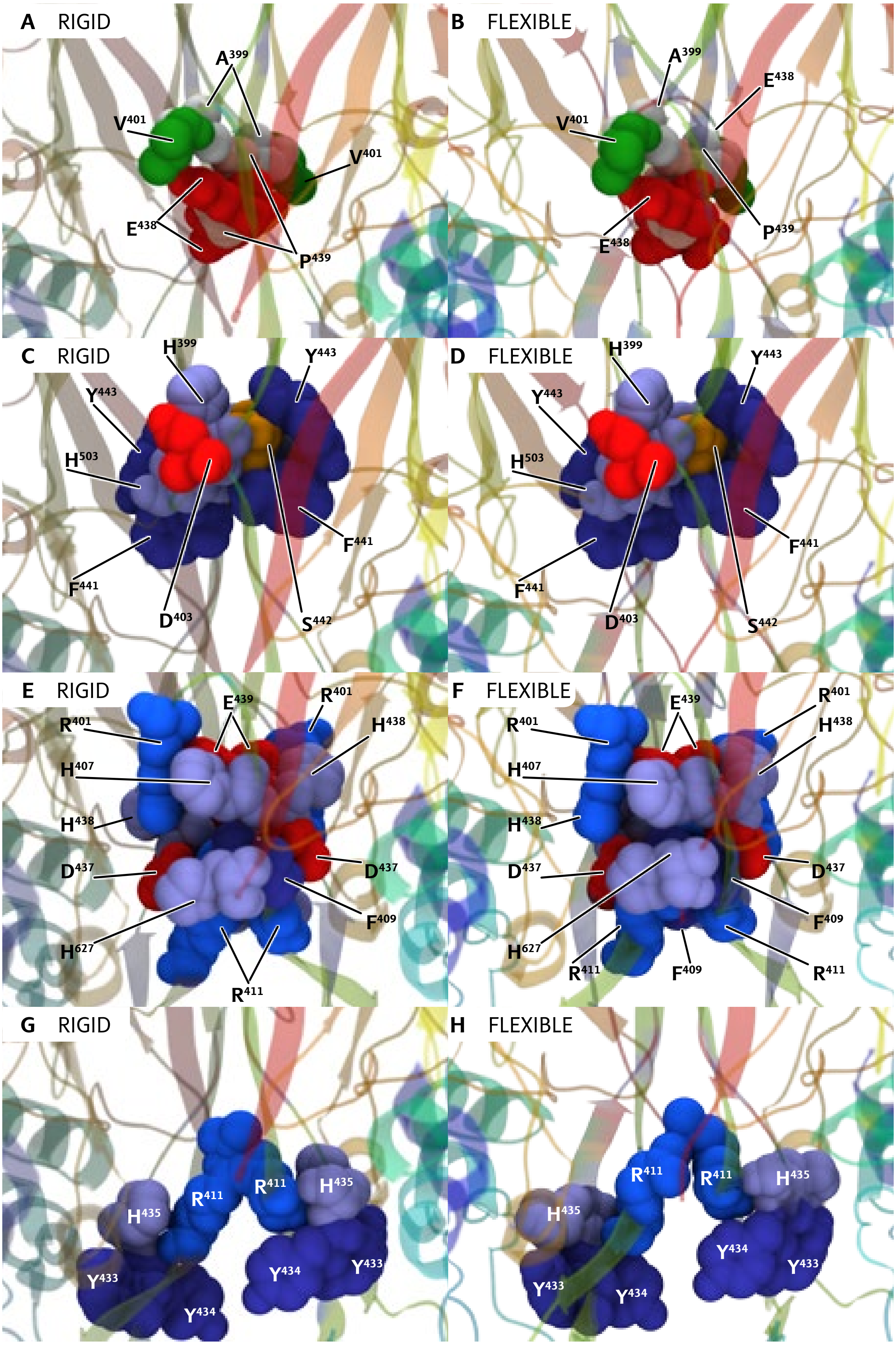


Fig. 111: The II–II interface – top, upper, middle and lower part. Left column depicts rigid-body fitting model, right column the flexible fitting result. Panel A and B: Side view of the top part of the interface. Panel C and D: Top view of the upper interface section. Panel E and F: middle chamber. Panel G and H: lower end of the interface.

4.6.1.2 The IV–VI interface

At a mass correlated threshold level, two bridges a visible (Fig. 112 A and B). Due to the imposed D2 point group symmetry, the IV–VI bridges are also present between subunits IV and V.

The 140HK141 loop is moved slightly closer to the glutamic acid residues (RMSD is 1,915 Å for H139 and 1,928 Å for K141. See Table 6 for complete list of RMSD values). Still the lower bridge in the reconstruction remains empty. It has been postulated that the negative charges without a bridging anion would prohibit binding at the upper site (Martin et al., 2007). Whether or not this interface is correctly modeled remains to be investigated.

Subunit	Residue	RMSD (Å)
IV	K380	1,986
	F387	2,661
	E443	3,509
	H445	2,040
VI	H139	1,915
	K140	1,928

Table 6: RMSD values of all participating residues of interface IV–VI. The values were calculated comparing all atoms of each residue individually.

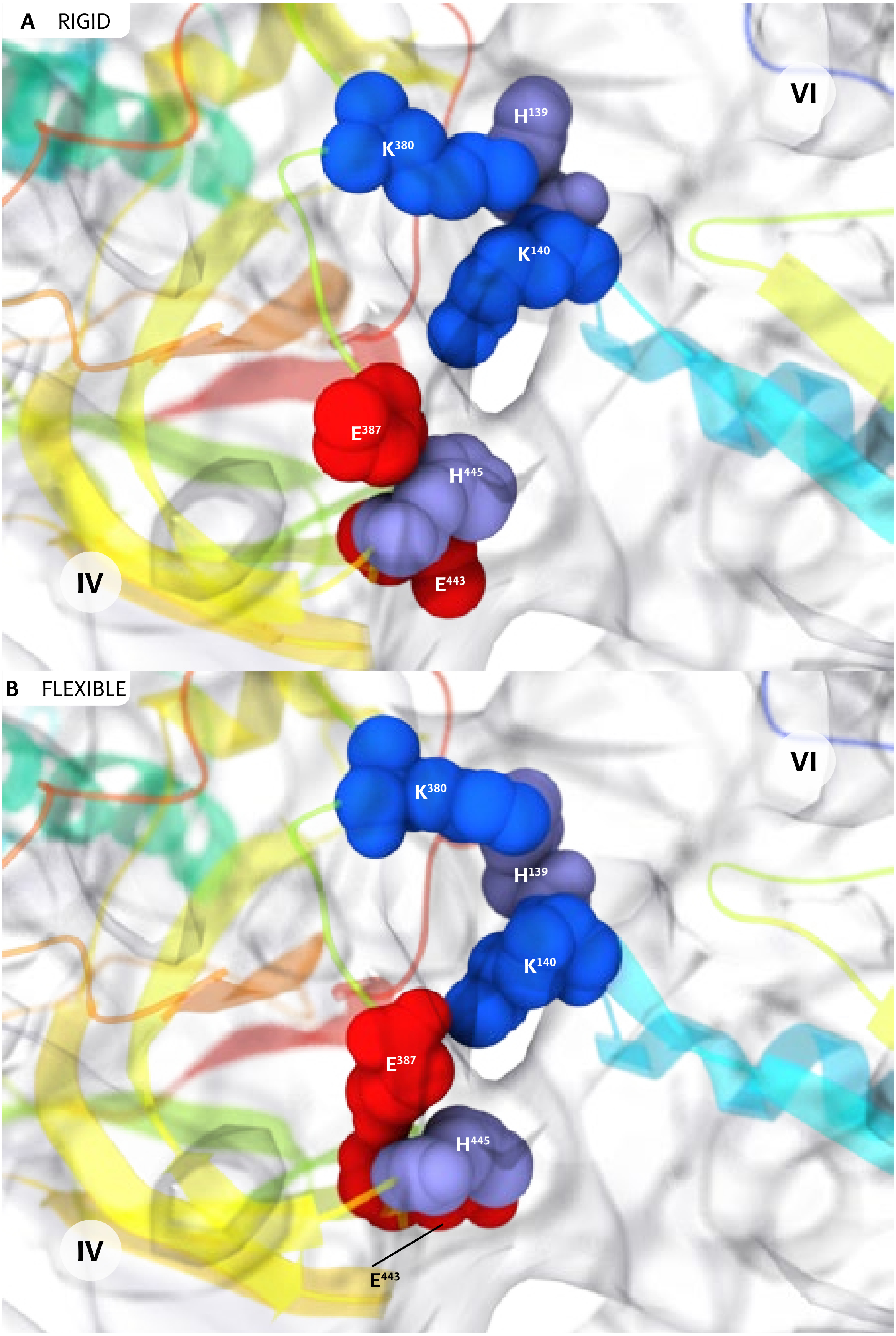


Fig. 112: The IV–VI interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay.

4.6.1.3 The II–IV interface

The II–IV interface consists of two separate contact points (Fig. 113). The first contact point, consisting of the N-terminal part of subunit II (domain #1) and the β3A→β3B loop in domain #3 of subunit IV, is compressed slightly by the flexible fitting. This results in a clash between T1 and D4 with K398.

The second contact point is opened up slightly, pulling the α-1.7→ β-1B loop away from the interface. The postulated salt bridge between E140 and R476 (Martin et al., 2007) in the β-3E→α-3.2 loop seems unlikely in the present model due a distance of ~14 Å.

Subunit	Residue	RMSD (Å)
II	T1	1,668
	L2	1,245
	H3	3,982
	D4	0,774
	H11	0,699
	F70	1,131
	E137	0,902
	F138	2,345
	D139	1,549
	E140	1,814
	H522	1,189
	H529	3,304

Subunit	Residue	RMSD (Å)
IV	K398	0,593
	H404	0,613
	F406	2,395
	H433	1,098
	F436	0,857
	R468	4,624
	R602	1,569
	T603	2,071
	P604	2,025
	S605	1,719
	Q606	1,650
	F621	4,382
	H623	2,422

Table 7: RMSD values of all participating residues of interface IV–VI. The values were calculated comparing all atoms of each residue individually.

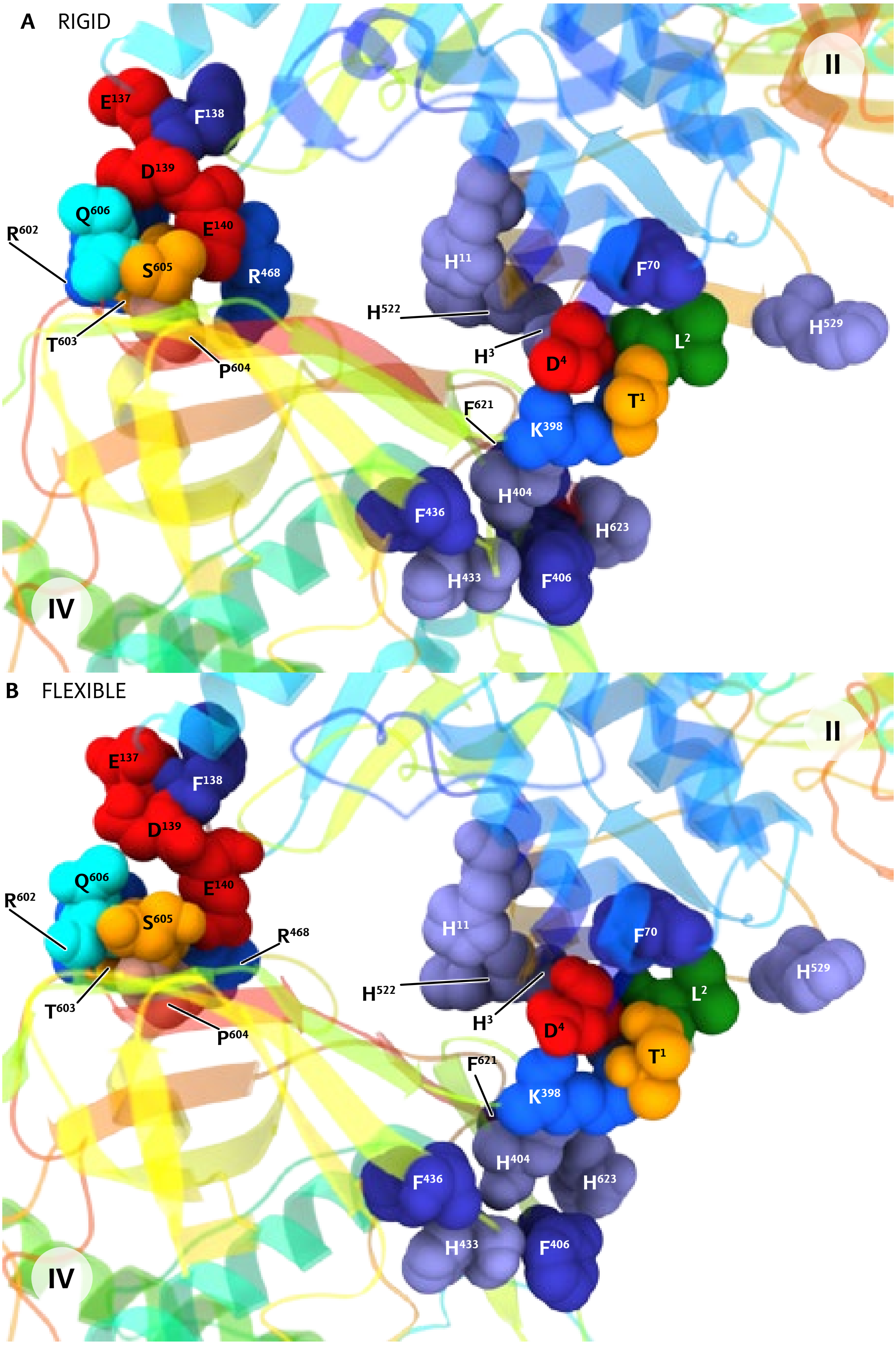


Fig. 113: The II–IV interface. Panel A: Rigid body fitting model. Panel B: Flexible fitting result.

4.6.1.4 The V–VI interface

Some clashing is present in the V–VI interface, both in the rigid body fitting and flexible fitting results (Fig. 114 A). Because the authentic sequence of subunit V is not known, the sequence of *Carcinoscorpius rotundicauda* was used instead. Also at the present resolution the reconstructed density map can not be used to gain further insight into the structure of the interface.

As previously described in Martin et al., 2007, a potential contact is the α-helix α1.5 (Fig. 115). Both helices come together in close proximity, giving rise to a number of contacts. A pair of histidine residues resides within a cavity at the center of the interface (H80/81, Fig. 114 B). This makes the interface a likely candidate after the II–II interface for allosteric interaction. A second contact area is possible at the loops following helix α1.1 (Fig. 115). Again, clashing can be observed in this area (Fig. 114 A). RMSD values are omitted since the interface remains ambiguous until a valid sequence of subunit V is available.

Two residues are changed in the new molecular model: H31 is substituted by a tyrosine residue and P35 is replaced by a histidine residue.

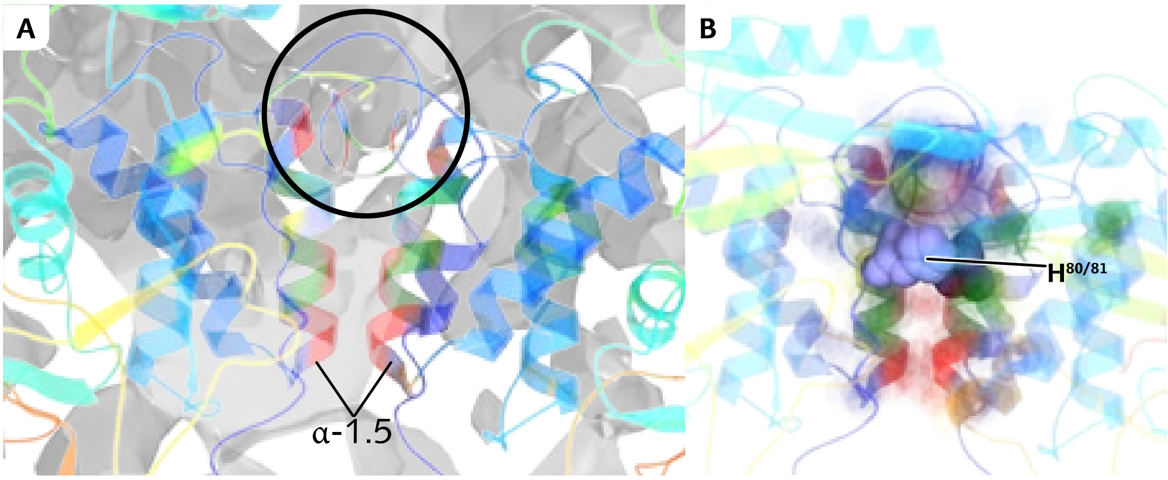
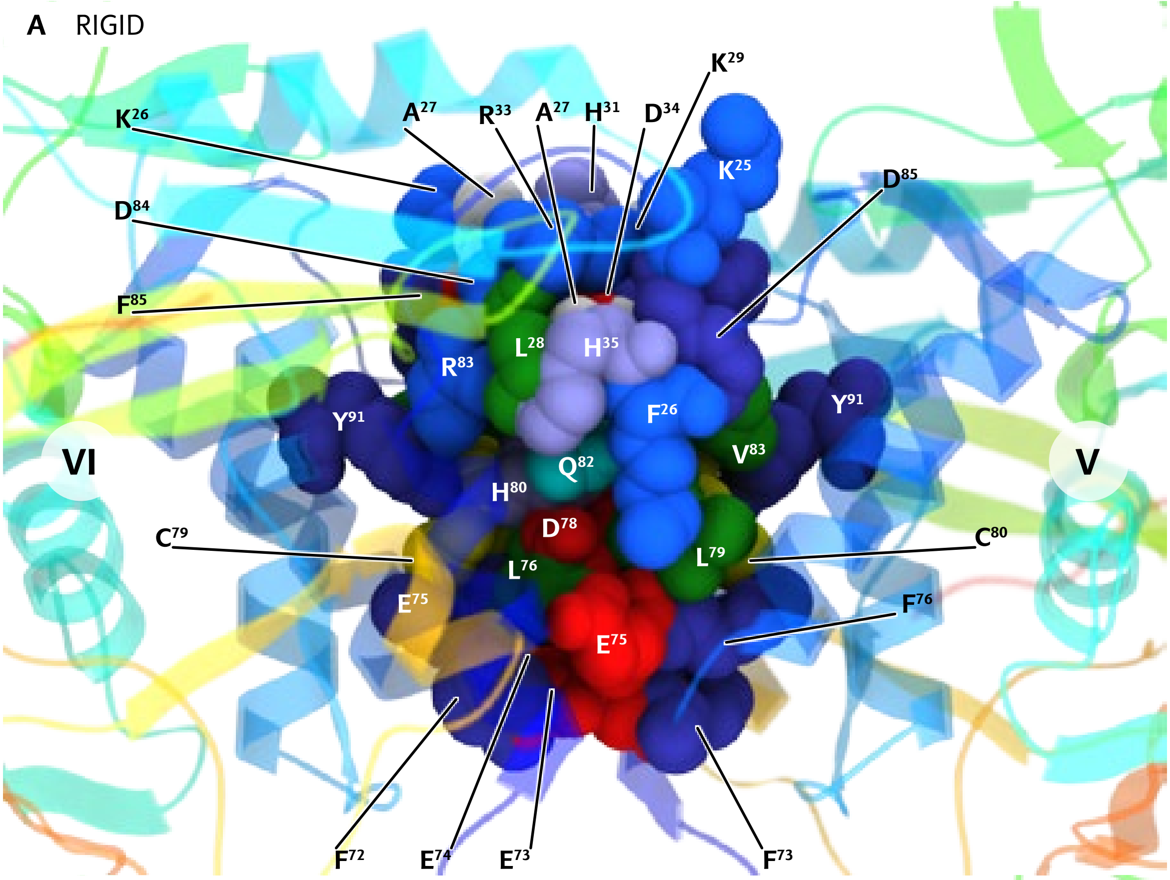


Fig. 114: The V–VI interface, cartoon depiction with reconstructed density map. Panel A: The interfacing α-helix 1.5 are in the center, clashing/crossing strands marked with circle. The density map is partly displayed with a slice sized similar to the depth of the interface. Panel B: Location of the interfaces central histidine residue pair.

A RIGID



B FLEXIBLE

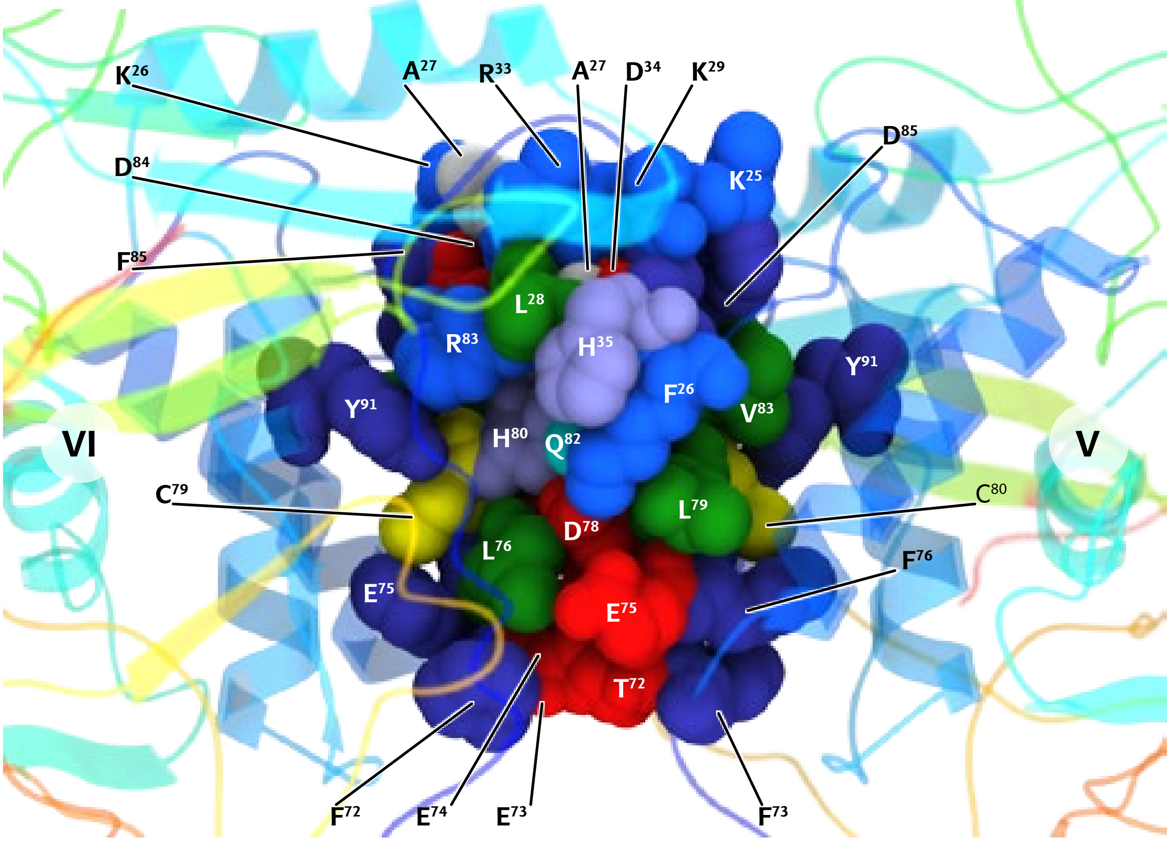


Fig. 115: The V–VI interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay.

4.6.2 Bridges between the two 2x6mers of the 4x6mer

Three types of interface between the two 2x6mers have been described (Fig. 116; Martin et al., 2007): two equivalent VI–IIIB/IV/V interfaces at the flip face, two equivalent V–VI interfaces and a small V–V interface, both at the flop face.

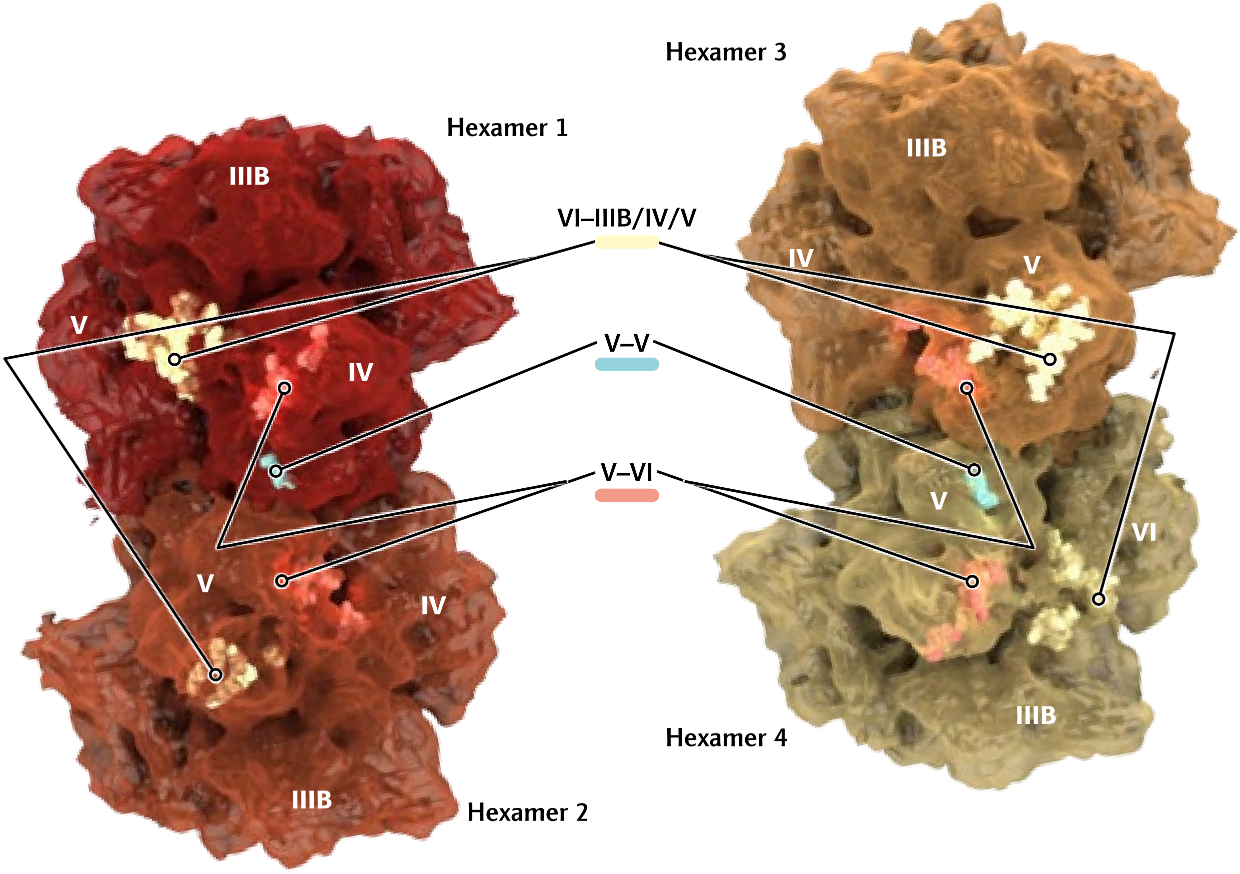


Fig. 116: Interface location between the hexamers of the 4x6mer. The hexamers are split at their plane of contact and rotated by 90° towards the viewer, thus exposing the contact surfaces.

4.6.2.1 The V–V interface

At the V–V interface interaction between the N-termini can be observed (Fig. 117). The bridge is visible in the reconstruction at a mass correlated threshold level. The flexible fitting increases the gap between the components by as much as 1,7 Å (α-carbon backbone distance. Residue RMSD see Table 8). Pair-wise flexible fitting is currently not supported by the available software. The need to use segmented single subunits for fitting results in loops being pulled further into the reconstructed volume. In the case of the V–V interface this deteriorates the quality of the molecular model.

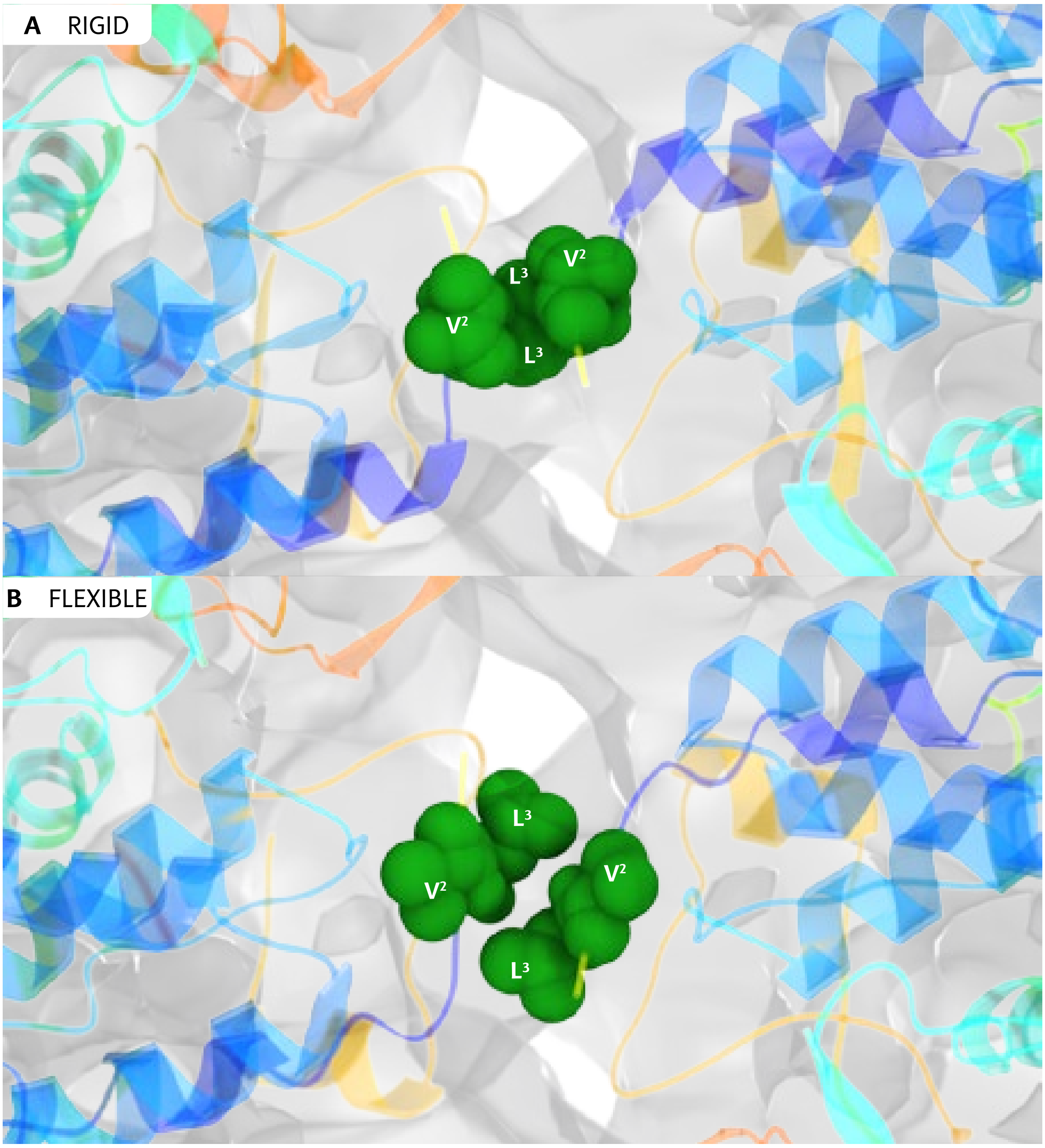


Fig. 117: The V–V interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay.

Subunit	Residue	RMSD (Å)
V	V2	2,354
	L3	2,360

Table 8: RMSD values of all participating residues of interface V–V. The values were calculated comparing all atoms of each residue individually.

4.6.2.2 The V–VI interface

The V–VI interface consist of two adjacent contact points (Fig. 118). The first contact occurs between the β2D→β2E loop of subunit VI and β3D→β3E loop of subunit V.

Although the sequence for subunit V has been switched to *C. rotundicauda* in the areas affecting the interface the residues are largely identical. 452TSK454 is substituted by 452SNR454, the second contact remains unchanged. Threonine 452 is substituted by a serine residue which only differs in that serine contains a hydrogen group instead of the methyl group in threonine (Betts & Russell, 2003). Serine 453 is replaced by a threonine. Lastly the lysine residue at position 454 is replaced by an arginine residue. Both changes are considered to be favored substitutions and functionally equivalent (Betts & Russell, 2003). Hydrostatic interaction can thus be confirmed for this section of the interface (Martin et al., 2007).

For the second contact the loop following helix α3.5 in subunit V meets the α2.2→α2.3 loop of subunit IV. Some clashes occur in the flexible fitting result. In particular K192 intersects with D572 and G571. Nevertheless with the authentic sequences applied an interaction between K192 and D572 is very likely. The contact is framed by two histidine residues (H567 and H573).

Subunit	Residue	RMSD (Å)
V	S453	1,654
	N454	1,810
	R455	3,176
	H567	1,669
	D568	0,847
	Q569	1,656
	V570	1,427
	G571	0,606
	D572	1,206
	H573	2,414

Subunit	Residue	RMSD (Å)
VI	E188	2,990
	V189	1,349
	I190	2,542
	G191	1,418
	K192	2,677
	O193	1,400
	D288	1,209
	E289	2,730
	N290	2,837

Table 9: RMSD values of all participating residues of interface V–VI. The values were calculated comparing all atoms of each residue individually.

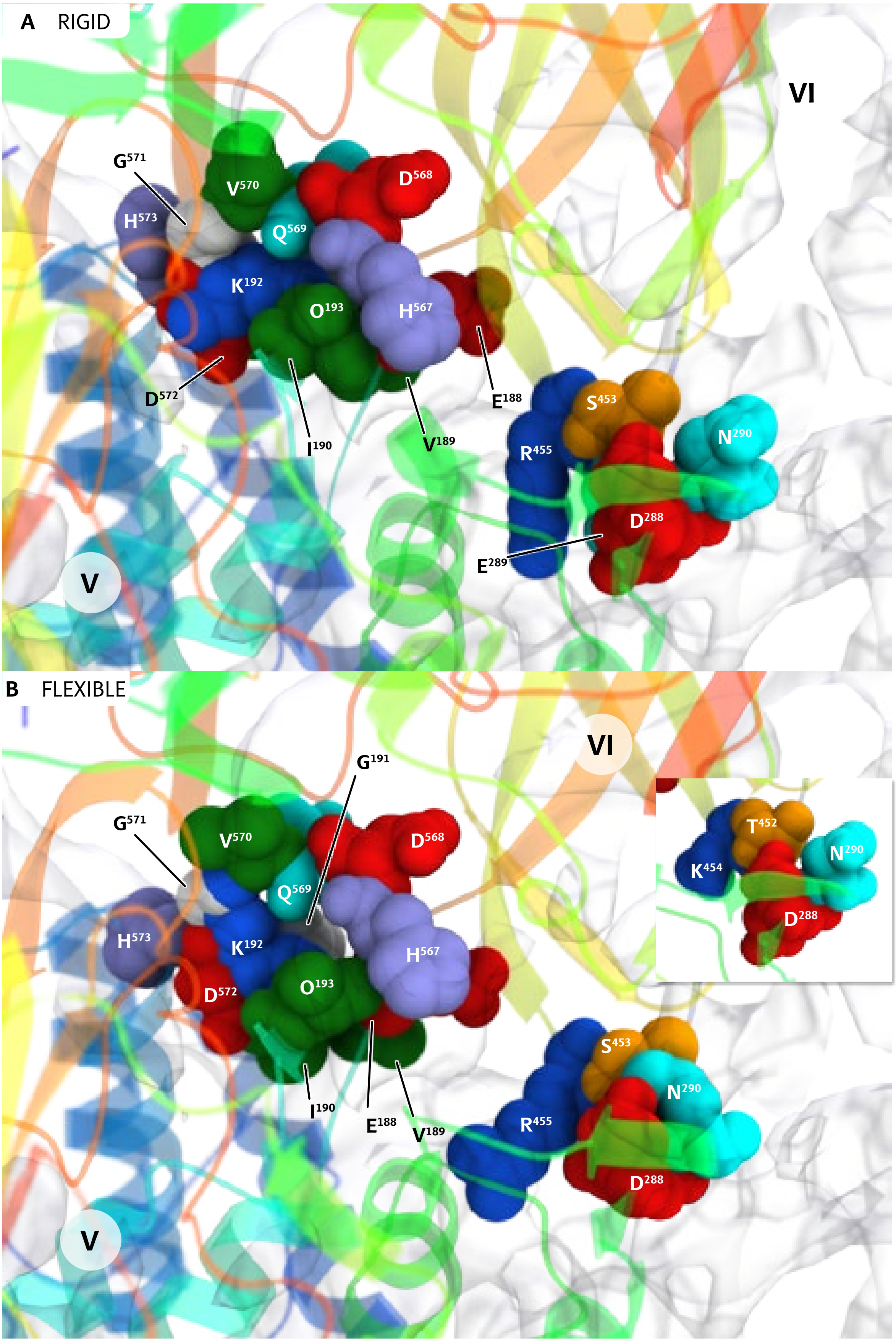


Fig. 118: The V–VI interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay. Inset: β2D→β2E loop with *Eurypelma* sequence from Martin et al., 2007.

4.6.2.3 The VI–IIIB/IV/V interface

Sequences from *C. rotundicauda* have been used for subunit V instead of *E. californicum* in Martin et al., 2007. For subunit IIIB a genuine *L. polyphemus* sequence has been used. There are some sequence changes relevant to the interface from subunit V (Fig. 119): 290NGS292 is replaced by NHH. A change from glycine to histidine is not favored. Within the context of an histidine-rich interface however this change is unlikely to prove detrimental to the function of the interface. The second change can be interpreted in a similar fashion: a serine residue is replaced by another histidine – again a disfavored substitution, albeit one not likely to impact the function of the interface. Lastly histidine H293 is replaced by glutamic acid E294, generally evaluated as a neutral change (Betts & Russell, 2003). In addition the *C. interruptus* sequence contains a histidine residue at position H293 which might serve a similar purpose.

The molecular model with the authentic sequence of subunit IIIB reveals additional changes to the interface: 288EGQE291 to HGHE (Fig. 121). The aspartic acid residue at position 288 changes to a histidine residue – a neutral change in the sense that a charged residue remains. Glutamine Q290 is replaced by another histidine residue, also considered to be a neutral change (Betts & Russell, 2003). Two more notable substitutions occur: H283 is replaced by L283, a highly disfavored change; and H287 by E287, a neutral change. It should be mentioned that there is also a neighboring histidine residue with H288 in the authentic sequence. Table 10 provides RMSD values for all contributing residues.

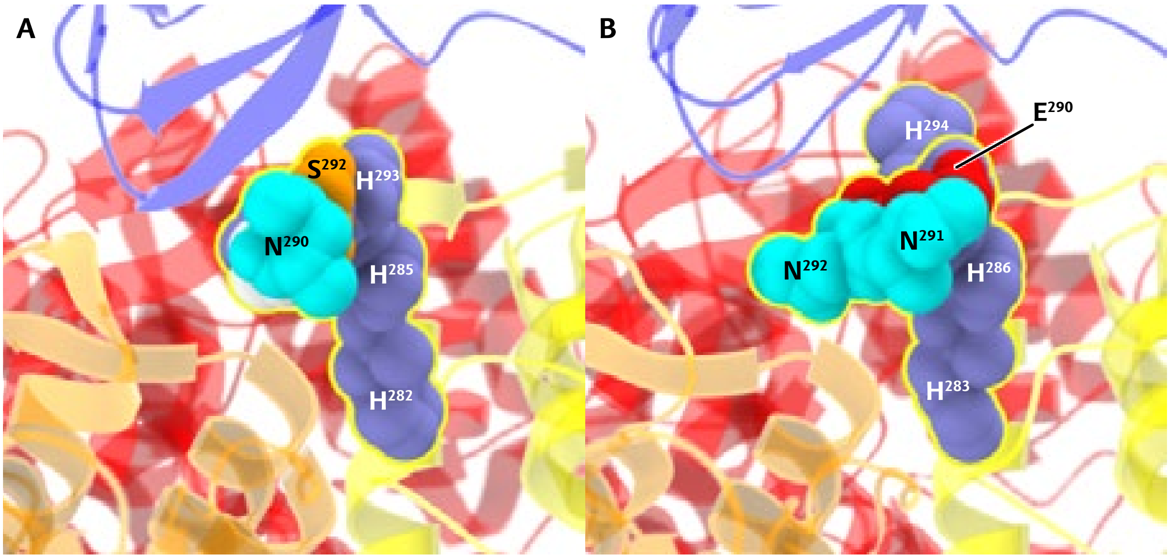


Fig. 119: Changes to the VI–IIIB/IV/V interface – new molecular model of subunit V. Panel A: Original *E. californicum* molecular model. Panel B: Revised molecular model with *C. rotundicauda* sequence and additional histidine residue H294.

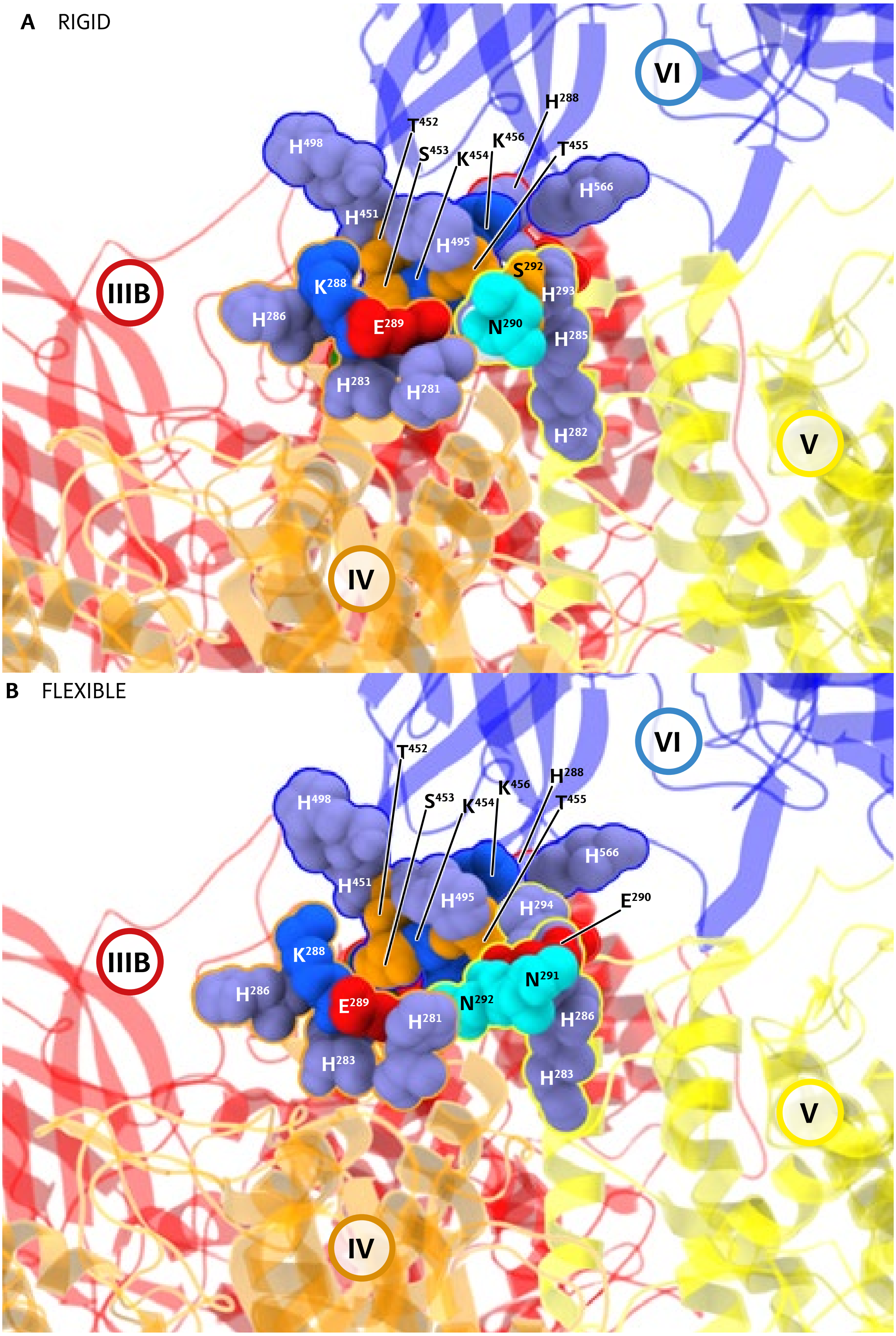


Fig. 120: The VI–IIIB/IV/V interface. Subunit secondary structure is colored uniformly for each subunit: VI: blue, IIIB: red, IV: orange, V: yellow Panel A: Rigid body fitting model. Additionally contributing residues are marked with a colored outline corresponding to their originating subunit. Panel B: Flexible fitting result.

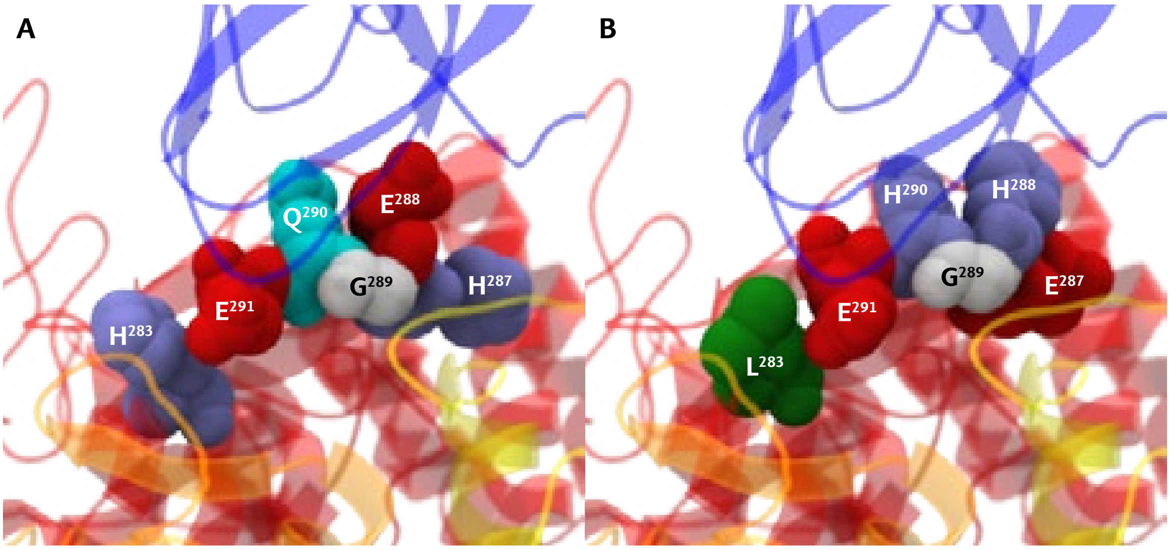


Fig. 121: Changes to the VI–IIIB/IV/V interface – new molecular model of subunit IIIB. A: Original *E. californicum* molecular model. Panel B: Revised molecular model with *C. rotundicauda* sequence.

Subunit	Residue	RMSD (Å)
IIIB	L283	1,883
	E287	2,143
	H288	4,906
	G289	2,071
	H290	2,613
	E291	2,543

Subunit	Residue	RMSD (Å)
V	E290	0,641
	N291	2,329
	N292	1,354
	I282	0,684
	G285	0,550
	H293	0,811

Subunit	Residue	RMSD (Å)
IV	K288	1,181
	E289	2,093
	H281	3,387
	H283	2,529
	H286	2,517

Subunit	Residue	RMSD (Å)
VI	T452	1,418
	S453	1,147
	K454	1,544
	T455	1,096
	K456	1,605
	H451	2,226
	H495	1,549
	H498	1,343
	H566	0,906

Table 10: RMSD values of all participating residues of interface VI–IIIB/IV/V. The values were calculated comparing all atoms of each residue individually.

4.6.3 Bridges between the two 4x6mers of the 8x6mer

Focusing on bridges connecting the two 4x6mers of the 8x6mer, four different types of interface were previously described (Fig. 122; Martin et al., 2007): The two central IV–IV bridges; and located around that central bridge four II–IV bridges, four IIIA–IIIB bridges and two IIIA–IIIA interfaces.

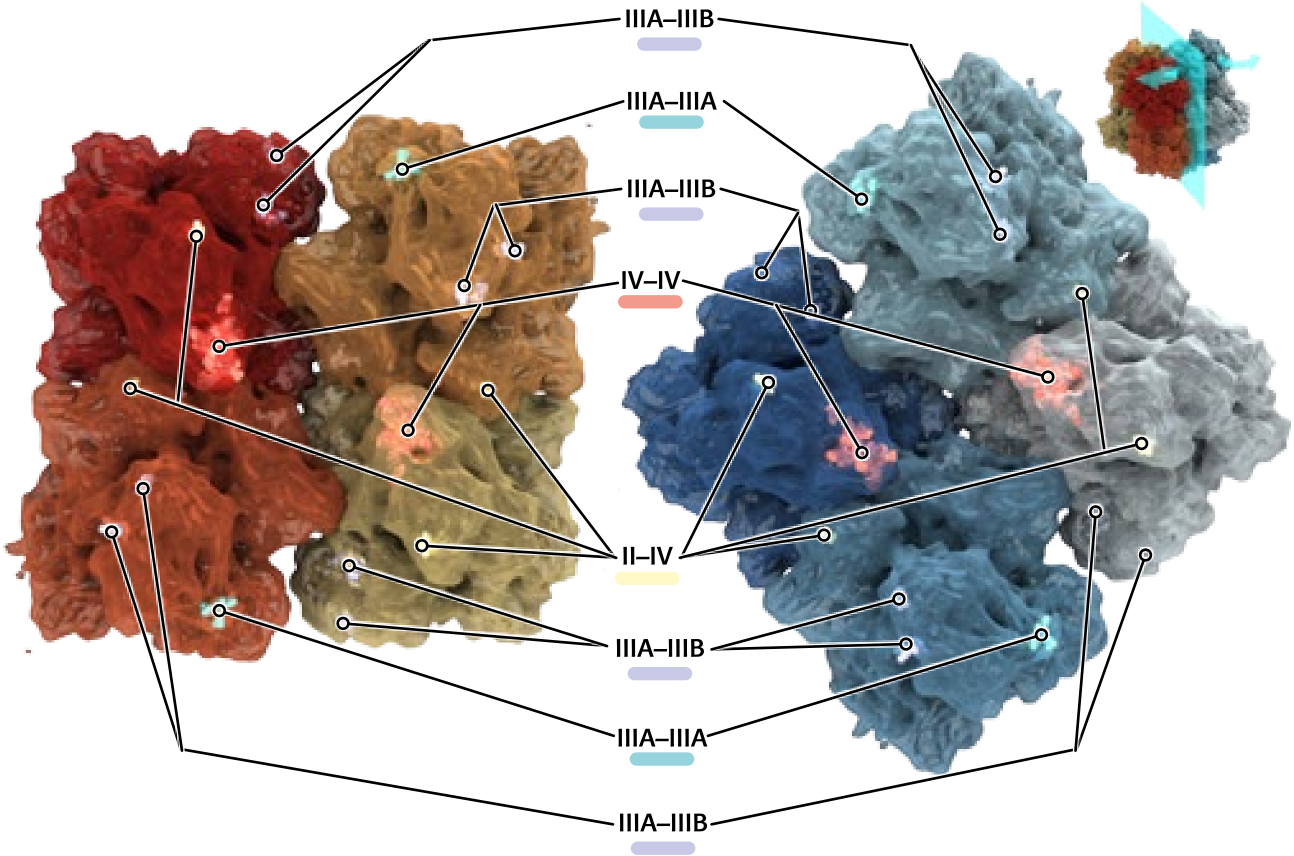


Fig. 122: Interface location between the hexamers of the 4x6mer. The hexamers are split at their plane of contact (inset in upper right corner) and rotated by 90° towards the viewer, thus exposing the contact surfaces.

4.6.3.1 The IIIA–IIIA interface

The IIIA–IIIA interface is formed between domains #1 and #3 (Fig. 123; Martin et al., 2007). It is formed by two copies of 136DK137 in the α1.7→β1B loop, which suggests the formation of salt bridges. The accompanying histidine residues form a ring with a wide opening suitable for a ligand. In both the rigid-body and flexible fitting models there is some slight overlapping of the 136DK137 residues. Overall the geometry of the interface stays intact with very modest changes (Table 11). The described chamber is somewhat smaller in the flexible fitting model due to the movement of the aspartic acid residues towards the histidine ring.

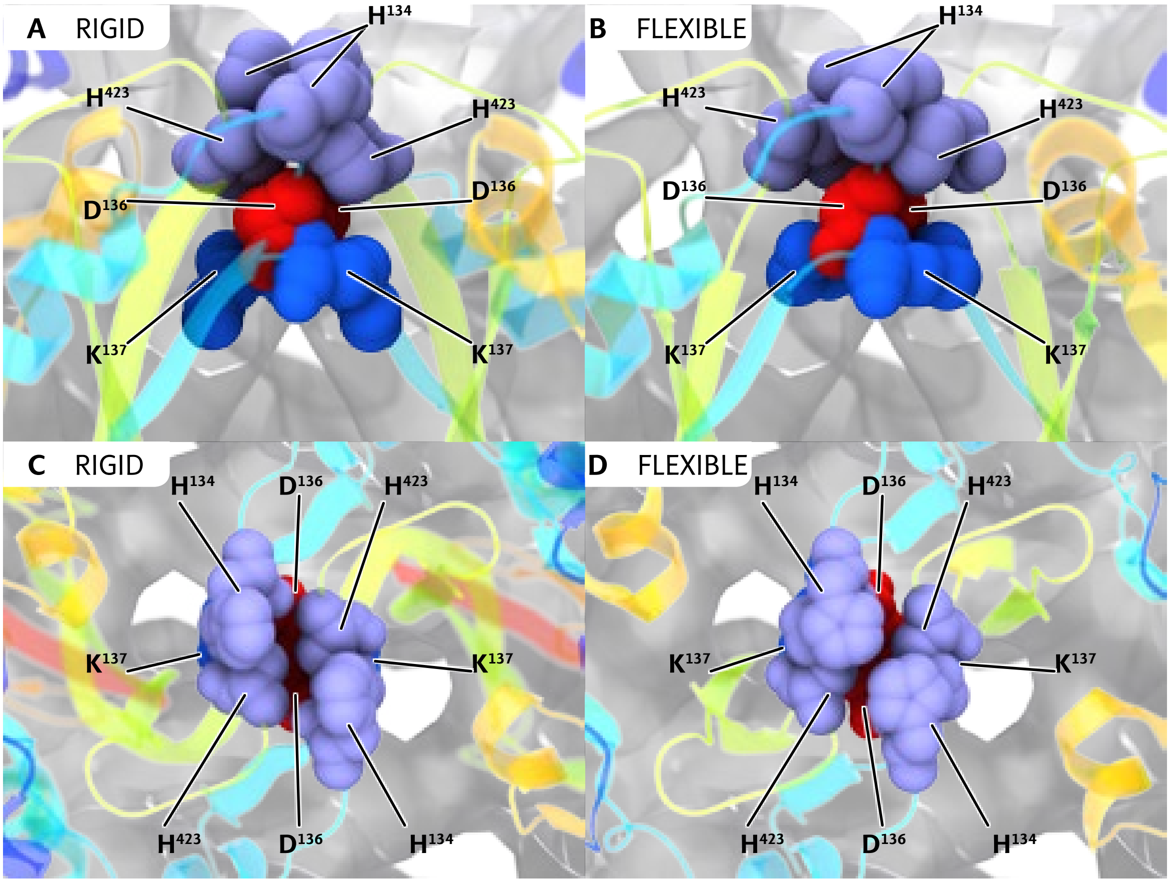


Fig. 123: The IIIA–IIIA interface. Panel A: side view of rigid body fitting model with reconstructed density map overlay. Panel B: side view of flexible fitting result with reconstructed density map overlay. Panel C: top view of rigid body fitting model with reconstructed density map overlay. Panel D: top view of flexible fitting result with reconstructed density map overlay.

Subunit	Residue	RMSD (Å)
IIIA	H423	1,240
	H134	1,645
	D136	2,016
	K137	0,918

Table 11: RMSD values of all participating residues of interface IIIA–IIIA. The values were calculated comparing all atoms of each residue individually.

4.6.3.2 The IIIA–IIIB interface

This interface has been described by Martin et al., 2007, with a sequence from *E. californicum* used for subunit IIIB. In the meantime the authentic sequence has been published (Rehm et al., 2012). The first linkage is described between both domains #3. Subunit IIIA contributes K576 in the loop downstream from helix α3.5. Its counterpart is D450, located in the β3D→β3E loop (Fig. 125A). However the aspartatic acid at position 450 is not present in the *L. polyphemus* sequence. It is replaced by a leucine residue (Fig. 124). The possibility for a hydrostatic bond is thus eliminated. This is considered to be a disfavored substitution (Betts & Russell, 2003) Furthermore there are no other charged amino acids in the near vicinity of L450 that might form such a bond.

A second contact might be possible between domain #3 of subunit IIIA and domain #2 of subunit IIIB. The β3D→β3E loop of subunit IIIB contributes 447SGS449 to E297 in the β2E→α2.4 loop (Fig. 125A). Position 297 is changed to a proline in the new model (Fig. 125B), again a disfavored substitution. However at position 296 the authentic sequence does contain a glutamic acid residue which might keep the contact at least partly working as originally described. The distance to S447 is 5,7 Å in the current model. The distance to S449 is 7,2 Å. For unchanged residues RMSD values between rigid-body and flexible fittings are provided in Table 12.

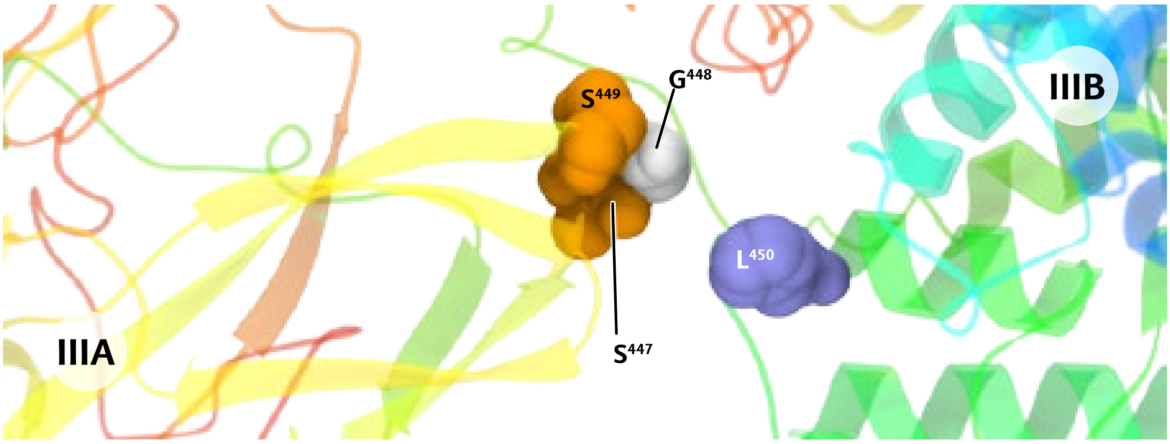


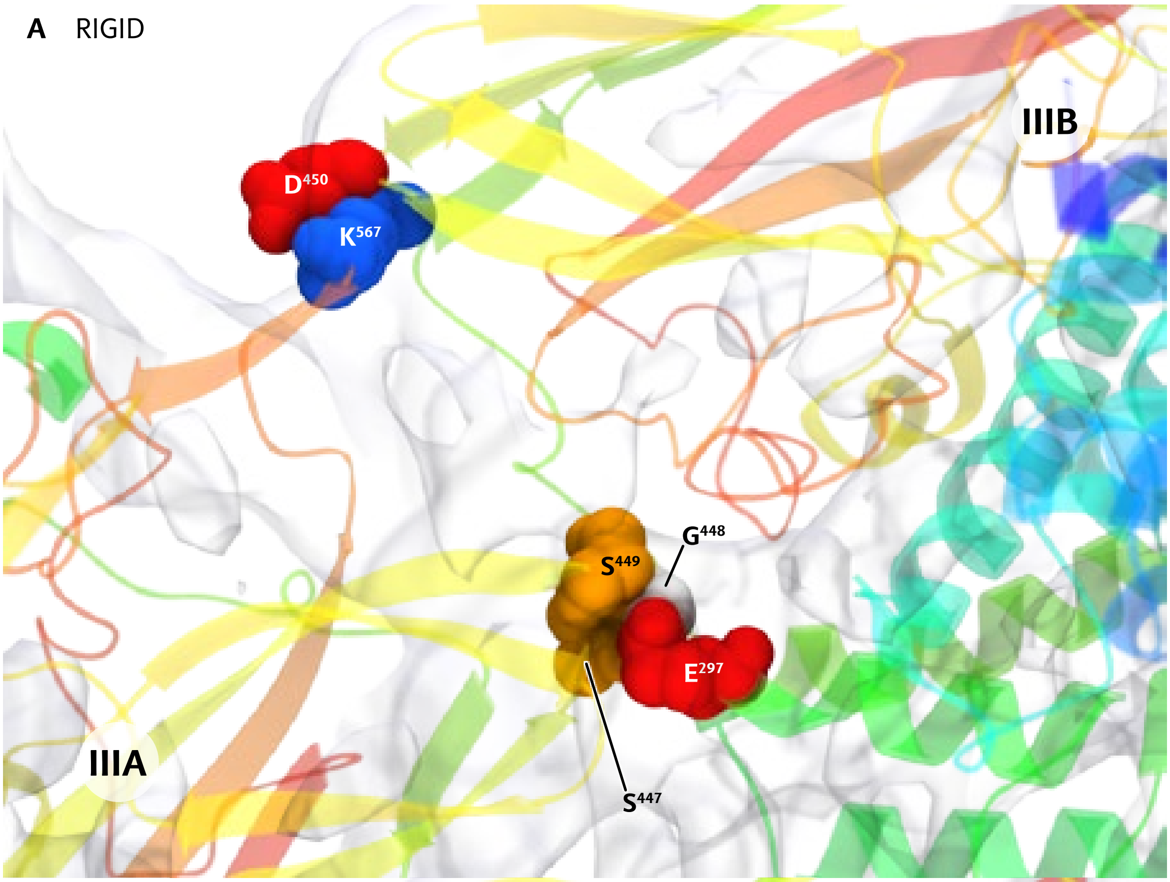
Fig. 124: The IIIA–IIIB interface. Rigid body fitting with authentic subunit IIIB sequence in place.

Subunit	Residue	RMSD (Å)
IIIA	S447	0,918
	G448	1,151
	S449	1,037
	K567	3,295

Subunit	Residue	RMSD (Å)
IIIB	P297	2,386
	L450	2,554

Table 12: RMSD values of all participating residues of interface IIIA–IIIB. The values were calculated comparing all atoms of each residue individually.

A RIGID



B FLEXIBLE

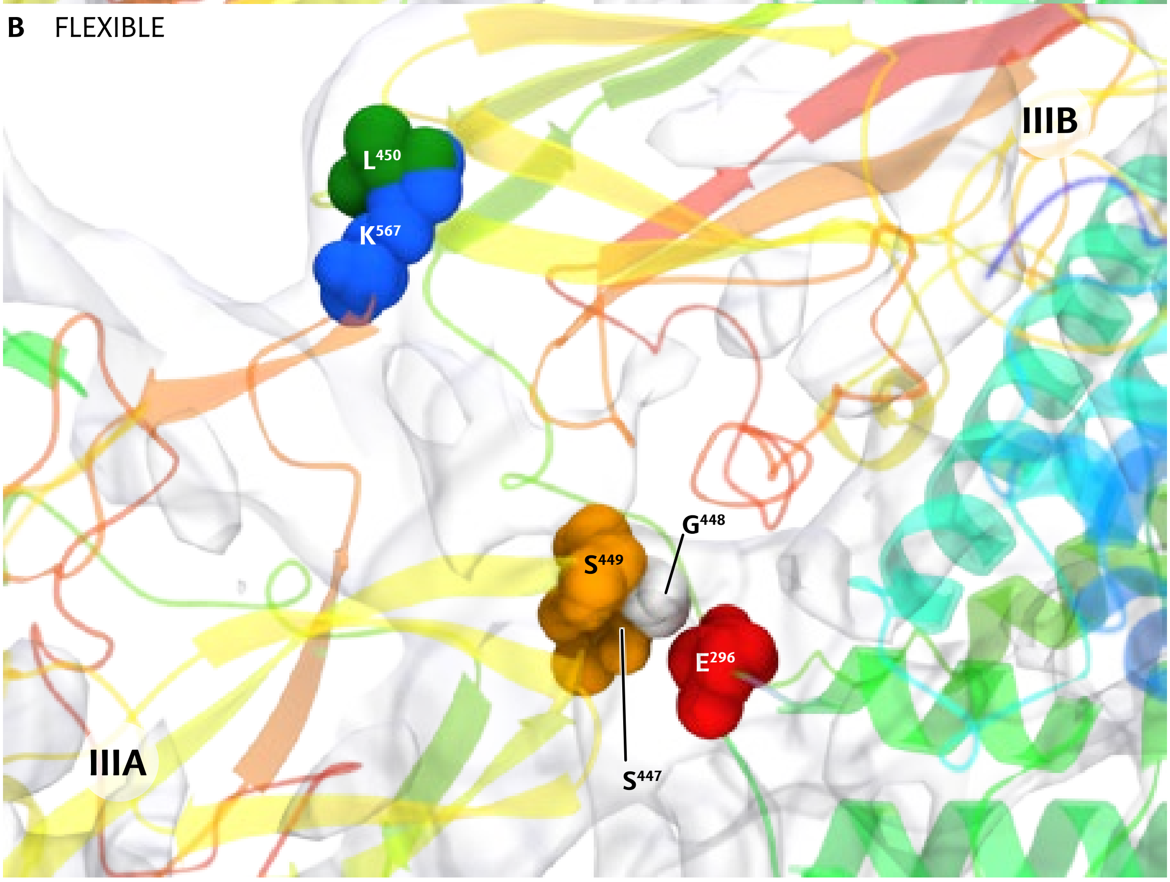


Fig. 125: The IIIA–IIIB interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay.

4.6.3.3 The II–IV interface

Both subunits α1.4→α1.5 loop forms the type II–IV interface (Fig. 126 A and B, Martin et al., 2007). Based on backbone distances interaction at this location seems likely. The lysine residue however remains to be oriented away from the aspartate residue. After flexible fitting the backbones are brought closer together by ~0,5 Å.

Subunit	Residue	RMSD (Å)
II	D69	1,272
IV	K67	1,869

Table 13: RMSD values of all participating residues of interface II–IV. The values were calculated comparing all atoms of each residue individually.

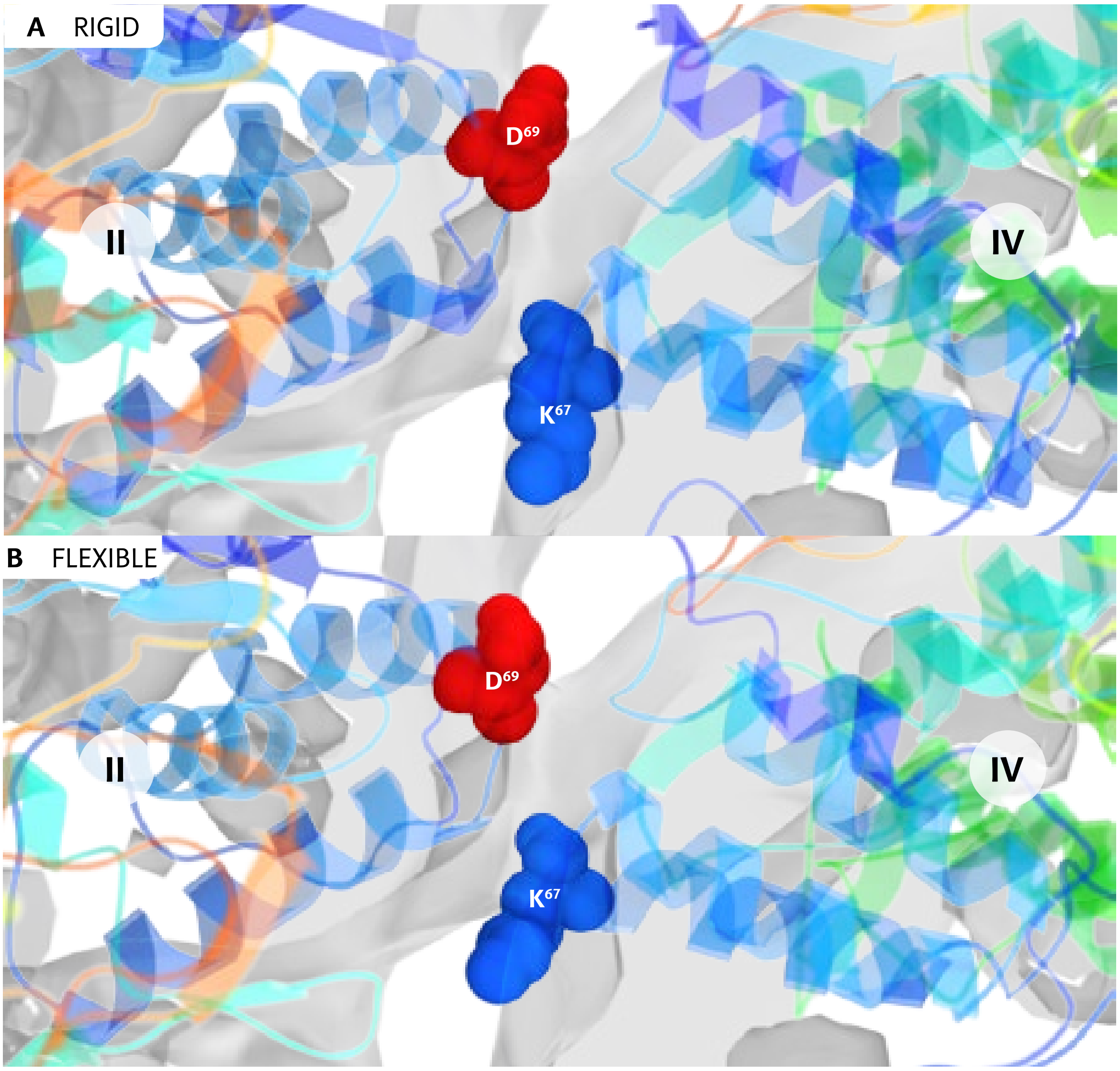


Fig. 126: The II–IV interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay.

4.6.3.4 The IV–IV interface

As described by Martin et al., 2007, this interface involves several contacts between the domain #3 of two type IV subunits (Fig. 128). It involves the β3D→β3E loop, the β3G and β3H strands as well as residues from α-helix 3.5 and the following loop. The secondary structure elements of the interface are arranged in an anti-parallel manner. There are some minor clashes in the model, H451 has minor contacts with W558, H485 and V449. A small intersection with the backbone occurs for H562.

The interface is altered slightly by the flexible fitting process. At the center two hydrophobic valine residues (V449, Fig. 127 A and B) are brought together. Hydrostatic interaction is likely to occur between the residues surrounding this core (D447, T448, K450). In the periphery bonding is likely to occur between D447 and K490 as well as H491 and E566. In this part of the contact E566 exhibits a rotation after flexible fitting away from its potential binding residue (Fig. 127 B). Indeed the RMSD value for E566 is unusually high with 5,061 Å (see Table 14 for a complete list of RMSD values).

The second part of the interface is made up of two tryptophan residues W558 (Fig. 127 C and D). They are flanked by four histidine residues (H451, H485, H489 and H562). The flexible fitting opens up this histidine ring, forming something more akin to a cleft as opposed to the configuration of the original model.

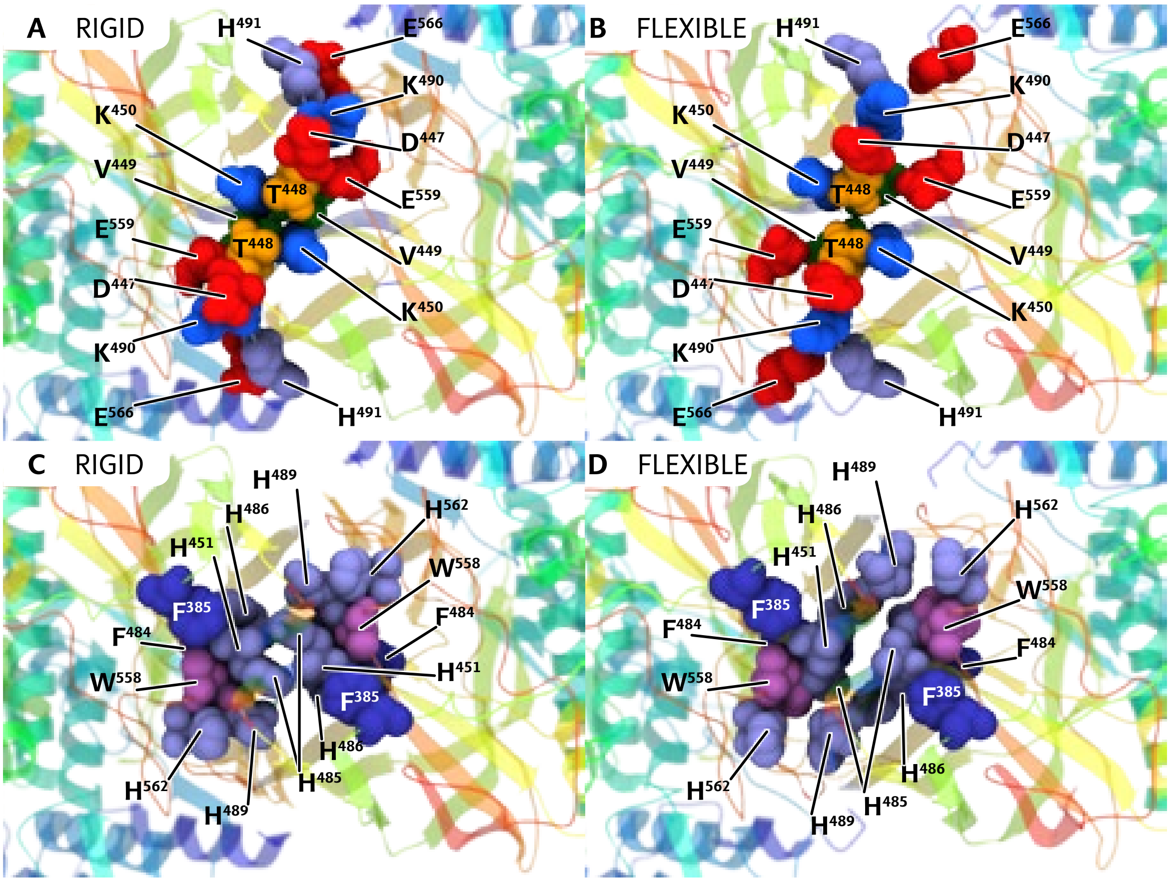


Fig. 127: The IV–IV interface. Panel A: part one of the interface – rigid body fitting model. Panel B: part one of the interface – flexible fitting result. Panel C: part two of the interface – rigid body fitting model. Panel D: part two of the interface – flexible fitting result.

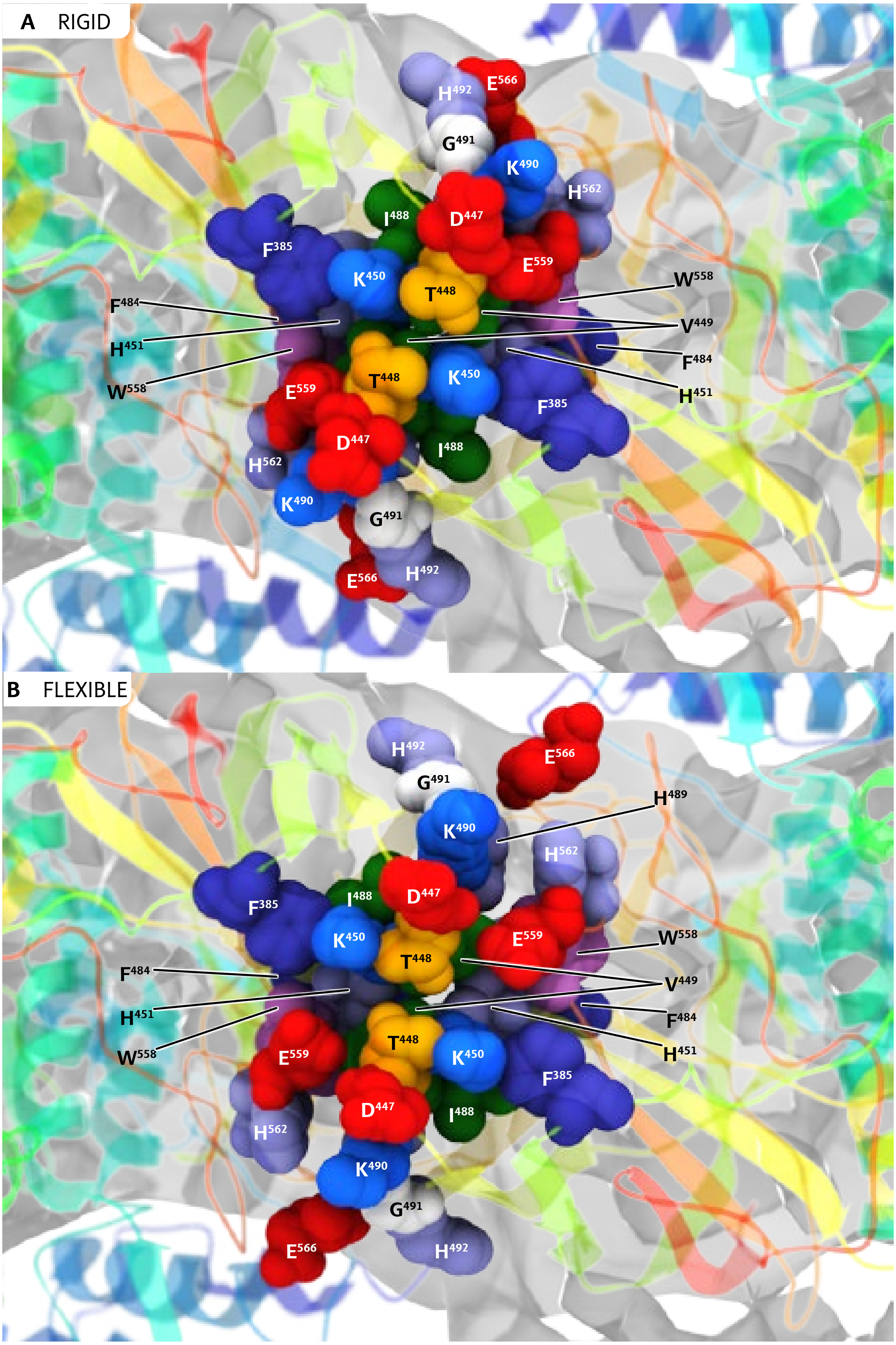


Fig. 128: The IV–IV interface. Panel A: Rigid body fitting model with reconstructed density map overlay. Panel B: Flexible fitting result with reconstructed density map overlay.

Subunit	Residue	RMSD (Å)
IV	F385	1,296
	D447	2,993
	T448	1,816
	V449	1,259
	K450	1,812
	H451	1,593
	F484	2,995
	H485	1,889
	H486	3,484
	G487	0,957
IV	I488	2,321
cont'd	H489	1,604
	K490	3,144
	G491	1,864
	H492	2,243
	W558	1,075
	E559	1,316
	H562	2,311
	E566	5,061

Table 14: RMSD values of all participating residues of interface IV–IV. The values were calculated comparing all atoms of each residue individually.

4.6.4 Flexible fitting summary

The process of flexible fitting has brought improvements to the molecular model. However it is not without trade offs. In general, for each subunit the molecular model was contracted and pulled slightly further into the reconstructed volume. Despite measures being taken to preserve existing secondary structure elements, in some parts distortions to α-helices and β-sheets could be observed. It should be mentioned that an additional energy minimization step was necessary to process the raw flexible fitting results. Even then, common quality indicators such as Z-scores only barely reached acceptable levels. Also of concern is the active site, which stayed largely intact. Improvements were usually noted in the fitting of α-helical components of domain #1. Less frequently in domain #2 α-helices were shifted into their respective volumes within the reconstruction. The β2D→β2E region was moved several times. This may be due to its exposed location at the periphery of the molecule.

Many α-helical elements are clearly resolved in the reconstruction, visible as columns or pillars. This provides a clear target for the fitting algorithm. The β-sheet rich domain #3 is less well resolved and thus provides less well defined structures as fitting targets. More commonly exposed areas at the periphery of the subunit were shifted. Well defined β-sheets require a resolution of 5 Å or better (Amunts et al., 2014) which is the limiting factor in this case.

To assess the fitting quality for loops, a second limitation needs to be taken into account: It stems from the need to segment the reconstruction into individual subunits. As each subunit is processed by itself, influences from neighboring subunits cannot be taken into account. This might be a drawback for the analysis of inter-hexamer contacts. As previously stated the flexible fitting algorithm was aware of secondary structure elements and prohibited from distorting them. Loop movement was far less restricted, leading to loops being pulled further into the reconstructed volume. In some cases, this effect proved detrimental to interfaces as contributing components were moved further apart. In part this effect is influenced by the size of the segmented volume that is used for the fitting algorithm.

An ideal approach would combine a high resolution 3D-reconstruction with the ability to process the entire molecule in one flexible fitting session. This is currently not supported by the available flexible fitting software.

D Summary

This work presents 3D-reconstructions of two arthropod hemocyanin molecules produced by single particle cryo-EM. The 3D-reconstruction of the 8x6mer hemocyanin of the horseshoe crab *Limulus polyphemus* was built upon a newly acquired dataset under oxygenated conditions and improved the previous results. The achieved resolution of 8,7 Å (0,5 criterion) allowed for further refinement of the existing molecular model by flexible fitting into the 3D density map. Also newly available additional authentic sequences of *Limulus polyphemus* hemocyanin as well as a more suitable source for homology models of subunits with unknown sequence remaining were used to update the molecular model and verify previously identified interface sites and in turn molecular mechanisms of cooperativity by chemo-mechanical signal transduction.

Compared to a conventional rigid body fitting, the secondary structure aware flexible fitting algorithm especially improved the overall fit in domain #1 of each subunit, which contains mainly well resolved α-helices. Domain #2 also contains α-helices as well as the active site. Changes in this region were generally less frequent with the active site left intact. Domain #3 is mainly comprised of β-sheets which lead to few discernible improvements as the observed level of detail contained within the reconstruction was insufficient to provide clear target masses for the fitting algorithm.

Secondly both coarse and fine structural details of the 4x6mer hemocyanin of the ghost shrimp *Callianassa truncata* were revealed by an unbiased reference-free reconstruction method yielding a resolution of 9,8 Å (0,5 criterion). In combination with available sequence data, a molecular model of the 4x6mer was constructed. This revealed a total of twelve contact points facilitated by four distinct types of interface. Most contacts are densely packed in the center of the 4x6mer.

The interfaces are the sites of interaction between the hexamers, both in terms of stabilization of the quaternary structure of the molecule as well as possible ways for chemo-mechanical signal transduction. Overall there are three subunits involved in contacts with subunit b of type *Ctr*Hc2 participating in every contact but one. This indicates its importance to the assembly of the 2x6mers and 4x6mer. Also, in combination with published low-resolution 3D-reconstrcutions of crustacean 2x6mers, insight into the evolutionary development of crustacean hemocyanin was gained.

E Zusammenfassung

Die vorgelegte Arbeit zeigt zwei 3D-Rekonstruktionen von Arthropoden-Hämocyaninen mittels Einzelpartikelanalyse nach Cryo-Elektronenmikroskopie. Die 3D-Rekonstruktion des 8x6meren Hämocyanins des Pfeilschwanzkrebses *Limulus polyphemus* konnte mit einem neuen Datensatz unter oxygenierten Bedingungen frühere Ergebnisse verbessern. Die erreichte Auflösung lag bei 8,7 Å (0,5 Kriterium), was eine Verfeinerung des existierenden molekularen Modells durch flexibles Einpassen in die 3D-Dichtekarte ermöglichte. Das molekulare Modell wurde darüber hinaus mit zusätzlichen authentischen Sequenzdaten sowie besser geeigneten Sequenzvorlagen für Homologiemodelle der Untereinheiten mit unbekannter Sequenz auf den neuesten Stand gebracht. Die zuvor postulierten Mechanismen der Kooperativität durch chemo-mechanischen Signalübertragung konnten damit bestätigt werden.

Verglichen mit einem konventionellen rigidem Einpassen des molekularen Modells in die 3D-Dichtekarte konnte der Algorithmus für das flexible Einpassen, welcher Sekundärstrukturelemente berücksichtigte, die Übereinstimmung in Domäne #1 jeder Untereinheit verbessern. Diese Domäne besteht hauptsächlich aus gut aufgelösten α-Helices. Domäne #2, welche ebenfalls viele α-helikale Bestandteile sowie das aktive Zentrum enthält, wurde weniger häufig verändert. Dabei blieb das Sauerstoffbindungszentrum intakt. Domäne #3 besteht zum Großteil aus β-Faltblättern, welche in der 3D-Rekonstruktion nicht ausreichend Details boten, um dem Algorithmus klare Massen als Ziele für die flexible Einpassung zu bieten.

Zweitens konnten sowohl der makromolekulare Aufbau als auch strukturelle Details mit der 3D-Rekonstruktion des 4x6meren Hämocyanins von *Callianassa truncata* mittels einer nicht durch Vorgaben beeinflussten Rekonstruktionsmethode erforscht werden. Die 3D-Rekonstruktion erreichte eine Auflösung von 9,8 Å (0,5 Kriterium). In Kombination mit vorliegenden Sequenzdaten konnte ein molekulares Modell des 4x6mers erstellt werden. Insgesamt wurden zwölf Kontaktstellen, aufgeteilt auf vier verschiedene Kontaktarten, gefunden. Fast alle Kontakte sind dicht gepackt im Zentrum des Moleküls zu finden.

Diese Kontaktstellen sind für den Zusammenhalt der Quartärstruktur und möglicherweise ebenfalls für eine chemo-mechanische Signalübertragung verantwortlich. Drei Untereinheiten sind in Kontakte involviert, wobei die Untereinheit b vom Typ *Ctr*Hc2 an drei der vier Kontakttypen beteiligt ist. Dies ist ein Zeichen für ihre Wichtigkeit bei der Assemblierung der 2x6 und 4x6mere. In Kombination mit publizierten, niedrig aufgelösten 3D-Rekonstruktionen von Crustaceen-2x6meren konnten neue Erkenntnisse über die evolutionäre Entwicklung der Crustaceen-Hämocyanine gewonnen werden.

F Bibliography

Amunts A, Brown A, Bai XC, Llácer JL, Hussain T, Emsley P, Long F, Murshudov G, Scheres SH, Ramakrishnan V. (2014): Structure of the yeast mitochondrial large ribosomal subunit. Science. Mar 28;343(6178):1485-9.

Arnold K, Bordoli L, Kopp J, Schwede T. (2006): The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22:195-201.

Arnold, P. (2012): 3D-Electron microscopy of protein complexes of different size and symmetry. Dissertation Johannes-Gutenberg-University Mainz

Bailly X, Vanin S, Chabasse C, Mizuguchi K, Vinogradov SN. (2008): A phylogenomic profile of hemerythrins, the nonheme diiron binding respiratory proteins. BMC Evol Biol. Sep 2;8:244.

Baldwin PR, Penczek PA. (2007): The Transform Class in SPARX and EMAN2. J Struct Biol. Jan;157(1):250-61.

Betts MJ, Russell RB. (2003): Amino acid properties and consequences of subsitutions. In Bioinformatics for Geneticists, M.R. Barnes, I.C. Gray eds, Wiley

Brenowitz M, Bonaventura C, Bonaventura J, Gianazza E. (1981): Subunit composition of high molecular weight oligomer: Limulus polyphemus hemocyanin. Arch Biochem Biophys. Sep;210(2):748-61.

Burmester T, Scheller K. (1996): Common origin of arthropod tyrosinase, arthropod hemocyanin, insect hexamerin, and dipteran arylphorin receptor. J Mol Evol. Jun;42(6):713-28.

Burmester T. (1999): Identification, molecular cloning, and phylogenetic analysis of a non-respiratory pseudo-hemocyanin of Homarus americanus. The Journal of Biological Chemistry, 274(19), 13217-13222.

Burmester T. (2001): Molecular Evolution of the Arthropod Hemocyanin Superfamily. Mol Biol Evol, 18(2), 184-195.

Burmester T. (2002). Origin and evolution of arthropod hemocyanins and related proteins. Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology, 172(2), 95-107.

Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. (2010): MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. Jan;66(Pt 1):12-21.

Decker H, Hellmann N, Jaenicke E, Lieb B, Meissner U, Markl J. (2007): Minireview: Recent progress in hemocyanin research. Integr Comp Biol. Oct;47(4):631-44.

Decker H, Ryan M, Jaenicke E, Terwilliger N. (2001): SDS-induced phenoloxidase activity of hemocyanins from Limulus polyphemus, Eurypelma californicum, and Cancer magister. The Journal of Biological Chemistry, 276(21), 17796-17799.

DeRosier DJ, Klug A. (1968): Reconstruction of three dimensional structures from electron micrographs. Nature, 217:130-134.

Dubochet J, Lepault J, Freeman R, Berriman JA, Homo JC. (1982): Electron microscopy of frozen water and aqueous solutions. Journal of Microscopy (128), S. 219–237.

Dubochet J, McDowall A. (1981): Vitrification of pure water for electron microscopy. Journal of Microscopy (124), S. 3–4.

Dunlap, J. (1999): Extraordinary horseshoe crabs. Minneapolis, Mn: Carolrhoda Books.

Ertas B, von Reumont BM, Wagele J, Misof B, Burmester T. (2009): Hemocyanin Suggests a Close Relationship of Remipedia and Hexapoda. Mol Biol Evol, 26(12), 2711-2718.

Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Stuart AC, Marti-Renom MA, Madhusudhan MS, Yerkovich B, Sali A. (2003): Tools for comparative protein structure modeling and analysis. Nucleic Acids Res., 31(13):3375-80.

Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. (2008): Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15:5.6.1-5.6.30

Fernandez JJ, Luque D, Castun JR, Carrascosa JL. (2008): Sharpening high resolution information in single particle electron cryomicroscopy. J. Struct. Biol., 164(1):170-5.

Fiser A, Do RK, Sali A. (2000): Modeling of loops in protein structures, Protein Science, 9:1753-1773.

French CE, Bell JM, Ward FB. (2007): Diversity and distribution of hemerythrin-like proteins in prokaryotes. FEMS Microbiol Lett. 2008 Feb;279(2):131-45. Epub 2007 Dec 12.

Fuseya M, Ichimura K, Yamamura T, Tachi'iri Y, Satake K, Amemiya Y, Kihara H. (1989) Dissociation and auto-oxidation of hemerythrin induced by SH-modification: a kinetic study. J Biochem. 1989 Feb;105(2):293-8.

Gaykema WP, Volbeda A, Hol WG. (1986): Structure determination of Panulirus interruptus haemocyanin at 3.2 A resolution. Successful phase extension by sixfold density averaging. Journal of Molecular Biology, 187(2), 255-275.

Gaykema WPJ, Hol WGJ, Vereijken JM, Soeter NM, Bak HJ, Beintema JJ. (1984): 3.2 [angst] structure of the copper-containing, oxygen-carrying protein Panulirus interruptus haemocyanin. Nature, 309(5963), 23-29.

Giomi F, Beltramini, M. (2007): The molecular heterogeneity of hemocyanin: Its role in the adaptive plasticity of Crustacea. Gene, 398(1-2), 192-201.

Goddard TD, Huang CC, Ferrin TE. (2007): Visualizing density maps with UCSF Chimera. J. Struct. Biol., 157(1):281-7.

Grant, T. (2007): Advances in Single Particle Electron Microscopy. Doktorarbeit. Imperial College London, London. Division of Molecular Biosciences.

Groneberg, R. (2011): Molekularbiologische Charakterisierung einer alpha-Typ Hämocyanin Untereinheit von Callianassa truncata (Crustacea). Diplomarbeit im Fachbereich Biolgie der Johannes-Gutenberg-University Mainz

Guex N, Peitsch MC. (1997): SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 18:2714-2723.

Hagner-Holler S, Kusche K, Hembach A, Burmester T. (2005): Biochemical and molecular characterisation of hemocyanin from the amphipod Gammarus roeseli: complex pattern of hemocyanin subunit evolution in Crustacea. Journal of Comparative Physiology B, 175(6), 445-452.

Harauz G, van Heel M. (1985): Direct 3D reconstruction from projections with initially unknown angles. In: Edzard S. Gelsema und Laveen N. Kanal (Hg.): Pattern recognition in practice II. Proceedings of an international workshop held in Amsterdam, June 19-21, 1985. Elsevier Science Pub. Co., S. 279-288.

Harauz G, van Heel M. (1986): Exact filters for general geometry in three dimensional reconstruction. Optik, 73:146–156.

Harris JR, Agutter P. (1970): A negative staining study of human erythrocyte ghosts and rat liver nuclear membranes. J. Ultrastruct. Res., 33(3):219-32.

Harris JR, Horne RW. (1991): Negative staining. Electron Microscopy in Biology (Harris, JR, ed.), 203-228.

Hartmann, H. & Decker, H. (2002): All hierarchical levels are involved in conformational transitions of the 4x6-meric tarantula hemocyanin upon oxygenation. Biochim. Biophys. Acta, 1601, 132–137.

Hazes B, Magnus KA, Bonaventura C, Bonaventura J, Dauter Z, Kalk KH, Hol WG. (1993): Crystal structure of deoxygenated Limulus polyphemus subunit II hemocyanin at 2.18 A resolution: clues for a mechanism for allosteric regulation. Protein Sci. Apr;2(4):597-619.

Hekkelman ML, Beek TAT, Pettifer SR, Thorne D, Attwood TK, Vriend G. (2010): WIWS: a protein structure bioinformatics Web service collection. NAR 38, W719-723

Higgins DG and Sharp PM. (1988): CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene, 73:237-244.

Hohn M, Tang G, Goodyear G, Baldwin PR, Huang Z, Penczek PA, Yang C, Glaeser RM, Adams PD, Ludtke SJ. (2007): SPARX, a new environment for Cryo-EM image processing. J Struct Biol. Jan;157(1):47-55.

Horseshoe crab life stages (2008). Online verfügbar unter http://www.ceoe.udel.edu/horseshoecrab/History/lifestages.html, last updated on 01.07.2008, verified on 13.09.2011.

Hubbard TJ, Blundell TL. (1987): Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. Protein Eng., 1(3):159-71.

Jiang N, Tan NS, Ho B, Ding JL. (2005a): Carcinoscorpius rotundicauda hemocyanin subunit I. Submitted (JUN-2005) to the EMBL/GenBank/DDBJ databases

Jiang N, Tan NS, Ho B, Ding JL. (2005b): Carcinoscorpius rotundicauda hemocyanin subunit IV. Submitted (JUN-2005) to the EMBL/GenBank/DDBJ databases

Johnson GT, Autin L, Goodsell DS, Sanner MF, Olson AJ. (2011): ePMV Embeds Molecular Modeling into Professional Animation Software Environments. Structure 19, 293-303.

Kota P, Ding F, Ramachandran S, Dokholyan NV. (2011): Gaia: automated quality assessment of protein structure models. Bioinformatics. 2011 Aug 15;27(16):2209-15

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. (2009): Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins; 77 Suppl 9:114-22.

Lamy J, Compin S, Lamy JN. (1983): Immunological correlates between multiple isolated subunits of Androctonus australis and Limulus polyphemus hemocyanins: an evolutionary approach. Arch Biochem Biophys. Jun;223(2):584-603.

Lamy J, Lamy J, Weill J, Bonaventura J, Bonaventura C, Brenowitz M. (1979): Immunological correlates between the multiple hemocyanin subunits of Limulus polyphemus and Techypleus tridentatus. Arch Biochem Biophys. Sep;196(2):324-39.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007): Clustal W and Clustal X version 2.0. Bioinformatics, 23:2947-2948.

Laughlin, R. (1983): The effects of temperature and salinity on larval growth of the horseshoe crab, Limulus polyphemus. Biol. Bull. 164(1), 93–103.

LeBarron J, Grassucci RA, Shaikh TR, Baxter WT, Sengupta J, Frank J. (2008): Exploration of parameters in cryo-EM leading to an improved density map of the E. coli ribosome. J. Struct. Biol., 164(1):24-32.

Lindahl SG. (2008): Oxygen and life on earth: an anesthesiologist's views on oxygen evolution, discovery, sensing, and utilization. Anesthesiology. Jul;109(1):7-13.

Lovell SC, Davis IW, Arendall WB III, de Bakker P, Word JM, Prisant MG, Richardson J, Richardson D. (2003): Structure validation by C-alpha geometry: phi, psi, and C-beta deviation. Proteins: Structure, Function, and Genetics, 50:437-450.

Ludtke SJ, Baldwin PR, Chiu W. (1999): EMAN: semiautomated software for high-resolution singleparticle reconstructions. J. Struct. Biol., 128(1):82-97.

Magnus KA, Hazes B, Ton-That H, Bonaventura C, Bonaventura J, Hol WG. (1994): Crystallographic analysis of oxygenated and deoxygenated states of arthropod hemocyanin shows unusual differences. Proteins. 1994 Aug;19(4):302-9.

Markl J, Decker H. (1992): Molecular Structure of the arthropod hemocyanins. In CH. P. Mangum, editor. Blood and Tissue Oxygen Carriers, volume 13 of Advances in Comparative and Environmental Physiology, pages 325-376. Springer Verlag Berlin.

Markl J, Stöcker W, Runzler R, Precht E. (1986): Immunological correspondences between the hemocyanin subunits of 86 arthropods: evolution of a multigene protein family. Biol Bull, 171(1), 90-115.

Markl J. (1986): Evolution and function of structurally diverse subunits in the respiratory protein hemocyanin from arthropods. Biol Bull, 171(1), 90-115.

Markl J. (1995): Blaues Blut – Struktur, Funktion und Evolution der Hämocyanine. Chemie in unserer Zeit 1, 6-18

Markl J. (2013): Evolution of molluscan hemocyanin structures. Biochim Biophys Acta. 2013 Sep;1834(9):1840-52. doi: 10.1016/j.bbapap.2013.02.020.

Martin AG (2006): 10A cryo-EM structure of Drosophila melanogaster hexamerin and Limulus polyphemus hemocyanin

Miller K, Elderd N, Arisaka F, van Holde K. (1977): Structure and function of hemocyanin from thalassinid shrimp, (115), 171-184.

Miller K, Pritchard A, Rutledge, P. (1976): Respiratory Regulation and the Role of the Blood in the Burrowing Shrimp Callianassa californiensis (Decapoda: Thalassinidea). Marine Biology, (36), 233-242.

Miller K. (1981): The effect of environmental variables on the structure and function of hemocyanin from Callianassa californiensis. J. comp. Physiol., (143), 261-267.

Mindell JA, Grigorieff N. (2003): Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol., 142(3):334-47.

Orlova EV, Dube P, Harris JR, Beckman E, Zemlin F, Markl J, van Heel M. (1997): Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 A resolution by electron cryomicroscopy and angular reconstitution. J. Mol. Biol., 271(3):417-37

Orlova EV, van Heel M. (1994): Angular reconstitution of macromolecules with arbitrary point group symmetry. Proceedings ICEM 13–Paris, Vol. 1, 507–508.

Pallavicini A, Negrisolo E, Barbato R, Dewilde S, Ghiretti-Magaldi A, Moens L, Lanfranchi G. (2001): The primary structure of globin and linker chains from the chlorocruorin of the polychaete Sabella spallanzanii. J Biol Chem. 2001 Jul 13;276(28):26384-90.

Penczek PA, Grassucci RA, Frank 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): 251-70.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. (2004): UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem., 25(13): 1605-12.

Presnell SR, Cohen FE. (1989): Topological distribution of four-alpha-helix bundles. Proc Natl Acad Sci U S A. Sep;86(17):6592-6.

R Core Team (2016): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Radermacher M, Wagenknecht T, Verschoor A, Frank J. (1987): Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J. Microsc., 146(Pt 2):113-36.

Radermacher M. (1988): Three-dimensional reconstruction of single particles from random and nonrandom tilt series. J. Electron Microsc. Tech., 9(4):359-94.

Ramachandran S, Kota P, Ding F, Dokholyan NV. (2011): Automated minimization of steric clashes in protein structures. Proteins. 2011 Jan;79(1):261-70.

Rehm P, Pick C, Borner J, Markl J and Burmester T. (2012): The diversity and evolution of chelicerate hemocyanins. BMC Evol. Biol. 12, 19 (2012)

Robert X and Gouet P.: Deciphering key features in protein structures with the new ENDscript server. Nucl. Acids Res. (1 July 2014) 42 (W1): W320-W324.

Rosenthal PB, Henderson R. (2003): Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol., 333(4):721-45.

Rost B, Sander C. (1996): Bridging the protein sequence-structure gap by structure predictions. Annu Rev Biophys. Biomol. Struct., 25:113-36.

Roxby R, Miller K, Blair DP, Van Holde KE. (1974): Subunits and association equilibriums of Callianassa californiensis hemocyanin. Biochemistry, 13(8), 1662-1668.

Royer WE Jr, Strand K, van Heel M, Hendrickson WA. (2000): Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. Proc Natl Acad Sci U S A. 2000 Jun 20;97(13):7107-11.

Sali A, Blundell TL. (1993): Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol., 234, 779-815.

Sanchez R, Sali A (1997): Advances in comparative protein-structure modelling. Curr. Opin. Struct. Biol., 7(2):206-14.

Satake K, Yugi M, Kamo M, Kihara H, Tsugita A. (1990): Hemerythrin from Lingula unguis consists of two different subunits, alpha and beta. Protein Seq Data Anal. 1990 Mar;3(1):1-5.

Savel-Niemann A, Markl J, Linzen B. (1988): Hemocyanins in spiders. XXII. Range of allosteric interaction in a four-hexamer hemocyanin. Co-operativity and Bohr effect in dissociation intermediates. J Mol Biol. Nov 20;204(2):385-95

Saxton WO, Frank J. (1977): Motif detection in quantum noise-limited electron micrographs by crosscorrelation. Ultramicroscopy, 2(2-3):219-27.

Schatz M, Orlova EV, Dube P, Jager J, van Heel M. (1995): Structure of Lumbricus terrestris hemoglobin at 30 A resolution determined using angular reconstitution. J. Struct. Biol., 114(1):28-40.

Schatz, Michael (1992): Invariante Klassifizierung elektronenmikroskopischer Aufnahmen von eiseingebetteten biologischen Makromolekülen. Autorenaufl. zur textgleichen Mikroed. Egelsbach, Köln, New York: Hänsel-Hohenhausen.

Scherbaum S, Ertas B, Gebauer W, Burmester T. (2010): Characterization of hemocyanin from the peacock mantis shrimp Odontodactylus scyllarus (Malacostraca: Hoplocarida). Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology, 180(8), 1235-1245.

Shuster, C.N., Jr. in Physiology and Biology of the Horseshoe Crabs: Studies on Normal and Environmentally Stressed Animals (eds. Bonaventura, J., Bonaventura, C. & Tesh, S.) 1–52 (Alan R. Liss, New York, 1982).

Smith SA, Berkson JM, Barratt RA. (2002): Horseshoe crab (Limulus polyphemus) hemolymph, biochemical and immunological parameters. Proc. Int. Assoc. Aquatic Anim. Med. 33, 101–102 (2002).

Solem, E. (2010): Molekularbiologische und proteinbiochemische Charakterisierung des Hämocyanins von Callianassa truncata (Crustacea). Diplomarbeit im Fachbereich Biolgie der Johannes-Gutenberg-University Mainz.

Stohr M. (2007): 3D Cryo-Elektronenmikroskopie des 4x6-Hämocyanins der Vogelspinne Eurypelma californicum unter Oxy- und Deoxybedingungen. Dissertation im Fachbereich Biolgie der Johannes-Gutenberg-University Mainz.

Sullivan B, Bonaventura J, Bonaventura C. (1974): Functional differences in the multiple hemocyanins of the horseshoe crab, Limulus polyphemus L. Proc Natl Acad Sci U S A. Jun;71(6):2558-62.

Tama F, Miyashita O, Brooks CL 3rd. (2004a): Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. J Struct Biol., 147(3):315-26.

Tama F, Miyashita O, Brooks CL 3rd. (2004b): Flexible multi-scale fitting of atomic structures into lowresolution electron density maps with elastic network normal mode analysis. J. Mol. Biol., 337(4):985-99.

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. (2007): EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol. Jan;157(1):38-46.

Taveau JC, Boisset N, Lamy J, Lambert O, Lamy JN. (1997): Three-dimensional reconstruction of Limulus polyphemus hemocyanin from cryoelectron microscopy. J Mol Biol. Mar 14;266(5):1002-15.

Thompson JD, Higgins DG, Gibson TJ. (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22:4673-4680.

Thompson RF, Walker M, Siebert CA, Muench SP, Ranson NA. (2016): An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. Methods. Feb 28.

Thon F. (1966): Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. Z. Naturforschung, 21a:476–478.

Touw WG, Baakman C, Black J, te Beek TA, Krieger E, Joosten RP, Vriend G. (2015): A series of PDB-related databanks for everyday needs. Nucleic Acids Res. 2015

van Heel M, Frank J. (1981): Use of multivariate statistics in analysing the images of biological macromolecules. Ultramicroscopy, 6(2):187-94.

van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A. (2000): Single-particle electron cryo-microscopy: towards atomic resolution. Q. Rev. Biophys., 33(4):307-69.

van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M. (1996): A new generation of the IMAGIC image processing system. In: J. Struct. Biol 116 (1), S. 17-24.

van Heel M, Keegstra W. (1981): IMAGIC: a fast, flexible and friendly image analysis software system. Ultramicroscopy, 7:113-130.

van Heel M, Schatz M. (2005): Fourier shell correlation threshold criteria. J Struct Biol., 151(3):250-62.

van Heel M. (1984): Multivariate statistical classification of noisy images (randomly oriented biological macro-molecules). Ultramicroscopy, 13(1-2):165-83.

van Heel M. (1987): Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy, 21(2):111-23.

van Holde KE, Miller KI, Decker H. (2001): Hemocyanins and invertebrate evolution. J Biol Chem. May 11;276(19):15563-6.

van Holde KE, Miller KI. (1995): Hemocyanins. Adv Protein Chem. 1995;47:1-81.

Volbeda A, Hol WG. (1989): Crystal structure of hexameric haemocyanin from Panulirus interruptus refined at 3.2 A resolution. J Mol Biol. 20;209(2):249-79.

Vriend G. (1990): WHAT IF: a molecular modeling and drug design program. J Mol Graph. Mar;8(1):52-6, 29.

Wang Z, Schröder GF. (2012): Real-space refinement with DireX: From global fitting to side-chain improvements. 1. Biopolymers. 2012 Sep;97(9):687-97.

Ward, JH. (1982): Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc., 58:236–244.

Zhu J, Penczek PA, Schröder R, Frank J. (1997): Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. J Struct Biol. Apr;118(3):197-219.

G Appendix

1 Abbreviations & Color Codes

1.1 Hemocyanin designations

*Cro*Hc	*Carcinoscorpius rotundicauda* hemocyanin
*Ctr*Hc1/2	*Callianassa truncata* hemocyanin type 1/2
*Lpo*Hc	*Limulus polyphemus* hemocyanin
*Pin*Hc(B)	*Panulirus interruptus* hemocyanin (type B)

1.2 Other abbreviations

1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
Å	Ångström
ATP	Adenosine triphosphate
CCF	Cross correlation function
cryoEM	Cryo electron microscopy
dpi	Dots per inch
DSSP	Define Secondary Structure of Proteins
EM	Electron microscope
ePMV	Extensible Python molecular viewer
FFT	Fast Fourier transformation
Fig.	Figure
FSC	Fourier shell correlation
g	Gravitational acceleration
GB	Gigabyte
GHz	Gigahertz
kDa	Kilo Dalton
kV	Kilo Volt
min.	Minutes
mM	(milli) Molar
MRA	Multi reference alignment
MSA	Multivariate statistical analysis
nm	Nanometer
pdb	Protein data bank
RMSD	Root mean square deviation
ROS	Reactive oxygen species
SNR	Signal to noise ratio
TIFF	Tagged image file format
TEM	Transmission electron microscope
VDW	Van der Waals radius

1.3 Amino acid abbreviations and color code

All depictions of amino acids employ a color code in addition to text labels. To help the viewers depth perception, the depictions are rendered using realistic soft lighting with shadows and depth of field. As a result colors might appear in slightly different shades depending on the lighting situation.

🟩	Isoleucine	ILE	I	🟩	Valine	VAL	V
🟦	Glutamine	GLN	Q	🟧	Threonine	THR	T
⬜	Glycine	GLY	G	🟪	Histidine	HIS	H
🟥	Glutamic Acid	GLU	E	🟪	Tryptophan	TRP	W
🟨	Cysteine	CYS	C	🟦	Phenyalanine	PHE	F
🟥	Aspartic Acid	ASP	D	⬜	Alanine	ALA	A
🟧	Serine	SER	S	🟨	Methionine	MET	M
🟦	Lysine	LYS	K	🟩	Leucine	LEU	L
🟫	Proline	PRO	P	⬛	Arginine	ARG	R
🟦	Asparagine	ASN	N	🟦	Tyrosine	TYR	Y

1.4 Secondary structure rainbow color scheme

Subunit chains are colored from N to C-terminus using a rainbow color scheme:

H Acknowledgments

Ich danke [redacted] für die Bereitstellung des Themas, seine immerwährende Unterstützung sowie die kritische Durchsicht meiner Arbeit.

Ich danke [redacted] für die Isolierung und Präparation der Hämocyaninmoleküle.

Ich danke [redacted] für die Cryo-Präparationen und Erstellung der EM-Bilder.

Ich danke [redacted] und [redacted] vom Max-Planck-Institut für Biophysik Frankfurt für die Bereitstellung und Unterstützung bei der Nutzung des Elektronenmikroskops.

Ich danke [redacted], [redacted] und [redacted] für die Bereitstellung der Sequenzdaten von *C. truncata*

Ich danke [redacted] für die Bereitstellung der SPARX-Rekonstruktion von *L. polyphemus* Hämocyanin.

Nicht zuletzt möchte ich allen aktiven und ehemaligen Mitgliedern der Arbeitsgruppe [redacted] für Ihre Unterstützung danken.

I Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig und nur mit Hilfe der angegebenen Personen und Mittel (Literatur, Apparaturen, Material) angefertigt zu haben. Bei den von mir durchgeführten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der Satzung der Johannes Gutenberg Universität Mainz niedergelegt sind, eingehalten.

Ort, Datum	Christoph Kühne