

*3D Strukturanalyse von Mollusken-  
hämocyaninen aus  
elektronenmikroskopischen Bildern*

**D i s s e r t a t i o n**  
zur Erlangung des Grades

**„ D o k t o r  
d e r N a t u r w i s s e n s c h a f t e n “**

am Fachbereich Biologie  
der Johannes Gutenberg Universität Mainz

Gatsogiannis Christos

Geb. am 13.05.1981 in Karditsa/Griechenland

Mainz, 2009

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 29.10.2009

*Für*





*"Nature does nothing without purpose or uselessly."*

Aristotle of Stagira (384–322 BC)







# I. Table of Contents

<b>I. Table of Contents .....</b>	<b>I</b>
<b>II. List of Figures .....</b>	<b>V</b>
<b>III. List of Tables.....</b>	<b>VII</b>
<b>A INTRODUCTION .....</b>	<b>1</b>
1. Oxygen transporters .....	1
2. Molluscan hemocyanins .....	5
2.1 Available X-ray structures of functional units .....	12
2.2 Medical relevance of molluscan hemocyanins.....	15
3. CryoEM and Single particle analysis .....	17
4. Objectives of the present work .....	18
<b>B MATERIALS AND METHODS.....</b>	<b>21</b>
1. Materials and equipment .....	21
1.1 Animals .....	21
<i>Megathura crenulata</i> .....	21
<i>Nautilus pompilius</i> .....	22
1.2 Electron microscopes .....	24
1.3 Computer-hardware.....	25
1.4 Further hardware .....	25
1.5 Software .....	26
2. Biochemical Methods.....	29
2.1 KLH1-purification.....	29
2.2 NpH-purification .....	29
2.3 Determination of hemocyanin concentration and absorption spectrum .....	29
2.4 Negative staining.....	30
3. Cryo-electron microscopy .....	30
3.1 Preparation of cryo-EM grids.....	30
3.2 Data collection in the Cryo-Electron microscope.....	30

KLH1 .....	31
NpH.....	31
4. Bioinformatical Methods.....	31
4.1 Scanning of micrographs.....	31
4.2 Particle selection.....	32
4.3 CTF-correction .....	33
4.4 Bandpass filter.....	35
4.5 Multiple Reference Alignment (MRA) .....	37
Translational alignment.....	38
Rotational alignment .....	38
4.6 MSA- Classification .....	40
MSA- multivariate statistical analysis.....	40
HAC-Classification and averaging.....	41
4.7 Orientation determination.....	43
Euler Angles.....	43
Angular reconstitution.....	45
Projection matching.....	48
4.8 3D Reconstruction.....	49
4.9 Iterative refinements.....	50
4.10 Visualization.....	53
4.11 Resolution assessment.....	53
4.11 Sharpening.....	54
4.12 3D-masking .....	55
4.13 Modelling the NpH polypeptide chain .....	57
4.14 Molecular modelling of the KLH1-polypeptide chains.....	60
4.15 Further methods.....	62
Alignment of molecular models .....	62
Generation of 3D-electron density maps from PDB files.....	62
Fitting of one density map into another.....	62
<b>C RESULTS AND DISCUSSION.....</b>	<b>63</b>
1. Overview .....	63

---

2. Comparative 11 Å structure of two molluscan hemocyanins from 3D cryo-electron microscopy .....	65
2.1 11 Å cryoEM structure of NpH and HtH1 .....	66
2.2 The decamer wall architecture.....	69
2.3 The architecture of the collar complexes.....	70
2.4 Pathway of the subunit .....	73
3. <i>Nautilus pompilius</i> Hemocyanin: 9 Å Cryo-EM structure and Molecular model reveal the subunit pathway and the interfaces between the 70 functional units .....	75
3.1 8.4 Å NpH decamer with imposed D5 symmetry .....	75
3.2 9.1 Å NpH-decamer with imposed C5 symmetry. ....	78
3.2.1 Subunit dimer and pathway of the subunit.....	78
3.3 Inter-FU interfaces .....	85
3.3.1 The morphological unit interfaces.....	86
3.3.2 The horizontal tier interfaces.....	86
3.3.3 The interfaces of the grooves .....	87
3.3.4 Arc-wall interfaces .....	89
3.4 Putative divalent cation-binding sites within the decamer .....	90
3.5 Putative N-glycosylation pattern of NpH .....	91
3.6 Model of the allosteric unit .....	94
4. Keyhole limpet Hemocyanin: 9 Å CryoEM Structure and Molecular Model of the KLH1 Didecamer reveal the Interfaces and Intricate Topology of the 160 functional units. ....	99
4.1 Cryo-EM structure and molecular model of KLH1.....	99
4.2 Molecular structure of the interfaces between functional units of the cylinder wall.....	106
4.2.1 Morphological unit interfaces (a↔b, d↔e, c↔f) .....	106
4.2.2 The horizontal tier interfaces.....	108
4.2.3 Minor groove interfaces .....	110
4.2.4 The major groove interfaces.....	113
4.3 Putative binding sites for divalent cations within the decamer .....	116
4.4 The decamer/decamer interfaces .....	118
4.5 Topology of the five arcs and the arc, arc/wall interfaces, hypothetical reassembly scenario	119
4.6 Interfaces within the Fu-h slab pentamer .....	121
4.7 Interfaces between the peripheral and the internal collar .....	123

4.8 Interfaces between peripheral collar and wall .....	123
4.9 Glycosylation pattern .....	123
4.9.1 N-glycosylation sites and possible consequences for assembly .....	125
4.10 Putative model of the allosteric unit .....	126
5. Conclusions and Outlook .....	129
<b>D SUMMARY .....</b>	<b>133</b>
<b>E ZUSAMMENFASSUNG .....</b>	<b>134</b>
<b>F APPENDIX.....</b>	<b>135</b>
1. Abbreviations .....	135
2. Abbreviations of hemocyanins .....	136
3. Abbreviation-code of amino acids.....	137
4. Database accession numbers .....	137
<b>G BIBLIOGRAPHY .....</b>	<b>139</b>
<b>H PUBLICATION LIST .....</b>	<b>149</b>
<b>I Danksagung – Acknowledgement - Ευχαριστία.....</b>	<b>151</b>
<b>Lebenslauf.....</b>	<b>153</b>
<b>Eidesstattliche Erklärung.....</b>	<b>155</b>

## II. List of Figures

Fig. 1: Oxygen transport proteins.....	2
Fig. 2 : Dinuclear copper Type 3 active site .....	3
Fig. 3: Quaternary structures of arthropod hemocyanins .....	4
Fig. 4 : TEM and 3D-Resonstructions of molluscan hemocyanins.....	7
Fig. 5 : Structural organisation of <i>Haliotis tuberculata</i> Hemocyanin Isoform 1 (HtH1).....	8
Fig. 6: Phylogenetic tree of the molluscan hemocyanins functional units.....	10
Fig. 7: TEM-Micrographs of oligomeric states of the two isoforms of <i>Haliotis tuberculata</i> hemocyanin (HtH21 and HtH2) .....	11
Fig. 8: Crystal structure of OdH-g, RtH2-e and KLH1- h .....	13
Fig. 9 : The giant keyhole limpet <i>Megathura crenulata</i> .....	22
Fig. 10 : <i>Nautilus pompilius</i> .....	23
Fig. 11 : Tecnai 30 FEG at the Johannes Gutenberg University of Mainz .....	24
Fig. 12 : “Beowulf” cluster .....	25
Fig. 13 : PRIMESCAN drum scanner.....	32
Fig. 14 : Block-convolution 2D-filter .....	33
Fig. 15: Contrast-Transfer-function CTF .....	34
Fig. 16: CTFFIND3 diagnostic output image .....	35
Fig. 17: Bandpass-filtering.....	36
Fig. 18: Bandpass-filtering with respect to noise reduction.....	36
Fig. 19 : Procedure of the MRA.....	38
Fig. 20: Example of a translational alignment of a target image against its reference. ....	39
Fig. 21: Example of a rotational alignment of a target image against its reference.....	39
Fig. 22: Example of a hierarchical ascendant classification (HAC) tree .....	42
Fig. 23: Typical class average images of the KLH1 didecamer and the corresponding original images.....	42
Fig. 24: IMAGIC 2D co-ordinate sytem .....	43
Fig. 25: Definition of the Euler angles in a common coordinate system.....	44
Fig. 26 : 2D projections of the same object represented in Fourier 3D space .....	45
Fig. 27: Sinogram.....	46
Fig. 28: Cross Sinogram correlation function (CSCF).....	47
Fig. 29 : Overlapping central sections in Fourier space.....	49
Fig. 30 Example of a density map.....	50

Fig. 31 (a): Single particle analysis by the angular reconstitution technique .....	51
Fig. 31 (b): Iterative refinements to the model by projection matching .....	52
Fig. 32: Sharpening, 3D-low pass filter and masking.....	56
Fig. 33 : Rigid-body fitting of the 20 initial molecular models of NpH-c into their respective density map .....	58
Fig. 34: Modelling of the linker peptides.....	59
Fig. 35: Surface representations of the present cryoEM structures in comparison to the previous corresponding highest resolution reconstructions .....	67
Fig. 36: Comparison of the 11 Å cryoEM structures of HtH1 and NpH .....	68
Fig. 37: Extracted morphological unit (slab) of the peripheral collar of HtH1 .....	71
Fig. 38: Comparison of NpH-g with the X-ray structure of OdH-g and FU-g from Benthocopus hemocyanin .....	72
Fig. 39: (legend on next page).....	73
Comparison of the X-ray Structure of OdH-g with the extracted cryo-EM density of a NpH wall FU.....	74
Fig. 40: 8.4 Å cryoEM structure of NpH with imposed D5 symmetry.....	76
Fig. 41: Revealing the architecture of the NpH-arc .....	77
Fig. 42: Fitting of OdH-g crystal structure into the NpH-cryoEM-structure.....	79
Fig. 43: 9.1 Å cryoEM structure of Nautilus hemocyanin and topology of the seven different types of FU .....	81
Fig. 44: Supportive evidence for the present model of the FU localization and the subunit dimer.....	82
Fig. 45: The individual Fus of NpH with their respective molecular models.....	83
Fig. 46: Topography of the inter-FU interfaces in the NpH cylinder wall.....	85
Fig. 47: Potential attachment sites for N-linked glycans in NpH.....	93
Fig. 48: N-glycosylation pattern of NpH .....	94
Fig. 49: Proposed model of the allosteric unit in NpH.....	96
Fig. 50: 9 Å cryoEM structure of KLH1 .....	101
Fig. 51: Comparative analysis of the cryoEM structure of NpH with the corresponding extracted cryoEM structure of KLH1.....	102
Fig. 52: CryoEM and crystal structure of the KLH1-h morphological unit.....	103
Fig. 53: Extracted subunit dimer of KLH1 .....	105
Fig. 54: Morphological unit interfaces localized in the cylinder wall of KLH1 .....	107
Fig. 55: Interfaces between the horizontal tiers in the cylinder wall of KLH1 .....	109

---

Fig. 56: CryoEM structures and molecular models of the FUs involved in the minor groove interfaces .....	111
Fig. 57: Molecular interfaces at the minor groove in the cylinder wall of KLH1 .....	112
Fig. 58 : Molecular structure of the interfaces a↔f and d↔d at the major groove in the cylinder wall of KLH1 .....	114
Fig. 59: Major groove interfaces a↔f/ d↔b of KLH1. ....	116
Fig. 60: Binding opportunities for bridging divalent cations. ....	117
Fig. 61: Structural organization of the peripheral collar .....	122
Fig. 62 : Hypothetical structural explanation for the inefficiency of KLH1 to form multidecamers .....	126

### III. List of Tables

Table 1: Putative attachment sites of N-linked glycan in NpH.....	92
Table 2: Putative attachment sites of N-linked glycans in KLH1 .....	124



# A INTRODUCTION

## 1. Oxygen transporters

Animals that live under aerobic conditions consume large amounts of O<sub>2</sub>, which is mainly used to sustain the production of ATP in the respiratory chain of the mitochondria. The constant supply of oxygen is crucial for the survival of the cells. In Protozoa and small Metazoa, simple diffusion is usually considered sufficient for the supply of the inner layers of the body with O<sub>2</sub>. Larger animals, however, demand a more efficient supply and require a variety of anatomical, physiological, and molecular adaptations that enhance the O<sub>2</sub> delivery to the cells and eventually to the mitochondria. These adaptations comprise respiratory organs, such as gills or lungs, circulatory systems, as well as the evolution of simple oxygen binding proteins into multi-subunit, circulating proteins which reversibly bind molecular O<sub>2</sub> for the purpose of transport or storage. Respiratory proteins enhance the O<sub>2</sub> transport capacitance of the body fluid, facilitate intracellular O<sub>2</sub> diffusion or enable O<sub>2</sub> storage for long- or short-term periods.

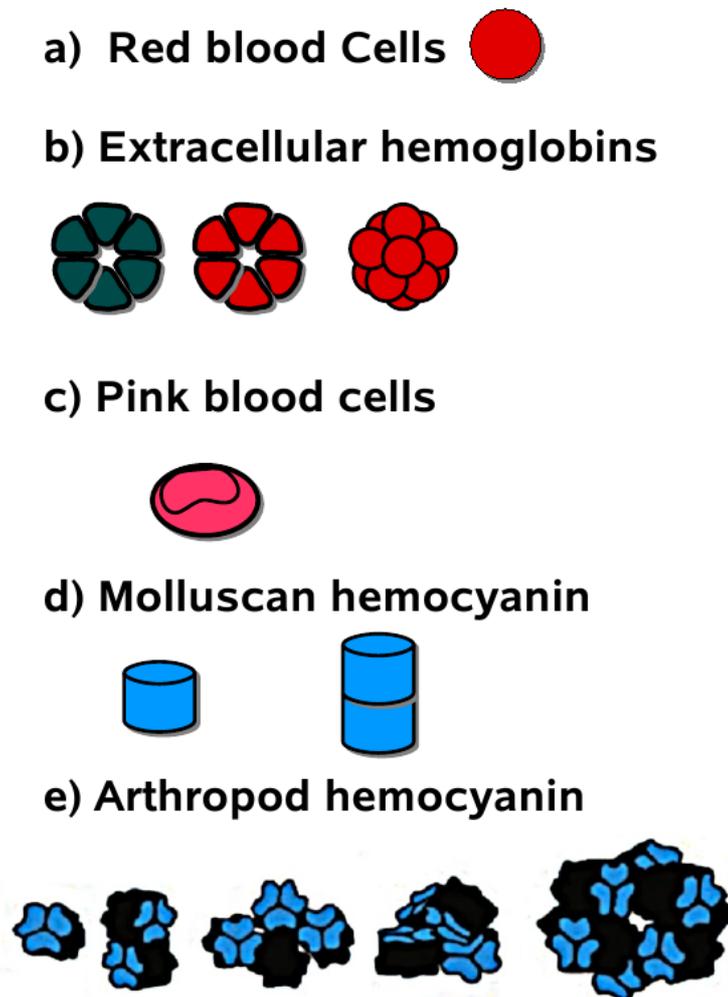
In the animal kingdom, three distinct types of metal-containing respiratory proteins evolved, in order to facilitate the circulatory transport of oxygen, which can be classified by their active sites: **(hemo-) globins (haem iron), hemerythrins (non haem di-iron) and hemocyanins (di-copper)** (as illustrated in Fig. 1).

Most familiar are the **hemoglobins (Hb)**, with representatives in almost every Phylum. Nowadays, deciphering the genome of several species allows one to draw the evolutionary tree of this protein going back 1800 million years, to a time when oxygen began to accumulate in the atmosphere. The first X-ray crystal structure of Hb was determined in the early 1960's (Perutz *et al.* 1960, Nobel Prize in chemistry). Human Hb is a tetramer with a molecular weight of ~64 kDa that is present at a high concentration in the red blood cells, the erythrocytes. Each subunit carries a heme group, which consists of a ferrous ion, coordinated by four nitrogens in the center of a porphyrin ring. In its deoxy form the iron ion of the heme is in a five-coordinate high-spin state, with imidazole serving as the axial ligand on the proximal side. Dioxygen binds at the sixth (distal) vacant coordination site. Binding of O<sub>2</sub> causes an in-plane movement of the Fe, resulting from shortening of the Fe-N distances, as well as conformational changes within the porphyrin macrocycle and the distal protein residues (Collman *et al.*, 2003).

Many annelids and other phyla have giant cooperative hemoglobins with molecular masses on the order of ~3.5 MDa, which are freely dissolved in the blood, rather than packaged in cells. These macromolecular assemblages are also known as **erythrocruorins**.

A variation of the hemoglobins are the giant, extracellular **chlorocruorins**, which are found in four marine polychaete families (Dewilde *et al.*, 2001). They contain an altered porphyrin ring, with a formyl substituting for the 3-vinyl group and consequently appear as greenish-red (Lamy *et al.*, 1996).

**Hemerythrins** are found in the erythrocytes of certain marine invertebrates. The oxygen binding site contains two iron atoms which, in contrast to hemoglobin, are covalently bound to the protein molecule (Loehr *et al.*, 1978). The two ferrous ions are joined by an oxo bridge and the carboxyl side chains of a glutamic acid and an aspartic acid residue. Five histidine residues complete the group of liganding amino acids (Holmes *et al.*, 1991).

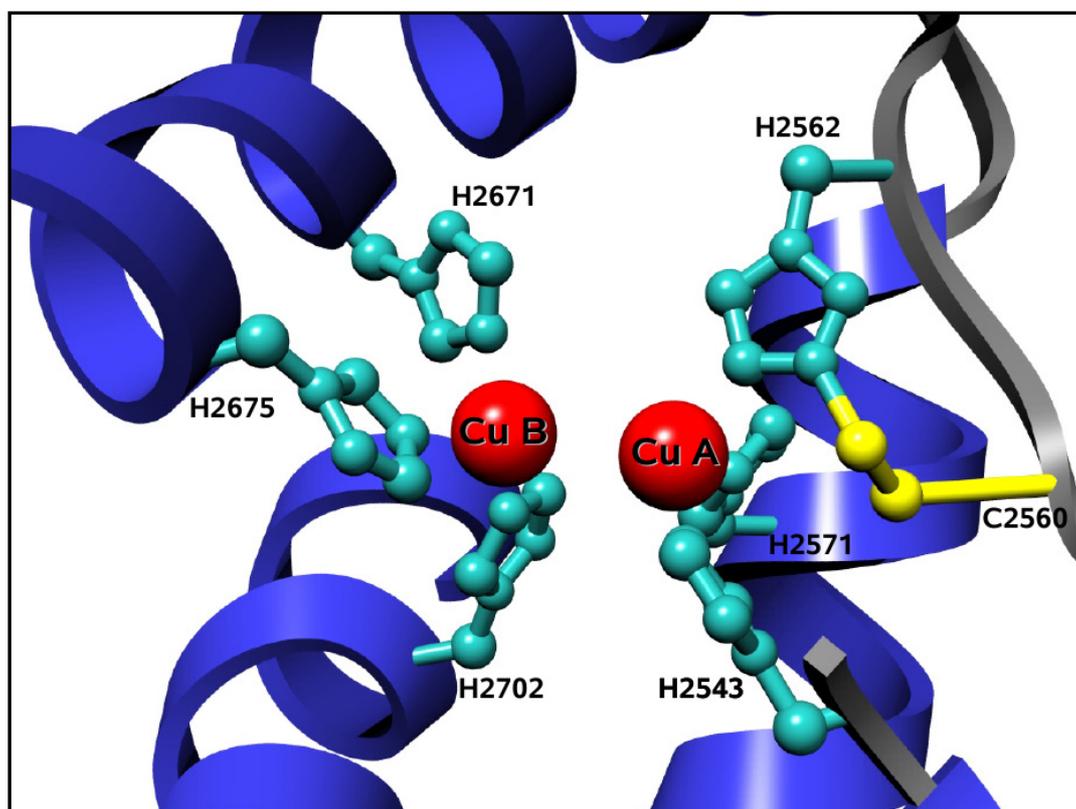


**Fig. 1: Oxygen transport proteins**

according to (Terwillinger, 1998) (modified). a) Cellular hemoglobins b) extracellular hemoglobins (annelid chlorocruorin and hemoglobin, arthropod extracellular hemoglobin) c) cellular hemerythrins d) molluscan hemocyanins (cephalopod, gastropod) e) arthropod hemocyanins (one-, two-, four-, six- and eight-hexamers). Models not drawn to scale.

**Hemocyanins** are binuclear type 3 copper proteins utilized by many molluscs and arthropods and freely dissolved in the hemolymph. The name hemocyanin does not refer to a heme group, but derives from Greek roots for blue blood (the Greek word for blood is *aima* or

haima). The oxygen-binding site involves a pair of copper atoms, which are in the Cu(I) state in the deoxy form but become Cu(II) upon oxygenation. This accounts for the blue color developed upon oxygenation. The active site is covalently bound to the protein through six histidine residues (see Fig. 2). Spectroscopic studies on the active site show similarities to the active site of a further group of copper containing proteins, the tyrosinases (Linzen, 1989).



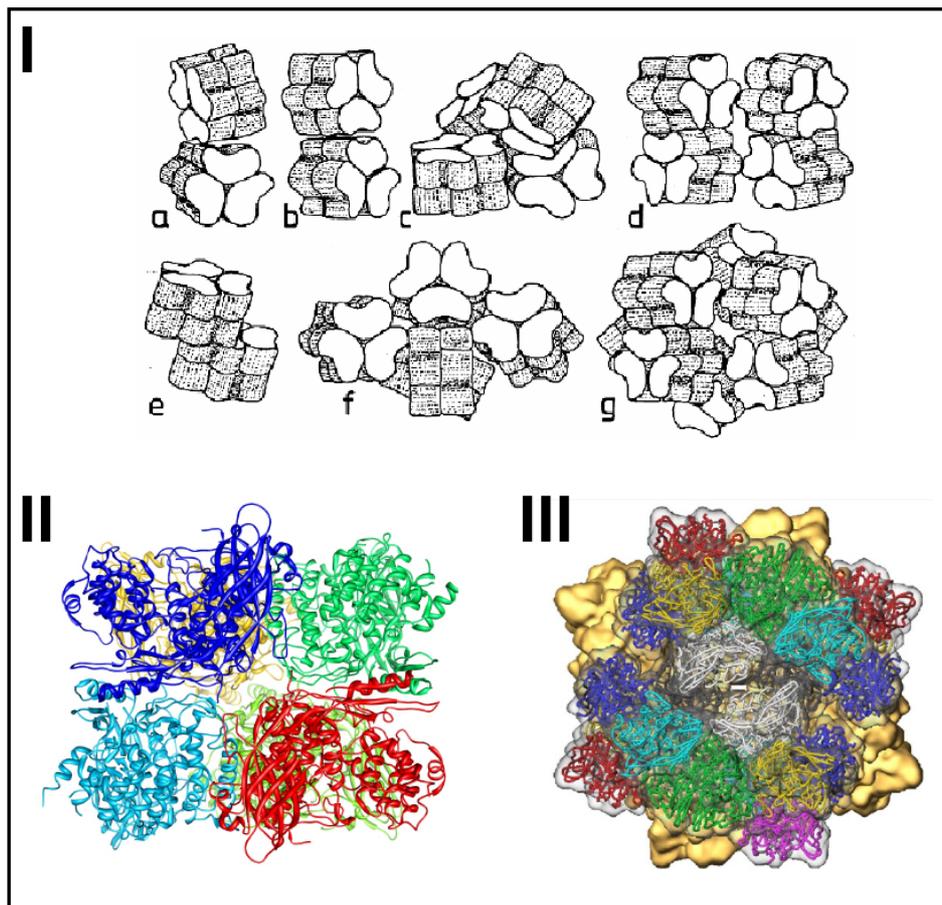
**Fig. 2 : Dinuclear copper Type 3 active site**

The structure shown here is taken from an X-ray-diffraction study of the molluscan *Octopus dofleini* hemocyanin (Cuff *et al.*, 1998) (PDB ID: 1JS8). Copper atoms are shown as red spheres and the residues of the active site are shown as ball and sticks and labeled. Six histidine residues (cyan) ligate the copper atoms (CuA and CuB). A thioether bridge is formed between a histidine ligand (H2562) and a nearby cysteine residue (C2560, yellow). This bond is absent in the active site of arthropod hemocyanins

Hemocyanins are always found as multimers and possess between 6 and 160 oxygen binding sites. Some of them display the highest molecular cooperativity observed in nature (Decker *et al.*, 2007), with Hill coefficients of more than 9 (Lowe, 1978).

However, arthropod and molluscan hemocyanins are profoundly different in quaternary structure and sequence. Although both use a copper pair for oxygen binding, their subunit sizes are different as are their subunit organization and aggregation states (Mellema and Klug, 1972) (see below). Primary sequence similarity is limited to an amino acid region involved in CuB organization. Because of these distinct differences, molluscan and arthropod hemocyanins are considered as different protein families (Markl & Decker, 1992).

Thus, it has been suggested that molluscan and arthropodan hemocyanins have evolved independently from a common ancestral monoclear copper protein (e.g. Drexel *et al.*, 1987). The hemocyanins of the arthropod phylum are built of multiples of hexamers consisting of 1, 2, 4, 6 and 8 such basic assemblies, depending on the class or species (see Fig. 3, I). The basic hexameric unit consists of bean shaped subunits organized in the form of two layers of trimers placed on top of one another (Herskovits, 1988), with approximately D3 point group symmetry. Each subunit has a molecular weight of 75 kDa and contains an oxygen binding site (for review, see Markl & Decker, 1992).



**Fig. 3: Quaternary structures of arthropod hemocyanins**

I) Schematic drawings of the quaternary structure of different arthropod hemocyanins (according to Markl and Decker, 1992) a – crustacean 2x6mer b – 2x6 of many spiders c – shrimps 4x6mer d – 4x6mer of many arachnids e – unusual 2x6mer hemocyanin found in *Squilla mantis* f – 6x6mer from *Scutigera coleoprata* g – 8x6mer from *Limulus polyphemus* II) x-ray structure of the 1x6mer from *Panulirus interruptus* (Volbeda and Hol, 1989) III) cryoEM structure and molecular model from *Limulus polyphemus* 8x6 Hemocyanin (Martin *et al.*, 2006) (structures in II and III are not shown to scale).

A X-ray structure of a 1x6 hemocyanin (*Panulirus interruptus*) (Gaykema *et al.*, 1984) (Fig 3, II), revealed for the first time a structure at the atomic level of the smallest hemocyanin known. Several decades later, a 10 Å cryoEM structure and molecular model of a 8x6 hemocyanin (*Limulus polyphemus*) (Martin *et al.*, 2006) (Fig 3, III) established for the first time the structural parameters of a multihexamer and also the molecular interfaces between

the eight hexamers. This was crucial for understanding how allosteric transitions are mediated between the different levels of hierarchy.

The present work is exclusively focused on molluscan hemocyanin, which will be presented in detail in the following section.

## 2. Molluscan hemocyanins

The basic structure of molluscan hemocyanins is a decamer with an external diameter of ~35 nm and a molecular weight of ~3.5 MDa (in some cephalopods) or ~4MDa (in the other molluscs). Hemocyanins are therefore among the largest proteins known in the nature.

Most molluscan hemocyanins (notably those from gastropods, chitons and protobranch bivalves) are based on a 400 kDa polypeptide chain which is subdivided into eight paralogous functional units (FUs, termed N' term-a-b-c-d-e-f-g-h-C' term)(e.g. Lieb *et al.*, 1999; Markl *et al.*, 2001). The FUs a to g have a molecular mass of ~ 50 kDa, whereas FU-h carries an additional C'-terminal tail of 100 amino acids and has a molecular mass of ~ 60 kDa. Each FU carries a binuclear copper (I) centre for oxygen binding. The eight FUs are interconnected by short linker peptides of 10 -15 aminoacids (Lang, 1988; Lang and van Holde, 1991). Several FUs carry also one or two binding sites for carbohydrate side chains. They are primarily bound by N-glycosylation to the polypeptide chain and can be recognized by the binding motives NXT or NXS (N for asparagine, X for any amino acid, S for serine and T represents threonine) (Forster and Davie, 1984; Keller *et al.*, 1999; Lieb *et al.*, 1999, 2000). The carbohydrate content of molluscan hemocyanins is between 2% and 9% (Idakieva *et al.*, 2004) and a variety of glycan side chains has been detected (Kurokawa *et al.*, 2002).

In some cephalopods, such as *Octopus* or *Nautilus*, the C'-terminal FU-h is missing (Miller *et al.*, 1998; Bergmann *et al.*, 2006). The subunit in this case has a molecular weight of ~350 kDa. Some other cephalopods, such as *Sepia* and *Loligo*, have also an 8 FU-subunit with special organization: in contrast to the gastropods, the additional FU is localized between FU-d and FU-e (Gielens *et al.*, 1983; Loncke *et al.*, 1990)

In all cases, ten subunits form a decamer and subunit dimerization is a characteristic feature. Dissociation/reassociation studies identified the subunit dimer as the major dissociation/reassociation intermediate (Siezen and van Bruggen, 1974). Furthermore, the reassociation steps subunit → subunit dimer → decamer require divalent cations such as magnesium and/or calcium (van Holde and Miller, 1995; Miller *et al.*, 1998).

In chitons and cephalopods only the single decamer is present, whereas by gastropods and protobranch bivalves two decamers are assembled face-to-face to form didecamers or even multidecamers by consecutive linear addition of decamers. The decamer → didecamer (→

multidecamer) transition can be controlled *via* the calcium/magnesium concentration level (Harris *et al.*, 1997) (see also Fig. 7).

The structure of hemocyanins has been studied by TEM since the introduction of the negative staining technique (van Bruggen *et al.*, 1960). In the electron microscope, molluscan hemocyanins produce 30 nm diameter circular profiles with a five-fold symmetry, called top views, and 19-38 nm high rectangular profiles, called side views. These views suggest that their 3D-structure is a hollow cylinder (Fig. 4a). In 1972, Mellema and Klug published the first 3D model of a hemocyanin didecamer, based on superposition of EM-images from negative staining (Fig. 4b). More than two decades later, several cryoEM-based reconstructions of didecamers and cephalopod decamer, in combination with immunoelectron microscopical data, established the following details, concerning the hemocyanin decamer:

The wall of the decamer can be described as a superposition of three layers or tiers. Each tier contains 20 functional units which encycle the cylinder. The wall FUs (FUs a to f) are arranged as pairs, described as morphological units. Furthermore, the molecule wall has the appearance of a right handed-helix and is engraved by five major and five minor oblique grooves. The subunits are antiparallel arranged as pairs and form five subunit-dimers. Ten copies of the FU-g form an internal collar complex at the level of the second wall tier. These functional units are organized as pairs and form five distinct collar morphological units, described as arcs.

In gastropods, five pairs of the additional FU-h form an outer collar complex at the level of one peripheral tier ("closed face"). These pairs are described as slabs. The opposite peripheral tier lacks this element ("open face"). Therefore the decamer in gastropods follows a C5 point group symmetry. Didecamerisation occurs by pairing two decamers at their open faces. These features are illustrated in Fig. 5.

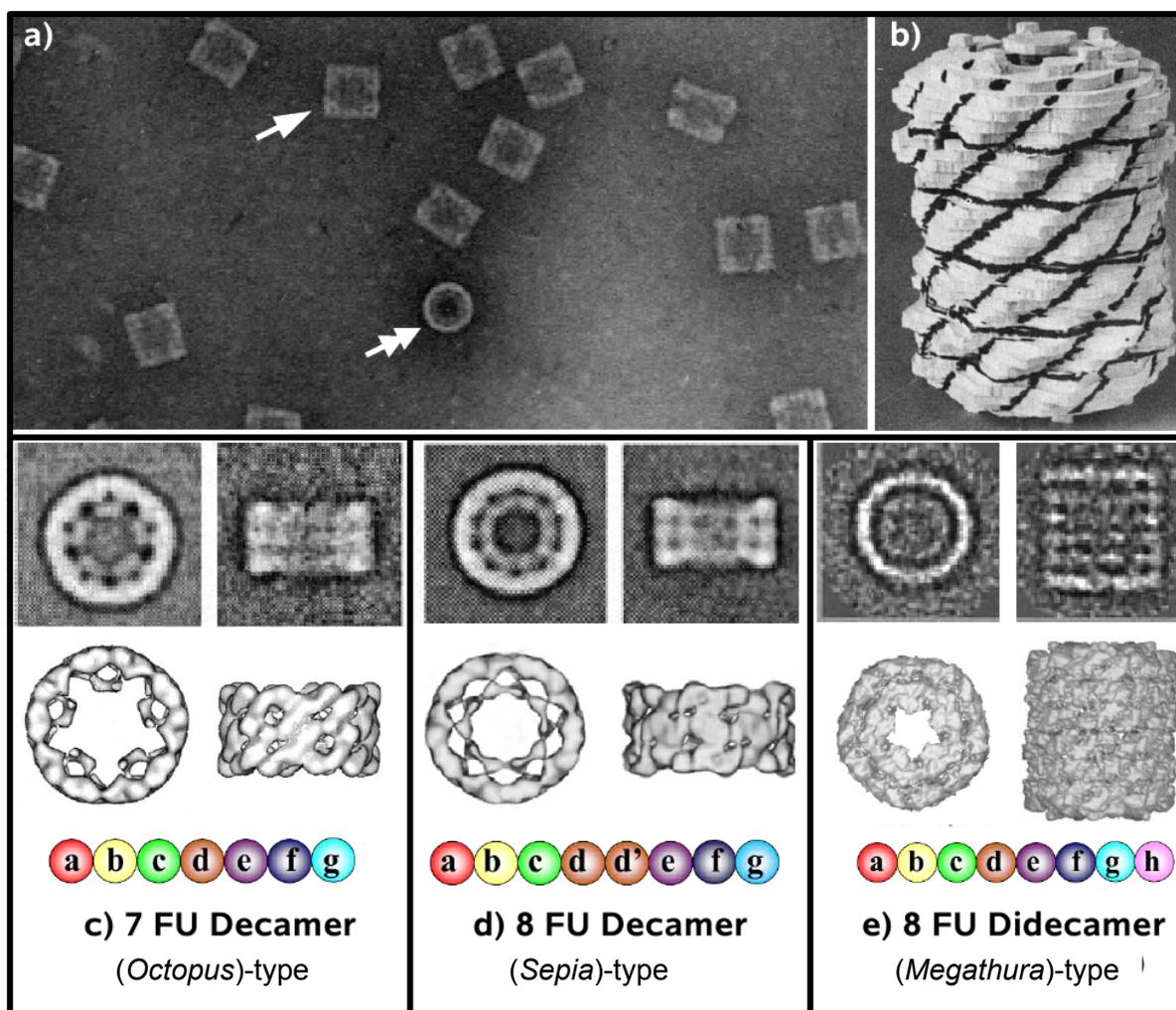
In Cephalopods, such as *Octopus*, the two FU-g copies (the arc) are symmetrically arranged and as mentioned above, the external collar complex (the five FU-h pair; slabs) is missing. Consequently, the decamer in this case follows a D5 point group symmetry.

In *Sepia*, each arc is flanked by two additional masses, in comparison to *Octopus* hemocyanin. The reason for this difference is the presence in *Sepia* of an additional FU (between FU-d and FU-e) which is absent in *Octopus* hemocyanin.

(Mellema and Klug, 1972; Wichertjes *et al.*, 1990; Lamy *et al.*, 1993; Lambert *et al.*, 1994; Lambert *et al.*, 1995; Orlova *et al.*, 1998; Lamy *et al.*, 1998; Boisset, 2000; Meissner *et al.*, 2000; Gebauer *et al.*, 2002; Mouche *et al.*, 2003).

In conclusion, concerning this wealth of electron microscopical and biochemical data on the primary (see below) and quaternary structure, one can define three different types of molluscan hemocyanins:

- the 7-FU decamer (*Vampyromorpha*, *Octopoda* and *Nautiloidea*) (*Octopus*-type)
- the 8-FU decamer (*Coleoidea*) (*Sepia*-type)
- the 8-FU didecimer (Gastropods) (*Megathura*-type) (see Fig. 4).

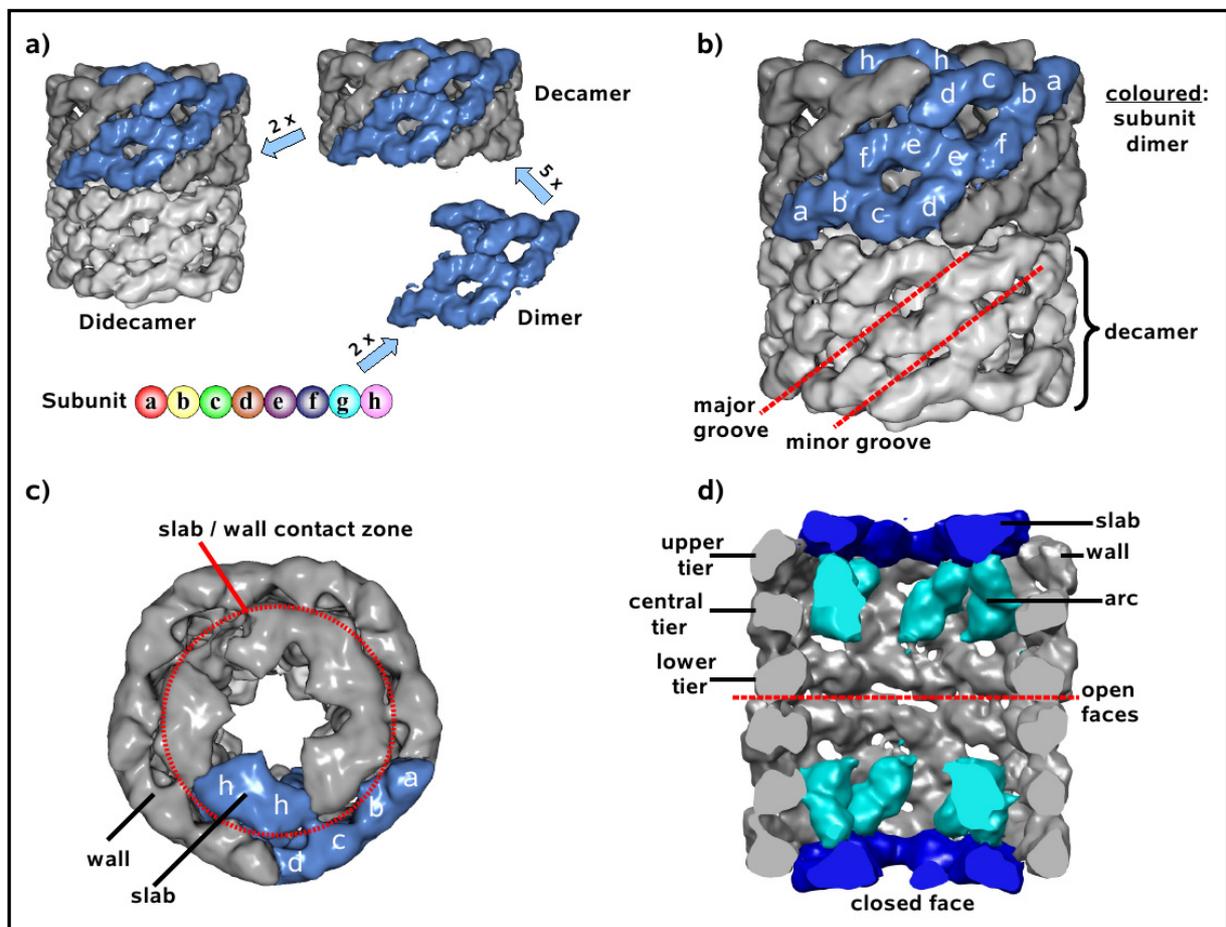


**Fig. 4 : TEM and 3D-Reconstructions of molluscan hemocyanins**

a) Micrograph of Keyhole Limpet Hemocyanin Isoform 1 (Harris and Markl, 1999). The didecimer is shown from the side as rectangle (arrow) and from the top as circle (double arrow) b) The first 3D-reconstruction and wooden model of a molluscan hemocyanin didecimer (Mellema and Klug, 1972) c-e) Class averages (top and side views) (upper row) and the corresponding 3D-reconstructions (lower row) from three different molluscan hemocyanin types. In cephalopods only the single decamer is present. (c-d). The subunit contains 7 (*Octopus*-type, c) or 8 FUs (*Sepia*-type, d). The additional FU in *Sepia* is visible in the opening of the cylinder. In gastropods (*Megathura*-type), two decamers are assembled face to face in order to form a didecimer (e). The subunit contains 8 FUs. The additional FU-h forms an external collar complex, which is absent in *Octopus*-type hemocyanin molecules. (Chiton hemocyanin is an asymmetrical decamer that resembles the decamers obtained by partial dissociation of gastropod (*Megathura*-type) hemocyanin molecules.) 3D reconstructions and class averages: c) Mouche *et al.*, 1999 d) Mouche *et al.*, 1999; Boisset and Mouche, 2000 e) Meissner *et al.*, 2000 (modified).

The quaternary structure of an 8 FU-didecimer from the abalone *Haliotis tuberculata* has been produced at ~12/14 Å from cryo-electron micrographs (Meissner *et al.*, 2000) (Fig. 4 e, 5 a-d). Similar studies have been performed in keyhole limpet (*Megathura crenulata*) hemocyanin (KLH) (Mouche *et al.*, 2003). Both 3D reconstructions presented an

improvement of the overall shape and of the details of the subunits and the interpretation and location of the subunit dimer, in comparison with previous published structures of molluscan hemocyanins (cephalopods and gastropods) solved at lower resolutions (15-40 Å). For the first time it was possible to define not only the shape of the morphological units, but also the shape of individual FUs within the morphological units. Furthermore, Meissner *et al.*, 2000, proposed two models for the path of the polypeptide chain, but only one was consistent with a study of the localization of FU-c (Gebauer *et al.*, 2002). The localization of the FUs abcdefgh within the structure of HtH1 has been adapted to KLH1, based on the fact, that KLH1 and HtH1 are orthologous isoforms (Mouche *et al.*, 2003).



**Fig. 5 : Structural organisation of *Haliotis tuberculata* Hemocyanin Isoform 1 (HtH1)**

a) The polypeptide chain is subdivided into eight functional units, termed FU-a to FU-h. Two subunits are arranged as antiparallel pairs, in order to form the subunit dimer. The decamer is formed from five subunit dimers; due to an oblique arrangement of the five subunit dimers, the cylinder wall appears as a right handed helix. The native molecule is a didecamer with two didecamers assembled face to face.  
 b) Side view c) top view and d) cut-open view of the didecamer. Note in b) the shape of the subunit dimer (coloured). Note in d) the architecture of the collar-complex. The outer collar complex consists of five slabs, representing FU-h pairs (blue). Ten copies of FU-g constitute the internal collar complex (arc; cyan). (according to Meissner *et al.*, 2000 and Gebauer *et al.*, 2002).

Furthermore, an isolated and stable 8 FU-subunit dimer from the chiton *Acanthochiton fascicularis* has been visualised by 3D electron microscopy (Harris *et al.*, 2004). The shape of the 3D-reconstruction can be compared with the HtH1 subunit dimer extracted from the 3D

reconstruction of the didecamer. However, single FUs within the two subunits could not be clearly resolved because of the limited resolution (~3nm). Thus, the authors could not define the exact arrangement of the functional units within the dimer. Although an updated and convincing model has already been proposed (Meissner *et al.*, 2000, Gebauer *et al.*, 2002), the precise subunit pathway remains still obscure.

A detailed structural comparison of the cephalopod and gastropod hemocyanins could not yet be performed, due to the limited resolutions of 3D reconstructions of cephalopod hemocyanins (< 18 Å, (Fig 4, c-d)). Such a comparison would be crucial, in order to draw conclusions for their unique structural features.

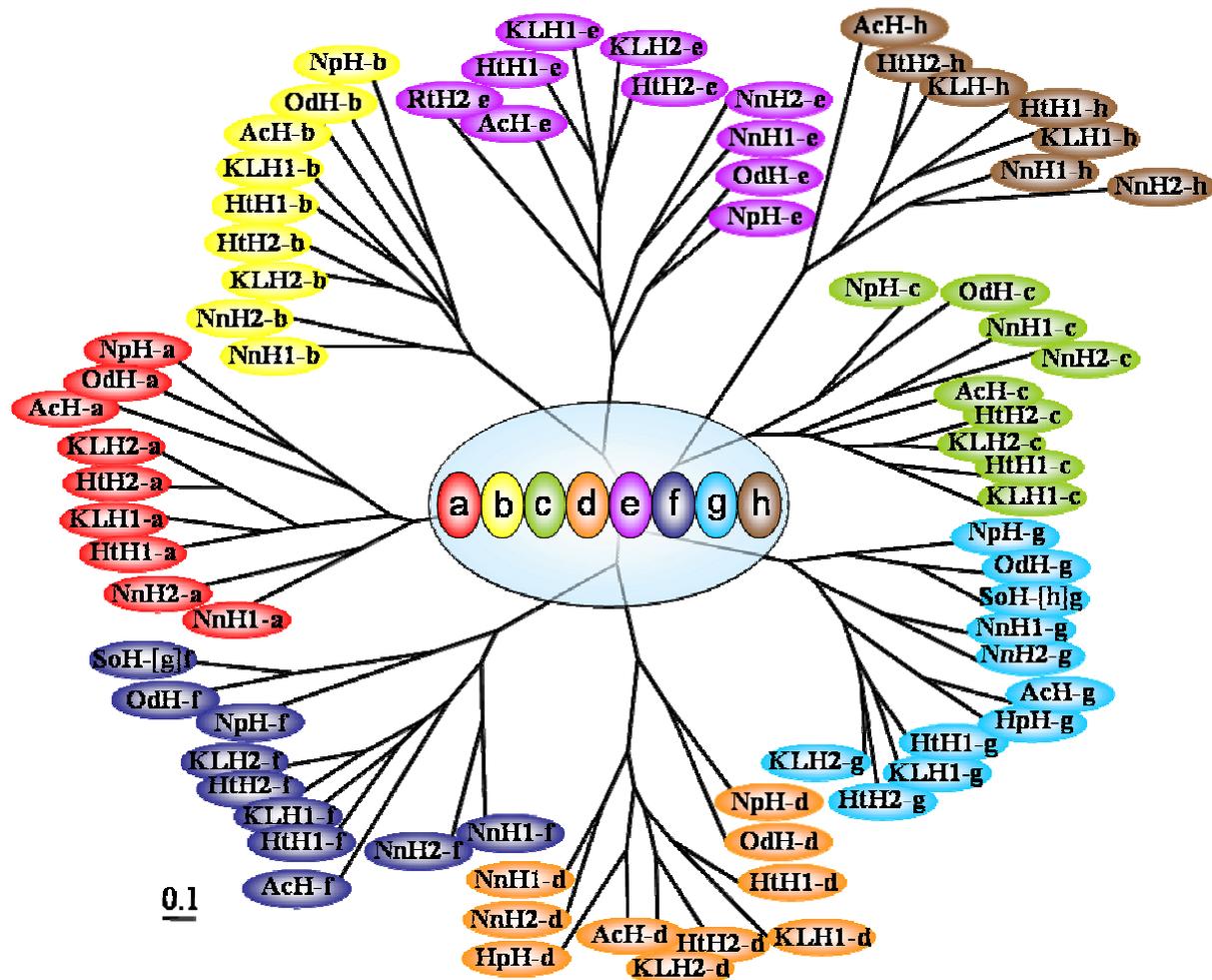
A number of polypeptide chains from different molluscan taxa has been completely sequenced, at both the cDNA and genomic level (Miller *et al.*, 1998; Lieb *et al.*, 2000, 2001, 2004; Altenhein *et al.*, 2002; Bergmann *et al.*, 2006, 2007).

From these data, it can be demonstrated that the hemocyanin polypeptide a-b-c-d-e-f-g of *Octopus* and *Nautilus* (Cephalopods, *Octopus*-type) is orthologous to the polypeptide fragment a-b-c-d-e-f-g from gastropods such as *Haliotis*, *Aplysia* and *Megathura*. The phylogenetic tree (Fig 6) shows that the eight different FUs evolved prior to the divergence of this animal phylum into different different classes. For example, all FUs-a from the different hemocyanin subunits, are more closely related to each other than they are to other FU-types from the same subunit. The sequence identity is ~55 % percent between orthologous FUs, but only ~45 % between paralogous FUs (Lieb *et al.*, 2000; Bergmann *et al.*, 2006). Therefore, corresponding FUs are likely to occupy the same positions and play the same architectural role within the quaternary structure of their respective hemocyanins. However, this has not yet been demonstrated directly.

The sequence data provide strong evidence that the eight different FUs, that form the subunit, evolved ca. 740 million years ago by three subsequent events of gene duplication and fusion from a precursor protein that contained a single copper active site (Decker *et al.*, 2007, Markl and Lieb, 2004). The fact, that within the cephalopoda the terminal FU-h is missing, is clearly an apomorphy. In *Sepia*, the additional FU evolved, by duplication of FU-d (Markl and Lieb, 2004, Loncke *et al.*, 1990).

Comparing the different species, it can be clearly shown, that several gastropod and bivalve hemocyanins are heterogeneous in that they consist of two (or even three) immunologically very distinct protein isoforms that are differentially expressed (Streit *et al.*, 2005).

These isoforms have arisen independently during the evolution of the various molluscan orders and classes (Lieb and Markl, 2004) and are sometimes difficult to distinguish from each other (Brouwer *et al.*, 1978).

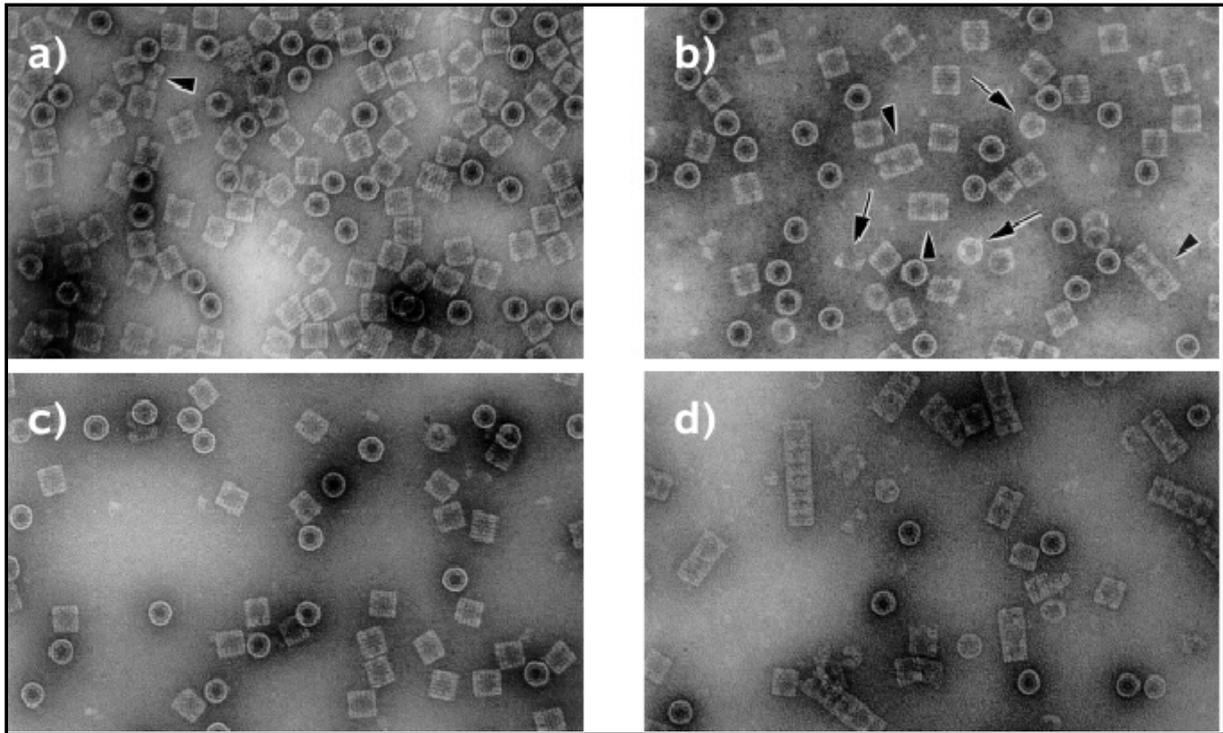


**Fig. 6: Phylogenetic tree of the molluscan hemocyanins functional units** (according to Decker *et al.*, 2007)  
 The eight different types of FU form eight distinct branches. NpH = *Nautilus pompilius* hemocyanin; OdH = *Octopus doffeini* hemocyanin; SoH = *Sepia officinalis* hemocyanin; NnH = *Nucula nucleus* hemocyanin (two isoforms); HtH = *Haliotis tuberculata* hemocyanin (two isoforms); KLH = keyhole limpet (*Megathura crenulata*) hemocyanin (two isoforms); RtH = *Rapana thomasiana* hemocyanin; AcH = *Aplysia californica* hemocyanin; HpH = *Helix pomatia* hemocyanin.

Two different hemocyanin isoforms are present in the giant keyhole limpet *Megathura crenulata* (Gebauer *et al.*, 1994, Söhngen *et al.*, 1997, Harris and Markl, 1999) (KLH1 and KLH2), in the abalone *Haliotis tuberculata* (Keller *et al.*, 1999, Lieb *et al.*, 1999, Harris *et al.*, 2000) (HtH1 and HtH2), in *Rapana Thomasiana* (Idiakieva *et al.*, 1993, Gebauer *et al.*, 1999) (RtH1 and RtH2) and also in the bivalve *Nucula nucleus* (Bergmann *et al.*, 2007) (NnH1 and NnH2).

The sequence identity between the two hemocyanin isoforms is around 65 %. The capability of these hemocyanins (notably from marine gastropods) to form didecamers, tridecamers and multidecamers is probably restricted to specific hemocyanin isoforms. Reassociation studies, performed using KLH1 and KLH2 and also HtH1 and HtH2 (Harris *et al.*, 1996; Harris *et al.*, 2000), demonstrated that the two distinct isoforms possess differing dissociation

properties (see Fig. 7). In purified samples of KLH1, KLH2, HtH1 and HtH2, the didecamer predominates, but tri- and longer multidecamers are detectable only in KLH2 and HtH2 (Harris *et al.*, 2000). The physiological functions of these various isoforms are still a matter of debate. Expression data of *Haliotis asasina* isoforms, demonstrate that they are differentially expressed during development, suggesting that the two isoforms play different physiological roles (Streit *et al.*, 2005).



**Fig. 7: TEM-Micrographs of oligomeric states of the two isoforms of *Haliotis tuberculata* hemocyanin (HtH21 and HtH2)**

a) Purified HtH1. Hollow cylindrical didecamers predominate (top- and side views). The arrowhead indicates the presence of a decamer b) Purified HtH2. Note the presence of didecamers together with a significant proportion of short multidecamers (arrowheads). Arrows indicate the presence of decamers  
 c) Reformation of HtH1 decamers into didecamers by the addition of 100mM concentrations of calcium and magnesium chloride. d) After addition of 100mM concentrations of calcium and magnesium chloride, HtH2 decamers form didecamers, multidecamers and short tubular polymers. (Harris *et al.*, 2000).

## 2.1 Available X-ray structures of functional units

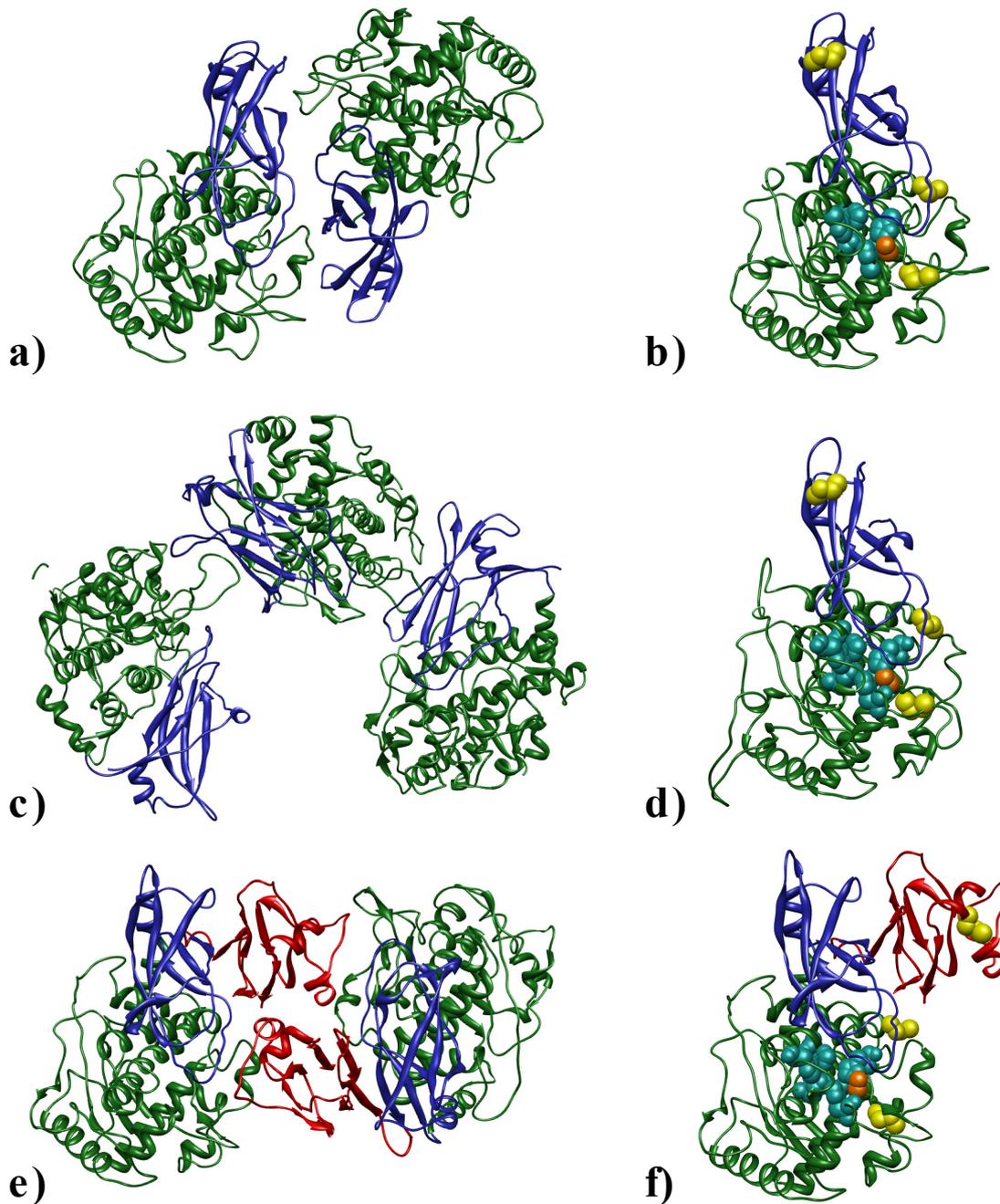
The first X-ray structure of an isolated functional unit of a molluscan hemocyanin was presented by Cuff *et al.* (1998). The carboxyl-terminal 47 kDa arc forming FU from *Octopus dofleini* hemocyanin (OdH-g) (Fig. 8 a,b), was solved at 2.3 Å resolution. The blue color of the crystals indicated an oxygenated state of hemocyanin in the crystal structure. OdH-g has two structural sub domains: **a largely  $\alpha$ -helical domain (core domain)** (which also contains the oxygen binding site) and **a five stranded  $\beta$ -sandwich domain** (see Fig. 8).

As already mentioned, the hemocyanin active site consists of six histidine residues which coordinate two copper atoms (CuA and CuB). The CuA- CuB distance is  $\sim 3.5$  Å. An unusual thioether bridge is formed between a cysteine residue and the second histidine of CuA. This bridge seems to be a general feature for functional units of molluscan hemocyanins and can contribute to the cooperative modulation of the oxygen affinity (Gielens *et al.*, 1997, Gielens *et al.*, 2007). This covalent bond occurs in a region devoid of regular secondary structure and at the core- $\beta$ -sandwich interface. Therefore, it appears to anchor this histidine residue to the active site. A strictly conserved leucine residue protruding from the  $\beta$ -sandwich domain contacts the second histidine of CuA at van der Waal's distance. This residue may have a significant role in detecting inter-domain motions and imparting that information directly to the active site, or vice versa (Cuff *et al.*, 1998). Further inter-domain interactions (such as salt bridges) could affect movements of the  $\beta$ -sandwich domain against the core domain, which are believed to modulate oxygen affinity. In addition, the core domain contains two disulfide bridges, close to the active site, which may add stability to the domain (Topham *et al.*, 1999). The first one stabilizes a  $\beta$ - $\beta$  hairpin loop, close to important active site residues of CuA, whereas the second one anchors a loop to an  $\alpha$ -helix, which provides two of the three histidines of CuB. A third disulfide bond, remote from the oxygen binding-site, connects two segments of the  $\beta$ -sandwich subdomain.

In addition, OdH-g forms dimers during crystal formation. The large interface between the two protomers involves a combination of hydrophilic and hydrophobic interactions. The residues involved in the dimer-interface are not strictly conserved (in contrast to the domain interface). Cuff *et al.*, proposed that this kind of pair-wise association might connect wall FUs from two neighboring subunits. However, this has not yet been demonstrated directly.

New insights have been provided by the X-ray- structure of the wall functional unit e of the prosobranch gastropod *Rapana thomasiana* (RtH2-e, Perbrandt *et al.*, 2003) (Fig. 8 c, d). The basic architecture of this FU is directly comparable with OdH-g. However, RtH2-e was crystallized in deoxygenated-form, which allowed a comparison of molluscan HC FUs in the deoxygenated and oxygenated state (deoxy-RtH2-e vs oxy-OdH-g). A prominent difference is a tunnel found in the structure of RtH2-e, leading from the protein surface to the active site. Most probably, this tunnel forms the entrance pathway for dioxygen molecules. It can be

supposed that the tunnel is closed after oxygen binding; however, no such tunnel was found in the structure of oxy-OdH-g.



**Fig. 8: Crystal structure of OdH-g (*Octopus dofleini*, (a, b)), Rth2-e (*Rapana thomasiana*, (c,d)) and KLH1-h (*Megathura crenulata* (e, f)).**

The N'-terminal  $\alpha$ -helical domain is shown in green, the  $\beta$ -sandwich domain in blue. KLH1-h contains an additional C-terminal cupredoxin-like domain, which is displayed in red. The crystals in case of OdH-g (a) and KLH1-h (b) contain dimers, while Rth2-e (c) forms trimers which further dimerize to form a regular cylindrical structure. The active site consists of six histidine residues (cyan spheres) which coordinate two copper atoms. A conserved cysteine (orange spheres) forms a thioether bridge with an active-site-histidine residue, which can contribute to the cooperative modulation of oxygen affinity. Cysteine residues involved in disulfide bridges are displayed as yellow spheres. The third disulfide bridge in the area of the  $\beta$ -sandwich domain is not present in KLH1-h. In contrast, KLH1-h contains an additional disulfide-bridge in the area of the cupredoxin-like domain. Furthermore, the loop 47-60 (arrow) has a completely different conformation in OdH-g and Rth2-e.

Another major difference is that six protomers (arranged in two trimers) of the wall FU Rth2-e form regular cylinders during crystal formation, while the arc FU OdH-g, as already mentioned, forms dimers. According to Perbandt *et al.* (2003), this can be explained by the different location of the respective FUs (wall FU vs arc FU) and suggests how the wall FUs are arranged in the native molluscan hemocyanin. However, cryoEM 3D-reconstructions of native molluscan hemocyanins could not yet confirm this FU-arrangement, due to their limited resolution. Furthermore, in contrast to OdH-g, Rth2-e contains a metal ion binding site, which connects the three protomers, located in the asymmetric unit. An additional difference relates to a loop (residues 47-60), which has a completely different conformation in oxy-OdHg. This loop is close to the active site and stabilized by a disulfide bridge. This bond is also in contact with the oxygen binding site related thioether-bridge. Thus, oxygenation could affect the conformation of this loop. In addition, there are two binding sites for glycan side chains within Rth2-e, in positions different from that of the single glycosylation site in OdH-g.

More recently a 4 Å X-ray structure of the slab-forming functional unit h from the gastropod *Megathura crenulata* became available (KLH1-h) (Jaenicke *et al.*, to be published) (Fig. 8 e, f). As already mentioned, the C'-terminal FU of gastropods, carries a C'-terminal-extension of 100 aminoacids, which is absent in other FU types. The crystal structure of KLH1-h revealed that the c-terminal extension of FU-h shows a cupredoxin like type 1 copperfold. However, the presence of an additional copper atom still has yet to be verified and requires atomic resolution.

Furthermore, similar to OdH-g, KLH1-h forms dimers during crystal formation, which might correspond to the slabs (outer collar morphological units composed of two copies of FU-h, see above), described in the 3D-reconstructions of native gastropod hemocyanins. However, a direct comparison of the x-ray-structure of the KLH1-h-dimer with a slab extracted from a 3D reconstruction of a gastropod didecamer has not been performed yet, because of the limited resolution of the available 3D-reconstructions.

In addition, the disulfide bridge in the area of the  $\beta$ -sandwich domain (conserved in OdH-g and Rth2-e) is not present in KLH1-h, and according to sequence data also in other FUs, and therefore may not play an essential role in oxygen binding (Topham *et al.*, 1999). An additional difference to the other two FUs relates to the location of the glycosylation sites.

However, the three crystal structures show a very high structural homology in the regions of the core and  $\beta$ -sandwich domain (RMSD values: 0.59 -0.64 Å). This was not surprising, since proteins that share more than 35% sequence identity (in this case the sequence identity between FUs is more than 40%), share similar structures (e.g. Launay and Simonson *et al.*, 2008). Moreover, these crystal structures can be taken as models for the other FUs, since

with more than 40% sequence similarity, the threshold of 35% sequence identity required for homology modelling, is exceeded.

## 2.2 Medical relevance of molluscan hemocyanins

Molluscan hemocyanins are not only of theoretical interest, but also of therapeutic and economic value (Herskovits and Hamilton, 1991; van Holde *et al.*, 1992). For more than 40 years, researchers have found that these giant extracellular respiratory proteins also have many other attributes. They are potent immunogens, which induce the synthesis of large amounts of specific antibodies (Herscovitz *et al.*, 1972). They enhance the host's immune response by interacting with T cells, monocytes, macrophages and polymorphonuclear lymphocytes (Tzianabos, 2000). A prominent example is KLH, which is known, to possess remarkable immunostimulatory properties (e.g. Curtis, 1970; Weigle, 1964). It appears to be a non-specific immune stimulant that induces both a cell mediated and humoral response. KLH was even introduced as a test for human immunocompetence in 1967 (Swanson and Schwartz, 1967).

The immune response also occurs after dissociation of the KLH didecamers into single subunits (Kippel *et al.*, 1991; Weigle *et al.*, 1964). Therefore, the large mass of native KLH is not the reason for its strong antigenicity. The significant factor is probably the multi antigenicity, provided by the 8 immunological different functional units found in each isoform, which each show a different glycosylation profile. These xenogen epitopes can challenge the immune system all at once and induce an enhanced immune response.

A safe KLH-subunit clinical product is commercially available (Immucothel®, biosyn Arzneimittel GmbH).

KLH has been used as a form of therapy for patients with bladder cancer for many years (Olson *et al.*, 1974, Jurincic *et al.*, 1988, Flamm *et al.*, 1990). It has been tested against mitomycin C chemotherapy in patients and was found superior in preventing bladder tumor recurrence, with no local or systemic side effects (Jurincic *et al.*, 1988; Riggs *et al.*, 2002). Wirguin *et al.* (1991), showed that native KLH contains Gal(beta 1-3)GalNAc-bearing oligosaccharides. Immunization with KLH induces the production of antibodies against this epitope, which cross-react with the Thomsen-Friedenreich (T) antigen, on the surface of bladder carcinoma cells. This could explain the effect of KLH immunotherapy in bladder carcinoma.

In addition, *in vitro* studies have previously shown that KLH inhibits cellular proliferation in human cancer cell lines of the breast, esophagus and prostate (Riggs *et al.*, 2002; Somasundar *et al.*, 2005). Further *in vitro* studies have shown that KLH alters tumoral cytokine production in both breast and pancreatic cancer, in part explaining its effectiveness in isolated culture systems (Riggs *et al.*, 2005). KLH has also been shown to decrease the

cellular proliferation of Barrett's adenocarcinoma and enhance apoptotic activity (Mc Fadden *et al.*, 2003; Vona-Davis *et al.*, 2004). Investigations of Somasundar *et al.* (2005) reported a significant reduction in cellular proliferation in a melanoma cell line *in vitro* (range 12% - 60% of the control), *via* early apoptotic pathways. More recently, significant growth inhibition of metastatic melanoma was also observed *in vivo*. These effects were significantly enhanced when KLH was administered in combination with conventional immunotherapy (IL-2, AIFN) (Riggs *et al.*, 2007).

Another prominent feature of KLH is a carbohydrate epitope that is cross-reactive with *Schistosoma mansoni* (Grzych *et al.*, 1987; Ko and Harn, 1987). *S. mansoni* is a trematode worm which causes the infectious disease Bilharziosis. With at least 200 million infected people, Bilharziosis constitutes one of the most severe problems of the tropics. Injection of KLH during infection induces a cross-reaction of anti-KLH-antibodies against the *Schistosoma*-glycolipids (Markl *et al.*, 1991). These antibodies can also be used for detection of the infection by ELISA testing. A terminal fucose residue linked to (α1→3) - position to N-acetylgalactosamine was found to be shared by schistosome glycoconjugates and KLH (Kantelhardt *et al.*, 2002). This epitope provides the basis for the application of KLH in serodiagnosis and as a potential vaccine against schistosomiasis.

In addition, KLH is also used as carrier for small molecules against which it has often proved to be difficult or impossible to raise polyclonal antibodies (reviewed by Harris and Markl, 1999).

In this context, it has been demonstrated that active immunization with a KLH immunoconjugate derived from the hapten GNC suppressed the psychostimulants effects of cocaine in rats and resulted in 80% decrease of brain cocaine levels compared with controls (Carrera *et al.*, 1995). Recently, a hapten-KLH anti-nicotine vaccine has been developed, which significantly decreased the nicotine levels in the brains of immunized rats (Carrera *et al.*, 2004). Another prominent example is Biovaxid, which is a patient-specific therapeutic cancer vaccine composed of ID conjugated to KLH. ID is a clonal immunoglobulin molecule, expressed on the surface of B-cells malignancies and can function as a tumor antigen. Currently, a phase III clinical trial is ongoing to determine the clinical benefit of Biovaxid (Lee *et al.*, 2007). KLH has also been used as a component of experimental synthetic minimal viral vaccines against AIDS (Naylor *et al.*, 1991) and papilloma virus (Meyer *et al.*, 1998). The purified mixture of KLH1 and KLH2 subunits for antigen/hapten conjugation is commercially available (Vacmun®, biosyn Arzneimittel GmbH).

Hemocyanins from other gastropods, such as the abalone *Haliotis tuberculata*, are considered to be possible substitutes for KLH as immunostimulants (Markl *et al.*, 2002). Recently, hemocyanin obtained from the Chilean gastropod *Concholepas Concholepas* was reported to possess adjuvant immunostimulatory effects, as well as a significant anti-tumor

activity against mouse bladder carcinoma cells (Oliva *et al.*, 2002, Moltedo *et al.*, 2006). The hemocyanin from the gastropod *Rapana thomasiana* (RtH) can also be used in different immunization protocols as an adjuvant or as a protein-carrier (Tchorbanov *et al.*, 2008). Furthermore, according to recent studies, RtH2 is a promising anti-herpes agent, especially against genital herpes virus (Genova-Kalou *et al.*, 2008).

Despite these advances and although the primary sequence of KLH is already available, accumulation of fundamental knowledge of its high resolution quaternary structure has been rather slow. A molecular model of KLH or another closely related molluscan hemocyanin is still lacking.

### 3 CryoEM and Single particle analysis

Structural characterization of macromolecules and multi-subunit complexes is an essential component in understanding complex biological processes. X-ray and NMR analysis are currently the predominant sources of high resolution structural information. Indeed the main advantage of X-ray crystallography is that of superior resolution, compared to TEM. The keys for successful X-ray analysis are to grow large, well-ordered crystals and to obtain isomorphous heavy atom derivatives for the initial phasing of the diffraction data. Finding suitable conditions can be a labor-intensive process as the conditions for successful crystal growth cannot be predicted. Given the difficulties of applying X-ray crystallography to large assemblies (such as hemocyanins) or heterogeneous proteins, a key role is currently played by cryo-electron microscopy (cryoEM) and single particle analysis of the digital images.

Using a small amount of purified protein, isolated vitrified molecules can be observed within the electron microscope and their projection images combined to produce a 3D reconstruction. No crystals need to be grown and thus the specimen preparation and data collection can be fast and simple. The individual molecules are not constrained by the crystal contacts, which may introduce artifacts, and can be imaged in near-native conditions as well as in different functional states. Furthermore, the “phase” problem is absent in EM because it produces projection images of the analyzed particles and not diffraction patterns.

Using the new generation of electron microscopes equipped with field emission gun electron sources, resolution levels up to 0.7 Å are theoretically possible. However, there are various experimental limitations concerning the resolution in the analysis of biological macromolecules. The single particle images are always noisy, because low levels of electrons are used during imaging, in order to avoid radiation damage and destruction of the “native” structural information. Furthermore, to increase the exposure dose and therefore the signal to noise ratio and also preserve the “native” structure of the protein, the sample is quickly vitrified in liquid ethane and kept at liquid nitrogen or helium temperatures. In order to

secure sufficient contrast of the particle over the background, the imaging takes place under focus.

To overcome this inherent noise problem, one is forced to image and combine a very large number of copies of the molecule and then average similar projections to improve the signal to noise ratio. Furthermore, the processing of thousands of images requires very substantial effort and computer power. In addition, the consequences of under focus and other optical aberrations which are introduced during imaging need also to be corrected.

With the new generation electron microscopes, the more powerful and sophisticated algorithms and software platforms and the increased speed of computing, single particle analysis has reached maturity. The strength of this method has already been demonstrated in a number of studies, from which the 3D structure of several macromolecular complexes has been determined at near –to atomic resolution ( $\sim 3.8 - 4.5 \text{ \AA}$ ) (Yu *et al.*, 2008; Zhang *et al.*, 2008; Ludtke *et al.*, 2008; Jiang *et al.*, 2008). At this resolution level, the amino-acid backbone can be traced.

However, due to instrumental or sample specific limitations, the vast majority of 3D-reconstructions obtained by cryoEM is currently limited to intermediate resolutions ( $6 - 15 \text{ \AA}$ ). In these cases, the combination of X-ray crystallography and cryoEM offers a unique opportunity to obtain high-resolution-models of the analyzed structures. Fitting of atomic models of assembly components or related structures into the EM density map may provide pseudo-atomic models for the whole assembly and thus significant insights into the structure, function and dynamics of single proteins and their complexes. The following study is based on such a hybrid approach.

(For reviews see: Orlova and Saibil, 2004; Zhou, 2008; van Heel *et al.*, 2000; Llorca, 2005; Thuman-Commike, 2001; Ruprecht and Nield, 2001).

## **4 Objectives of the present work**

In the last decade, a detailed understanding of how molluscan hemocyanins are organized, from quaternary to primary structure, has been gained. In a series of recent publications, the complete gene sequence and primary structure of various hemocyanins from different molluscan classes has been presented. The crystal structures of single functional units have provided a detailed view of the steric arrangement of the polypeptide chain, elucidating in part the relationship between structure and function, and due to high sequence homologies they can be taken as models for other FUs. Furthermore, 3D-cryoEM-reconstructions of multi-subunit-assemblies have become increasingly detailed through the past decade. However, the intermediate resolutions and the lack of secondary structure elements made it impossible to determine the exact orientation of the FUs within the 3D-reconstructions.

Therefore, docking of FUs (available X-ray structures or homology models) within the lower resolution EM density maps was unfeasible and molecular models could not be constructed. Thus, although complete sequence information is now available from several molluscan hemocyanins, many details of the quaternary structure are still unclear, including the topology of the 10 subunits within the decamer.

The main goal of the present work was to provide a superior resolution 3D-reconstruction (>10 Å resolution) of a molluscan hemocyanin, which should allow high quality molecular fitting and construction of a reliable molecular model, leading to a better understanding of the complex folding of the subunits and the arrangement of the FUs within the subunit dimer.

As first approach, our attention turned to the hemocyanin from the cephalopod *Nautilus pompilius* (NpH). Recently, the primary structure of NpH became available (Bergmann et al., 2006). According to these data and earlier TEM-studies, it was clear that NpH is an *Octopus*-type hemocyanin (decamer, based on a 7 FU subunit (a-b-c-d-e-f-g)), and therefore resembles the basic structure of molluscan hemocyanins.

The second aim of this study was to provide a sub-nanometre reconstruction and the first molecular model of a gastropod hemocyanin didecamer, based on a 8 FU-subunit (a-b-c-d-e-f-g-h). In this context, we focused on the structure of keyhole limpet hemocyanin (KLH), in particularly KLH isoform 1. As already mentioned, KLH has been applied for several decades in immunological basic research and also has been successfully used in clinics. Although the primary sequence of KLH and a low-resolution cryoEM-structure are already available (Orlova et al., 1997; Mouche et al., 2003; Lieb & Markl, 2004), many fundamental aspects of its quaternary structure are still unclear.

Moreover, comparison of highly resolved reconstructions and corresponding molecular models of KLH and NpH should substantially clarify the difference between the cephalopod hemocyanin decamer and the more complex gastropod hemocyanin didecamer.

In addition, a secondary goal of this study was to refine the available cryoEM-structure of HtH1, solved within our group at ~14 Å resolution (Meissner et al., 2000). HtH1 and KLH1 are closely related orthologous isoforms and at 14 Å their quaternary structures are almost indistinguishable (Mouche *et al.*, 2003; Meissner *et al.*, 2000). A detailed comparison of their quaternary structures at higher resolution might reveal minor differences, but also demonstrate the reliability of the method used for reconstructing these proteins in our laboratory.



## B MATERIALS AND METHODS

### 1. Materials and equipment

#### 1.1 Animals

##### *Megathura crenulata*

*Megathura crenulata* (Sowerby, 1852), commonly known as the giant keyhole limpet, is classified as follows:

Phylum: Mollusca

Class: Gastropoda

Subclass: Orthogastropoda

Superorder: Vetigastropoda

Superfamily: *Fissurelloidea*

Family: *Fissurellidae*

Genus: *Megathura*

Species: *Megathura crenulata*

The natural habitat of *M. crenulata* is the pacific coast of California and Mexico. Keyhole limpets adhere to rock surfaces with their large muscular flattened foot and live in the low-tide line, between 3 and 12 m of depth, but they are also common on breakwaters. They live as herbivores or scrapers of small detritus (algae, tunicates, sponges).

The source of their name is because of a single large opening in the center of the shell's radiating stripes (keyhole), which allows waste products and also sperm and eggs to exit. The internal organs of keyhole limpets are also modified to permit an exit channel through the body, leaving at the top. Another characteristic of the keyhole limpets is the outer mantle, which completely covers the shell, leaving only the keyhole visible. This tissue may be of cream, grey or black color and can also be striped. The shell is usually light tan in color and has a rough texture.

The animals used for this work were kindly donated by the Biosyn Company, Fellbach, Germany. The animals were maintained in a 300 l aquarium at a constant water temperature of 17°C with alternating 12-hr periods of faint light and complete darkness in the Institute of Zoology at the University of Mainz. The snails were fed once weekly with zucchini and maintained good appetite, health and mobility.



**Fig. 9 : The giant keyhole limpet *Megathura crenulata***  
 (Source: Kai Büchler, left picture and www.diver.net, right picture)

### ***Nautilus pompilius***

*Nautilus pompilius* (Linnaeus, 1758) is classified as follows:

Phylum: Mollusca

Class: Cephalopoda

Subclass: Nautiloidea

Order: Nautilida

Family: *Nautilidae*

Genus: *Nautilus*

Species: *Nautilus pompilius*

Nautiluses are the most ancient lineage among extant cephalopods and represent the only living members of their subclass, and therefore are considered to be “living fossils”. Among the recent cephalopods, only the nautiloids have an external shell, forming a geometrical spiral and consisting of an organic matrix of calcium and other mineral compounds. The shell is divided into chambers, which are connected by the siphuncle, a strand of living tissue enclosed in a chitin tube (Westermann *et al.*, 2004). The animal occupies only the outer, most recently formed, chamber.

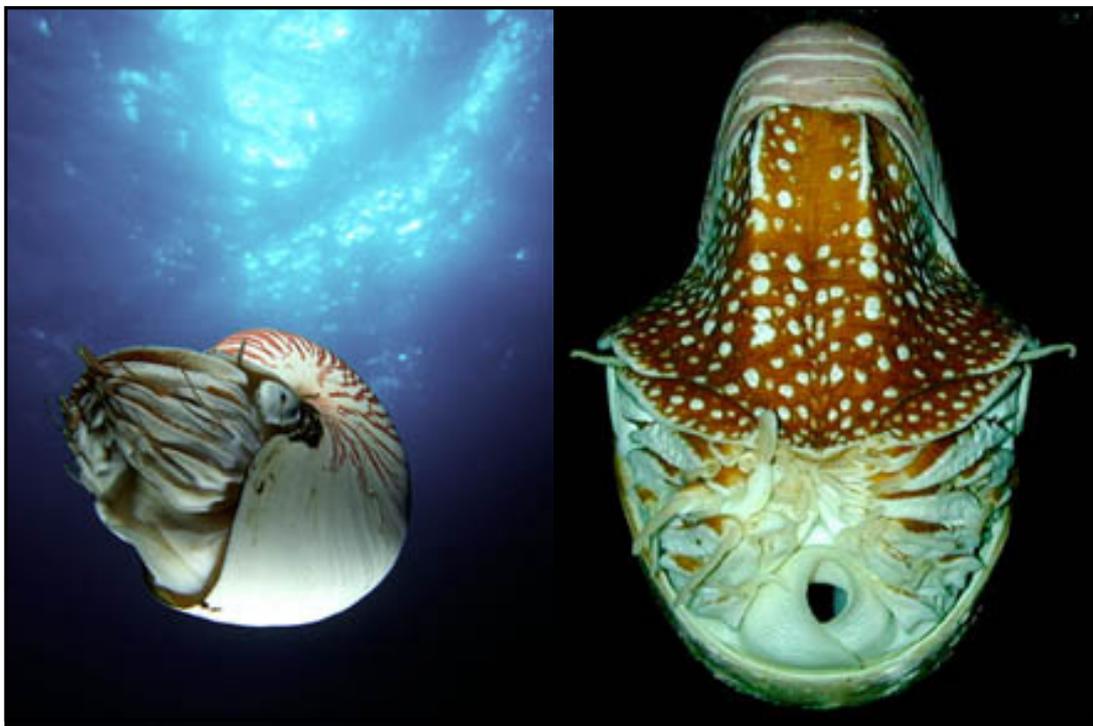
Nautiluses are found in the waters of the tropical western Pacific. They inhabit the deep slopes of coral reefs and are capable of tolerating dramatic changes in both pressure and temperature (depth range: 450-90 m). However, a depth range of ~300m can be exceeded only for short periods of time (Saunders and Ward, 1987). Buoyancy is controlled by pumping fluid in and out of the chamber along the siphuncles using an osmotic mechanism.

When water is pumped out, the animal adjusts its buoyancy with the gas contained in the chamber.

In order to swim, *Nautilus* makes use of “jet-propulsion”. They alternately pull water into the mantle cavity within the shell and blow it out through the muscular siphon beneath the tentacles. The muscular siphon is flexible and can direct the jet of water, thereby allowing *Nautilus* to swim forwards, backwards, or sideways.

Nautiluses feed on shrimps, small fishes and crustaceans, which are captured by their tentacles. The tentacles are quite different from them of coleoid cephalopods (squids, octopuses). Furthermore, unlike other cephalopods, *Nautilus* lacks a solid lens. The simple “pinhole” lens can probably permit detection of light and dark.

At the beginning of this study, hemolymph was kindly provided by [REDACTED] (AG Schipp, Justus-Liebig-University, Giessen, Germany). In addition, 3 specimens of living *Nautilus pompilius* were purchased from the Zoopalast Company, Wiesbaden-Biebrich, Germany. The animals were maintained in a 300 l aquarium in a closed sea water system at a constant water temperature of 17°C with alternating 12-hr periods of faint light and complete darkness in the facilities of the Institute of Zoology at the University of Mainz. The animals were fed with frozen fish once weekly.



**Fig. 10 : *Nautilus pompilius***

(Source: [http://tolweb.org/tree\\_group=Nautilidae&contgroup=Cephalopoda](http://tolweb.org/tree_group=Nautilidae&contgroup=Cephalopoda) (a) ©1996 Michael Vechionne (b) © Mark D. Norman (licence: <http://creativecommons.org/licenses/by-nc/3.0/>)

## 1.2 Electron microscopes

During this study three different transmission electron microscopes were used:

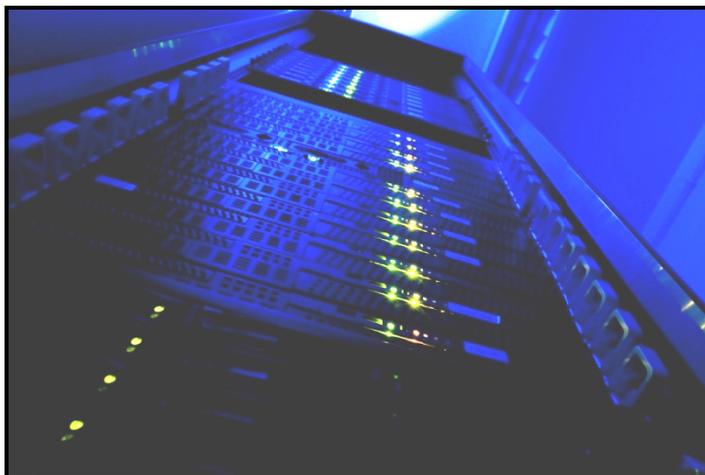
- Tecnai 30 FEG (F30) cryoTEM equipped with a field emission gun (FEG) (operated at an accelerating voltage of 200 or 300kV) (Jonannes Gutenberg University Mainz)
- Tecnai 20 FEG (F20) cryoTEM (200kV), equipped with a FEG (Johannes Gutenberg University of Mainz)
- Tecnai 12 LaB<sub>6</sub> (T12) cryoTEM (120kV), equipped with a LaB<sub>6</sub> filament (Jonannes Gutenberg University Mainz)



Fig. 11 : Tecnai 30 FEG at the Johannes Gutenberg University of Mainz  
(Picture by ██████████)

### 1.3 Computer-hardware

Image processing was performed on a “Beowulf” cluster with a 64 bit AMD Operon 2,4 GHz dual core processor (4 GB RAM) PC as headnode and twenty 64 bit AMD Opteron 2.2 GHz dual core dual processor blades (4GB RAM) (=80 processors) as nodes. The Beowulf cluster was purchased at transtec AG ([www.transtec.de](http://www.transtec.de), Tübingen, Germany).



**Fig. 12 : “Beowulf” cluster**  
(Picture by ████████ ef ████████)

Visualization and interpretation of the data and all non-parallel processes were performed on a Intel Core 2 Quad Processor at 2.4 GHz (4GB RAM) PC, equipped with a nvidia graphic card (8800GTX, PCIe, 768MB DDR3). This PC was acquired from Dell GmbH ([www.dell.de](http://www.dell.de), Frankfurt, Germany).

For demonstration and interpretation purposes, the structures were also visualized with a Cn 3D monitor (SeeReal Technologies, [www.seereal.com](http://www.seereal.com), Dresden, Germany). The monitor was connected to an Intel Core 2 Quad Processor at 2.4 GHz (4GB RAM) PC, equipped with an nvidia quadro graphic card. This PC was acquired from Dell GmbH.

### 1.4 Further hardware

A drumscanner type PRIMESCAN (Heidelberger Druckmaschinen AG, Heidelberg, Germany) was used for digitalization of the micrographs. The scanner was operated from a PowerMac 64 with a 400 MHz Processor (512MB RAM).

Solid 3D-models were produced by rapid prototyping using a Z Corporation 3D printer (Z 401) ([www.zcorp.com](http://www.zcorp.com), Boston, USA). Other apparatus is described elsewhere.

## 1.5 Software

**Linocolor** (6.0.12) is a software package designed to operate the PRIMESCAN drumscanner and was purchased along with the scanner.

**IMAGIC-5** is a modular image analysis software package aimed primary at single particle reconstruction. Typical operations are: fourier transforms, 2D (3D) alignments, multivariate statistical analysis, angular reconstitution and 3D reconstruction from 2D images. The various modules will be explained in detail in the following sections. The software was purchased at Image Science Software Gmbh ([www.imagescience.de](http://www.imagescience.de), Berlin, Germany) (Van Heel and Keegstra, 1981; van Heel *et al.*, 1996).

**EMAN 1.7** is similar to IMAGIC-5, a suite of scientific image processing tools with a particular emphasis to TEM structure determination. It provides functions for automatic and manual particle selection (boxer), Fourier processing, real space filters, alignment, classification, 3D reconstruction, projection etc. EMAN 1.7 is an open source software package and was acquired at <http://blake.bcm.tmc.edu/eman/> (Ludtke *et al.*, 1999).

**CTFFIND3** is a computer program that determines defocus and astigmatism in images of untilted specimens. The determination of lens defocus and astigmatism is needed to correct the measured data for the contrast transfer function (CTF) of the electron microscope. This software is freeware and was acquired at the website [http://emlab.rose2.brandeis.edu/grigorieff/download\\_ctf.html](http://emlab.rose2.brandeis.edu/grigorieff/download_ctf.html) (Mindell and Grigorieff, 2003).

The **ClustalW** and **ClustalX** programs carry out automatic biologically meaningful multiple alignments of divergent protein or DNA sequences. ClustalW uses a simple text menu system, which is portable to all computer systems, whereas ClustalX features a graphical user interface, aiding the interpretation of the alignments. Clustal W and X are freely available and were acquired at the website <http://www.clustal.org/download/> (Higgins and Sharp, 1988; Thompson *et al.*, 1994; Thompson *et al.*, 1997; Larkin *et al.*, 2007).

**Genedoc** is a software package for visualizing, editing and analyzing multiple sequence alignments of protein and nucleic acid sequences. Genedoc 2.6 is a freeware and was acquired at the website <http://www.nrbsc.org/gfx/genedoc> (Nicholas *et al.*, 1997).

**Modeller** (8v2, 9v1, 9v4) is an automated program for comparative modeling of protein 3D-structures and their assemblies by satisfaction of partial restraints. Comparative modeling predicts the 3-D structure of a given protein sequence (target) based primarily on its

alignment to one or more proteins of known structure (templates). Modeller was downloaded free of charge at the website [http://salilab.org/modeller/download\\_installation.html](http://salilab.org/modeller/download_installation.html) (Sali and Bundell, 1993; Fiser *et al.*, 2000; Eswar *et al.*, 2008).

**SWISS-MODEL** is a fully automated protein structure homology-modelling web server (<http://swissmodel.expasy.org/SWISS-MODEL.html>) (Guex and Peitsch, 1997; Arnold *et al.*, 2006).

**APSSP** is an advanced protein secondary structure prediction web server (<http://imtech.res.in/raghava/apssp/>), which allows prediction of the secondary structure of proteins from their amino acid sequence (Raghava, 2002).

**Molprobit** is a web server offering quality validation for atomic resolution structures of proteins, nucleic acids and complexes. It provides detailed all-atom contact analysis of any steric problems within the molecules and Ramachandran and rotamer distributions. This service is available free to all users at <http://molprobit.biochem.duke.edu> (Lovell *et al.*, 2003).

**PROCHECK** (v. 3.5.4) is a protein structure validation program, which provides an idea of the stereochemical quality of all protein chains in a given molecular model. It allows highlighting of regions which appear to have unusual geometry and provides Ramachandran distributions. PROCHECK is a freeware and was obtained from the website <http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html> (Laskowski *et al.*, 1993).

**SITUS 2.4** is a modular program package for the rigid body fitting of atomic resolution structures into lower resolution density maps, using a variety of fitting strategies. Situs 2.4 is freeware and was downloaded from the website <http://situs.biomachina.org/> (Wriggers *et al.*, 1999).

**MOLREP 9.3** is an automated program for molecular replacement and is part of the CCP4 programs suite. MOLREP is freeware and was obtained from the website <http://www.yz-bl.york.ac.uk/~alexei/molrep.html> (Vagin and Teplyakov, 1997, 2000).

**NMFF** is a software package that enables flexible fitting of large atomically detailed structures into electron density maps from cryoEM and other lower resolution methods. NMFF is available for download free of charge at the website <http://mm-tsb.org/software/nmff.html> (Tama *et al.*, 2004a; Tama *et al.*, 2004b).

**eINémo** is a web server (<http://www.igs.cnrs-mrs.fr/elnemo/>) for computing the low frequency normal modes of proteins. One major application of normal modes is the identification of potential conformational changes (Suhre and Sanejouand, 2004).

**USCF Chimera** is a highly extensible, interactive molecular graphics system, which provides both a graphical menu/window interface and a command line. Chimera is divided into a core, that provides basic services and visualization, and extensions, that provide most higher level functionality. For example, extensions include Volume Viewer, for display and analysis of density maps; Multalign Viewer, for showing multiple sequence alignments; Movie, for replaying molecular dynamics trajectories; Fit in Map, for fitting atoms into a density map or one map into another; Minimize Structure, for energy minimization of protein structures; Rotamers, for viewing and evaluating amino acid side chain rotamers; Color Zone, to color surfaces to match selected atoms and split the volume data by the resulting color zones and Movie Recorder, to capture image frames and assemble them into a movie file.

USCF Chimera is freely available for noncommercial use and was downloaded at the website <http://www.cgl.ucsf.edu/chimera/download.html> (Pettersen *et al.*, 2004; Goddard *et al.*, 2007).

**Amira – Resolve RT** (v 4.1) is a professional, general purpose, multifaceted tool for 3D data visualization, processing, analysis and presentation. The software was downloaded from <http://www.tgs.com> and the licence was purchased at TGS Europe (Düsseldorf, Germany).

**Coot** is a molecular graphics application with a particular emphasis on crystallographic macromolecular model-building, manipulation (idealization, real space refinement, manual rotation/translation, rigid body fitting, mutation, rotamers) and validation (Ramachandran plots, etc...). Coot is a freeware and was acquired at the site <http://www.ysbli.york.ac.uk/~emsley/coot/>.

**POV-Ray** is a ray-tracing program, used for 3D-image generation. Povray 3.6, was implemented in the OS SUSE Linux 10.3

**Operating systems** The Beowulf cluster was run on the Debian 3.1 free operating system. The PCs used for visualisation and non-parallel processes were run on SUSE LINUX operation system (versions 9 -10.3) or Windows XP Professional SP 2.

## 2. Biochemical Methods

### 2.1 KLH1-purification

KLH1 was purified as described in Harris *et al.*, 1995. In short, hemolymph was removed by inserting a sterile hollow needle (1mm diameter) into the pedal blood sinus. Prior to bleeding, the animal was immobilized by cooling on ice. Hemocytes were removed by centrifugation (Sorvall RC-5B, Du Pont Instruments, Frankfurt) at 800 g for 30 min at 4 °C.

The hemocyanin sedimented by centrifugation at 30000 g/ 4°C for ~17 h (Airfuge Ultracentrifuge, Beckman Optima L-70, Munich). The hemocyanin pellets were resuspended in a low-salt pH 7.4 stabilizing buffer (0.05 M Tris, 0.15 M NaCl, 5 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>). Overnight dialysis against 2 % ammonium molybdate at pH 5.7 resulted in a selective dissociation of KLH2 into its subunits, while KLH1 remains intact as didecamers and clusters of didecamers under these conditions. Separation of both KLH isoforms was then achieved by pelleting intact KLH1 at 30000 g for 4 h, leaving KLH2 subunits in the supernatant. The KLH1-pellet was then dissolved in stabilizing buffer.

The purification of KLH1 from *M. crenulata* hemolymph, was performed in our laboratory by

██████████.

### 2.2 NpH-purification

In order to allow periodic extraction of hemolymph from the same source animals, a non-lethal method was developed in our laboratory. In short, prior to bleeding, the animal was immobilized by cooling on ice. The hemolymph (~1 ml) was withdrawn from the region of the gills of a living *Nautilus pompilius* with a syringe. The animal was then returned to the aquarium for recovery.

The hemolymph was then centrifugated (Sorvall RC-5B, Du Pont Instruments, Frankfurt) at 1000 g for 5min. The hemocyanin molecules were then highly enriched from cell-free hemolymph by a single ultracentrifugation (126000 g/ 4°C for ~12 h), as described in Bergmann *et al.* (2006). The hemocyanin pellets were then dissolved in stabilizing buffer.

### 2.3 Determination of hemocyanin concentration and absorption spectrum

To determine the concentration of native hemocyanin an absorption spectrum was produced from 220 nm to 600 nm in an Ultraspec 3100 Pro UV/Visible Spectrophotometer. For the native hemocyanin, two characteristic absorption maxima occur, at 280 nm and 340 nm, respectively. The first maximum is typical for proteins, due to the absorption by amino acids with aromatic rings. An OD of 1 corresponds to a protein concentration of ~ 1mg/ml. The oxygenated active centre of native hemocyanin (Cu-O<sub>2</sub>-Cu-complex) has a typical absorption

maximum at 340 nm (van Holde, 1967) and therefore provides evidence for the quantitative presence of hemocyanin, when measured at this wavelength.

## **2.4 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. A single drop (5 $\mu$ l) 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  $\mu$ l droplets of water. The grid was then covered with a 10  $\mu$ l 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.

This technique was used to control the assembly-status of KLH1 after purification.

## **3. Cryo-electron microscopy**

### **3.1 Preparation of cryo-EM grids**

In our laboratory, a plunging apparatus for respiratory proteins has been developed (██████████) that enables their incubation and shock-freezing in different gaseous environments in an automated process.

In short, single droplets of the purified protein (7-8  $\mu$ l) (0.4 – 0.7 mg/ml) were applied to freshly glow-discharge treated holey carbon films, on copper EM grids. Excess fluid was directly blotted and the copper grids rapidly plunged into liquid ethane. In order to guarantee a defined oxygenation state of the hemocyanin molecules, the preparation before plunging and vitrification was performed in an atmosphere chamber. A constant stream of gas inside the chamber provided stable and controlled conditions during plunging. To obtain oxy-KLH1, aliquots were exposed to of 25% O<sub>2</sub> and 75% N<sub>2</sub>, prior to shock freezing. NpH was prepared in deoxygenated state, using a 100% N<sub>2</sub> gas stream.

This fully automated vitrification device guaranteed reproducible sample freezing and a high sample throughput. Details of the apparatus will be published elsewhere. CryoEM grids were stored under liquid nitrogen.

### **3.2 Data collection in the Cryo-Electron microscope**

Cryo-electron microscopic study of the vitrified grids was performed within 12h of preparation by ██████████. To maintain the low temperature and prevent the contamination of the grids with ice crystals, the vitrified cryoEM-grids were transferred into a cooled cryo-electron

microscope using a Model 626 single tilt cryo-transfer holder from Gatan (Gatan GmbH, Munich, Germany, [www.gatan.com](http://www.gatan.com)). The quality of the ice was assessed by manual inspection. Only grids with an even distribution of molecules were selected for data collection.

The cryo-electron micrographs were recorded as focal pairs. The focal pair technique is commonly used in TEM to compensate for low contrast micrographs. An initial image is collected at low dose, close to focus, followed by a second image of the same area further under focus. The first image contains higher resolution information at lower contrast, whereas the second far from focus image has high contrast but less reliable high-resolution information (Ludtke and Chiu, 2003). The far from focus images were used for picking particles, which were only poorly visible on the corresponding close-to-focus micrographs.

### **KLH1**

CryoEM of KLH1 was performed on a Tecnai F20 TEM equipped with a FEG and operating at 200 kV accelerating voltage. Images were recorded as focal pairs under low-dose conditions using Kodak So-163 film (8.3 x 10.2 cm) at 50,000x nominal magnification. The defocus was set between 0.8 and 4  $\mu\text{m}$ . Micrographs were developed according to the manufacture's recommendations, using full strength Kodak D19 developer.

### **NpH**

Images were recorded on Kodak SO-163 film at a magnification of 49,000x using a Tecnai F30 FEG electron microscope at an accelerating voltage of 200 kV. The images were recorded as focal pairs under low dose conditions and the defocus was set between 0.8 and 6  $\mu\text{m}$ . The negatives were developed in full strength Kodak D19 developer, according to the manufacture's recommendations

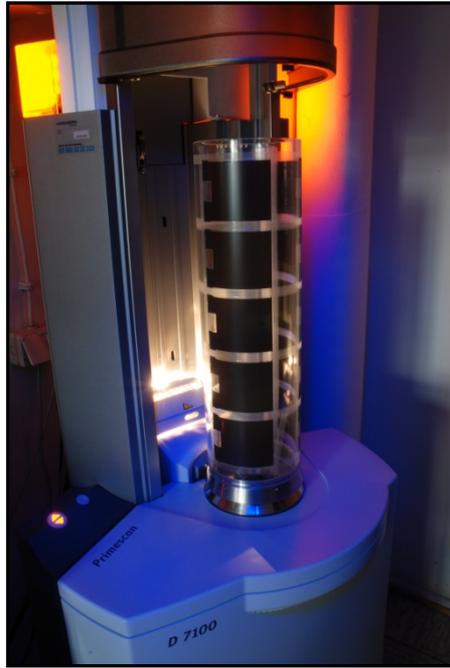
## **4. Bioinformatical Methods**

### **4.1 Scanning of micrographs**

Selected micrographs were scanned using a drum scanner (PRIMESCAN) and saved as 8 bit gray scale TIFF-files (images contain 256 possible shades of gray). The KLH1 micrographs were scanned with 6.2  $\mu\text{m}$  step size with a pixel corresponding to 1.24  $\text{\AA}$  on the micrograph. NpH micrographs were scanned with 9.1  $\mu\text{m}$  step size, with a pixel corresponding to 1.86  $\text{\AA}$  on the micrograph.

According to sampling theory, a continuous signal can be represented faithfully by a set of discrete samples if the signal is sampled at a rate of at least twice the signal's highest spatial frequency. In digital image processing, numerical errors make it necessary to use at least a

factor of 3 instead (LeBarron, 2008). With a sampling rate of 1.24 Å (KLH1) and 1.86 Å per pixel (NpH), the highest resolution possible from the raw micrographs is 3.7 Å and 5.6 Å respectively.

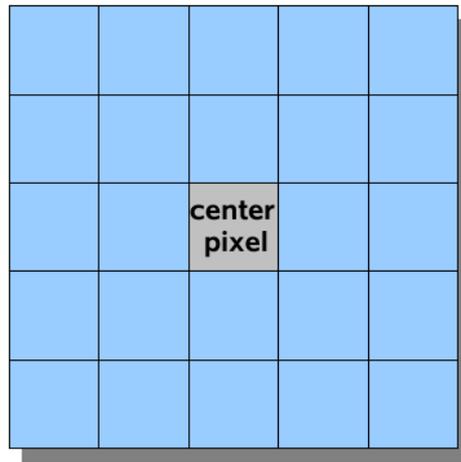


**Fig. 13 : PRIMESCAN drum scanner**  
(Picture by ██████████)

## 4.2 Particle selection

The images of the recorded micrographs, were decimated by  $\frac{1}{2}$ , and a low-pass block-convolution-2d-filter (IMAGIC-5, module block-2D) was applied with a linear size of convolution area of 5. This filtering operation produces an output image by averaging all pixels within the block-filter window. The larger the size of the window, the stronger the low-pass filter effect. The decimated, filtered-images, were used for picking particles, which were only poorly visible on the raw-micrographs.

Single particles were selected semi-automatically in the far from focus images, with the module boxer from the EMAN 1.7 software package. Then the particles of the lower contrast close-to-focus decimated micrographs were determined by mapping the boxes from the far-from-focus to the corresponding close-to-focus decimated micrographs, also using the automated process (Focal Pair Autoalign) included in the module Boxer. The coordinates of the particles were then multiplied by 2 to compensate for the initial decimation and each particle was windowed from the original raw-micrographs. The KLH1 and NpH particles were contained in arrays of 450x450 and 256x256 pixels respectively.



**Fig. 14 : Block-convolution 2D-filter**

A 5x5 convolution area (linear size: 5) was used for the two dimensional block convolution of the micrographs

### 4.3 CTF-correction

Image formation in the transmission electron microscope results from a combination of sample-induced elastic and inelastic electron scattering. The elastically scattered electrons that pass through the objective aperture produce the phase contrast that contains most of the structural information in cryo-EM of unstained specimens; the amount of amplitude contrast from these specimens is usually very small.

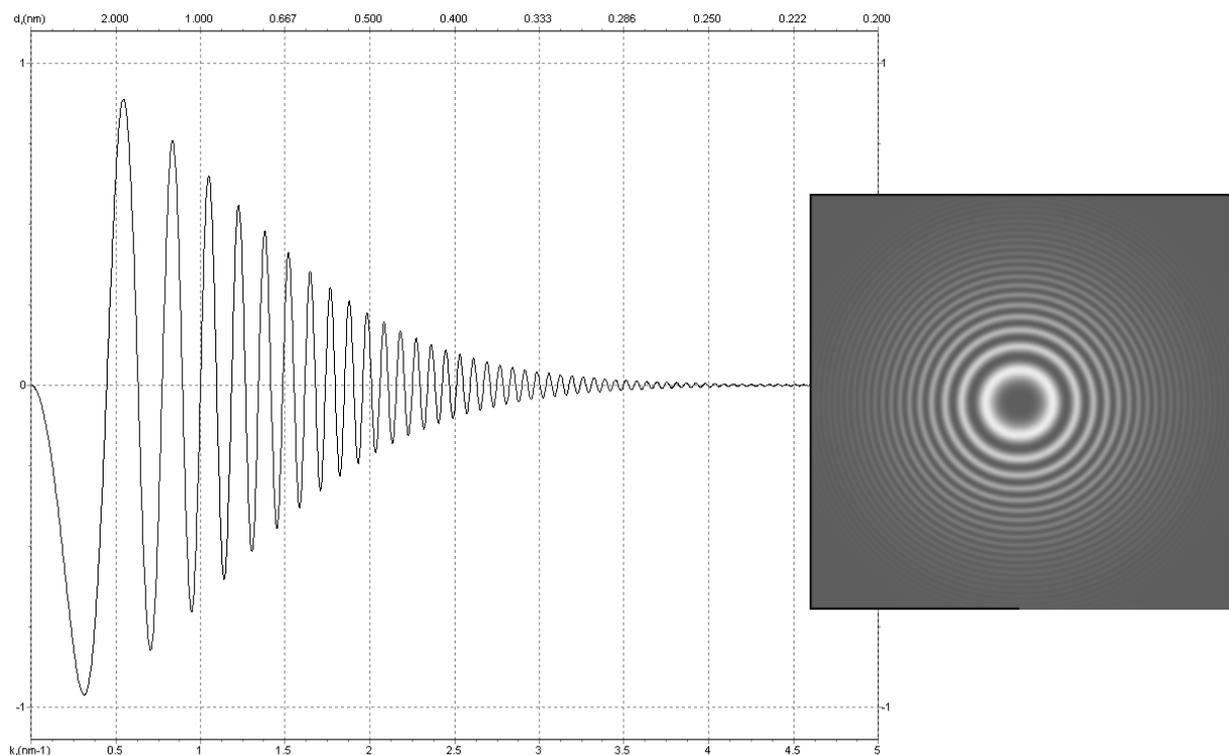
Unstained biological specimens are often well approximated as being weak phase objects and usually must be recorded with a relatively high amount of defocus in order to generate an adequate amount of phase contrast at low spatial frequencies. This however leads to a systematic alteration of the image data.

The image that is produced by the EM corresponds, unfortunately, to the convolution of the ideal projection image with a point spread function (PSF). The PSF is usually described by its Fourier transform that is commonly called contrast transfer function (CTF), which involves both phase- and amplitude contrast components.

The shape of the CTF depends on defocus, astigmatism, electron wave length (defined by accelerating voltage) and spherical aberration coefficient (for review see Wade, 1992). The effect of the CTF is to modulate the various frequencies of the signal and introduce spatial frequency dependent oscillations into the Fourier space representation of the image (see Fig. 15). These effects can be readily observed in the power spectrum, which exhibits a series of concentric ripples called Thon rings (Thon, 1966). The overall shape of the pattern is determined by the amount of axial astigmatism in the objective lens of the microscope. The CTF oscillates between -1 and +1 and the exact locations of the zero crossings (no contrast

transfer, information is lost) depend on the defocus. An accurate knowledge of defocus and astigmatism is therefore crucial for the correct determination of the image phases.

Furthermore, phase reversals occur with increasing rapidity at higher spatial frequencies and higher-resolution features become badly corrupted. The amplitude is attenuated towards higher resolution by an approximately Gaussian envelope function, as a result of a variety of factors (finite electron source size, energy spread of the beam, drift, etc.). The point where the envelope function damps the CTF to zero corresponds to the information limit of the microscope.

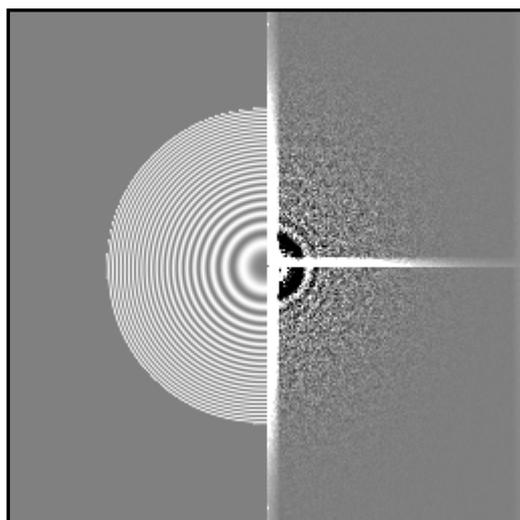


**Fig. 15: Contrast-Transfer-function CTF**

An example CTF generated using the program CTFEXPLORER. 1D-CTF profile, left picture, and 2D-CTF, power spectrum, right picture. The following parameters were used: Microscope Tecnai F20, defocus - 2000 nm;

In order to achieve highly resolved cryoEM structures, a restoration operation must be performed to correct for oscillations in the CTF at higher resolution.

In this study, the correction was performed by flipping the phases of the negative peaks of the CTF. The estimation of the defocus and astigmatism parameters of the micrographs was performed with the program CTFIND3. The diagnostic output images of the defocus estimation created by this program (see Fig. 16), were also used to evaluate the quality of the micrographs and to discard micrographs with astigmatism and/or drift (elliptical or hyperbolic thon rings).



**Fig. 16: CTFIND3 diagnostic output image**

An example diagnostic output image of a NpH-micrograph, showing the average background-subtracted power spectrum on the right, and the fitted CTF on the left.

The CTF correction was performed with the module *transfer* from the IMAGIC-5 software and applied to each extracted particle, based on the defocus parameter of its corresponding micrograph. No correction was performed for the envelope function.

It should be mentioned, that image restoration is never ideal, since the CTF has zero crossings and part of the information is lost. To compensate for this, images were collected at a series of different defocus settings. The goal was that the resulting CTF (after merging the ctf-particles from different micrographs) covers the whole Fourier space without information gaps.

#### 4.4 Bandpass filter

The raw CTF corrected images of the individual macromolecules were placed one after the other in a single huge file. This dataset was then bandpass-filtered in order to reduce the influence of unimportant or disturbing low and high spatial frequencies, for the subsequent reconstruction steps.

High spatial frequencies beyond the resolution limit mainly represent noise and hence can be suppressed. On the other hand, low-spatial frequencies (often associated with density-ramps etc.) may seriously disturb alignment procedures using cross-correlation functions.

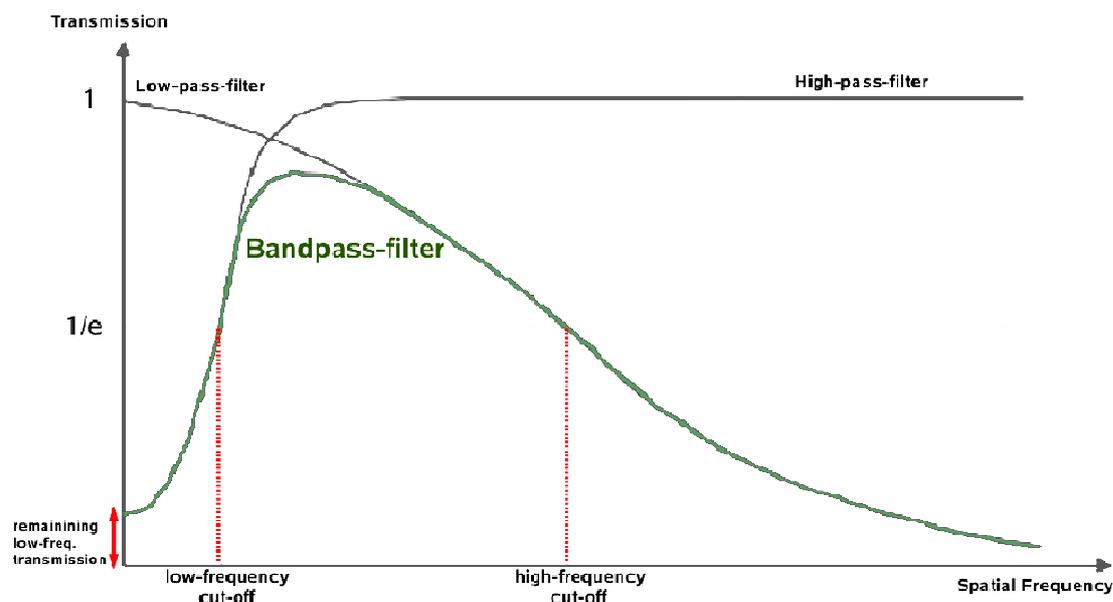
The applied band-pass filter is a double Gaussian filter in Fourier space. It is a product of a narrow Gaussian low-pass and a wide Gaussian high-pass filter.

The cutoffs were determined as follows:

High-pass cut-off:  $2 \times \text{Pixel Size } (\text{\AA}) / \text{particle size } (\text{\AA})$

Low-pass cut-off:  $2 \times \text{Pixel Size } (\text{\AA}) / \text{particle size } (\text{\AA})$

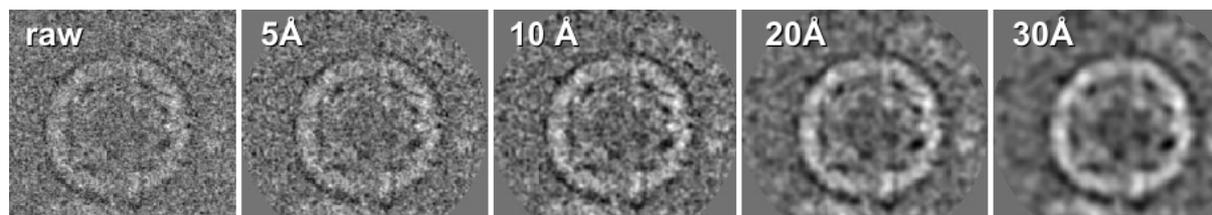
The high-pass cut-off is associated with the low-pass filter. Spatial frequencies higher than this frequency limit are gradually suppressed. The low-pass cut-off is associated with the highpass-filter, which gradually cuts off spatial frequencies lower than this value. The effects of band-pass-filtering at various high-frequency-cut-offs (low-pass-filter) are demonstrated in Fig 18. During the final refinements, the high-frequency cut-off was gradually shifted towards the high spatial frequencies.



**Fig. 17: Bandpass-filtering**

The bandpass-filter is the product of a Gaussian low-pass and a Gaussian high-pass filter.

In addition, after filtering, a soft edged circular mask was applied to all images. The densities in the output images outside the mask were set to zero, in order to reduce the influence of the surroundings of the molecules. Furthermore, within the mask, the data were normalized to a zero average density of gray values and an arbitrary variance of three. Band-pass-filtering and normalization was performed with the module *incore-prepare-filter* from the IMAGIC-5 software package.



**Fig. 18: Bandpass-filtering with respect to noise reduction**

An image of a KLH1-didecamer shown from the top as circle, band-pass-filtered at frequency limits of 450 (low-frequency-cutoff) and 5, 10, 20, 30 Å (high-frequency cutoff). The applied 2D-circular-mask has a radius of 0.99.

## 4.5 Multiple Reference Alignment (MRA)

The mixed dataset of the preprocessed (CTF-corrected, band-pass filtered) randomly orientated particles was aligned using a direct alignment algorithm (CCF) with respect to a large number of reference images.

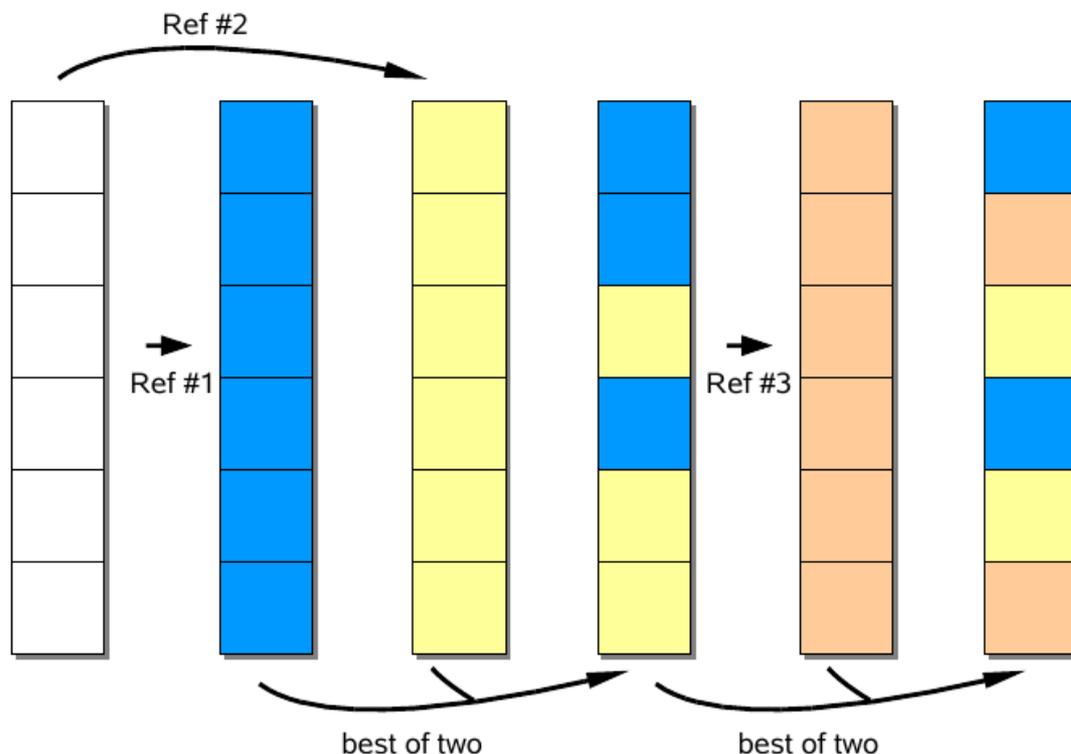
The aim of particle alignment is to modify the rotational and translational position of the noisy particles within the image such that all particles are positioned similarly within their individual images. This step is crucial for averaging identical projections in order to improve the signal to noise ratio of the images (see below).

In order to obtain reference images, usually a starting model is required. This 3D-model is reprojected into all directions (module *threed-forward*, IMAGIC-5), with an angular step determined by the resolution and the size of the data set. The set of the preprocessed images is then cross-correlated with the whole set of reference projections, so that the images to be aligned resemble the corresponding reference images.

However, 3D-Models give very different projections, with differing levels of contrast. Therefore, it is essential to normalize the reference images, otherwise the cross-correlations will be biased towards the references with the most contrast. This task was performed with the module *arithmetic-with-image* (option *threshold*) of the IMAGIC-5 software package. In addition, class averages may also be used as references for a multi-reference-alignment (see below).

In detail, in the first step of a MRA, all images of individual macromolecules are aligned with respect to reference no. 1, and a correlation coefficient is associated to each aligned image. Then the images are aligned to reference no. 2 and again correlation coefficients are associated. Aligned images with the best alignment quality (highest correlation coefficient, with respect to reference no. 1 or no. 2) are kept and the result is a merged data set with differently aligned images. The mixed dataset is then aligned to reference no. 3 etc. and this procedure is repeated for all references. (See Fig. 19). Aligning thousands of particles to thousands of reference images during MRA-procedures needs most of the computing resources during single particle analysis.

During the alignment procedure three orientation parameters of the image need to be found with respect to its corresponding reference: two translational parameters (X, Y) and the rotation angle. Rotational alignment finds the in-plane Euler rotation which aligns the single particle image with the most-similar reference image. The translational alignment finds the shifts which align the reference image with the single particle image. Although these processes are independent from each other, they are both used iteratively to optimize the alignment. Usually, each alignment iteration consists of a rotational alignment followed by a translational one.



**Fig. 19 : Procedure of the MRA**

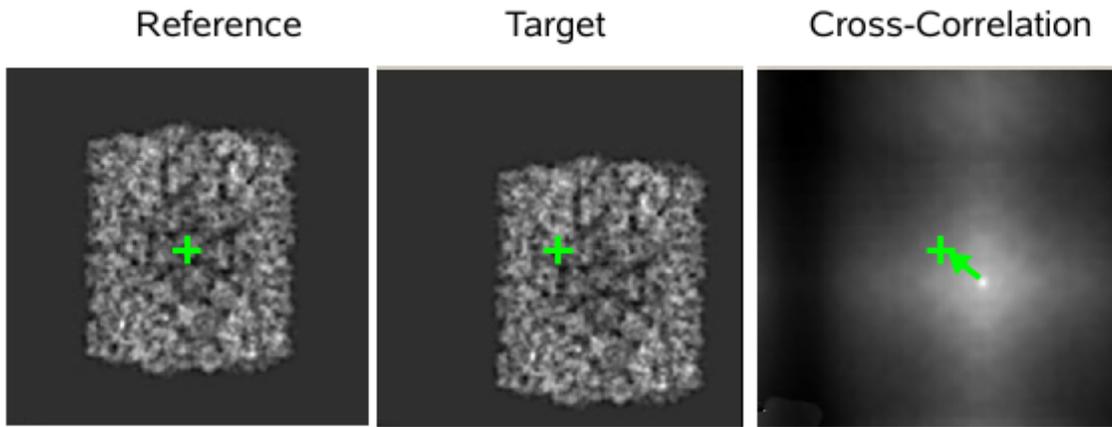
The resulting output dataset contains the aligned images associated with the highest correlation coefficients (see main text) (according to van Heel *et al.*, 2000 (modified)).

### **Translational alignment**

The experimental and reference image are initially Fourier Transformed (2D). The FT's are then multiplied and the result, a cross correlation function between both images (CCF, Saxton and Frank, 1977), is then transformed back to real space. The real space CCF contains a well defined peak in an offset position, shifted to the image center (Fig. 20). This shift of the peak can be described as a vector, which corresponds to the vector by which the experimental image appears to be shifted with respect to the reference image. Therefore, the search for the two translational parameters ( $X$ ,  $Y$ ) is thus reduced to the search for a sharp peak in the CCF.

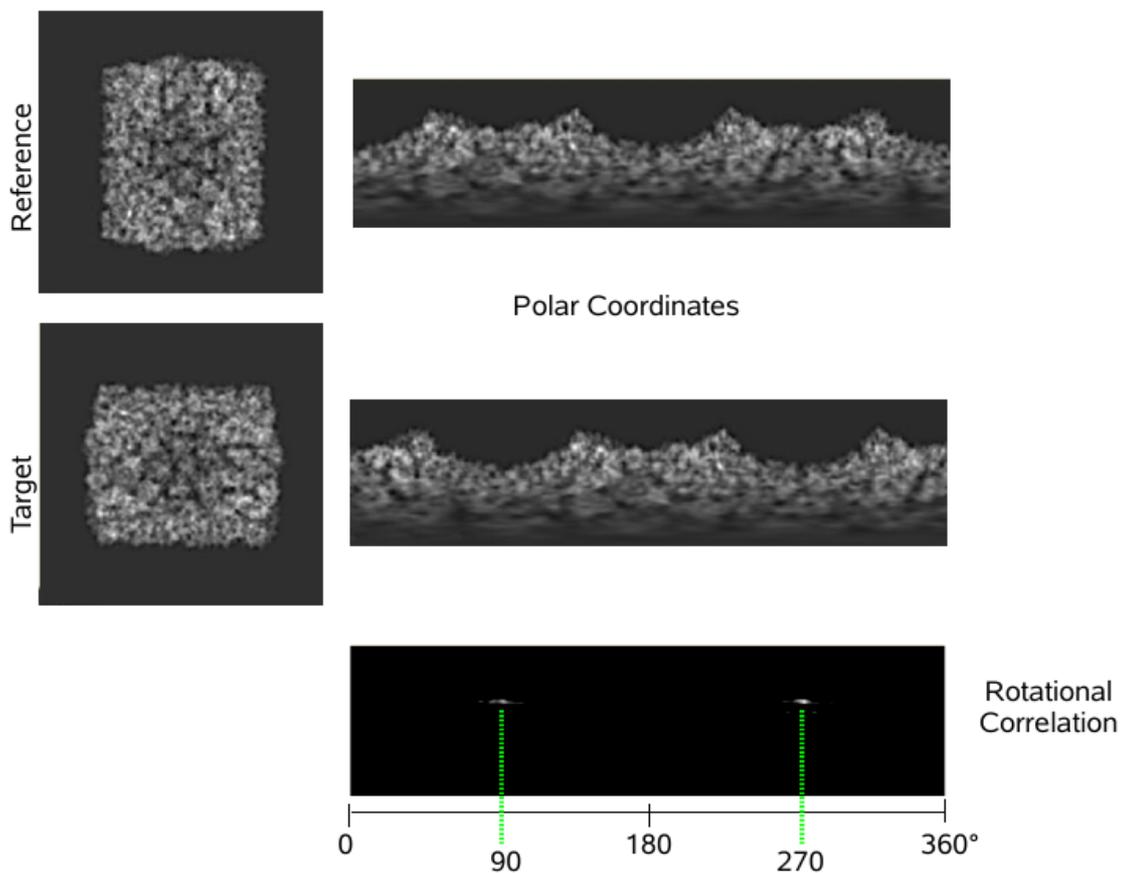
### **Rotational alignment**

Initially, experimental image and reference are displayed in cylindrical coordinate system and 1D-Fourier-Transformed. The result, a rotational correlation function (RCF, or angular correlation), is then 1D-reverse transformed. The resulting image in real space is the RCF plotted against the rotation angle. The RCF shows a maximum (or maxima, molecules with symmetry), which corresponds to the rotation angle by which the experimental image appears to be rotated in relation to the reference image (Fig. 21). Therefore, the search for the rotation angle is thus reduced to the search for a sharp peak in the RCF.



**Fig. 20: Example of a translational alignment of a target image against its reference.**

The real space-CCF is the product of the multiplication of the 2D-FFTs of target and reference image, back-transformed in real space. The CCF has a sharp peak that is shifted by a vector (green arrow) from the center of the image (green cross). The target is shifted in relation to the reference image according to this vector. (The images of target and reference shown here correspond to a side view reprojection of the final KLH1-reconstruction. The target image was shifted with the module rotate-image (option shift) of the IMAGIC-5 software package).



**Fig. 21: Example of a rotational alignment of a target image against its reference.**

Reference and target image (side view reprojection of the final KLH1-reconstruction) are represented in a cylindrical coordinates system (polar coordinates or Cyl. Coos). The real space RCF is the product of the multiplication of the 1D-FFTs of the Cyl. Coos and subsequent back-transformation. The RCF plotted against the rotation angle shows in this case two maxima (90 and 270°) (because of the object's symmetry) which correspond to the rotation angle by which the target was rotated (module rotate-image, IMAGIC-5) in relation to the reference image.

The multiple reference alignment (rotational and translational) was performed with the module *multi-reference-alignment* from the IMAGIC-5 software package.

#### **4.6 MSA- Classification**

Due to the very high radiation sensitivity of biological macromolecules, low levels of electrons are used during imaging. This, in turn, implies that the images are very noisy and hence the signal to noise ratio (S/N) low. To increase the S/N and therefore the statistical reliability of the images, it is necessary to apply single particle classification and averaging techniques.

First, the aligned particles are classified to identify groups of particles that represent the same views of the macromolecule. Then, all images belonging to a class are averaged, improving thereby the S/N, which is proportional to the number of averaged images. The accuracy of image alignment prior to classification is crucial, since it increases the accuracy of classification by positioning each particle similarly.

#### **MSA- multivariate statistical analysis**

In order to identify the different molecular views of the aligned particles and analyse the mixed population of images, a multivariate-statistical-analysis in the form of the eigenvector-eigenvalue analysis was applied (van Heel and Frank, 1981; van Heel, 1984). In the MSA technique, images are represented as points in a hyper-dimensional space with as many dimensions as there are pixels in one image. The measured density value of each pixel is used as a coordinate of the point along the corresponding axis of the hyper-dimensional space. Thus, a single point in the MSA-hyperspace fully describes a 2D molecular image. A mixed dataset of particles with different molecular views will form a cloud of points in the MSA-hyperspace, in which similar images will be close to each other and different images will be far apart. Furthermore, the cloud may be structured, consisting of separate sub-clouds corresponding to different classes of molecules. This is often the case especially from stained specimens, since negative staining tends to induce preferred orientations of the molecules on the carbon support film.

The coordinate system of the hyper-space is then rotated and shifted so that the first axis of the new coordinate system points in the direction of maximum inter-image variance. The second axis, perpendicular to the first, describes the maximum remaining inter-image variance etc. The new coordinated system is therewith adapted to the shape of the data cloud. The first unit vectors (max. 69) (eigenvectors) fully characterize the major differences

in the dataset and the remaining axes may be disregarded since they represent rather noise than structural information.

This results in a large reduction of data since each molecular image can be expressed in merely max. 69 coordinate values, rather than in the typical  $450 \times 450 = 202500$  pixel grey values (KLH1-images) that originally described the image.

The eigen vectors in the new coordinate system correspond to single points in hyper-space and represent therefore images, which are called "eigen images". The significance of the eigen vectors is reflected by their corresponding eigen values.

As already mentioned, eigen images account for the most representative differences in the electron density distribution within the analyzed set of images. In some cases, analysis of eigenimages has been instrumental in the determination of the symmetry components within oligomeric protein complexes (Dube *et al.*, 1993).

The MSA procedure was performed with the module *msa-run* from the IMAGIC-5 software package. Furthermore, during the MSA-procedure, a circular mask was applied to all images. This mask contained zeros in those areas of the images which did not participate in the eigenvector calculation (surroundings of the molecules). The MSA-mask was created with the module *test-image* from the IMAGIC-5 software package.

### **HAC-Classification and averaging**

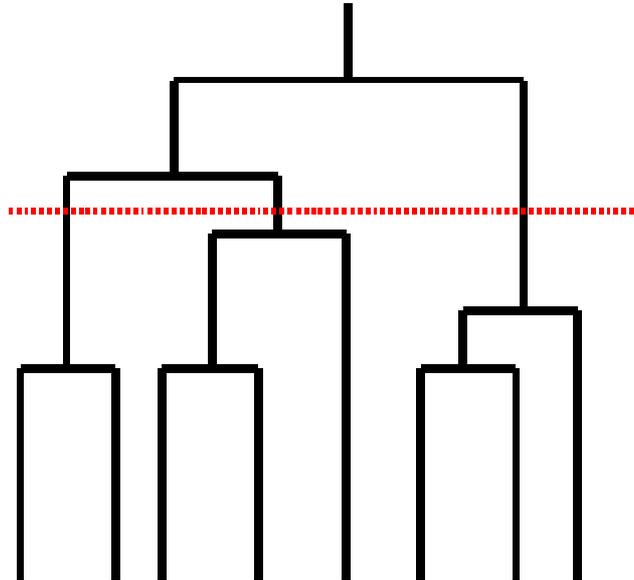
The MSA-analyzed datasets were then classified with an automatic hierarchical ascendant classification (HAC) combined with a moving elements refinement (van Heel, 1984; van Heel *et al.*, 2000). The optimal classification can be defined as that classification in which the inter-class variance between the centers of mass of the class averages is maximal and the total intra-class variance is (therefore) minimal. HAC procedures start with as many classes as there are images and then merge two classes at a time to form larger classes. At the end of the HAC, all images are members of a single huge class. During this procedure, the algorithm follows the criterion of the lowest possible increase of the intra-class variance (Ward, 1982).

The HAC can be represented by a "tree", in which the merging of classes can be followed (Fig. 22). The number of classes is defined by the user and the "tree" of the history of class mergers is cut at the appropriate level, in order to obtain the predefined number of classes.

This primary classification is then refined by allowing members of all classes to migrate to other classes, following the variance criterion. This refinement procedure (moving elements) is performed iteratively until the classes remain stable.

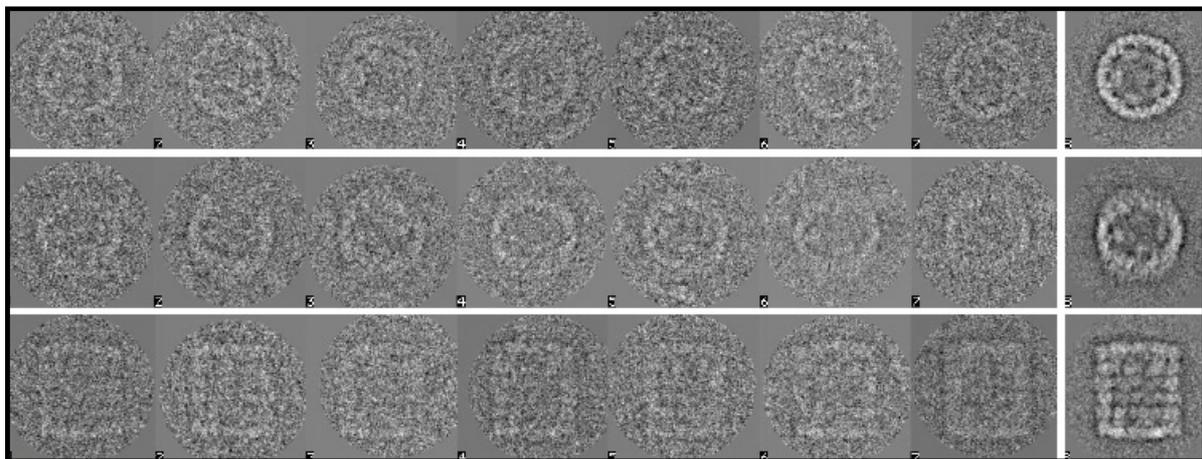
Finally, the members of each class are averaged, improving the S/N. Ideally, all images averaged, apart from the noise, should correspond to identical molecular views.

MSA-HAC-classification was performed with the module *msa-classify*, averaging of class-members with the module *msa-sum*, both from the IMAGIC-5 software package.



**Fig. 22: Example of a hierarchical ascendant classification (HAC) tree**

The procedure starts with as many classes as there are images. Two classes that are closest together in terms of the Ward criterion are merged into a larger class. The dotted line represents the level of the requested number of classes (according to Schatz, 1992 (modified)).



**Fig. 23: Typical class average images of the KLH1 didecamer and the corresponding original images**

Right: Three typical class averages obtained after MSA-classification of the KLH1-dataset (end-on-view, tilted-view, side view). Left: The corresponding original images (CTF-corrected, band-pass-filtered and aligned). The original images were extracted from the dataset with the module *msa-extract* from the IMAGIC-5 software Package. Note the improvement of the S/N in the class averages.

## 4.7 Orientation determination

Depending on the available data, single particle orientation determination can be performed by following one of three general approaches:

- conical tilting (Radermacher *et al.*, 1987; Radermacher, 1988)
- angular reconstitution (van Heel, 1987)
- projection matching (Penczek *et al.*, 1994)

When the conical tilting method is not available (as in this study) the angular orientations of the macromolecular images are *a priori* unknown.

In order to find the unknown Euler angles of the randomly orientated particles, the technique of angular reconstitution is applied, since it places no restriction on the type of the data available as it determines orientations using properties of the projection images. This approach is used to obtain a set of initial angle assignments that generate a first 3D-model.

With a starting model from the angular reconstitution approach or from a similar structure, the structure can then be refined by projection matching, which enables a more accurate angular assignment of protein projections. Both techniques will be presented in detail in the following sections.

### Euler Angles

Single molecules have six degrees of freedom: three translational ones; X, Y, and Z and three rotational ones corresponding to three “Euler angles”;  $\alpha$ ,  $\beta$  and  $\gamma$ . X, Y and  $\alpha$  are the “in-plane” degrees of freedom, whereas Z,  $\beta$  and  $\gamma$  are the out-of-plane degrees of freedom.

The image of the object created in the EM is a projection along the Z direction (projection axis). If we assume that the EM provides an ideal projection along Z, all images should share the same image plane (same values for Z).

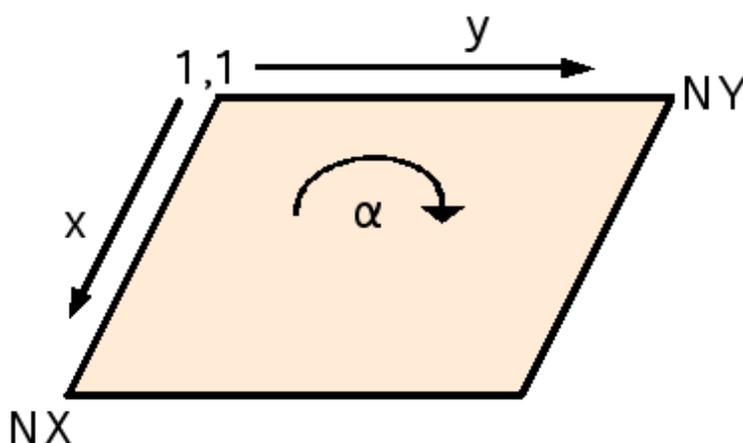


Fig. 24: IMAGIC 2D co-ordinate system

The coordinates  $X$  and  $Y$  describe the position of the object within the image plane. The IMAGIC co-ordinate system is a right handed system with its origin in the top-left corner of the image, whereas the length of the lines is  $NY$  and the number of lines is  $NX$  (Fig. 24). The center of the object after translational alignment operations should be given by:

$$(NX/2+1, NY/2 +1).$$

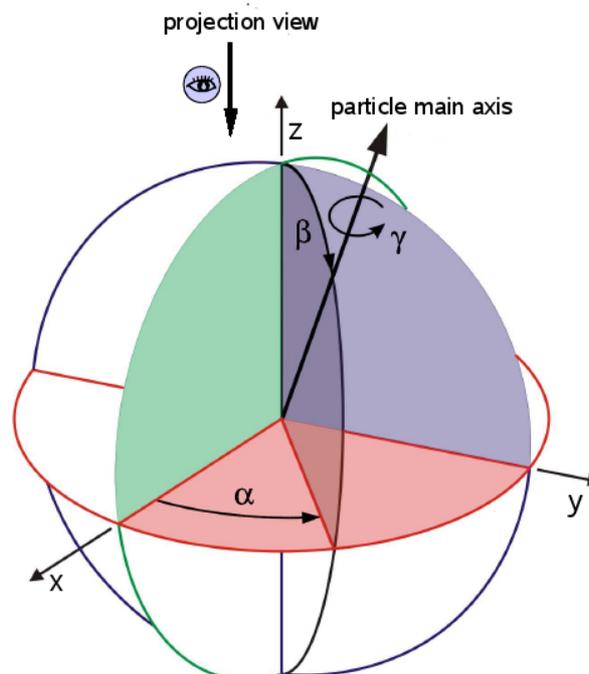
The  $\alpha$  angle describes the rotation of the object about the projection axis  $Z$ , within the image plane (see Fig. 25). This degree of freedom can be removed after rotational alignment.

Thus, after removing the in-plane degrees of freedom by alignment operations, only the out-of-plane rotations ( $\beta$  and  $\gamma$ ) remain unknown.

The  $\beta$  angle is the angle of the tilting of the particle main axis relative to the projection axis  $Z$ . The particle main axis (or molecule vector) is the molecule axis with the highest symmetry. An end-on-view (top – view) of the molecule has  $\beta = 0^\circ$ , whereas a side-view has  $\beta = 90^\circ$ .

The  $\gamma$  angle represents a rotation around the particle main axis. Molluscan hemocyanins have five-fold symmetry and therefore identical views recur every  $72^\circ$  of rotation around the molecule's main axis.

When the particle main axis coincides with the projection axis  $Z$ , the  $\alpha$  and the  $\gamma$  angle rotate in the same plane (for review see Heymann *et al.*, 2005).



**Fig. 25: Definition of the Euler angles in a common coordinate system**  
(According to Stohr, 2007, modified) (For details see main text).

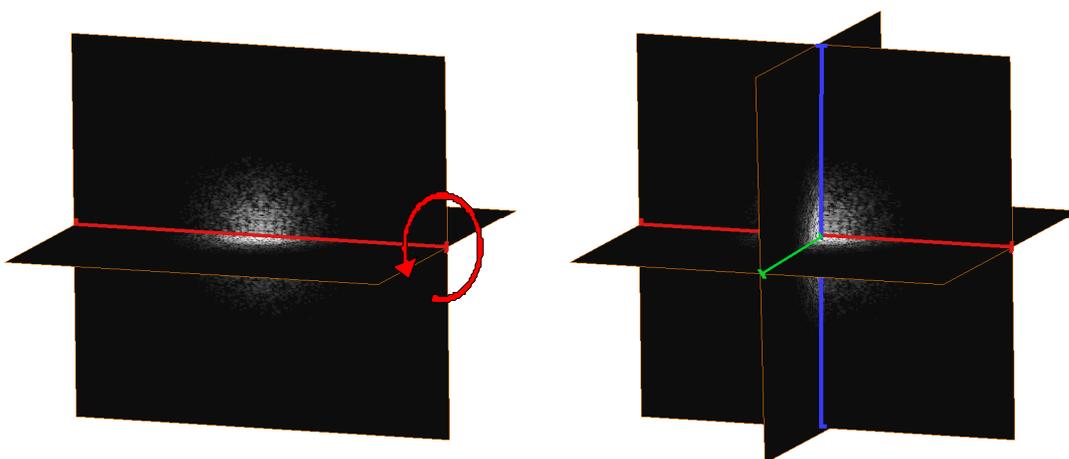
The out-of-plane rotations  $\beta$  and  $\gamma$  of individual molecules provide the 3D information and must be identified before one can perform a 3D reconstruction.

### **Angular reconstitution**

In the angular reconstitution procedure, orientation angles are defined using “common” lines, an algorithm that can find the angular relation between projections without additional input. The algorithm requires a high S/N to diminish false solutions and therefore the input projections are class averages obtained by MSA-classification.

The technique is based on the common line projection theorem stating that two 2D projections of the same object always have a 1D line projection in common (Van Heel, 1987). This theorem is the real-space equivalent of the Fourier-space common lines theorem (DeRosier and Klug, 1968).

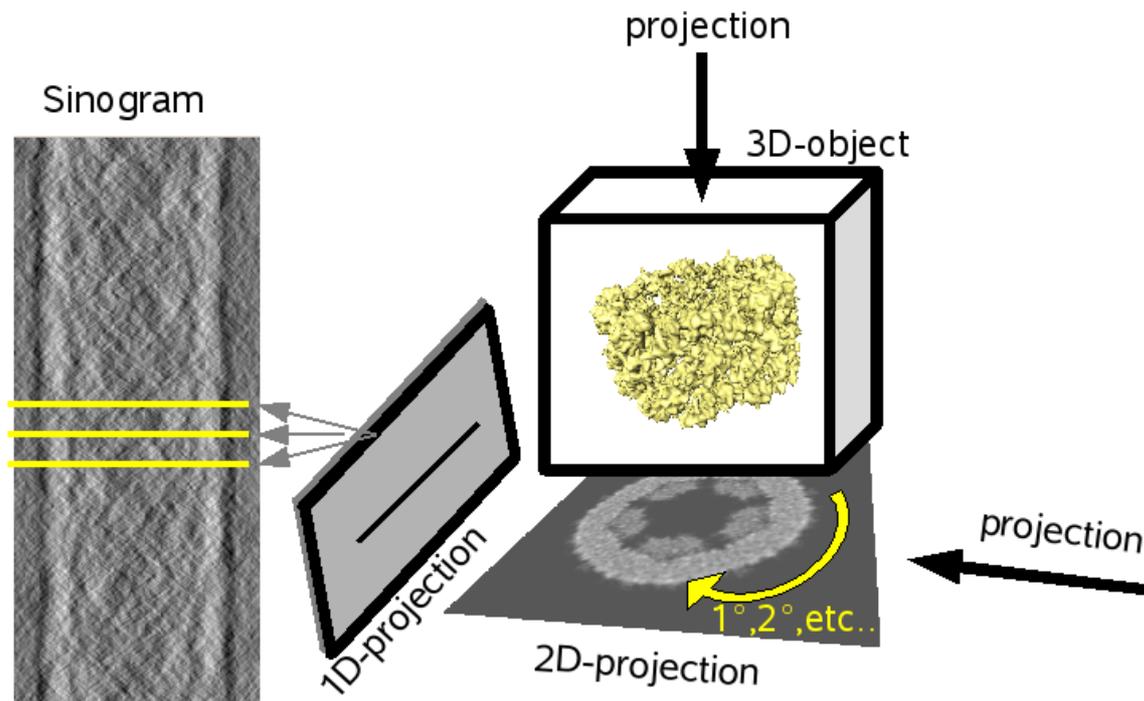
The relative Euler-angle orientations of a set of class averages can be determined from the angles between such common-line projections. In order to solve the orientation problem for asymmetric particles, at least three different projections are required, because the first two 2D projections, which correspond to two central sections in 3D Fourier space, can freely rotate around their common line (Fig. 26). A third central section in the system gives two more common lines and fixes the spatial orientation (van Heel, 1987).



**Fig. 26 : 2D projections of the same object represented in Fourier 3D space**

Left picture: The common line of two 2D projections does not fully fix their spatial orientation, since a rotation of the central sections around their common line (red line) remains possible. Right picture: With a third 2D projection, 2 more common lines are introduced (shown in green and blue) and the relative spatial orientation of all three central sections is fixed.

To find the relative orientations, all line projections (sinograms) of the 2D images (class averages) need to be found. A sinogram is calculated by summing the 1D projection lines of the 2D image, after rotating that image over 0, 1, 2° etc. (Fig. 27).



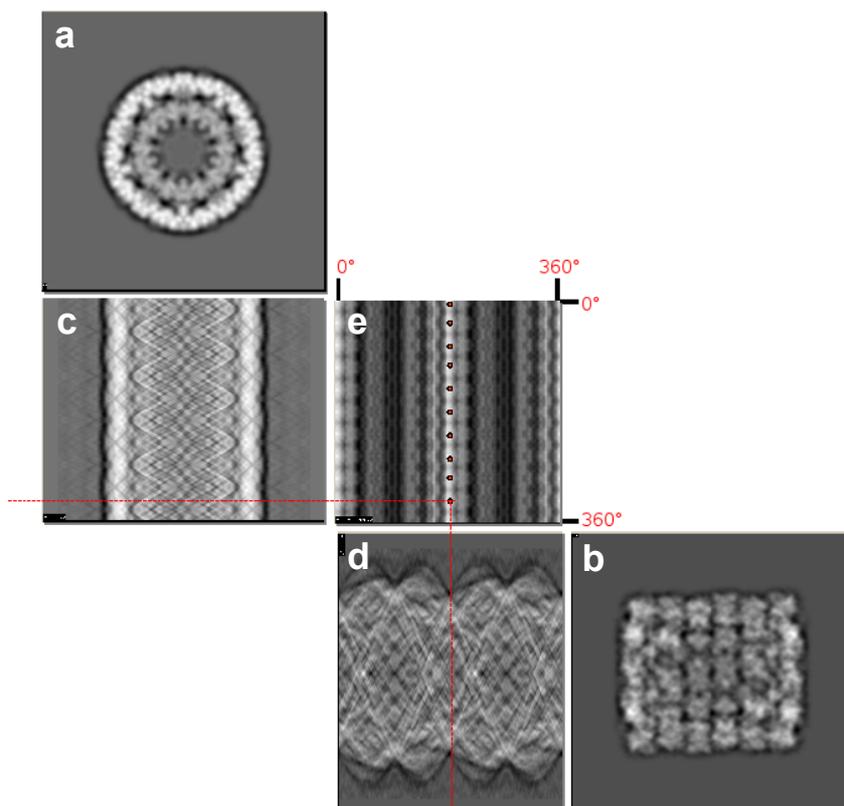
**Fig. 27: Sinogram**

The sinogram of a 2D-projection image of a 3D-object is the collection of all line projections through the 2D projection image. Each line in a sinogram represent a one-dimensional projection of the 2D-image at radial angles ranging uniformly from 0 (top line) to  $2\pi$  (bottom line).

Then, the sinograms of two class averages are compared line-by-line in a sinogram correlation function (SCF) (Fig. 28). The SCF shows a maximum at a position corresponding to a pair of shared line projections (common line). Sinograms of asymmetric particles show only one peak, whereas symmetric particles show a number of corresponding symmetry-related peaks (i.e. for a symmetric object with D5 symmetry, two different 2D projection share 10 common lines).

The introduction of a new projection image, to be included in the set that has already been assigned an Euler-angle orientation, is performed over all possible Euler-angles orientations within the asymmetric triangle for the given point group symmetry (Schatz *et al.*, 1995). During this search, the normalized standard deviation of peak heights among all corresponding peaks serves as an internal consistency check.

The asymmetric triangle is the unique part of the sphere. The coordinates describing the surface of the unit are the two Euler angles  $\beta$  and  $\gamma$ . For a particle with D5 point group symmetry, i.e. KLH1, the asymmetric triangle extends from the "North Pole" to the equator ( $0 < \beta < 90$ ) and covers 1/5 of this northern hemisphere ( $72^\circ$ ).



**Fig. 28: Cross Sinogram correlation function (CSCF)**

**a, b:** reprojection images of a KLH1-reconstruction **c, d:** the “sinogram” of the projection image is the collection of all line projections through the 2D projection image (from top to bottom) after rotation of the image over angles ranging from  $0^\circ$  to  $360^\circ$  **e:** CSCF of projections images **a, b**. The value of each pixel in the SCF is the correlation coefficient of a line in one sinogram with a line in the other sinogram. Both horizontal and vertical axes run from 0 to  $360^\circ$ . The ten correlation peaks (red dots) indicate the optimal relative Euler angle orientations of these two projections assuming D5 symmetry (according to van Heel *et al.*, 2000, modified).

The first cryoEM structure can be used to create a set of images by reprojection within the asymmetric triangle (“anchor set”) (Schatz, 1995, Orlova and van Heel, 1994). In this case, Euler-angle-determination is performed by using exclusively the CSCF of each class with respect to these anchor set reprojections. The anchor set procedure for the assignment of euler angles offers several advantages in comparison to the Euler-angle assignment with respect to other classes. The reprojections are perfectly consistent with each other, have a much reduced noise level compared to class averages and their euler angles are known exactly. The final euler angle assignment is usually performed with an anchor set of some 30 reprojections and angular search interval of only  $1^\circ$  within the asymmetric triangle.

However, the absolute handedness of the object cannot be determined by the angular reconstitution technique. Nevertheless, at resolution below  $15 \text{ \AA}$ , one is able to compare the 3D-cryoEM reconstruction with hopefully available X-ray-data. From such a comparison, the handedness of the structure can be inferred.

The search of the Euler-angles (relative orientations) of selected class averages was performed with the module *angular-reconstitution* from the IMAGIC-5 software package.

### **Projection matching**

In the projection matching procedure, orientation angles of single particles are defined based on cross correlation values by comparison with projections of preliminary volumes that act as templates of known angles. Within each cycle of refinement the reconstructed volumes and their projections are improved, so that angular assignment is also iteratively improved.

The SPIDER software package directly utilizes the single particles with defined Euler angles to reconstruct the volume, whereas the EMAN software package classifies and averages similar views to produce a 2D average with improved S/N (single particles, which were aligned against the same template projection are averaged, thereby improving the S/N). The class averages then constitute the input to reconstruct the 3D volume. Thus, in this case, the creation of class averages is not based on MSA and HAC, and Euler angles assignment is not based on the common lines (as in the IMAGIC-reconstruction strategy), but on cross correlation coefficients.

Besides these differences in the general approaches between IMAGIC-5 and the other platforms (for review see Llorca, 2005), a “projection matching” strategy can be perfectly carried out using the flexible modules provided by IMAGIC in combination with shell scripts.

To do so, the routine of 3D reconstruction from the EMAN software package was adapted to our IMAGIC-5 environment (Martin, 2006). Upon this procedure, [REDACTED] (Johannes Gutenberg University of Mainz, Institute of Zoology) developed a C-program, called *create-mra-classes* (Stohr, 2007). This program was later replaced by a refined Pearl program called *r0c-extreme* ([REDACTED]). Both programs generate class averages with defined Euler angles, based on Fourier correlation values, using modules from the IMAGIC-5 software package, in an automated process.

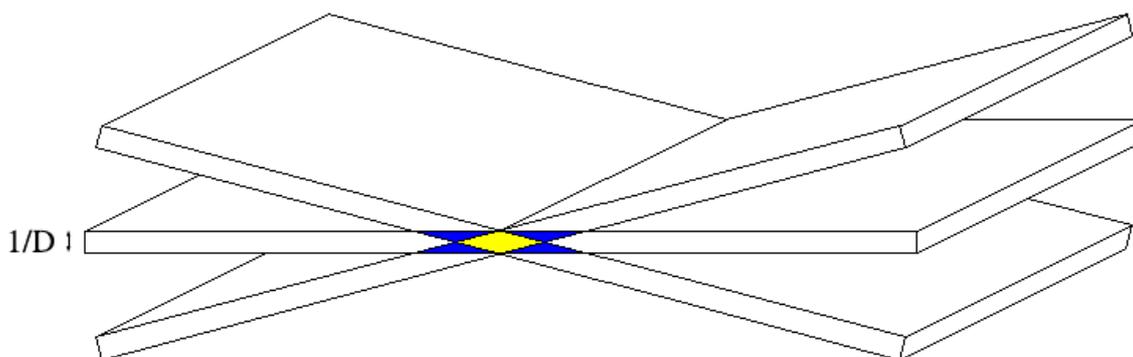
In short, the single particles data set is aligned against thousands of reprojection images (templates) that cover the asymmetric triangle uniformly (multi-reference-alignment, see section 4.5). Images that were aligned against the same reference should correspond to similar views and thus can be averaged, in order to improve the S/N. During this procedure, averaging is improved by discarding “bad” particles with low correlation coefficients. Then the Euler-angles of the template-reprojections are assigned to the corresponding class averages. The disadvantage of the “projection-matching” method is the potential model bias in the assignment of angles.

In this study, projection matching was exclusively applied after a 3D reconstruction with medium resolution (15-12 Å) was achieved by the angular reconstitution procedure. In the angular reconstitution approach, angles come direct from the data and model bias in the assignment of angles is therefore greatly reduced.

## 4.8 3D Reconstruction

Following orientation determination, the image data are combined in a computational task known as three-dimensional-reconstruction that computes the three-dimensional-structure of the object. This task is performed with a method known as “exact filter back projection” (Harauz and van Heel, 1986; Radermacher, 1988; Schatz *et al.*, 1995), a process similar to inverse Fourier transform.

The 3D reconstruction is gained by back-projecting the class averages along their assigned Euler angles. However, simply adding the data in Fourier-space to perform a 3D-reconstruction would end in over-weighting the low spatial frequencies (see Fig. 29). To avoid this over-weighting and down-weight over-populated class averages, a specific filter is computed for each projection in the reconstruction.

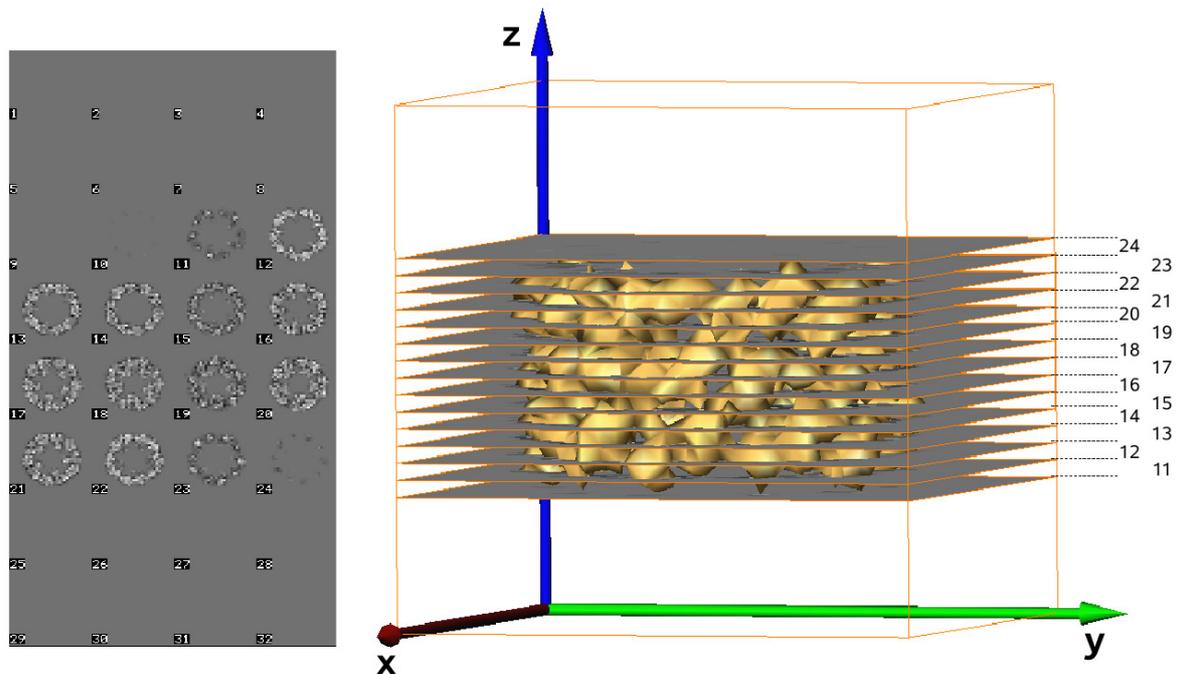


**Fig. 29 : Overlapping central sections in Fourier space according to van Heel *et al.*, 2000 (modified)**

A 2D projection of the 3D object corresponds to a central section of the 3D Fourier transform of the object (De Rosier and Klug, 1968). The 3D-reconstruction is performed by filling the Fourier space with central sections. The central sections have a thickness of  $1/D$ , where  $D$  is the largest linear dimension of the 3D object, and overlap close to the origin of Fourier space (yellow and blue areas) The overlapping areas correspond to low spatial frequencies in the 3D Fourier space.

The resulting gray scale 3D-density map is calculated on a grid with the sampling ( $\text{\AA}/\text{Pix}$ ) of the original data. For example, in the case of NpH, where 2D-projections (class-averages) have a box size of 256x256 pixel and sampling of 1.86  $\text{\AA}/\text{Pix}$ , the corresponding 3D-density map contains 256 horizontal sections with a size of 256x256 pixel, each of 1.86  $\text{\AA}$  thickness (Fig. 30).

With the help of visualization software, volume rendering is then performed and an isosurface of the 3D-object is generated.



**Fig. 30 Example of a density map**

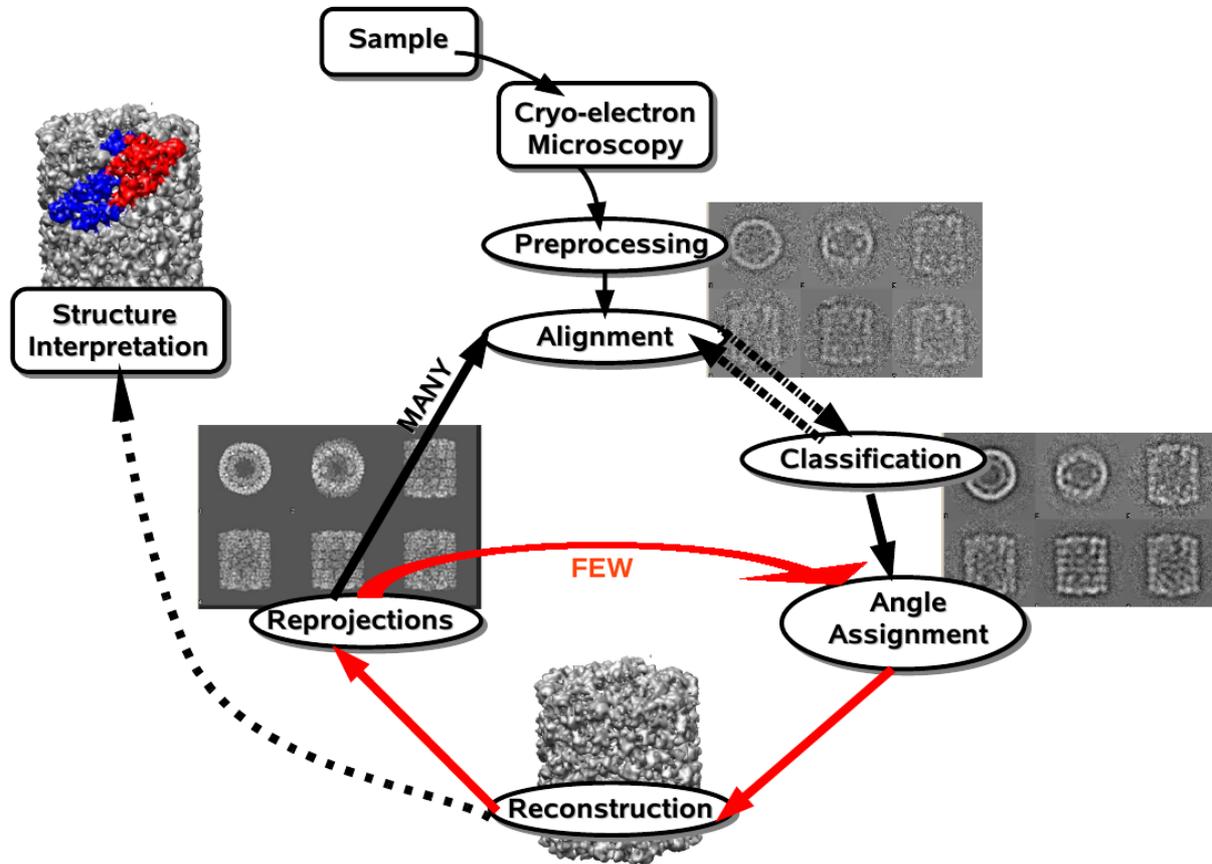
Left: 32x32x32 density map of NpH hemocyanin, shown as 14 sections each of 14.9 Å thickness. Right: rendered three-volume with AMIRA 4.1 indicating the 14 sections (No. 11-24) describing the 3D-object. The sections are along the main symmetry axis of the molecule.

#### 4.9 Iterative refinements

As already mentioned, the whole process of single particle analysis is a constant iterative refinement cycle. Beginning with an optimized sample preparation, a large number of micrographs is generated using a high resolution cryoEM (200kV, FEG). Selected micrographs are then digitized at high resolution (1 -2 Å / Pix) and subsequently particles are selected. After CTF-correction and band-pass filtering (preprocessing), the particles from the micrographs are reference-free aligned. The class averages obtained by MSA and HAC serve as reference images for a first multi-reference alignment. After stabilization of the class averages, the Euler angles are assigned and a first cryoEM-structure is calculated.

The 3D model is then refined as follows. Reprojected 2D-images from the latest model are used as references for a multi-reference alignment of the dataset of bandpass-filtered, CTF-corrected-images. Following MSA and HAC, particles that represent the same views of the macromolecule are recognized and grouped into statistically significant classes. The classes are then compared to the reprojections from the previous model ("anchor set"), using the angular reconstitution technique. Class averages with assigned Euler angles are used to calculate the next cryoEM-structure. The procedure of alignment, MSA, HAC and angular reconstitution is iteratively applied to refine the 3D results. Poor class averages are

discarded before the next round of iteration. The refinement cycle during single particle analysis by the angular reconstitution technique is depicted in Fig. 31 (a).



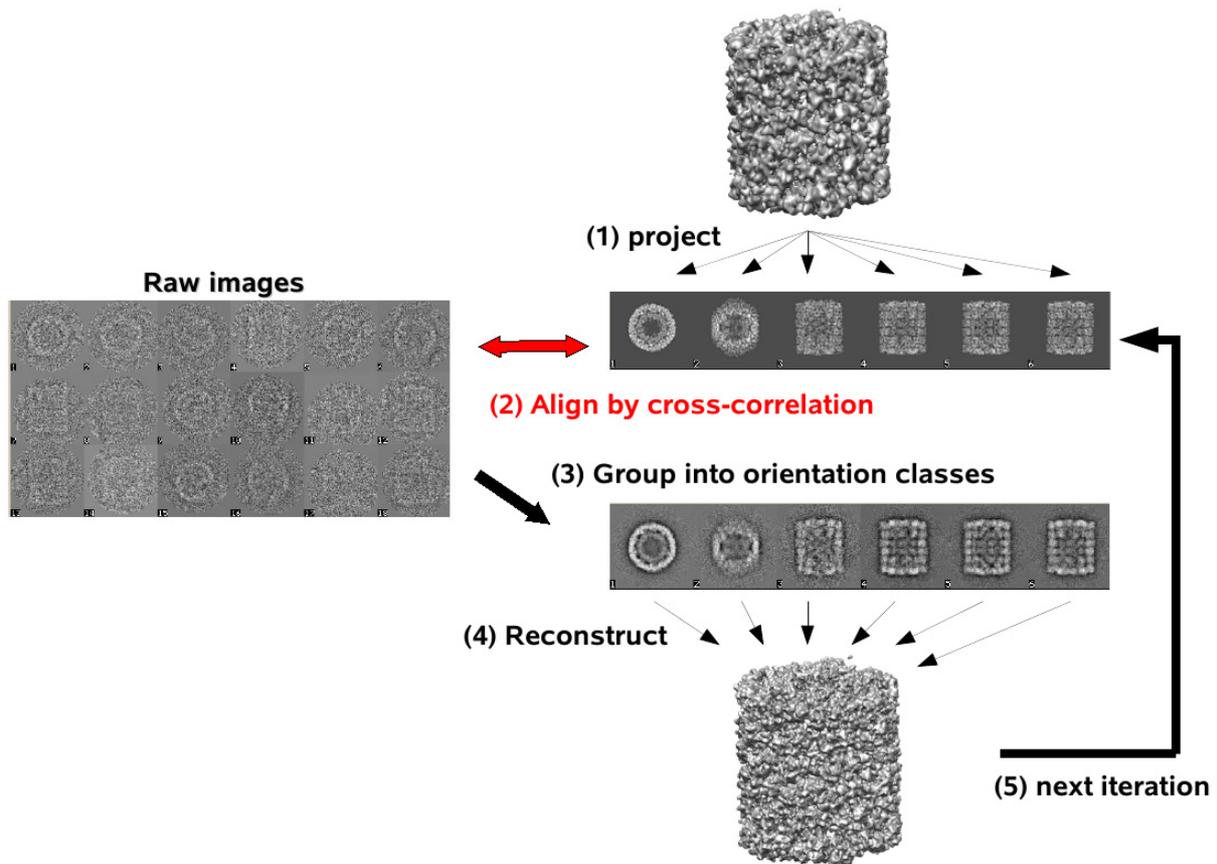
**Fig. 31 (a): Single particle analysis by the angular reconstitution technique**

The overview presented here shows data from the KLH1 project. After sample preparation (top left; the explanation proceed clockwise), a large number of micrographs is generated and subsequently digitized. Single particles are extracted from the micrographs and pretreated (CTF-correction, Band-pass filtering) to then be aligned (MRA). Class averages obtained by MSA classification can serve as references for new rounds of MRA. Stable class averages are assigned Euler angles per se and these Euler Angles allow a first 3D-reconstructon to be calculated. A set of few reprojections (“anchor set”) is used to refine the Euler-angle assignments and a new refined 3D reconstruction can thus be calculated. The refined reconstruction is then used to create many reprojections, which are then used for a new round of MRA etc. The process is iterated until the results are stable. A stable cryoEM-structure allows structure interpretations (according to van Heel et al., 2000) (modified).

After achieving medium resolution, the technique of projection matching is applied to further improve the cryoEM-structure. This method can also be directly used if a closely related structure is available. The 3D model is reprojected into all directions and the set of raw images is cross-correlated with the whole set of reprojections to refine the position and orientation parameters. Each raw image is assigned the orientation of the reference giving the highest correlation coefficient. The aligned images are grouped and averaged and these class averages are then used to create a cryoEM-structure with improved resolution. This process is iterated until the resulting density maps converge. A stable cryoEM-structure

allows structure interpretations. The basic refinement procedure used in projection-matching technique is shown in Fig. 31 (b).

It should be noted that single particle analysis is far from a fully automated method that does not require user intervention. Evaluation of the output results and decisions during processing are completely user dependent.



**Fig. 31 (b): Iterative refinements to the model by projection matching**

A 3D structure obtained by the angular reconstitution approach or an appropriate available reference structure (e.g. structure of a closely related molecule) is projected into a set of different orientations that cover the asymmetric triangle uniformly (step 1). Each raw image is then aligned to each reference image and is assigned the Euler angles of the one giving the highest correlation coefficient (step 2). Raw images with the same Euler angles are grouped and averaged (step 3). These class averages with assigned Euler angles are used to create an improved map. This procedure is then iterated (step 5) and once the resulting 3D reconstruction has stabilized, one starts the interpretation process (according to Saibil, 2000) (modified).

## 4.10 Visualization

Visualization and further analysis of the 3D-volumes and molecular models (see below) was done within the software packages AMIRA 4.1 and CHIMERA. Molecular graphics images were produced using AMIRA 4.1 and CHIMERA, partly in combination with the rendering software POV-Ray (implemented in SUSE Linux).

Solid carbohydrate polymeric models of the cryoEM-structures were constructed by rapid prototyping using a Z Corporation 3D printer (Z 406), directly using the digital data sets created by the IMAGIC software as input (after converting them to the VRML format with CHIMERA).

During this procedure, the 3D-printer lays down successive layers of powder and in this way builds up the solid model from a series of cross sections. The sections are fused automatically to create the final shape. The solid models allow one to physically interact with the molecules in real space and provide complete information and also new insight.

## 4.11 Resolution assessment

The resolution can be determined by measuring the Fourier shell correlation (FSC) between two 3D reconstructions, each based on half of the available class averages (Harauz and van Heel, 1987; Orlova *et al.*, 1997). FSC measures the normalized cross-correlation between two 3D reconstructions as a function of spatial frequency.

The resolution can be determined as the reciprocal of the spatial frequency at which the FSC-curve has dropped to half its maximum value (**0.5-criterion**).

Another criterion is to take the value at which the curve crosses a baseline curve for a random correlation with the appropriate symmetry. According to the **3 $\sigma$ -criterion** threshold criterion (Orlova *et al.*, 1997), the resolution can be determined as the reciprocal of the spatial frequency at the intersection of the FSC function with a function that represents three times the standard deviation of random noise ( $3\sigma$ ) corrected for the molecular point-group symmetry. Therefore, this criterion indicates the resolution level at which one has collected information significantly above the noise level (van Heel and Schatz, 2005).

According to the newly introduced **1/2-bit criterion** (van Heel and Schatz, 2005), the resolution can be determined as the reciprocal of the spatial frequency at the intersection of the FSC function with the 1/2-bit information threshold curve, which is calibrated to approximately yield resolution values comparable to resolution values in use in X-ray-crystallography. The 1/2-bit criterion indicates the resolution level at which enough information has been collected for interpretation.

However, there remains controversy as to which threshold correlation value to use. In the end the important question is what structural features are visible in the density map. For example, at  $\sim 9 \text{ \AA}$  resolution long  $\alpha$ -helices can be seen as cylinders of density. Therefore, a visual inspection of the cryoEM structure is necessary for a reasonable assessment of the resolution.

In this study the resolution values were determined by deriving the FSC 1/2-bit criterion from the FSC, which delivered values which conform to the observed details in the final density maps (see results and discussion). The FSC  $3\sigma$  criterion yields values that are rather too optimistic, whereas the 0.5-criterion approach is clearly too conservative.

The FSC was calculated with the module *Fourier-shell-correlation* from the IMAGIC-5 software package.

#### 4.11 Sharpening

The high-resolution information may be missing from the representation even though the information is present in the data. Contrast loss is caused by experimental (specimen movement, charging, radiation damage etc) and computational factors (inaccurate determination of the orientation parameters, etc.). The combined effect of all these factors is traditionally modeled by a Gaussian decay of structure factors with a temperature factor or B-factor. This amplitude fall-off affects the high resolution components, thereby making the density map look apparently smooth (Rosenthal and Henderson, 2003). Therefore it is necessary to restore (“sharpen”) the suppressed components in order to unravel high resolution molecular features concealed in density maps. Contrast restoration is crucial, especially at subnanometer reconstructions, in order to resolve secondary structure elements.

A common method for contrast restoration is to emphasize the high frequency details by suppression of over-represented low frequency components, using an appropriate high-pass filter (amplitude scaling) (module *threed-filter*, IMAGIC-5 software package).

Another standard restoration procedure is to sharpen the experimental maps by applying a Gaussian function with an inverse B-factor. Automated B-factor determination and sharpening can be performed with the newly introduced program called EM-BFACTOR (Fernandez *et al.*, 2008). Both procedures do, however, also amplify high resolution noise. A common solution is based on low-pass filtering at the nominal resolution of the map, thus eliminating structural details beneath this cut off.

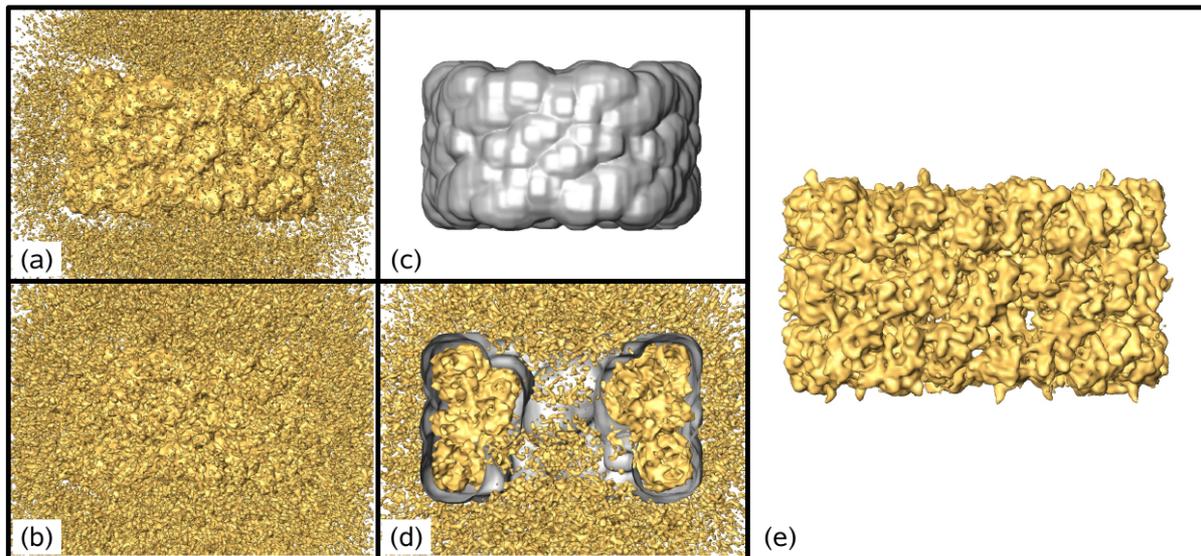
Sharpening was performed after a nominal resolution level of  $< 12 \text{ \AA}$  was achieved. The “sharpened” density maps were low-pass-filtered at their nominal resolution with the module *proc3D* from the EMAN 1.7 software package or alternatively with the module *threed-filter*, option *low-pass*, from the IMAGIC-5 software package.

## 4.12 3D-masking

In order to remove background noise present in the sharpened cryoEM structure and also in both half-dataset reconstructions, a broad soft-edged mask was applied to each volume. The use of sharp-edged masks was avoided, since it is known that such masking introduce artificial similarities between the two half volumes one compares, especially in the very high frequency components.

First, with the module *create surface* from the AMIRA 4.1 software package, a surface object was generated containing the set of currently visible triangles at the estimated threshold. Then, with a lasso selection mechanism (port: draw, AMIRA 4.1) all triangles representing background noise were highlighted and subsequently removed. The resulting surface was back-converted into an isosurface with the module *ScanConvertSurface*. A maximum filter was then applied to the entire resulting 3D-volume, which replaces the value of a pixel by the largest value of neighboring pixels covered by a  $N \times N$  mask. The effect of this filter is to increase the volume, by adding pixels to its boundary, whereas the size of the bounding box remains the same. This enlarged volume was then multiplied with the original density map, which resulted in a noise-free density map. This volume was then further enlarged with a maximum filter and a edge preserving smoothing filter was subsequently applied. This filter smooths out the difference between grey levels of neighboring voxels but contrary to the Gaussian filter, it preserves the edges. This volume was then binarized (density values above a specified threshold value are set to 1 else they are set to 0) using the module *arithmetic-with-image*, option: *binarize*, from the IMAGIC-5 software package. In order to smear out edges of the 3D volume, a Gauss filter was then applied (module *threed-filter*, option: Gauss, IMAGIC-5). The resulting volume was used as a 3D-mask, in order to remove unwanted background noise from the sharpened, low-pass filtered density map. To do so, the density map was multiplied with the 3D-mask using the module *add-image* from the IMAGIC-5 software package.

The steps of sharpening, filtering at the nominal resolution and 3D-masking are depicted in Fig. 32.



**Fig. 32: Sharpening, 3D-low pass filter and masking**

(a) Side view of a original 3D-density map of the NpH (b) the same density map after sharpening and low-pass filtering at a resolution level of 9Å. Note the amplification of high resolution noise. (c) Side view of the corresponding 3D-mask. For details see main text (d) Cut-open view of c and b. Note that the 3D-mask is clearly broader than the cryoEM structure and does not contain sharp edges (e) The resulting noise-free density map, generated after multiplication of b and c. The effect of sharpening (enhancement of molecular details) is now clearly visible.

#### 4.12 Comparative molecular modeling

Comparative or homology protein modeling uses experimentally determined protein structures (templates) to predict the conformation of another protein that has a similar amino acid sequence (the target) (Rost & Sander, 1996). This approach is possible because a small change in the protein sequence usually results in a small change in its 3D structure (Hubbard & Blundell, 1987).

Homology modelling was performed using the software modeller 8v2-9v1, by satisfaction of spatial restraints. The method consists of three stages:

- (1) Alignment of the sequence to be modeled with related protein structures.  
(Sequence alignments were performed with the software clustalw)
- (2) Extraction of spatial restraints on the sequence using the alignment and
- (3) Satisfaction of the restraints to obtain the 3D-model.

The spatial restraints include:

(i) Homology-derived restraints on the distances and dihedral angles in the target sequence, extracted from its alignment with the template structures.

(ii) Stereochemical restraints such as bond length and bond angle preferences, obtained from the CHARMM-22 molecular mechanics forcefield.

(iii) Statistical preferences for dihedral angles and non-bonded interatomic distances, obtained from a representative set of known protein structures (Eswar *et al.*, 2003).

Accuracy of the resulting molecular models tends to increase with the target-template sequence identity. A protein sequence that has at least 35% identity to a known structure can be modelled with an accuracy approaching that of a low resolution X-ray structure or a medium resolution NMR structure (Sanchez and Sali, 1997). With more than 40% (-50%) identity when comparing the X-ray structures of OdH-g (PDB-ID: 1JS8) and RtH-e (PDB-ID: 1LNL) with the FUs from NpH and KLH1, this threshold is clearly exceeded. Moreover, the high conservation areas correspond to secondary structure elements of the available X-ray structures.

The models of the FU types a, b, c, d, e, f and g from NpH were therefore built using the 3D-structures of RtH-e and OdH-g as templates. The models of the FU-types a, b, d, d, e, f, g and a fragment of FU-type h from KLH1 were also built using the two X-ray structures as templates. For each FU-type 10-20 different models were calculated.

#### 4.13 Modelling the NpH polypeptide chain

After identifying the subunit pathway (see results and discussion; Gatsogiannis *et al.*, 2007), the seven densities of the final 9.1 Å cryoEM structure were interactively extracted with the Amira4.1 software, using a tool called *Segmentation Editor*.

Segmentation was done by first selecting the corresponding voxels and then assigning these voxels to a particular material. The labels were stored in a label field. The label field contains for each voxel a number indicating which material the voxel belongs to. From the label field a surface can then be reconstructed using the SurfaceGen module. With the module *ScanConvertSurface* a single material (in this case a FU) was then picked out.

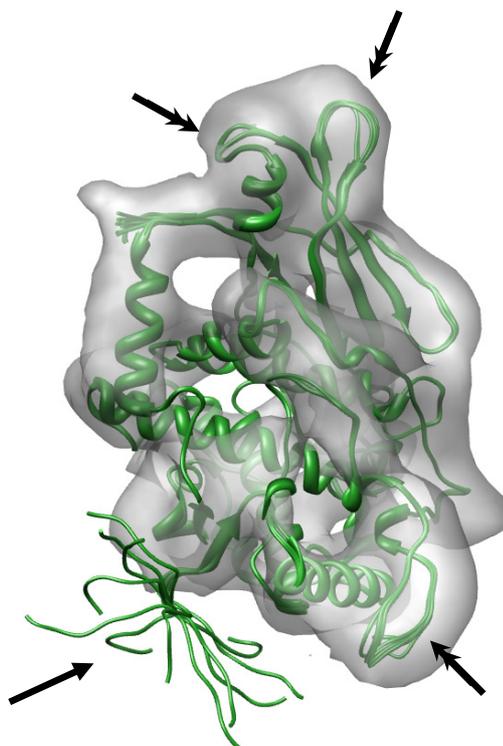
The resulting isosurface was used as a basis to generate a 3D-mask for a single FU as described in section 4.11. The NpH-density map was then multiplied with the 3D-mask with

the module *add-image* from the IMAGIC-5 software package in order to extract the density of a single FU.

Alternatively, the 7 FUs were extracted with the *color zone / split map to color* modules of the CHIMERA software which yielded comparable results in a significantly shorter time.

In a first step, each FU-model was fitted into its respective density within the whole density map with the CHIMERA software (module: *Fit in Map*). Each FU was assigned a different color. Then all atoms of the 7 different FU-types were selected and with the Color Zone module, surface triangles within a specified distance (5-10 Å) of the selected atoms were colored respecting the color of the selected atom within this cutoff distance. Then, with the *Split Map* module a volume was created for each distinct color zone and for the “uncolored” data, which was then saved in a separate file.

For each FU a ranking of the 20 models was performed by fitting all of them into their respective extracted FU density with the MOLREP 9.2 software; the model with the highest correlation was then chosen. It should be mentioned, that that the spatial orientations of  $\alpha$ -helices and  $\beta$ -strands were almost identical in the 20 models and in the two existing crystal structures, used as templates. The main differences were restricted in the orientation of large exposed loops and especially in the folding of the linker peptides connecting functional units, since no template was available. The RMSD between the 20 different models was lower than 0.3 Å (see Fig. 33).



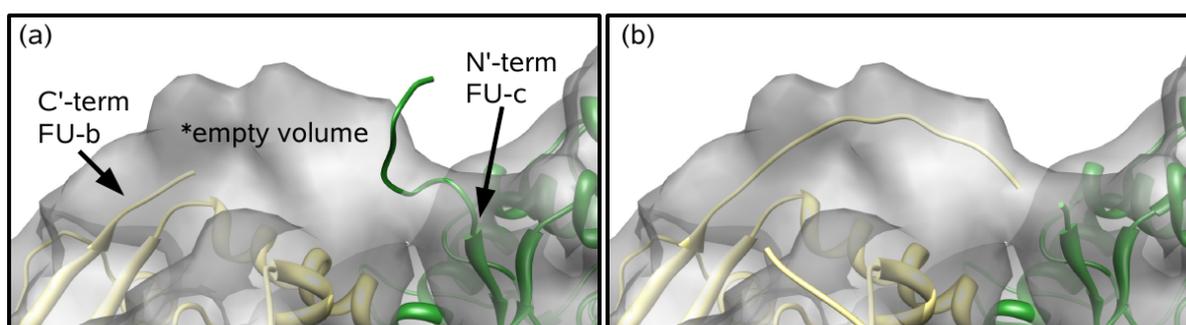
**Fig. 33 : Rigid-body fitting of the 20 initial molecular models of NpH-c into their respective density map**  
Note the similarity between the 20 different molecular models. Variable regions are limited to linker peptides (arrow) and large exposed loops (double-arrows).

The seven selected molecular models were then rigid-body fitted into their respective cryoEM densities using the 6-D FFT-accelerated exhaustive search methodology of the module COLORES (COrrelation based LOw RESolution docking) of the SITUS 2.2 software package, employing a Laplacian filtering and an off-lattice Powell optimization. The improved molecular model of the decamer was then assembled in chimera. Editing of the resulting PDB files (i.e. setting the chain ID, residue numbers etc) was done with the PDB-mode for the text-editor EMACS (<http://stein.bioch.dundee.ac.uk/~charlie/soft-ware/pdb-mode/pdb-mode.html>).

The rigid body fits revealed that loop refinements would be necessary to improve the models and avoid some overlap of side chains. A first refinement was done, by using this first model of the subunit chain as a multimeric template to remodel the subunit chain itself.

This approach corrected some overlap of side chains and improved the overall stereochemical quality of the molecular model (see below).

Loop refinements were performed by MODELLER 8v2. This procedure was also used for a rough modelling of the linkers connecting the functional units (see Fig. 33) since no template was available. Various secondary structure prediction servers (e.g. APSSP; see section A.1.5) confirmed that all “linker” residues have indeed zero strand and helix probabilities. For each of the six different linkers, 40 possible models were then calculated. Models of linker peptides fitting empty volumes or bridges between functional units (see Fig. 34) were then chosen.



**Fig. 34: Modelling of the linker peptides**

a) Fitting of the initial models of FU-c (green) and FU-b (yellow) into their corresponding density masses, showing the unmodelled b-c linker peptide in the vicinity of an empty volume between the corresponding termini b) the final molecular model of the linker peptide connects both termini and clearly fits into otherwise empty mass visible in the cryoEM structure.

Estimation of the quality of the various fitting stages was done by computing the cross-correlation coefficient between the extracted FU density maps and the corresponding (resolution lowered) molecular models using the module COLACOR (Combined off-lattice

correlation) from the SITUS 2.2 software package. In this case the off-lattice Powell optimization was deactivated and due to the high-resolution of the density map (<10 Å) we preferred to use a standard linear cross correlation.

The stereochemical quality of the generated molecular models was checked at all relevant stages with the PROCHECK software. Its outputs (Ramachandran plots) give an assessment of the overall quality of the structure as compared with defined structures of the same resolution.

The quality of the final molecular models was validated using MolProbity. The Molprobity server performs an all-atom contact and geometric analysis of molecular structures. In particular, atoms with substantial van der Waal's overlaps are considered to be errors in the model. Furthermore, Ramachandran and rotamer outliers flag very unusual conformations of the protein backbone or sidechains, respectively, and therefore highlight regions that may need further investigation. Moreover, deviation of C $\beta$  atoms from ideal positions is a measure sensitive to incompatibilities between sidechain and backbone and reveals bond angle distortions.

In order to reduce the number of diagnosed potential errors, the geometry of the subunit-model was improved using the Coot software package. To do so, an automatically generated chart of the detected potential errors was imported in Coot. The geometry of amino acids which were identified both for rotamer and for C $\beta$  deviation was then optimized. This procedure was performed only in regions with >2-3 errors, taking in account the backbone independent sidechain library of the Richardson group (Lovel *et al.*, 2000) and "dynamic" Ramachandran plots (i.e. when one edits the molecule (such as by moving the coordinates of some of the atoms) the Ramachandran plot is updated to reflect the changes).

It should be mentioned that these advanced improvements were only required in some variable loop regions, whereas the highly conserved secondary structure elements yielded high quality values from the beginning.

Possible movements of the subunit chain (and subunit dimer) were predicted by normal mode analysis using the web interface EInemo. Flexible fitting using normal mode analysis according to the NMFF procedure was tried but did not improve the correlation coefficients and the conformations significantly; it was therefore omitted.

#### **4.14 Molecular modelling of the KLH1-polypeptide chains**

Comparative modeling of KLH1-a to KLH1-g and of the corresponding fragment of KLH1-h (core/  $\beta$ -sandwich domain), was performed by the MODELLER 9v4 software using the crystal structures of OdH-g (Cuff *et al.*, 1998) and RtH-e (Perbrandt *et al.*, 2003) as templates. Due to the "low" resolution of the crystal structure of KLH1-h, the orientation of the side chains of

the additional cupredoxin domain remained completely unsolved. Homology modeling with structurally comparable cupredoxins is critical because of low sequence identities and was therefore omitted.

For each FU-type 10 different models were calculated. The resulting homology models of the segment of FU-h showed an overall high-similarity with the available low-resolution X-ray structure (RMSD: 0.6 Å). The major differences were located in large exposed loops. To take in account the backbone-folding of these loops, a chimeric template was built, consisting of the secondary structure-elements of the best homology model of the FU-h segment and the loops of the X-ray structure. Subsequently, 10 models of the FU-h segment were calculated using this chimeric-template and the best model according to Mol-Probity was then chosen. The final model of the FU-h segment showed an overall higher quality in comparison to the X-ray structure, especially in the geometry of the b-sandwich domain and also a lower RMSD (~0.4 Å). The molecular model of the FU-h segment was then combined with the X-ray structure of the cupredoxin domain, in order to obtain a complete molecular model of FU-h. As mentioned above, the side chains of the additional cupredoxin domain remain unresolved. The final FU models of KLH1-a to KLH1-g were also selected based on their stereochemical quality, which was determined using Mol-probity. The molecular models of the eight FUs were then semi-automatically rigid-body fitted into their respective FU-density using the module *Fit-in-Map* from the Chimera software and the first molecular model of the subunit dimer was then assembled. This approach yielded the same results (and corr. coeff.) with an automated rigid-body-fitting procedure using the 6-D FFT-accelerated exhaustive search methodology of the module COLORES of the SITUS 2.2 software package.

Linker peptides were modeled by MODELLER 9v4 using the corresponding inter-FU linkers of the NpH molecular model as templates. In some cases, loop refinement was performed by MODELLER 9v4 in order to increase the correlation with their corresponding densities. The modelling of the two additional g-h linker peptides was performed by the method described for the NpH-linkers (see Fig. 33).

The resulting molecular models of the 2 subunit conformers were then refined by using the first model of the subunit dimer as a multimeric template to remodel the subunit dimer itself. This approach corrected some overlaps of side chains between FUs and improved the stereochemical quality of the molecular model. In some interfaces between FUs, a few rotamers were corrected in CHIMERA for an improved fit with bridges in the density map; only rotamers with high probability values according to the Dunbrack rotamer library (Dunbrack, 2002) were accepted. For quality control of the various fitting stages, FU densities were extracted with the “color zone/split map” module of the CHIMERA software, and correlation coefficients were computed using the module COLACOR from the SITUS

2.2.1 software package. The complete molecular model of the didecamer was then assembled in CHIMERA.

Rigid Body Fitting of glycan side chains in empty masses of the density map (which were not occupied by the molecular model) associated with potential attachment sites for N-linked glycans, was performed manually in CHIMERA. The glycan side chains were obtained from the crystal structure of OdH-g (Cuff *et al.*, 1998). Flexible fitting using normal mode analysis according to the NMFF procedure was tried but did not improve the correlation coefficients and the conformations significantly.

#### 4.15 Further methods

##### **Alignment of molecular models**

Molecular models of single FUs were compared to each other using the module MatchMaker from CHIMERA. MatchMaker superimposes structures by first constructing a sequence alignment and then performing a least squares fit to superimpose the aligned residue pairs. (Meng *et al.*, 2006). Such an alignment was also used to measure the RMSD between molecular structures. The Multalign Viewer was used to view and manipulate the sequence alignments.

##### **Generation of 3D-electron density maps from PDB files**

3D density maps of the molecular models were generated using the module pdb2mrc from the EMAN 1.7 software package. Alternatively, 3D-volumes were created with the command *THREED-MODEL-MOLECULE* (IMAGIC-5) using a “plt” coordinates file. The PDB-files were converted to the IMAGIC “plt” coordinates file using the module *IMPORT-PDB-TO-PLT*. Simulated maps from atoms were also created with CHIMERA (module Fit in Map).

These maps were usually used for a visual comparison with the cryoEM-structures. A simulated surface of the molecular model of the NpH-decamer was also used as reference for the final refinement rounds of the cryoEM structure (see Gatsogiannis *et al.*, 2007).

##### **Fitting of one density map into another**

To allow a detailed visual comparison between various cryoEM-structures (and/or simulated density maps), map-in-map fitting was performed with the *Fit-in-Map* module from the Chimera software. Alternatively, fitting of one density map into another was realized with the module *AffineRegistration* from the AMIRA software.

# C RESULTS AND DISCUSSION

## 1. Overview

In this study, the structures of three extracellular respiratory proteins from mollusks have been investigated:

- *Nautilus pompilius* hemocyanin (NpH)
- *Haliotis tuberculata* hemocyanin Isoform 1 (HtH1)
- Keyhole limpet (*Megathura crenulata*) hemocyanin isoform 1 (KLH1)

The results of the research reported in this thesis have been published in three peer reviewed articles:

[1] Comparative 11 Å structure of two molluscan hemocyanins from 3D cryo-electron microscopy

Meissner U, [Gatsogiannis C](#), Moeller A, Depoix F, Harris JR, Markl J  
Micron 2007; 38(7):754-65. doi: 10.1016/j.micron.2006.11.005

[2] *Nautilus pompilius* hemocyanin: 9 Å cryoEM structure and molecular model reveal the subunit pathway and the interfaces between the 70 functional units

[Gatsogiannis C](#), Moeller A, Depoix F, Meissner U, Markl J  
J Mol Biol. 2007 Nov 23; 374(2):465-86. doi: 10.1016/j.jmb.2007.09.036

[3] Keyhole limpet hemocyanin: 9 Å cryoEM structure and molecular model of the KLH1 didecamer reveal the interfaces and intricate topology of 160 functional units

[Gatsogiannis C](#), Markl J  
J Mol Biol. 2009 Jan 23; 385(3): 963-983. doi: 10.1016/j.jmb.2008.10.080

This chapter is subdivided in three sub-chapters, which consist of each peer reviewed publication and a section summarizing the most important findings and providing supplementary data. Throughout the next sections these publications are referred by the corresponding numerals.



## 2. Comparative 11 Å structure of two molluscan hemocyanins from 3D cryo-electron microscopy

A wealth of electron microscopical data on the quaternary structure of molluscan hemocyanins are now available and also hemocyanin sequences from different molluscan phyla and crystal structures for two different FU-types are known. These data in combination with previous lower resolution structures (e.g. Lambert *et al.*, 1994a, b, c; Mouche *et al.*, 1999, 2003; Meissner *et al.*, 2000) have established the following details:

- The hemocyanin hollow cylindrical decamer consists of a wall and an internal collar complex.
- The cylinder wall consists of 10 copies of the subunit segment containing FUs a-b-c-d-e-f.
- In cephalopods the collar is built from ten copies of FU-g (an exception is *Sepia* hemocyanin; this hemocyanin will not be further addressed here).
- In gastropods and chitons the decamer resembles the native cephalopod hemocyanin decamer, but with an off-set collar at one edge of the decamer. In this case, the collar complex additionally contains 10 copies of FU-h.

(for a more detailed review see section A.2 and Introduction in [1], [2] and [3]).

The decamer is the basic oligomeric assembly of native molluscan hemocyanin. However, many details of its molecular architecture still remain unclear. Although several models for the topology of the wall segment a-b-c-d-e-f of the subunit and the architecture of the subunit dimer (the major assembly and disassembly intermediate of molluscan hemocyanin) have already been proposed (Meissner *et al.*, 2000; Gebauer *et al.*, 2002; Lambert *et al.*, 1994) the orientation of the subunits within the wall of the decamer and the convoluted pathway of the elongated subunit remain obscure.

To approach this, firstly a 3D reconstruction from cryo-electron micrographs of the hemocyanin molecule from the cephalopod *Nautilus pompilius* was performed (recently our group also solved its complete amino acid sequence; Bergmann *et al.*, 2006). The NpH decamer resembles the basic structure of molluscan hemocyanins and is therefore probably an adequate molecule to be used as a molluscan hemocyanin prototype.

Indeed, sequence alignments demonstrated that the polypeptide a-b-c-d-e-f-g of *Nautilus* shows a high sequence identity to the polypeptide chain of another Cephalopod (*Octopus dofleini*) (Miller *et al.*, 1988) and both sequences are also orthologous to the polypeptide fragment a-b-c-d-e-f-g from gastropods such as *Haliotis* (Lieb *et al.*, 2000) or *Megathura* (Lieb & Markl, 2004).

However, as already mentioned, gastropod hemocyanins have an additional FU (FU-h) attached to the common polypeptide fragment a-b-c-d-e-f-g, which forms in 10 copies the additional off-set collar. It is still unclear if this additional collar complex resembles the only difference between cephalopod and gastropod hemocyanins. Previous low resolution cryoEM structures suggested a similar wall, but an overall different collar architecture. To-date, a detailed comparison between a cephalopod and a gastropod hemocyanin has not been performed.

To approach this, we decided to perform in parallel to the NpH-reconstruction, and our attempt to solve the basic structure of molluscan hemocyanins, to refine the available “low” resolution cryoEM structure of the gastropod hemocyanin HtH1 (Meissner *et al.*, 2000).

Here we present the first detailed structural comparison between a cephalopod and a gastropod hemocyanin (published in 2006 [1]).

## 2.1 11 Å cryoEM structure of NpH and HtH1

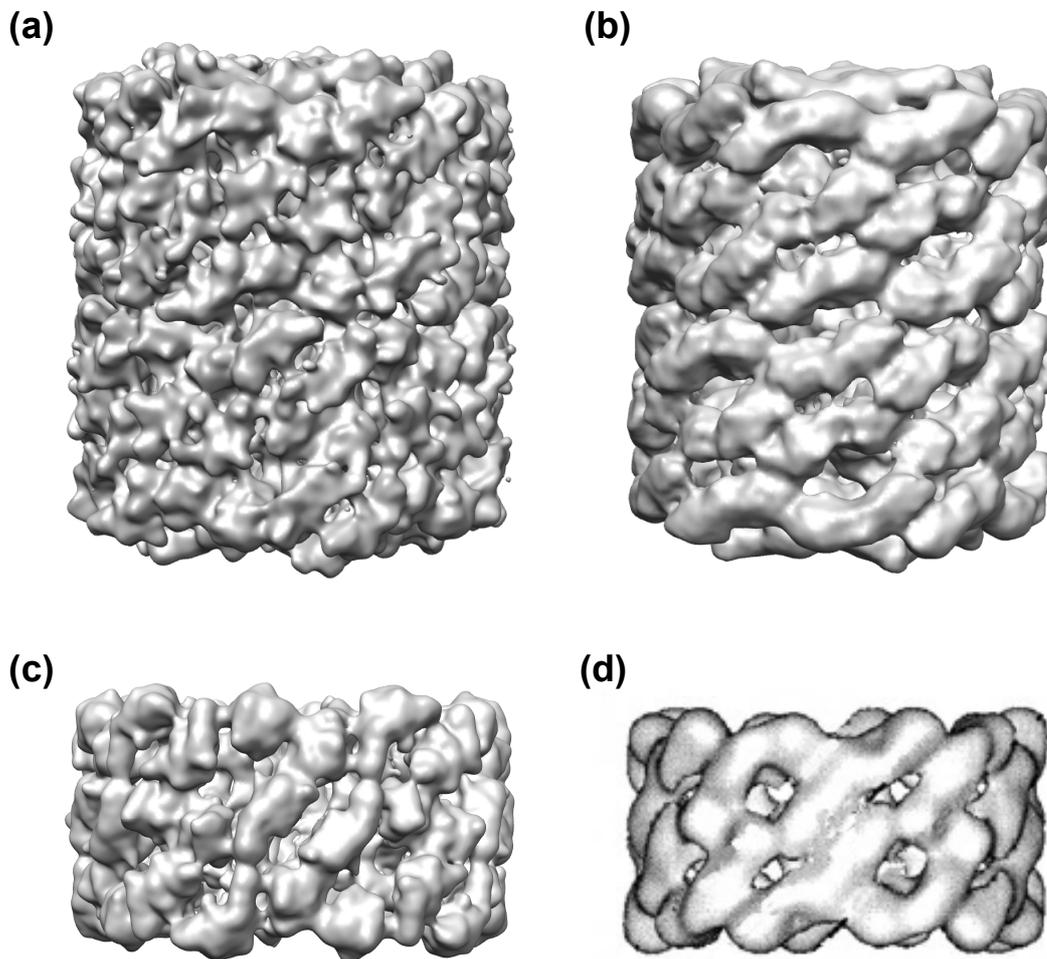
My first attempt to reconstruct NpH (Gatsogiannis, 2005; diploma thesis) using the angular reconstitution technique, achieved a resolution of  $\sim 15$  Å (1/2-bit-criterion) (data not shown). This cryoEM structure was used as a starting model for the projection matching technique. After several refinement rounds, a 10.9 Å reconstruction was achieved with imposed D5 symmetry (for more details see Material and Methods in [1] and Gatsogiannis, 2005).

Furthermore, using the previously published  $\sim 14$  Å cryoEM structure of HtH1 as an initial reference, with a refined image dataset and improved methodology, a cryoEM structure of HtH1 at 10.6 Å was produced (for more details see Material and Methods in [1] and Meissner *et al.*, 2000).

The obtained reconstructions of HtH1 and NpH represent the highest resolution reconstructions of single particles so far achieved for a gastropod and a cephalopod hemocyanin respectively. Furthermore, both reconstructions are directly comparable, at approximately the same resolution.

The new data confirm and extend the results from previous lower-resolution structures (Lambert *et al.*, 1994a; Mouche *et al.*, 1999, 2003; Meissner *et al.*, 2000) (see Fig. 35) and accordingly allow a more detailed interpretation. In comparison to the previous 14 Å reconstructions of HtH1 and KLH1 (Meissner *et al.*, 2000; Mouche *et al.*, 2003), the difference in resolution implies  $\sim (14/10.6)^3$  or 2.3 times more information in the new HtH1

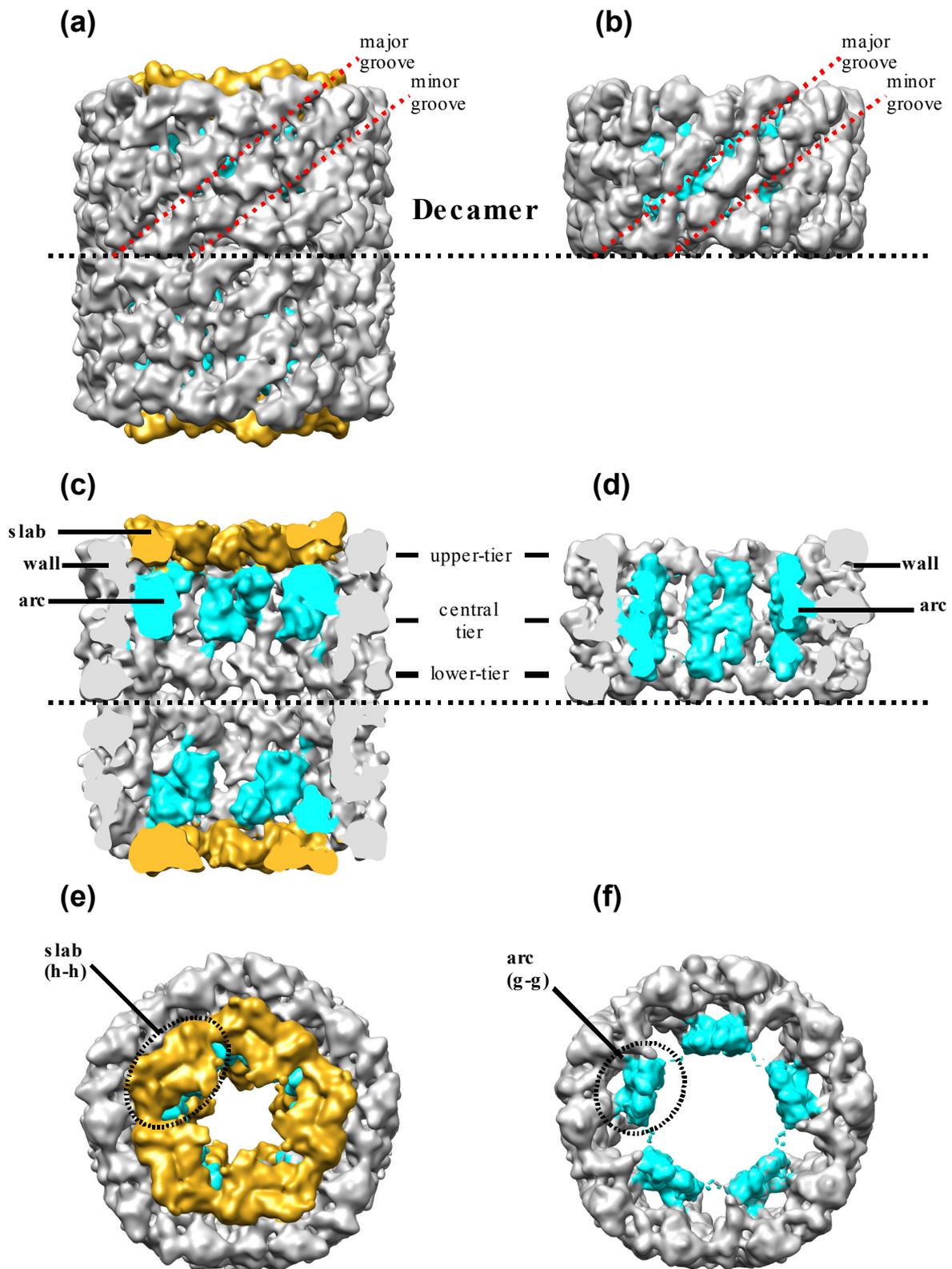
reconstruction. Similarly, the NpH reconstruction contains 13.5 times more information in comparison to the 26 Å reconstruction of *Benthoctopus* hemocyanin (Mouche *et al.*, 1999).



**Fig. 35: Surface representations of the present cryoEM structures in comparison to the previous corresponding highest resolution reconstructions**

a) Side view of the 11 Å cryoEM structure of HtH1 b) Side view of HtH1 at 14 Å (Meissner *et al.*, 2000) c) Side view of the present 11 Å 3D reconstruction of NpH d) Side view of the closely related hemocyanin from the cephalopod *Benthoctopus* hemocyanin at 26 Å (Mouche *et al.*, 1999). In comparison with the previous published structures, the present 3D reconstructions show superior morphological details.

The present cryoEM-reconstructions of NpH and HtH1 show all structural features that characterize a typical decameric cephalopod hemocyanin and a didecameric gastropod hemocyanin, respectively. To give a better overview, these features and for a first comparison, both reconstructions are depicted in Fig. 36.



**Fig. 36: Comparison of the 11 Å cryoEM structures of Hth1 and NpH**

(a) Side view (c) cut-open view and (e) end-on view of the Hth1 didecimer. (b) Side view (d) cut-open view and (f) end-on view of the NpH decamer. The wall of the hemocyanin molecules is shown in grey and the central collar complex in cyan. The additional peripheral collar (FU-h) of Hth1 is shown in orange. For more details see main text.

Briefly, the NpH-decamer is organised as a hollow cylindrical oligomer, which consists of an external wall (shown in grey in Fig. 36) and an internal collar complex (shown in cyan).

The external wall can be described as a superposition of three well defined layers or tiers. Each tier contains 20 functional units which encircle the cylinder. Furthermore, the molecule wall has the appearance of a right handed-helix and is engraved by five major and five minor oblique grooves. Ten copies of the FU-g form an internal collar complex at the level of the central wall tier. These functional units are organized as pairs and form five distinct collar morphological units, described as arcs (dashed circle in Fig. 36 (f)).

The HtH1 decamer resembles the NpH-decamer, but with an additional collar (FU-h shown in gold) at one edge of the decamer (closed face), yielding an asymmetrical arrangement. The native structure, the didecamer, is an assembly of two decamers at their open faces. The additional collar consists of five morphological units, called slabs (dashed circle in Fig 36 (e)). Each slab contains two copies of FU-h.

## 2.2 The decamer wall architecture

A detailed analysis and comparison of both 3D reconstructions revealed that the wall architecture of the two decamers is very similar. Indeed, the arrangement of the masses within the wall of both hemocyanin decamers is almost identical.

The wall FUs are anti-parallel and are arranged in pairs, described previously as morphological units (Orlova *et al.*, 1997). The wall of each decamer consists of 30 morphological units (=60 FUs).

At 11 Å, the shape of individual FUs within the morphological units is clearly defined and six different types of wall FUs (corresponding to FUs a-b-c-d-e-f) are for the first time individually structurally discernable. However, the present 11 Å cryoEM structures did not allow us to identify the pathway of the subunit and therefore confirm the FU-topology as proposed previously (Meissner *et al.*, 2000; Gebauer *et al.*, 2002) (see Fig.5 (d) in [1]). Although we were able to identify the 6 different wall FU-types (corresponding to FUs a to f), their location still remains unclear.

Therefore, we used a new structurally defined nomenclature for the six different wall FUs, which is independent of their position in the subunit polypeptide sequence FU-a to FU-f, and are now termed FU-1 to FU-6 (see Fig. 4 and 5 in [1]).

The morphological units are formed by the FU-s 1-2, 3-4 and 5-6. Each peripheral tier of the wall contains 5 copies of the morphological units 1-2 and 3-4 (=10 morphological units, =20 FUs). Due to the dyad symmetry, the FU sequences in the peripheral tiers run anti-parallel to each other. The five-fold repeating unit of the peripheral tiers is the tetrad 1-2-3-4.

The central tier contains 10 copies of the morphological unit 5-6 (=20 FUs). In this case, the repeating unit is the tetrad 6-5-5-6 (see Fig. 4 in [1]).

The extracted density maps of single wall FUs have been found to be strikingly similar in both hemocyanins (see Fig. 7 in [1]). Moreover, corresponding wall FUs show a strong resemblance with respect to their orientation and typical views within the molecule and also their contacts with adjacent FUs, which correlates with the amino acid sequence data. The sequence identity is ~55 % percent between orthologous FUs, but only ~45 % between paralogous FUs (Lieb *et al.*, 2000; Bergmann *et al.*, 2006). Hence, it can be directly shown (at least for the wall FUs) that corresponding FUs occupy the same positions and play the same architectural role within the quaternary structure of their respective hemocyanins. This suggests that all molluscan hemocyanins have a strictly similar structural organization at the level of the decamer wall and possess the same intricate path for the six wall-forming FUs.

In contrast, the collar architecture in both molluscan hemocyanins appears to be substantially different.

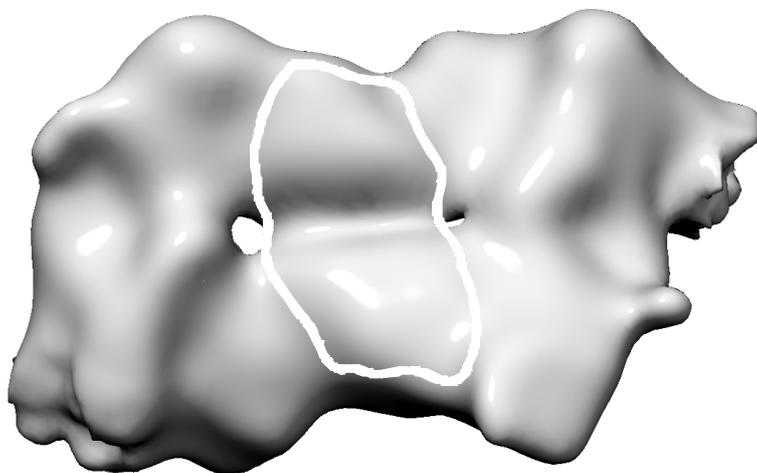
### 2.3 The architecture of the collar complexes

As already mentioned, the collar complex of HtH1 consists of central collar and a peripheral collar (Fig. 36 (c)). The central collar consists of five arcs, each being FU-g pairs. The peripheral collar is composed of five slabs, each being FU-h pairs (see Fig. 8 in [1]). FU-h contains 100 amino acids more than the other seven FU types (Lieb *et al.*, 2000).

The present reconstruction provides superior structural definition of the peripheral collar because the two copies of FU-h can now for the first time be identified within the morphological units, termed the slabs. The ten FU-h copies have a more defined shape and are clearly arranged as anti-parallel pairs to form five slabs. The five slabs of each decamer are particularly clear in the end-on view (Figure 7a in [1]; Fig 36 (e)).

The large contact region (see curved white field in Fig 8d [1] and Fig. 37) within the slab clearly corresponds to two copies of the unique extension of FU-h that extends this FU C'-terminally by ~100 aminoacids (Lieb *et al.*, 2000).

As for HtH1, NpH also contains a central collar consisting of five arcs (FU-g pairs) (see Fig. 8 in [1]; Fig. 36 (d)), but the peripheral collar is lacking. The biological significance of this apomorphy is obscure. Moreover, a direct comparison of the collar complexes reveals that the two arcs differ in their relative position and also in shape.



**Fig. 37: Extracted morphological unit (slab) of the peripheral collar of HtH1**

The two FU-h copies within the slab are anti parallel arranged. Note the large contact region (curved white margin) between the two FUs.

In HtH1, the arc is attached to the central tier and is directly connected to the slab, *via* two well defined bridges (see Fig. 9b in [1]). One of the two FU-g copies constituting each arc, is shifted towards the slab, yielding an asymmetrical arrangement (see Fig. 9 in [1]). This arrangement is in complete agreement with previous lower resolution reconstructions. Apparently, both FU-g copies are arranged in such a way that their C'-termini point contact the N'-termini of the adjacent FU-h pair (see Fig. 9 in [1]). Furthermore, the shape of both FU-g copies is compatible with the X-ray structure of OdH-g (Cuff *et al.*, 1998).

In NpH, the two FU-g copies of the arc show an arrangement with dyad symmetry. The symmetrical arc appears to be tetrameric and has an elongated, flattened shape. These results are in contrast with previous low resolution reconstructions, showing a symmetrical dimeric arc.

The central part of the arc contains 4 masses which correspond in size and overall shape with the 2 copies of OdH-g, which consist of 2 copies of the small six-stranded anti-parallel  $\beta$ -sandwich C-terminal domain (yellow dashed circle in Fig. 38) and 2 copies of the large  $\alpha$ -helical copper binding N-terminal domain (blue dashed circle).

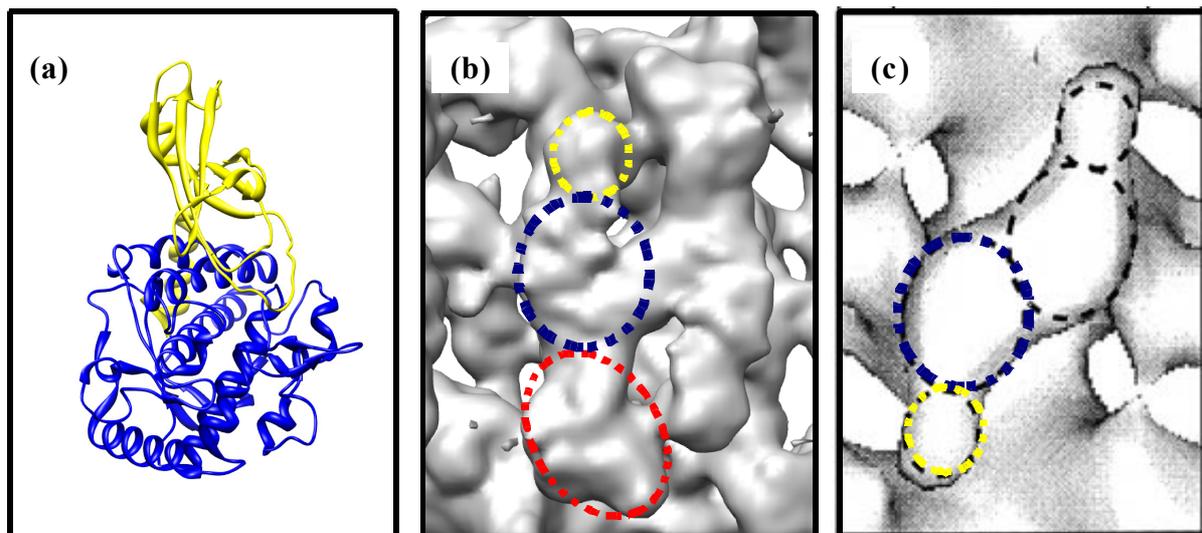
Moreover, the central part of the arc shows a similar structure to the 26 Å reconstruction of the closely related *Benthoctopus* hemocyanin (Mouche *et al.*, 1999). However, it appears to be the enantiomeric version of the arc of this hemocyanin (see Fig. 38).

If the two additional masses (red dashed circle) are indeed part of FU-g, then each FU-g would have approximately a molecular mass of ~65 kDa. This value is ~18 kDa (28%) higher than the molecular mass of an individual FU-g deduced from the amino acid sequence data (46,614 Da).

One possibility is that this additional mass is a symmetry induced artifact. However, an independent reconstruction with imposed C5-symmetry confirmed its existence. It should be also mentioned that the previous 3D-reconstructions of cephalopod hemocyanins were also performed with imposed D5-symmetry.

We also speculated that this additional mass is probably an associated 18 kDa protein, but a careful re-evaluation of the data from the previous biochemical study (Bergmann *et al.*, 2006) did not confirm this.

Another possibility is that a specific glycosylation pattern might account for the observed abnormality of NpH-g. Nevertheless, the presence of a small protein attached to NpH-g could not at that time be excluded.



**Fig. 38: Comparison of NpH-g with the X-ray structure of OdH-g and FU-g from Benthoctopus hemocyanin**  
 a) X-ray structure of OdH-g at 2.3 Å resolution (N-terminal- and C-terminal domain shown in blue and yellow respectively) (Cuff *et al.*, 1998). (b) Orientation of the N- (blue dashed circle) and C-terminal domains (yellow dashed circle) within an arc of the 11 Å cryoEM structure of NpH. Note the additional associated density mass (red dashed circle) (c) Orientation of the N- and C-terminal domains within an arc of the 26 Å cryo-EM structure of Benthoctopus hemocyanin (according to Mouche *et al.*, 1999; modified).

## 2.4 Pathway of the subunit

There are two main reasons why the path of the subunit remains obscure even at 11 Å despite the fact that its C'-terminal FU is known (FU-g for NpH, FU-h for HtH1).

- The arcs FUs (FU-g) do not have direct contact with specific FUs of the wall. The arcs in both hemocyanins are attached to the wall through a previously undescribed structural element of the wall, now termed the anchor (see Fig. 4; Fig. 10 in [1]). The central part of the anchor has the shape of a vertical rod with a terminal bifurcation, one leading to FU-5 and FU-6, and the other one to FU3-and FU-4. At 11 Å the individual contribution of these four wall FUs to the anchor structure remains puzzling and the further path of the subunit remains uncertain. Thus, in spite of their 11 Å resolution, the current reconstructions do not yet allow a precise identification of all individual connections between functional units and definition of the complex twisting of the complete polypeptide chain.
- Homology modeling of the known amino acid sequences of NpH and HtH1 using the two available X-ray structures as templates, should enable rigid-body fitting of comparative molecular models within the 11 Å density maps and a resulting pseudo-atomic model should reveal the elongated pathway of the subunit. Unfortunately, docking of X-ray data and molecular models within the extracted cryoEM-maps of single FUs yielded ambiguous results. Due to the lack of definable secondary structure elements and the ovoid shape of FUs, the rigid body fittings were not unique (see Fig. 39).

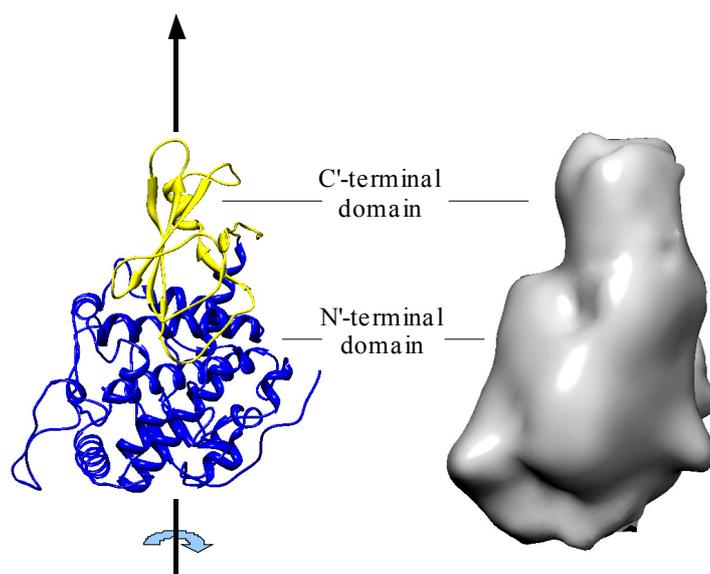


Fig. 39: (legend on next page)

**Comparison of the X-ray Structure of OdH-g with the extracted cryo-EM density of a NpH wall FU.**

The comparison allows identification of the N- and C'-terminal domains in the cryoEM density but uncertainty remains about the rotation around the FU long axis.

---

Nevertheless, at sub-10 Å resolution or better, one can expect to observe not only individual domains, but also secondary structure within the masses now defined as FUs, which would allow high quality molecular fitting. This approach should then enable us to define the location of individual subunits and FUs within the quaternary structure of the hemocyanin molecule and solve the basic structure at (pseudo)atomic resolution.

### **3. *Nautilus pompilius* Hemocyanin: 9 Å Cryo-EM structure and Molecular model reveal the subunit pathway and the interfaces between the 70 functional units**

The comparative cryoEM-3D analysis at 11 Å resolution (above) has shown that the wall of *Nautilus* hemocyanin is very similar to the wall of a gastropod hemocyanin (HtH1) and the collar architecture is significantly different [1]. However, reliable fitting of comparative molecular models into the 11 Å cryoEM-structures failed and the pathway of the subunit was still unclear. Our subsequent study published in 2007 [2] extends our previous work on *Nautilus* hemocyanin and provides the first superior resolution 3D reconstruction of a molluscan hemocyanin. From advanced molecular modeling and rigid-body-fitting, we generated an all atom cryoEM/crystal structure hybrid model of NpH, which provides a fundamental understanding of the structure and function of molluscan hemocyanins in general.

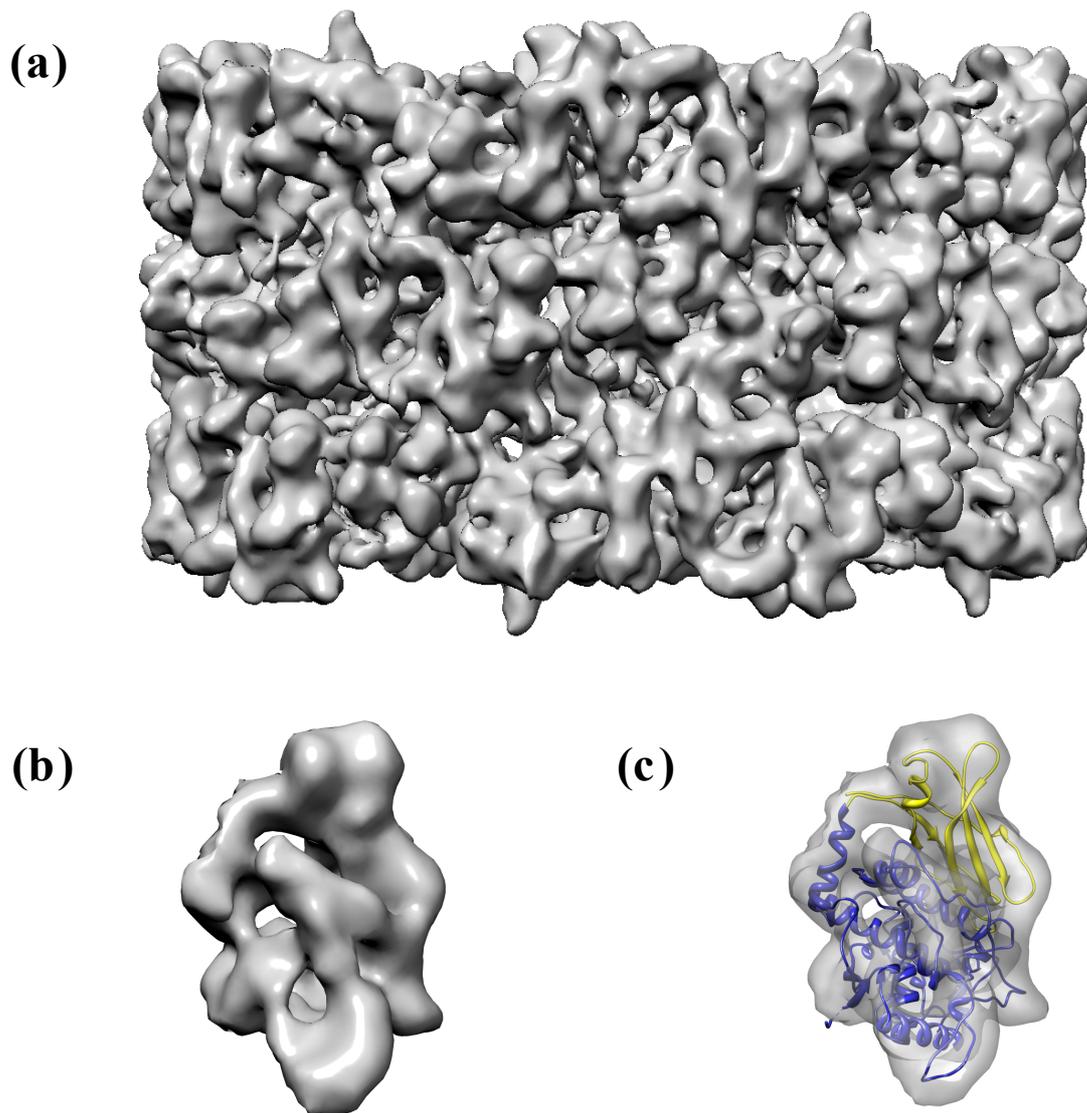
#### **3.1 8.4 Å NpH decamer with imposed D5 symmetry**

In order to improve the resolution of the 11 Å NpH reconstruction, we generated a new image dataset of superior quality, consisting of 18,875 low defocus images of single hemocyanin molecules, extracted from 63 micrographs.

The single particles were submitted to several refinement rounds, using the previous 11 Å structure as reference and taking the imposed D5 symmetry into account (for details see Material and Methods in [2]).

The resulting 3D reconstruction reached a resolution of 8.4 Å. At this resolution level, well separated  $\alpha$ -helices could for the first time be resolved and we were able to perform high-quality rigid body fitting of the available X-ray-structures into the density map (see Fig. 40).

However, the architecture of the five arcs still remained puzzling. Each arc remained less-well resolved and still contained two additional masses (component “x”) associated with the “FU-g dimer”, and completely unpredicted from biochemical analysis (see Fig. 41 (f) and Fig. 2 in [2]).

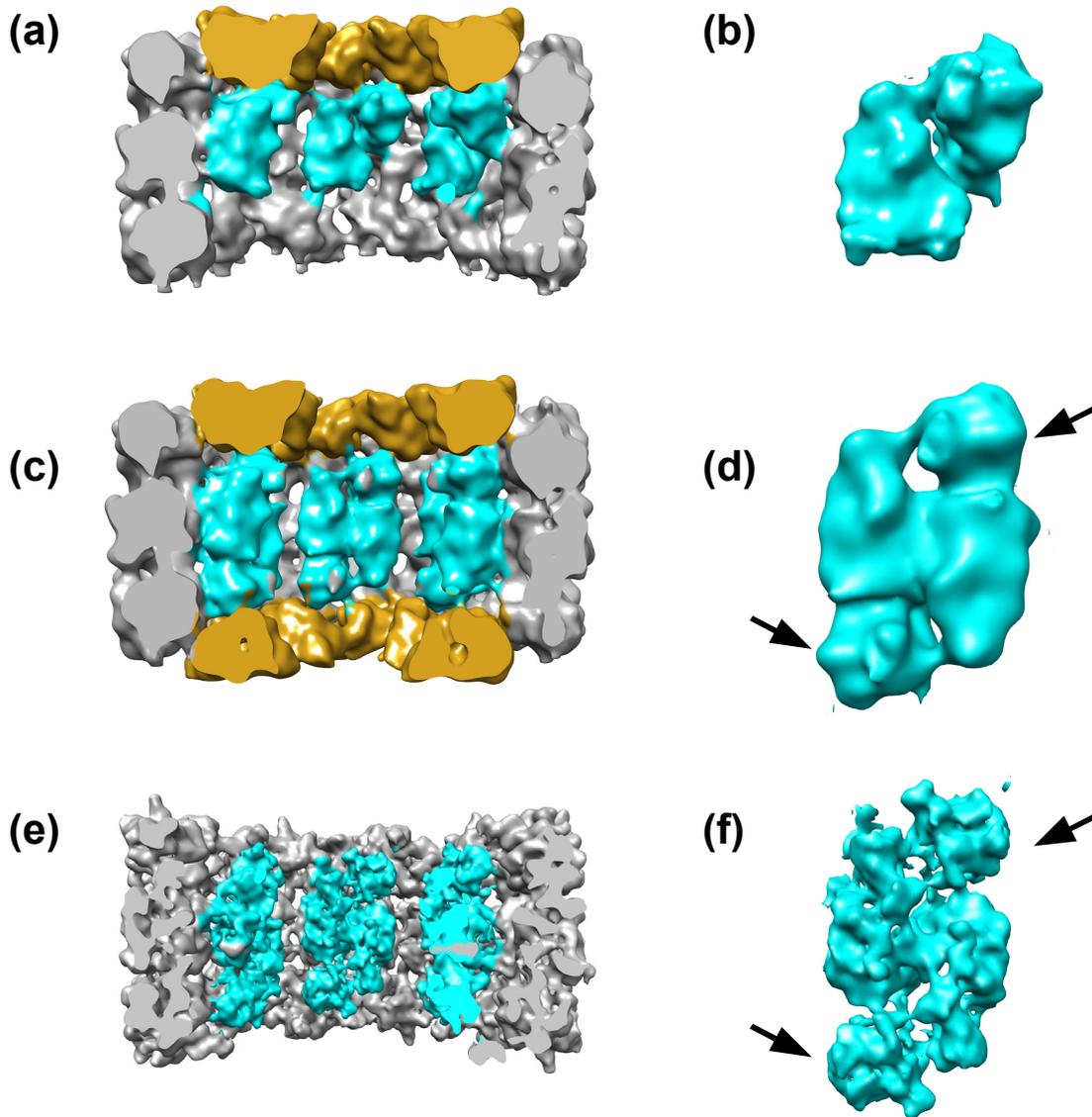


**Fig. 40: 8.4 Å cryoEM structure of NpH with imposed D5 symmetry**

a) Side view of the decamer. Note the superior morphological details in comparison to the previous 11 Å cryoEM structure (see Fig. 35 c, 36 b and Fig. 2 in [1]). b) Extracted density map of a single wall FU c) rigid body fitting of the crystal structure of the wall FU RtH-e into the density map.

Nevertheless, we found out that the strange appearance of the arc surprisingly resembles a structure obtained by fusing *in silico* the asymmetric arc of HtH1 with its flipped copy (see Fig. 41 (d)). It becomes clear, that the imposed D5 symmetry caused arc duplication in the cryoEM structure of NpH. Thus, the symmetrical arc is a symmetry artifact and “component X” simply does not exist. Furthermore, this result suggests that the NpH-molecule consists of a wall with perfect D5 symmetry and an asymmetric internal collar complex, which

corresponds to the internal collar complex described in gastropod hemocyanins (see [1]). Consequently, the quaternary structures of NpH and HtH1 are more similar than previously expected ([1], van Holde and Miller, 1995).



**Fig. 41: Revealing the architecture of the NpH-arc**

**a)** 11 Å structure of the HtH1-didecamer. Note that the HtH1 didecamer is the half structure of the native didecamer and due to the presence of the asymmetric arc (shown in cyan) and peripheral collar (shown in orange) in part asymmetric **b)** extracted arc of (a) **c)** HtH1 decamer with imposed D5 symmetry. Note that the applied symmetry causes duplication of the central (cyan) and peripheral collar (orange), whereas the wall architecture remains the same. **d)** Extracted arc of (c) **e)** 8.4 Å structure of the NpH-decamer with imposed D5 symmetry **f)** Extracted arc of (e). Note that the NpH arc (f) resembles exactly the “artificial” arc (d), extracted from the HtH1 structure with the wrong symmetry. Thus, the symmetrical tetrameric arc is an artifact of the imposed D5 symmetry and component X (arrows in d and f) does not exist.

However, it still remains a mystery, that all earlier 3D reconstructions of cephalopod hemocyanins, obtained by applying a D5 symmetry (Lamy *et al.*, 1993; Lambert *et al.*, 1994a, 1995; Mouche *et al.*, 1999) do not show a tetrameric arc architecture, as the D5-reconstruction of NpH, but a symmetrical dimeric arc, lacking a component X (see also Fig.

38). There might be a specific situation concerning the arc architecture in *Nautilus* hemocyanin that differs from the properties in other 7-FU cephalopods, such as *Octopus*, but this would be in contrast with the primary sequence and biochemical data, which suggest an overall similar architecture.

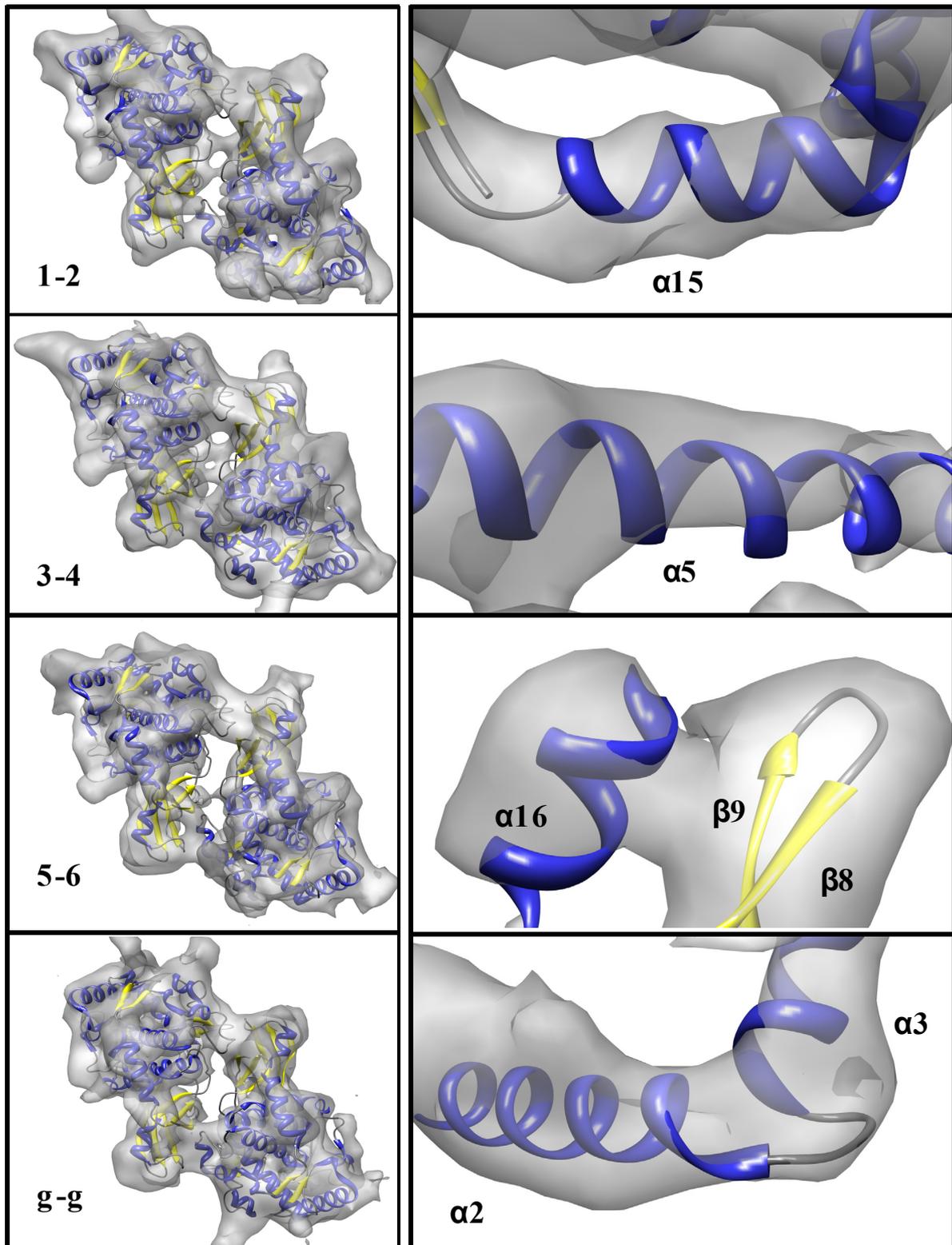
Using point group C5 symmetry in the reconstruction process and an asymmetric decamer as initial reference (see Material and Methods in [2]), a 9.1 Å cryo-EM structure of NpH was ultimately achieved.

### **3.2 9.1 Å NpH-decamer with imposed C5 symmetry.**

The principal architecture and parameters of the wall of NpH have already been discussed in the previous 11 Å cryoEM structure. The new C5 reconstruction additionally reveals that the collar is shifted towards one of the cylinder openings. The constellation of the two FU-g copies strikingly resembles the arc described in gastropod hemocyanins (see [1]). Moreover, the superior resolution of the present density map allows a convincing fit of molecular models into the cryo-EM structure and has ultimately revealed what we now consider to be the correct pathway of the FUs within the folded subunit.

#### **3.2.1 Subunit dimer and pathway of the subunit**

Individual FUs are discernible at lower resolution (see [1]), but the present cryoEM structure reveals them in greatly improved detail. Moreover, the superior resolution of the map allowed us to detect that all FU-pairs forming morphological units (see also [1]) correspond in constellation to the two protomers in the OdH-g crystals (Cuff *et al.*, 1998) (see Fig. 42). In the wall morphological units, the two protomers are associated as a pseudo-dyad symmetry (the two associated FUs are different), whereas by the arc morphological units (g-g) it is a perfect dyad symmetry. Based on the rigid-body-fitting of the OdH-g dimer into the density map of NpH, a first pseudo-atomic model of the whole decamer was then assembled, which allowed an approximate identification of the C'- and N'-termini of the seven different FU-types. Nevertheless, the complex architecture of NpH and the long linker peptides, which encompass 12-20 amino acids, required careful topological analysis to trace the subunit pathway, from FU-a to FU-g.



**Fig. 42: Fitting of OdH-g crystal structure into the NpH-cryoEM-structure**

Left-row: rigid-body fitting of the X-ray structure of the FU dimer described for the Octopus FU-g, into the four different extracted morphological units of NpH (1-2, 3-4, 5-6, g-g). The morphological units are shown in approximately the same orientation. Right row: The rigid-body fitting allows identification of helical (shown in blue) and strand (shown in yellow) segments in the 3D structure. The best representation of these fine details is visualized in surface renderings of the map, enclosing 50% of the expected molecular volume.

In our previous study (see [1]) we introduced a structurally defined nomenclature for the six different wall FUs. The 4 different FU types of the peripheral wall tiers were designated as FU-1 to FU-4, whereas the two central FU types as FU-5 and FU-6.

For the topology of the seven different FU-types, defined within the present study, we consider that only one model of the subunit pathway, in particular  $3 \rightarrow 4 \rightarrow 1 \rightarrow 5 \rightarrow 6 \rightarrow 2 \rightarrow g$ , remains as a reasonable possibility.

This assessment takes the following observations into account:

- the positions of the N'- and C'-termini in the first pseudo-atomic model (see above)
- the potential length of the 6 different inter-FU linker peptides (see Table 3 in [2])
- the central collar of the *Octopus dofleini* hemocyanin decamer, which was shown to be formed from FU-g (van Holde and Miller, 1995), corresponds structurally to the arc of NpH. This is in complete agreement with the available primary sequence data (Keller *et al.*, 1999; Bergmann *et al.*, 2006)
- according to available immunolocalization data from closely related hemocyanins, FU-a, FU-b and FU-c are localized in both outer tiers (Lambert *et al.*, 1994a) (Lamy *et al.*, 1998; Gebauer *et al.*, 2002)

(for more details see section “the pathway of the subunit” in [2])

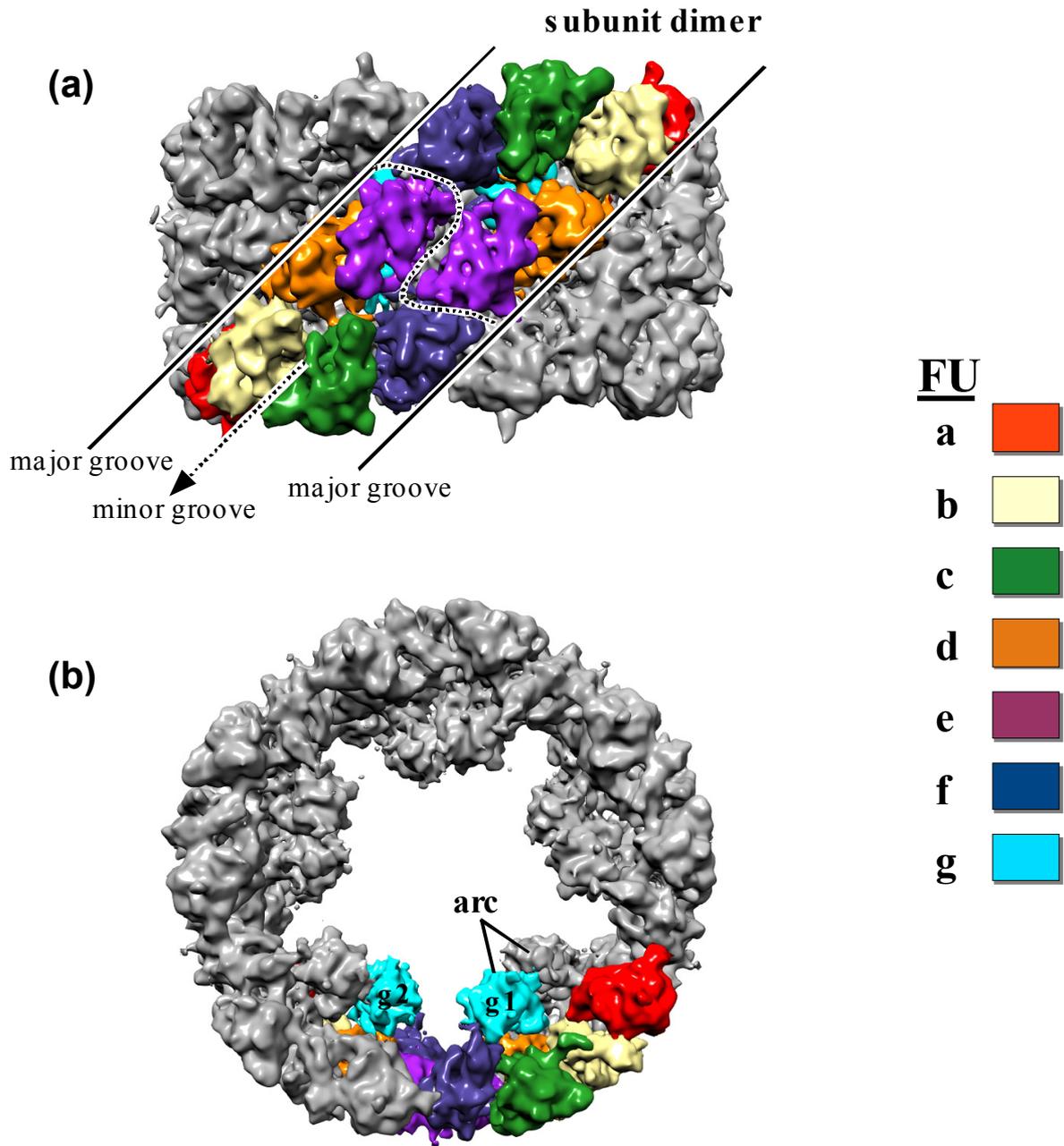
In our advanced model, the FUs **a** (3), **b** (4), **c** (1), **f** (2) are located in the external tiers, FUs **d** (5) and **e** (6) are located in the central tier and the last one (FU-g) in the collar (see Fig. 43 and Fig. 2 in [2]).

In each subunit dimer (the basic repeating unit of the decamer) the two polypeptide chains run in opposite (anti-parallel) direction. As a result of this arrangement of the subunit dimer, the internal tier contains two functional units from each polypeptide chain (2 x FU-d and 2x FU- e) and the external tiers each contain four FUs (FU-a, FU-b, FU-c, FU-f) from the same polypeptide chain.

It should be noted, that the same subunit folds its tail (FU-g) into two different conformations (termed here (FU-g1 and FU-g2)), whereas the two FU-g copies are not directly connected. Consequently, the arcs do not link the two polypeptide chains of the dimeric subunit but bridge neighboring subunit dimers.

Five subunit dimers linked by five arcs assemble the whole molecule. The arcs cross the major grooves of the molecule, which “separate” the wall segments of the five subunit dimers.

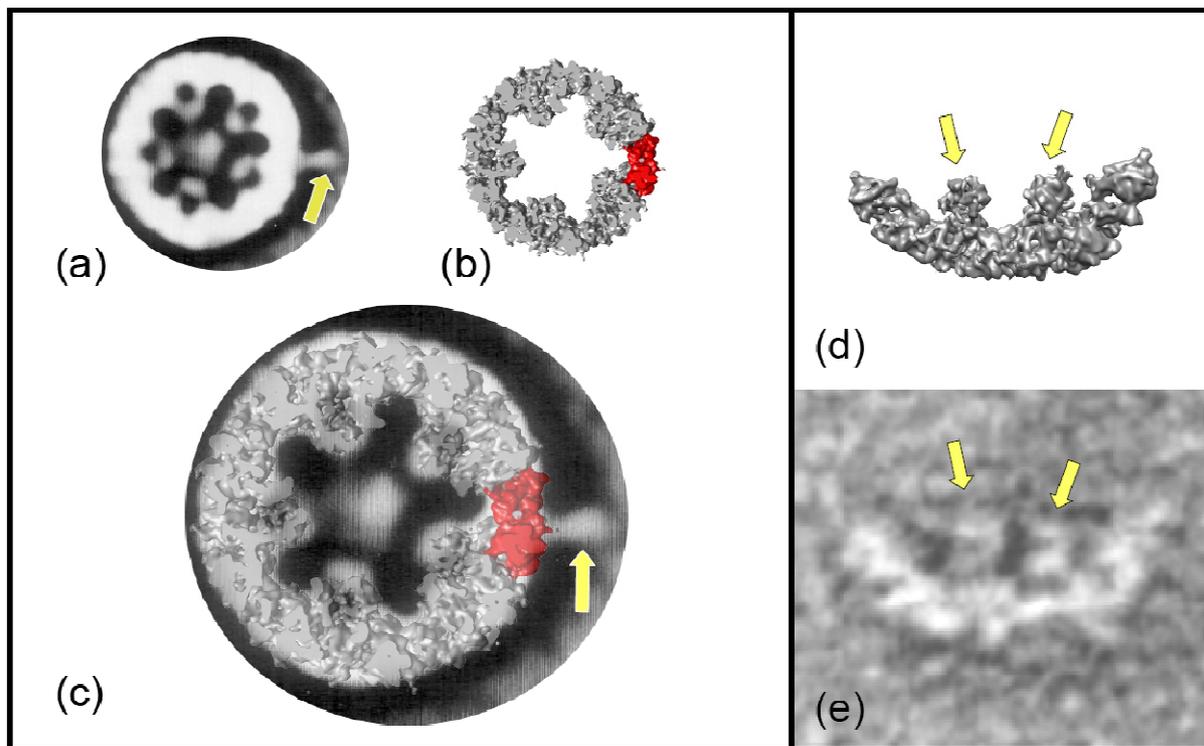
Although similar in shape, the proposed subunit dimer for *Haliotis* (Meissner *et al.*, 2000) is excluded for *Nautilus* hemocyanin, because it does not reasonably comply with the current structural considerations, based on the high resolution cryoEM structure and the initial molecular model. In contrast, the present results fit the subunit dimer previously proposed for *Octopus* hemocyanin (Lambert *et al.*, 1994a). However, the topology of the functional units within the dimer is not compatible with the model of Lambert *et al.*.



**Fig. 43: 9.1 Å cryoEM structure of *Nautilus* hemocyanin and topology of the seven different types of FU**  
 a) Side view and b) end-on view of the 3D map, with one of the five subunit dimers highlighted in colour. The dashed line indicates the border of the two constituent subunits. Note in (b), that the arc is composed from two copies of FU-g (g1 and g2), which stem from adjacent subunit dimers. For more details see main text.

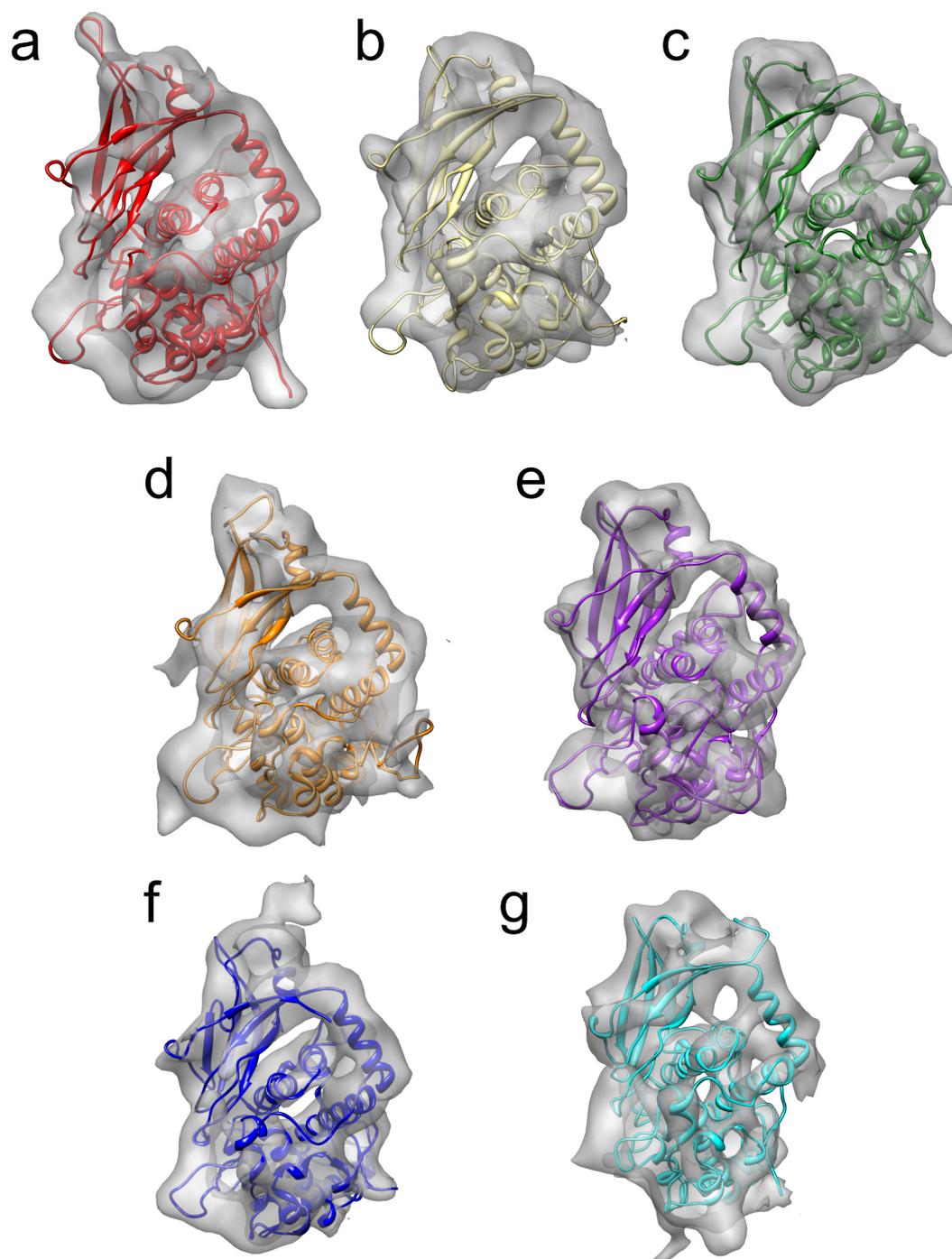
The present model for the subunit dimer and pathway is further verified by the following observations:

- the proposed model is compatible with the parallelogram and “man-in-boat” aspects seen in electron micrographs of the *Octopus* hemocyanin subunit dimer (van Holde *et al.*, 1991) (see Fig. 44 (d) and (e)).
- an independent additional argument for the assignment of FU-6 as FU-e, comes from a comparison of the current reconstruction with *Octopus* hemocyanin tagged with a Fab fragment against FU-e (Lambert *et al.*, 1994) (see Fig. 44 a, b and c)
- the resulting model elucidates several as yet unexplained protrusions and bridges in the cryoEM-structure (see below; see Fig. 2 (b) (arrow) , Fig. 4 d in [2] and Fig. 47)



**Fig. 44: Supportive evidence for the present model of the FU localization and the subunit dimer**

(a) End-on view average image of negatively stained immunocomplex composed of an *Octopus dofleini* hemocyanin molecule and an anti-OdH-e Fab fragment (arrow) (Lambert *et al.*, 1994) (b) End-on-view of the NpH reconstruction with the upper tier removed. From previous immuno-localization data it is undisputed that FU-e is located in the internal tier (see Lambert *et al.*, 1994, Fig. 6b). Two adjacent copies of FU-e are shown in red (c) Superposition of (a) and (b), using the collar contours as reference, confirms the assignment of this FU (FU-6) as FU-e (d) End-on view of the extracted NpH subunit dimer. Arrows indicate the arc FUs g1 and g2 (“man-in-boat” aspect). According to the present model, the two FU-g copies within a subunit dimer are not directly connected. Consequently, the arc morphological stabilizes the assembly of two adjacent subunit dimers. (e) End-on view of an *Octopus* hemocyanin subunit dimer (van Holde *et al.*, 1991). The isolated subunit dimer shows two internal blobs, which obviously correspond to FUs g1 and g2.



**Fig. 45: The individual FUs of NpH with their respective molecular models**

The FUs are shown in approximately the same orientation. Note the quality of the rigid-body-fitting and the similarity of the 7 different FU types, which conforms to the primary sequence data. The threshold of surface representations encloses 0.75 EMV.

On the basis of this model for the subunit-pathway and using the available X-ray-structures as templates, comparative models of single FUs were generated which were then rigid-body-fitted into their respective cryoEM-densities. The models were then refined, in order to

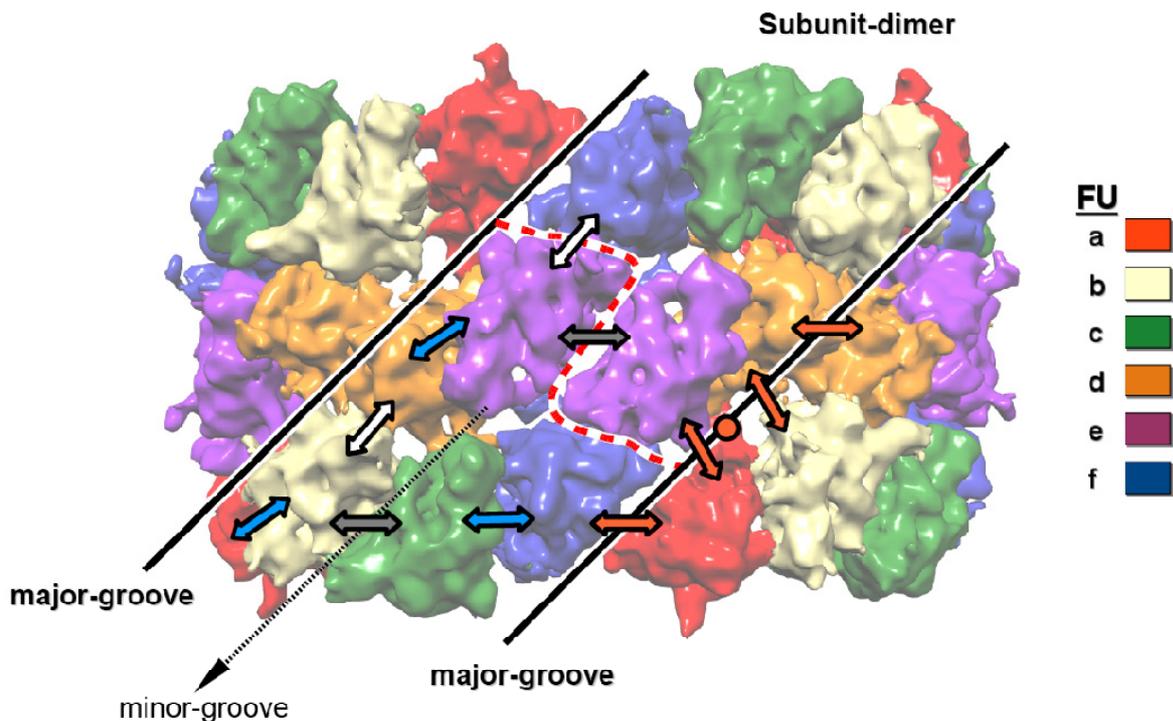
improve the fitting of several loops and linker peptides into appropriate cryoEM-densities (see Material and Methods in [2], also section B.4.13). The quality of the docking was measured (see Table 1 in [2]) and is directly visible in Fig. 45. Moreover, evaluation of the protein geometry of the seven FUs models (see Table 2 in [2]), certified their high stereochemical quality.

This hybrid approach enabled us to construct a reliable molecular model of the entire decamer (encompassing 29,020 amino acids) and thus present the first all-atom model of a molluscan hemocyanin. The molecular model of this giant invertebrate respiratory protein allowed us to understand further critical structural features of the molecule, such as:

- the various FU-FU interfaces, which can be now assigned to few aminoacids (candidates for allosteric signal transfer)
- localization of the putative attachment sites for N-linked glycans
- localization of the putative binding sites of divalent cations, which are required for assembly

### 3.3 Inter-FU interfaces

The molecular model of NpH reveals 12 different types of interfaces between FUs localized in the cylinder wall of NpH, two different non-covalent bridges between wall and arc and the  $g1 \leftrightarrow g2$  interface which links two adjacent subunit dimers. All these contacts can be described at the amino acid level and interpreted in terms of their possible contribution to allostery. For better orientation throughout this section, the topography of the wall interfaces is shown in Fig. 46 (also in Fig. 6 in [2]), the amino acid sequence of NpH is given in Fig. 5 in [2] and a sketch of the basic structure of molluscan hemocyanin FU is shown in Fig. 4 in [3]. The amino acids involved in these fifteen interfaces are summarized in Table 4 in [2].



**Fig. 46: Topography of the inter-FU interfaces in the NpH cylinder wall**

Wall fragment of NpH with the 6 different wall FUs highlighted in colour. Blue, white, grey and orange double arrows indicate the morphological unit, the horizontal tier, the minor groove and the major groove inter-FU interfaces respectively. The orange circle indicates the major groove interface  $a \leftrightarrow d$ , which corresponds to the previously introduced structural feature of the molluscan hemocyanin wall, termed the anchor [see 1]). In addition to the 12 types of interface indicated here, there are three interfaces associated with the collar: another morphological unit interface ( $g1 \leftrightarrow g2$ ) forming the five arcs, and two types of collar-wall interface ( $g1 \leftrightarrow d$  and  $g2 \leftrightarrow d$ ).

### 3.3.1 The morphological unit interfaces

The six wall FUs a-b-c-d-e-f are arranged pair-wise in an anti-parallel manner as three morphological units (ab, de, cf) thereby defining the morphological unit interfaces  $a \leftrightarrow b$ ,  $d \leftrightarrow e$  and  $c \leftrightarrow f$ . Moreover, two FU-g (g1 and g2) copies from adjacent dimers are also pair-wise associated to form the arc morphological unit. These four pairs correspond in constellation to the two protomers in the OdH-g crystals.

Interestingly, anti-parallel arrangement and formation of dimers, appears to be a well conserved general feature in molluscan hemocyanins, not only on the structural level of subunits, but also of single FUs.

The observed arrangement of three protomers in the crystal structure of the wall FU Rth-e (Perbandt *et al.*, 2003) is clearly incompatible with our cryoEM-structure. Thus, the building blocks of the hemocyanin decamer are indeed FU-dimers and not trimers (or trimers and dimers). Probably, the crystallization process of Rth-e may have forced the assembling protomers in an artificial helical symmetry.

Cuff *et al.* (1998), suggested that the pair-wise association observed in OdH-g probably connects wall FUs from two adjacent subunits. However, the present molecular model reveals that the three wall morphological unit interfaces connect FUs exclusively from the same subunit. The FU-connecting linker peptides are long enough to allow the formation of OdH-g like hetero-FU dimers.

The morphological unit interface  $g1 \leftrightarrow g2$  has already been described for the two protomers in the crystal structure of OdH-g and a similar situation is encountered not only in the  $g1 \leftrightarrow g2$  interface of *Nautilus*, but also in the three wall morphological unit interfaces (described in detail in [2]; see Fig. 7 in [2]).

Nevertheless, the wall morphological unit interfaces show in addition several histidine residues which appear as promising candidates for mediating cooperativity between the two adjacent FUs. These and further specific structural differences in the four morphological unit interfaces, suggest that they may be differently involved in allosteric force transfer. Indeed, earlier studies on oxygen binding by *Octopus* hemocyanin indicate heterogeneity of function within the 7FU-subunit (Miller, 1985).

### 3.3.2 The horizontal tier interfaces

The wall of the molecule can be described as a superposition of three well defined layers or horizontal tiers. Each tier contains 10 morphological units which are consecutively arranged around the ring. The peripheral tiers are formed by five copies of the tetrad (ab-cf). The central tier contains 10 copies of the morphological unit de.

The peripheral and central tiers are interconnected by a common form of interface that occurs as two types:  $b \leftrightarrow d$  and  $e \leftrightarrow f$  (described in detail in [2]; see Fig. 8 (a) and (b) in [2]). The interface  $b \leftrightarrow d$  connects two FUs from the same subunit whereas interface  $e \leftrightarrow f$  links the two polypeptide chains of the dimeric subunit.

As a common feature, both interfaces contact long  $\alpha$ -helices (helix  $\alpha 4$ ) from both adjacent FUs, and these  $\alpha$ -helices carry an active site histidine at their distal end. This arrangement might transfer forces between the two neighboring active sites during the cooperative oxygen-binding-process, by dislocating the long  $\alpha$ -helices.

Moreover, both interfaces exhibit a central cluster of charged amino acids which forms a central cavity. In interface  $b \leftrightarrow d$  this cavity is formed from positively charged residues and has a diameter of 4-5 Å. Therefore, it appears as promising candidate for binding an anion as allosteric regulator.

In  $d \leftrightarrow e$  interface, however, the cavity is less dominated by basic residues. This indicates that the two interfaces might play distinct roles in allosteric interaction.

### 3.3.3 The interfaces of the grooves

The cylinder wall shows two kinds of alternating right-handed oblique grooves, termed the major and the minor groove. The minor groove is located within the subunit dimer, whereas the major groove separates the wall segments of neighboring subunit dimers (see Fig. 43, 46).

The minor groove is crossed by 5 different bridges. Two of them correspond to inter-FU contacts ( $b \leftrightarrow c$ ,  $e \leftrightarrow e$ ) and the rest to linker peptides connecting FUs ( $b \rightarrow c$ ,  $c \rightarrow d$ ,  $e \rightarrow f$ ).

The  $b \leftrightarrow c$  interface is located in the peripheral tier and connects FUs from the same subunit. In the central tier, the minor groove is bridged from the interface  $e \leftrightarrow e$ , which links two subunits from the same subunit dimer (both interfaces are described in detail in [2]; see Fig. 8 (c) and (d) in [2]).

In both types of interfaces the two FUs are assembled in an anti-parallel manner at their core domains, showing a central window and massive contacts in the periphery.

Interface  $b \leftrightarrow c$  offers various possibilities for chemical bonding, in addition to the covalent  $b \rightarrow c$  linker. Moreover, the amino acid constellation of the massive contacts in the periphery should be considered in discussions of a possible allosteric signal transfer *via* the minor groove. Communication between the two adjacent active sites appears to be possible *via* the invariant PYWDW motif by mutually dislocating helix  $\alpha 10$ .

The situation in interface  $e \leftrightarrow e$  appears to be simpler. Although allosteric signaling could be possible also by dislocating helix  $\alpha 10$ , the strong salt bridges suggest, that the role of this

interface is to provide strong “glue” between the two adjacent subunits within the subunit dimer.

The major groove is crossed by five different bridges which correspond to inter-FU contacts ( $a \leftrightarrow f$ ,  $d \leftrightarrow d$ ,  $a \leftrightarrow e$ ,  $d \leftrightarrow b$  and  $a \leftrightarrow d$ ). All of them provide connection across the major groove between two adjacent subunit dimers (the major groove interfaces are described in detail in [2], see Fig. 9, in [2]).

Interface  $a \leftrightarrow d$  corresponds to the previously introduced structural feature of the molluscan hemocyanin wall, termed the anchor [see 1]. The molecular model of NpH, reveals that this prominent wall bridge, is formed between FUs a and d by their joining  $\beta 8$ - $\beta 9$  loops stabilized by disulfide bridges. In NpH-a, the  $\beta 8$ - $\beta 9$  loop is enlarged by three aspartic acids, which are absent in other FU types. The cryoEM structure of FU-a also shows a prominent protrusion in this region that is lacking in the density maps of the other FU types. Hence, the aspartic acid triplet yields a convincing explanation for this protrusion and provides a further strong argument in favor of the suggested subunit pathway. The anchor structure would remain rather obscure if another FU type was to occupy this position .

The aspartic acid triplet of FU-a offers strong bonding opportunities with polar and basic residues in the opposing  $\beta 8$ -  $\beta 9$  loop of FU-d. Thus, the anchor may play an important role as a connecting element between two adjacent dimers.

The interfaces  $a \leftrightarrow f$  in both peripheral tiers and  $d \leftrightarrow d$  in the central tier are structurally equivalent and connect FUs at their  $\alpha$ -helical domains. At both interfaces, the long loop  $\alpha 5$ - $\alpha 6$  offers various opportunities for chemical bonding and interaction between the two active sites might be possible by dislocating helix  $\alpha 8$ . Interface  $d \leftrightarrow d$  shows in addition four central histidine residues which appear as promising candidates for mediating cooperativity between the two adjacent FUs.

The interfaces  $a \leftrightarrow e$  and  $d \leftrightarrow b$  also appear as candidates for allosteric interaction. In this case, a loop following strand  $\beta 11$  in the  $\beta$ -sandwich domain of FU-a and FU-d, contacts the  $\alpha 3$ - $\alpha 4$  loop in the  $\alpha$ -helical domain of FU-b and FU-e respectively. The  $\alpha 3$ - $\alpha 4$  loop by both FU-types is stabilized by a disulfide bridge and carries a strictly conserved cysteine, which forms a thioether bridge to an active site histidine. Thus, allosteric interaction might be possible by oxygenation-dependent movements of the  $\beta$ -sandwich domains of FU-a and FU-b, which might transmit forces to the  $\alpha 3$ - $\alpha 4$  loop of FU-b and FU-e and the cysteine residues forming a thioether bridge with an active site histidine. This thioether bridge might be responsible for modification of the behavior or properties of the active site histidine (Gielens *et al.*, 1997; Gielens *et al.*, 2007).

Interestingly, this constellation within both interface types resembles the protomer-protomer contact in the Rth-e crystal structure (Perbandt *et al.*, 2003). However, there is also one major difference:

- the rotation of the two Rth-e protomers against each other around the interface longitudinal axis shows a  $\sim 40^\circ$  deviation in comparison to our molecular model and is incompatible to the NpH-cryoEM-structure.

Thus, although the basic association pattern of the protomers in the crystal structure of Rth can be found in the native molecule, the molecules appear to be constrained by the crystal contacts.

In summary, five major groove interfaces (including the anchor) and the  $g1 \leftrightarrow g2$  interface connect adjacent subunit dimers. In contrast, only two contacts (minor groove interface  $e \leftrightarrow e$ , horizontal tier interface  $e \leftrightarrow f$ ) link the subunits within the subunit dimer. The present results agree with the observation, that in *Nautilus* and *Octopus*, the subunit dimer is a rather transient dissociation/ reassembly intermediate (Bonaventura *et al.*, 1981; van Holde *et al.*, 1991). However, the present molecular model of NpH also reveals that the subunit dimer of NpH might be further stabilized by divalent cations such as  $Ca^{2+}$  and  $Mg^{2+}$  (see below).

### 3.3.4 Arc-wall interfaces

The collar complex is constituted by five discrete asymmetric arcs consisting of FU-g pairs (morphological units). As already mentioned, the two FU-g copies of each arc represent two different conformations, designated as  $g1$  and  $g2$ .

The one FU-g copy (termed  $g2$ ) is positioned at the level of the central tier whereas the other FU-g copy (termed  $g1$ ) is shifted toward one of the peripheral tiers (see Fig. 2(d) in [2]). Each FU-g of the arc is covalently linked to the wall *via* the  $f \rightarrow g$  linker peptide. Apparently, the very long  $f \rightarrow g$  linker peptide (18 amino acids) allows both positions.

The arc is primarily attached to the wall *via* two distinct  $g \leftrightarrow d$  interfaces ( $g1 \leftrightarrow d$ ;  $g2 \leftrightarrow d$ ) (both interfaces are described in detail in [2], see Fig. 10 (a) and (b) in [2]). Both alternative interfaces appear not only as fixation points (they provide several opportunities for electrostatic bonding), but also as possible sites of allosteric interaction. In both cases, allosteric signal transfer might occur by dislocating helix  $\alpha 10$ , thereby influencing an active site histidine.

We suggest that the intriguing observation, that two subunits of the same type show a different conformation of their C-terminal FU within the subunit dimer and consequently different arc-wall interfaces, derives from the need to form the arc morphological units. As already mentioned, FU-dimerization and formation of OdH-g like FU-dimers appears to be a well conserved feature (see above). The morphological units of the wall are formed between

FUs of the same subunit, whereas the arc morphological units are formed between FUs of adjacent subunit dimers.

According to our hypothesis, within an isolated subunit dimer both subunits show identical conformations and the standard conformation of FU-g might be the conformation of FU-g1. However, during oligomerization and as soon as two subunit dimers join each other, five FU-g copies, orientated towards the same cylinder opening, shift to g2 position in order to associate with the alternating five FU-g copies and form the OdH-g like FU dimers, which connect the adjacent subunit dimers. This dislocation produces the asymmetric position of the collar complex within the NpH-decamer (see Fig. 10 (c) in [2]).

As already mentioned, the constellation of the two FU-g copies strikingly resembles the arc described in gastropod hemocyanins (see above; see [1]). However, in gastropod hemocyanin, dislocation of FU-g to FU-g2 position may be additionally forced by the formation of the slab morphological units between two FU-h copies. However, at that time, it was still unclear, if the slab morphological unit connects two subunits within the same subunit dimer or two subunits between adjacent subunit dimers. Therefore, for gastropod hemocyanins, it remains uncertain if arc-formation takes place as soon as two subunit dimers join each other or as soon as two subunits associate to form the subunit dimer. Nevertheless, it is fascinating, that NpH obviously follows the same reassembly scenario with the 8-FU type hemocyanins, even in the absence of FU-h.

In order to prove our reassembly scenario, a 3D-reconstruction of an isolated subunit dimer might be necessary. However, as already mentioned, in *Nautilus* (and other cephalopods), the subunit dimer is a rather transient dissociation/reassembly intermediate, compared to other molluscan hemocyanins. Despite many attempts, we failed to produce a stable subunit-dimer. Nevertheless, an isolated and stable 8 FU-subunit dimer from the chiton *Acanthochiton fascicularis* has already been visualized by 3D electron microscopy (Harris *et al.*, 2004). Unfortunately, single FUs within the two subunits could not be clearly resolved because of the limited resolution (~3nm). Thus, it may be worthwhile to re-study the structure of isolated subunit-dimer of *Acanthochiton* or a gastropod hemocyanin by high-resolution cryoEM.

### 3.4 Putative divalent cation-binding sites within the decamer

*Nautilus* hemocyanin reassembles from subunits in the presence of Calcium and Magnesium cations (Bonaventura *et al.*, 1981). Furthermore, similar studies on other molluscan hemocyanins have shown that stabilizing the subunit dimer requires  $\text{Ca}^{2+}$  and/or  $\text{Mg}^{2+}$  (van Holde & Mille, 1995; Harris *et al.*, 2004). In the present study, several carboxyl residues that

are likely to stabilize the quaternary structure by bridging divalent cations such as  $\text{Ca}^{2+}$  and/or  $\text{Mg}^{2+}$  have been identified (see Fig. 10 (d) in [2]).

Three putative binding sites for divalent cations could be detected at the horizontal tier interfaces. Two of them are located within the same subunit ( $b \leftrightarrow d$ ,  $c \leftrightarrow d$ ), whereas the third one ( $e \leftrightarrow f$ ) links the two polypeptide chains of the dimeric subunit.

The minor groove interface  $b \leftrightarrow c$  contains also two putative binding sites for divalent cations, connecting two FUs from the same subunit. The interfaces between adjacent subunit dimers lack such opportunities, with the exception of the central major groove interface  $d \leftrightarrow d$ .

Thus, the vast majority of divalent cation bonding obviously occurs within the subunit dimer. These results fit the previous reassembly studies showing enrichment of the subunit dimer following addition of divalent cations (Harris *et al.*, 2004) and provide the structural basis for these observations.

### 3.5 Putative N-glycosylation pattern of NpH

Molluscan hemocyanins are glycoproteins. The quantity of hemocyanin-bound polysaccharides can be up to 9 % of the whole molecular mass (van Holde, 1992) and a variety of glycan side chains has been identified (Stoeva *et al.*, 1999; Kurokawa *et al.*, 2002). Carbohydrate chains are primarily bound by N-glycosylation to the polypeptide chain. The linkage occurs by covalent binding to the nitric amide of an asparagine residue. The glycosylation motives consist of the sequences NXT, NXS, or NXC (Forster and Davie, 1984), where X can be substituted by any amino acid, except proline. The N-glycosylation *via* NXC can only take place if the cysteine (C) is not involved in the formation of a disulfide bridge (S-S). Although N-linked oligosaccharides are generally quite different, a common core structure can be defined. This uniform element consists of a penta-saccharide, containing three mannose and two N-acetylglucosamine (GlcNAc) residues.

The amino acid sequence, determined within our working group (Bergmann *et al.*, 2006) shows 14 potential attachment sites (NXT/NXS) for N-linked glycans in NpH (see Table 1). NpH-c lacks such sites, following the scheme known from other molluscan hemocyanins, whereas the other FU-types show either one or two of them. In NpH-a, the potential binding site ( $^{403}\text{NPT}^{405}$ ) might be inaccessible due to the central proline. This leaves 13 potential sites per subunit. It should be noted that 2 of the 13 potential sites are localized in linker peptides, which is an exception among the molluscan hemocyanins so far sequenced (Bergmann *et al.*, 2006).

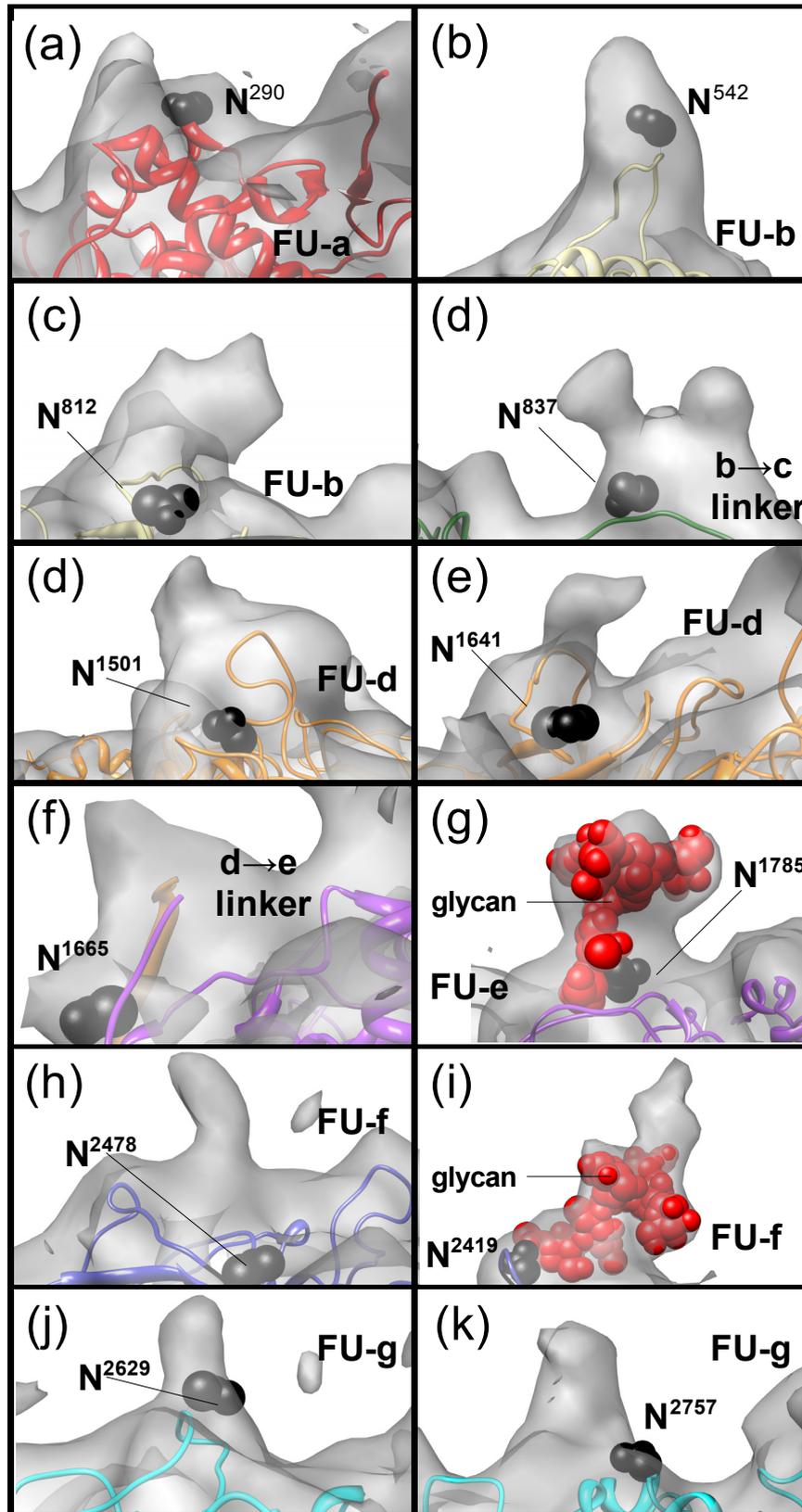
The present cryoEM structure and molecular model of NpH additionally reveal the potential distribution of N-glycan binding sites. Indeed, 12 of the 13 potential binding sites (N290, N542, N812, N837, N1501, N1641, N1665, N1785, N2419, N2478, N2629, N2757) are

associated with protruding masses of the density map, that are not occupied by the molecular model and are likely to represent carbohydrate trees (see Fig. 47). Only N392 in FU-a is not associated with such a protrusion.

	<b><u>potential N-glycosylation sites</u></b>
NpH-a	N290; N392
a→ b linker peptide	-
NpH-b	N542;N812
b→ c linker peptide	N837
NpH-c	-
c→ d linker peptide	-
NpH-d	N1501; N1641
d→ e linker peptide	N1665
NpH-e	N1785
e→ f linker peptide	-
NpH-f	N2419;N2478
f→ g linker peptide	-
NpH-g	N2629; N2757

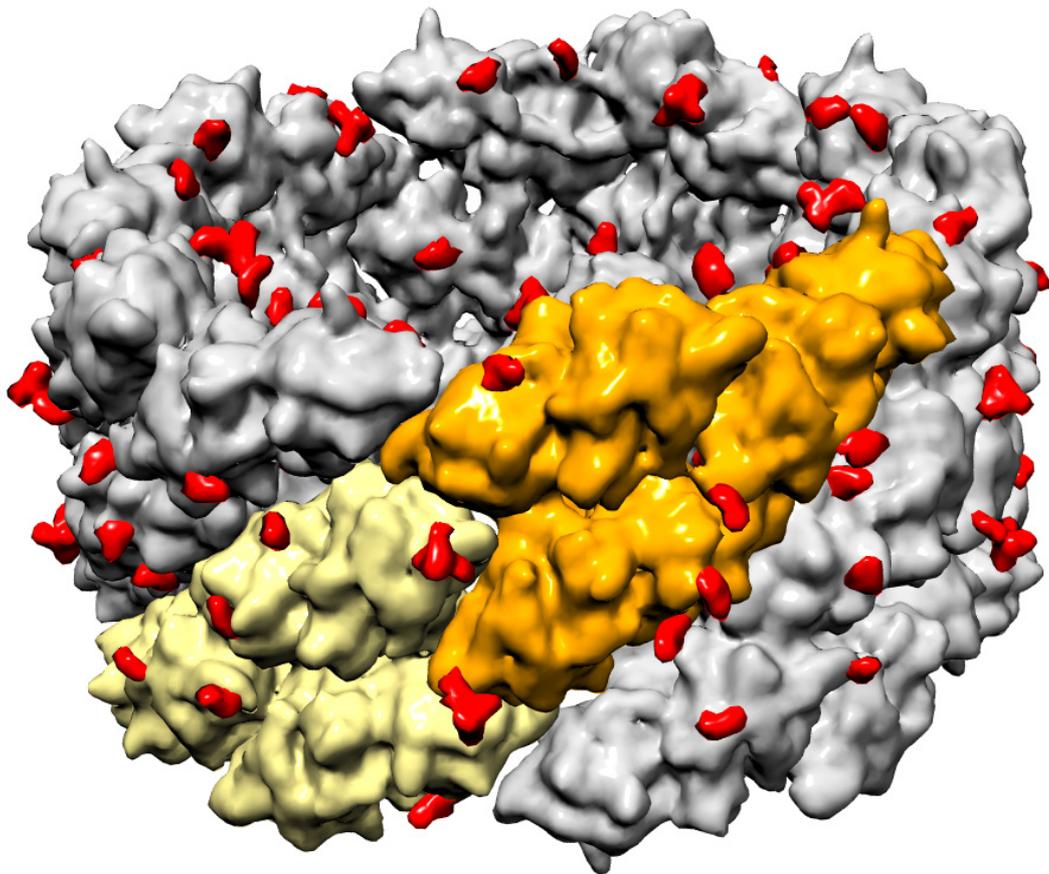
**Table 1: Putative attachment sites of N-linked glycan in NpH**

The most prominent protrusions of the density map are associated with N1785 (Fig. 47 (g)) and N2419 (Fig. 47 (i)) and are likely to represent a diantennate carbohydrate tree. Indeed, a convincing docking of such a tree extracted from the crystal structure of FU-g from the closely related *Octopus dofleini* (Cuff *et al.*, 1998) was achieved. Thus, this giant invertebrate respiratory protein appears to be almost evenly tagged by a total of 120 sugar trees (see Fig. 48). Some project from the side wall (N392, N542, N812, N1501, N1641, N1665) and some protrude from the rim into the cylinder openings (N290, N837, N2419, N2478). Two pairs of sugar trees (N2629; N2757) might point from each arc into the cylinder lumen.



**Fig. 47: Potential attachment sites for N-linked glycans in NpH**

(a-k) Asparagines of NXT/NXS motives and associated by protruding masses that are not occupied by the molecular model. In (g) and (i), short glycans (red) obtained from the crystal structure of OdH-g have been docked.



**Fig. 48: N-glycosylation pattern of NpH**

Tilt view of a 9-Å volume calculated from the molecular model of the NpH decamer, tagged with putative carbohydrate trees at 120 positions that correlate with empty masses in the cryoEM structure (red). In comparison to the “downfiltered” molecular model, the cryoEM structure (see Fig. 41 (a)) shows more details, suggesting that the estimated resolution might be rather pessimistic.

### 3.6 Model of the allosteric unit

Oxygen binding by *Nautilus pompilius* hemocyanin and the closely related *Octopus dofleini* hemocyanin has been carefully studied (Bonaventura *et al.*, 1981; Miller, 1985; Conelly *et al.*, 1989). For description of their oxygen binding behavior the classical MWC model is efficient (Miller, 1985). The binding curves by *Octopus* can be fitted accurately by the standard MWC model with an allosteric unit of seven sites (van Holde *et al.*, 2000) and from their structural similarities the same is expected for NpH. The number of seven, suggest that all seven types of FUs (a to g) are involved with one copy.

Thus, this might lead one to expect that a single subunit is identical with the allosteric unit. However, the multi-site subunits, when isolated, are not cooperative (van Holde *et al.*, 1995; Lambert *et al.*, 1994a). It may be that contacts with other subunits somehow establish constraints that permit the formation of distinct T and R states (van Holde *et al.*, 2000).

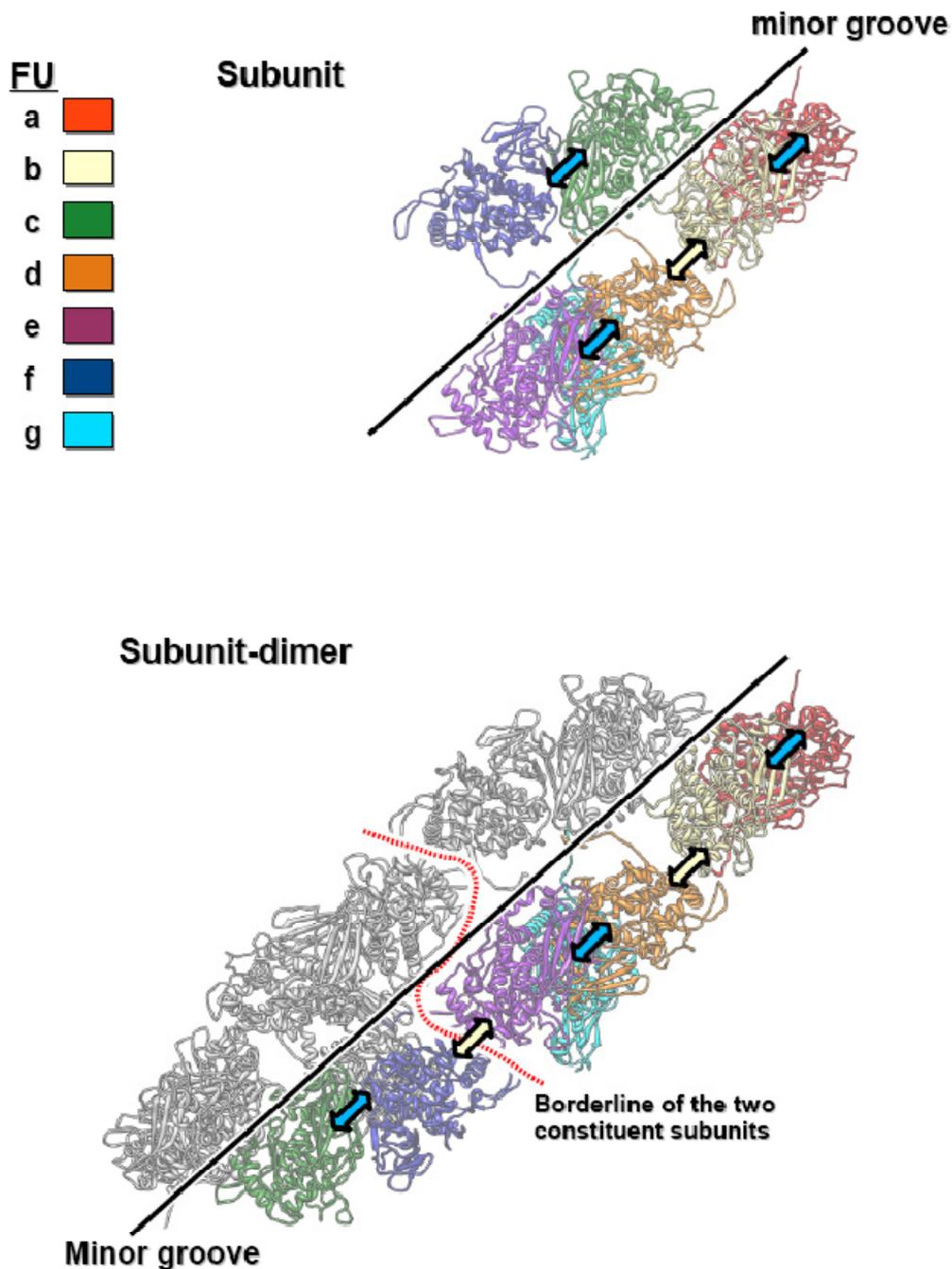
If so, is the allosteric unit identical with the subunit, when incorporated in the higher order assembly? Is the allosteric unit perhaps a different combination of seven FUs? Is the whole decameric structure required, or will more limited contacts (as within dimers of subunits) be sufficient to confer the allosteric integrity responsible for cooperative oxygen binding? What is the minimal structure for allosteric behavior?

Based on the first molecular model of a molluscan hemocyanin, presented within this study, most of the inter-FU interfaces appear as promising candidates for allosteric signal transfer. However, in terms of a direct site-to-site communication the morphological unit interfaces of the wall (a↔b, d↔e, c↔f) and the horizontal tier interfaces (b↔d, e↔f) are firm candidates for centers of allostery. This suggests that the allosteric unit contains the 6 FUs that are located between the major and the minor groove. This straight row of six densely packed wall FUs resembles a wagon train, which is a standard model of linear impulse transmission. To reach the determined number of 7 active sites for the allosteric unit, we might include one FU-g copy. As already mentioned, the arc-wall interfaces appear not only as fixation points, but also as possible sites of allosteric interaction.

According to our model, the 7-site allosteric unit is formed when two subunits associate to form the subunit dimer. This would explain why an isolated subunit is not cooperative. The 7 FUs stem from two different subunits within the same subunit dimer (see Fig. 49).

Indeed, Klarman *et al.* (1975) reported that isolated dimers of the subunit of the hemocyanin of the gastropod *Levantina hierosolima* exhibited cooperative oxygen binding. However, this is not a general phenomenon among molluscan hemocyanins. In *Sepia* hemocyanin (8FU-type cephalopod hemocyanin), cooperativity is correlated with the presence of the decamer. Contacts between FUs in the dimer are not sufficient to impose cooperative oxygen binding (van Holde *et al.*, 2000). However, *Sepia* hemocyanin has a multi-site-subunit with special organization: compared to the other cephalopods, it contains an additional FU; in contrast to gastropods, this additional FU is localized between FU-d and FU-e (Gielens *et al.*, 1983; Loncke *et al.*, 1990).

Therefore, analysis of oxygen binding by dimers of subunits by 7-FU type hemocyanins of cephalopods (such as NpH or *Octopus*) (basic assembly of molluscan hemocyanins) should provide a more critical test, in context of the allosteric behavior of isolated subunit dimers. Unfortunately, as already mentioned, in *Nautilus* and *Octopus* such dimers are found only as transient species during association or dissociation.



**Fig. 49: Proposed model of the allosteric unit in NpH**

A single subunit contains four of the five most promising candidates of inter-FU interfaces (the three morphological unit interfaces (blue double arrows) ( $a \leftrightarrow b$ ,  $d \leftrightarrow e$ ,  $c \leftrightarrow f$ ) and 1 horizontal tier interfaces (yellow double arrow) ( $b \leftrightarrow d$ )) for allosteric signal transfer. However, these contacts are not sufficient to impose cooperative oxygen binding; isolated subunits are not cooperative. As soon as two subunits associate to form the subunit dimer, the allosteric unit is formed, which encompasses six densely packed wall FUs. The allosteric unit functionally links three consecutive morphological unit interfaces (blue double arrows) and 2 inserted horizontal tier interfaces (yellow double arrows). In order to reach the experimentally determined number of 7 active sites for the allosteric unit, we included one copy of the arc FU-g (the one which is non-covalently linked to cylinder wall). The attraction of the present model is the wagon-train-like linear transmission of allosteric effects. However, there is evidence that allosteric interaction might additionally occur *via* the minor and major grooves and therefore alternative allosteric unit models might be forthcoming from future research.

---

In addition, according to comparative SAXS analysis data, upon oxygenation, the 8FU-type keyhole limpet hemocyanin isoform 1 (from the gastropod *Megathura crenulata*) shows a massive conformational change which is explained as a twist along the molecule's main symmetry axis. Moreover, the helical grooves (major and minor grooves) become less pronounced (see Fig 2(b) in Hartmann *et al.*, 2004).

Thus, in addition to the proposed strong allosteric interaction within the 7-site allosteric unit, a weaker (and therefore hitherto overlooked) allosteric effect *via* the grooves could be possible. In this case, allosteric interaction during oxygen binding would encompass two or more of such 7FU allosteric units. As already mentioned, major grooves separate adjacent subunit dimers and allosteric interaction is additionally also possible *via* the g1↔g2 arc interfaces, which connect adjacent subunit dimers.

Thus, the entire decamer assembly might additionally enable the multi-site compartments defined here as allosteric units to interact *via* the major groove and arc interfaces. Critical re-evaluation of function data in terms of the Nested MWC model might be justified in order to seek support for such a higher-ordered interaction.



#### **4. Keyhole limpet Hemocyanin: 9 Å CryoEM Structure and Molecular Model of the KLH1 Didecamer reveal the Interfaces and Intricate Topology of the 160 functional units.**

The 9 Å cryo-EM structure and first molecular model of a cephalopod hemocyanin decamer (NpH; see section C. 3, see also [2]) answered many hitherto unsolved questions concerning the quaternary structure of molluscan hemocyanin. Here we extend our study on molluscan hemocyanins and present the first sub-nanometer cryo-EM structure and complete molecular model of the more complex gastropod hemocyanin didecamer (KLH1), which was published in 2008 [2]. KLH is obtained from the Californian keyhole limpet *Megathura crenulata* and is intensively utilized in immunological research and clinics, as an immunoactivator and tumour vaccine carrier; it occurs in two molecular isoforms termed KLH1 and KLH2 (see [3]; see also A.2).

The present data on KLH1, together with the corresponding data on NpH, allow for the first time a detailed structural comparison of the seven-FU cephalopod hemocyanin and the eight-FU gastropod hemocyanin at pseudo-atomic resolution, which elucidates several long-standing questions about their unique structural features.

##### **4.1 Cryo-EM structure and molecular model of KLH1**

With D5 point symmetry applied in the reconstruction process, a 9.1 Å cryoEM-structure of KLH1 was ultimately achieved from 4762 high-quality particles, obtained from 98 negatives. The present reconstruction represents the highest resolution reconstruction of single particles achieved for a gastropod hemocyanin.

The present high resolution cryoEM structure of KLH1 is in complete agreement with the previously published structure of this protein (Mouche *et al.*, 2003) and other gastropod hemocyanin didecamers (Meissner *et al.*, 2000; also see [1]) at lower resolution. The principal architecture and parameters of the gastropod didecamer have already been discussed in the 11 Å cryoEM structure of HtH1 (see [1]) and the previous low resolution cryoEM structure of KLH1 (Mouche *et al.*, 2003) (see Fig.2 in [3] and Fig. 50).

KLH1 and HtH1 are orthologous isoforms; indeed at the resolution level of 11 Å, their cryoEM structures are almost indistinguishable.

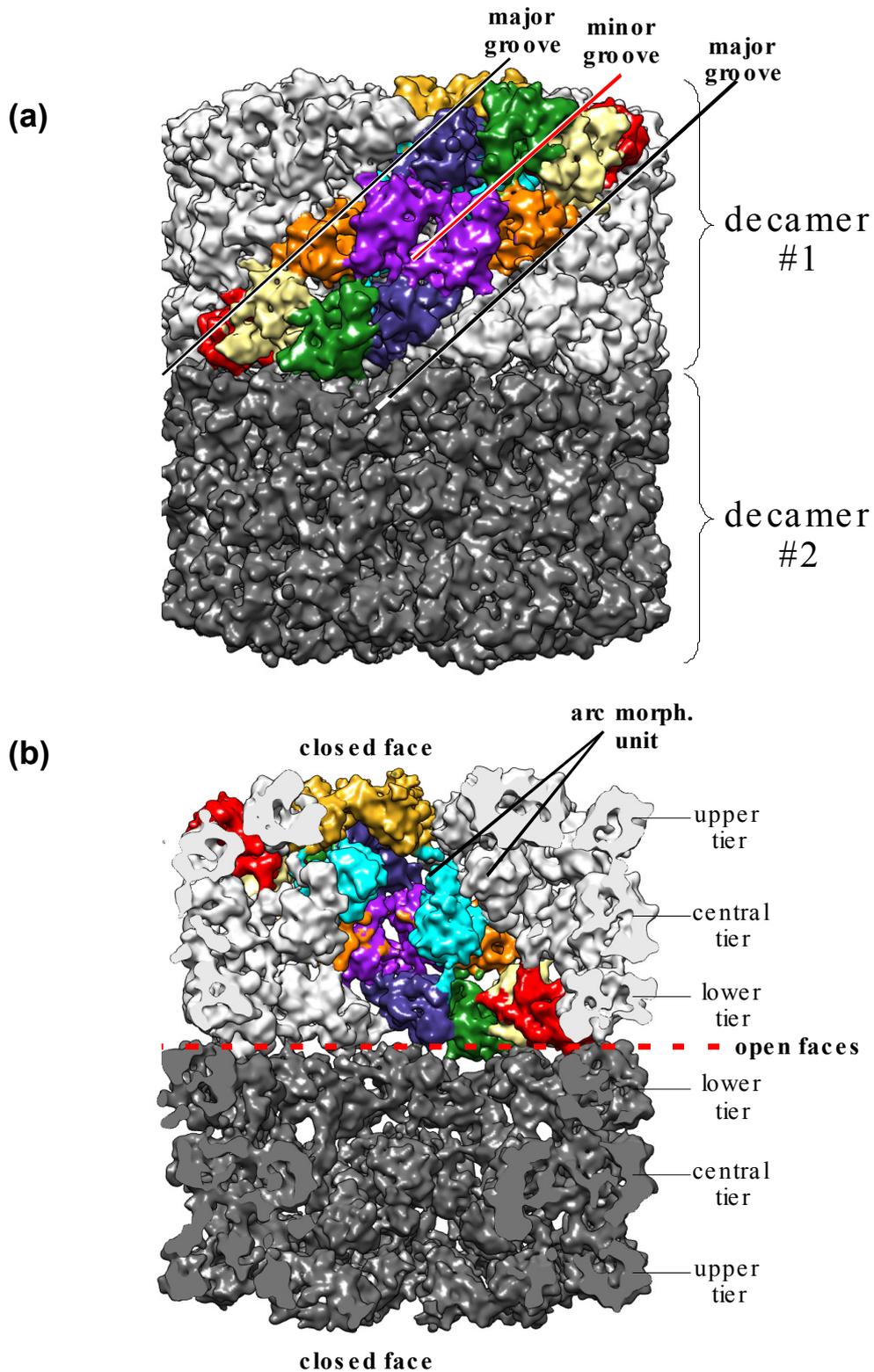
At 9 Å, the 8 types of FUs (wall FUs KLH1-a to -f, arc FUs KLH1-g, slabs FUs KLH1-h) are clearly discernible. The present cryoEM structure reveals them in greatly improved detail. The extracted cryoEM densities of the eight different FUs, viewed in the same orientation, are closely comparable, but also show some specific differences, especially for FU-h and

FU-a (see Fig. 3 (a) in [3]). The topology of the 8 different FUs within the didecamer was determined by the strategies developed for NpH (see [2]).

Overall, the definition of structural features is clearly superior to those in previous studies of gastropod hemocyanins (e.g. Meissner *et al.*, 2000; Mouche *et al.*, 2003); in particular, for the first time, one can define individual domains and secondary structure within the FUs of a gastropod hemocyanin. As already demonstrated for NpH, this allowed high quality rigid body fitting of the modelled tertiary structures of the seven 340 kDa FUs (walls FUs a-f, arc FU-g) and the recently obtained crystal structure of the 400 kDa collar-FU-h, into their respective cryoEM densities (see Fig. 3 in [3]), resulting in a molecular model of the entire didecamer.

Homology modelling of the different FU-types has been performed on the basis of their known sequence and the published high resolution X-ray structures (for more details see section B.4.14 and material and methods in [3]). Linker peptides of the subunit fragment a-b-c-d-e-f-g were modelled using the corresponding NpH inter-FU linker as templates. This was considered to be justified, since a comparative analysis of the molecular models and cryoEM structures of KLH1 and NpH revealed that equivalent FU types are connected *via* similar bridges in the density map.

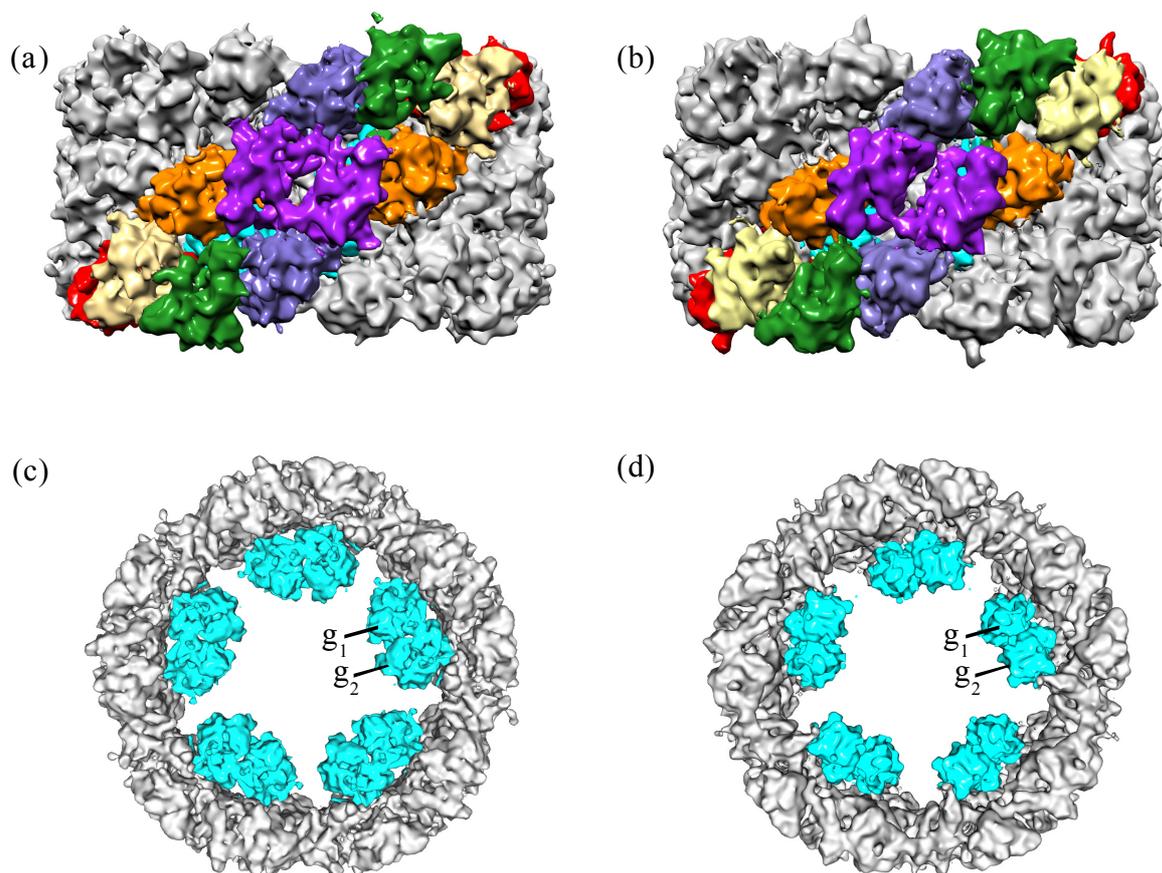
As in NpH, the wall and arc FU types occur as morphological units (a-b, c-f, d-e, g-g); these FU-pairs strikingly correspond in constellation to the two protomers in the OdH-g crystals. The 7 common FU types of types of NpH and KLH1 (FUs a to g) are topologically identical in both hemocyanins; their shape, orientation and connections are closely comparable.



**Fig. 50: 9 Å cryoEM structure of KLH1**

a) Side view b) cut open view of KLH1. The native molecule is a didecamer, with two decamers assembled at their open faces. Each decamer can be described as a superposition of three tiers, with an off-set internal collar at the level of the upper tier (“closed face”). The decamer consists of five subunit dimers assembled as a hollow cylinder. Note the shape of the subunit dimer (highlighted in colour; FU-a: red; FU-b: yellow; FU-c: green; FU-d orange; FU-e: purple; FU-f: blue; FU-g: cyan; FU-h: gold). Also note that the wall segment of the subunit-dimer corresponds to the oblique segment between two alternating major grooves. The arc morphological unit is indicated; note that this FU-dimer connects two adjacent subunit-dimers *via* the major groove.

All in all, the cryoEM structure of NpH is almost identical to the corresponding KLH1-decamer, when the slab pentamer is removed (see Fig. 51). The similarity is indeed greater than expected and the structure conservation in the two hemocyanins that diverged more than 520 million years ago is remarkable. Minor differences will be discussed in the next sections.



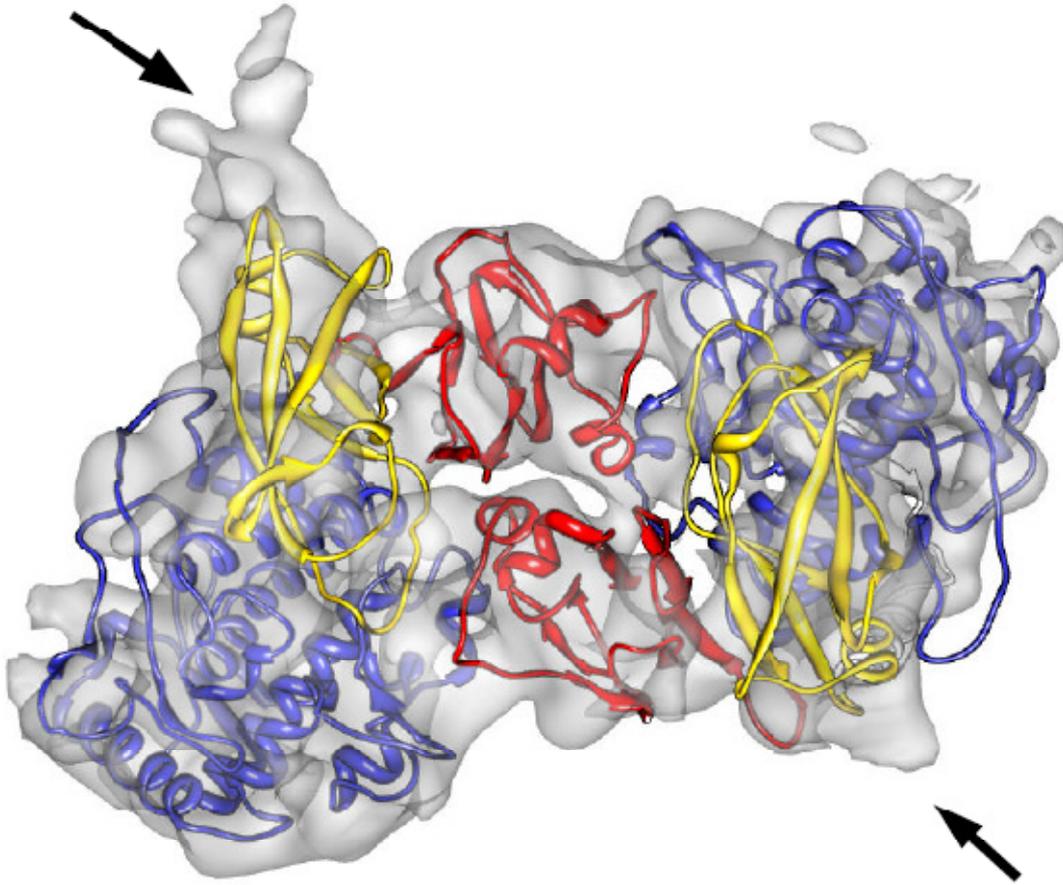
**Fig. 51: Comparative analysis of the cryoEM structure of NpH with the corresponding extracted cryoEM structure of KLH1**

(a) Side view and (c) of top view of one extracted KLH1 decamer, with the slab pentamer removed to reveal similarity with the NpH whole molecule, shown in (b) and (d). In (a), (b), equivalent FUs, of a subunit dimer, are highlighted in colour. In (c), (d) the five arc morphological units, which contain the FU-g conformers 1 and 2, are indicated.

Thus, although NpH and KLH1 differ significantly in their quaternary structure (decamer vs didecamer; seven vs eight FUs), the anti-parallel subunit arrangement and the pathway and 3D-structure of the subunit segment a-b-c-d-e-f-g, determined within our previous study for NpH, is also true for gastropod hemocyanins, with the additional information that its peripheral collar (that is lacking in cephalopod hemocyanins) is composed of 10 copies of FU-h.

The ten copies of FU-h forming the slab pentamer are also arranged as pairs (slabs). We have discovered that these pairs strikingly resemble, in constellation, the recently obtained 4

Å X-ray structure of the FU-dimer described for KLH1-h (Barends *et al.*, 2008) (see Fig. 52; see also section A.2.1.; for more details see section 4.2.6).



**Fig. 52: CryoEM and crystal structure of the KLH1-h morphological unit**

The N'-terminal  $\alpha$ -helical domain is shown in green, the  $\beta$ -sandwich domain in blue and the additional C-terminal cupredoxin-like domain is displayed in red; arrows: empty spaces in the cryoEM-density map that are not occupied by the molecular model and are likely to represent carbohydrate side chains (see also section 4.2.9 and Fig. 10 in [3]).

As already mentioned, the KLH1-decamer is in part asymmetric, due to the off-set slab-pentamer at one edge of the decamer. Therefore, to facilitate description, we introduced a new nomenclature for the two subunits forming the subunit-dimer: conformer-1 (a1-b1-c1-d1-e1-f1-g1-h1, positioned at the closed face of the decamer) and conformer-2 (a2-b2-c2-d2-e2-f2-g2-h2, open face) (see Fig. 2, Fig. 3 in [3]). Both conformers have the same amino acid sequence, but some of their interfaces with neighbouring FUs are different.

The asymmetry of the subunit dimer, derives from the fact, that the same subunit folds its last two FUs (FU-g, FU-h) into two different conformations (termed here FU-g1, FU-h1 and FU-g2, FU-h2), whereas the two FU-g (g1 and g2) copies are not directly connected. Consequently, as in NpH, the arcs do not link the two polypeptide chains of the dimeric subunit, but bridge neighbouring subunit dimers.

Surprisingly, the molecular model of the present cryoEM structure did not immediately allow us to determine how the subunit pathway continues in both conformers from FU-g to FU-h. It was still unclear, which FU-h copy of each slab belongs to conformer 1 and which to conformer 2.

However, a careful topological analysis based on:

- the positions of the N'- and C'-termini of FUs g and h in the preliminary model
- the potential length of the g-h linker peptides
- potential correspondence of g-h linker-peptides with inter-FU bridges in the cryoEM-structure

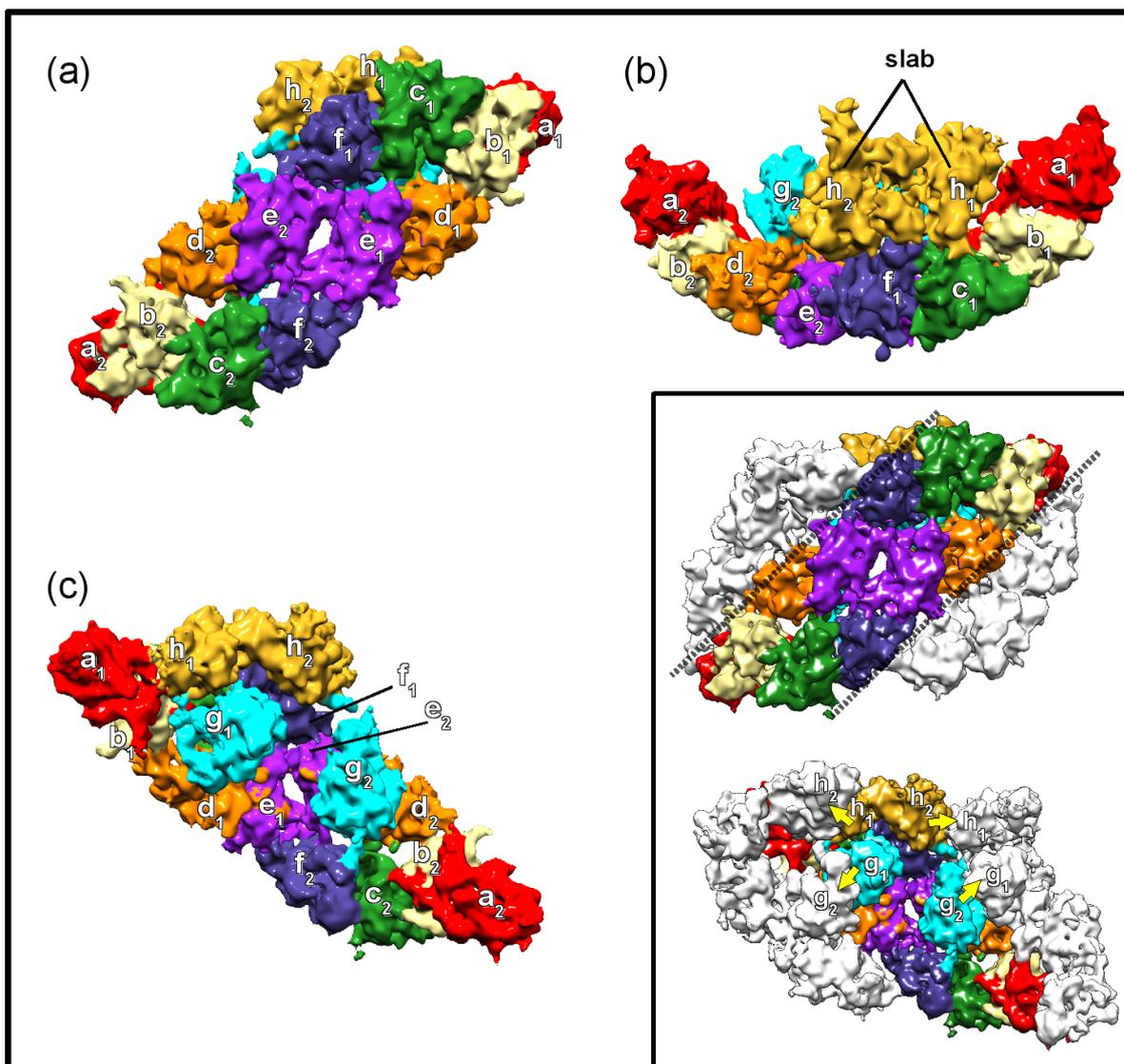
(For more details see section *Topology of KLH1-h in the slab pentamer and the subunit pathway* in [3]).

revealed that the FU-hvcopy with the tip of its  $\beta$ -sandwich domain contacting the wall corresponds to FU-h conformer 1 and is covalently linked to g1; the alternative FU-h copy, where the tip of its  $\beta$ -sandwich domain is directed into the cylinder lumen, corresponds to FU-h conformer 2 and is covalently linked to g2 (see also Fig. 3 and Fig. 9 (b) in [3]). FU-h1 and FU-h2 are anti-parallel in order to form the slab morphological unit, (see also Fig. 52) which links two subunits from the same subunit dimer.

The extracted subunit dimer shows the typical “man-in-the-boat” structure when viewed from the top and a “melon-slice” structure when viewed from the side, as seen in electron micrographs of purified subunit dimers of 8 FU-type hemocyanins (Siezen and van Bruggen, 1974; Harris *et al.*, 2004). Furthermore, it fully confirms the subunit dimer model that we previously proposed from NpH ([2]) (see Fig. 53; see Fig. 3 b, c, d in [3]).

Due to the additional structural reinforcement provided by the FU-h slab, the KLH1 subunit dimer appears to be more compact than the NpH subunit dimer, suggesting greater stability. This is in full agreement with previous studies on molluscan hemocyanins, showing successful purification of subunit dimers, only by hemocyanins containing the additional FU-h (gastropods, chitons; (Siezen and Van Bruggen, 1974; Harris *et al.*, 2004)). The subunit dimer of 7-FU-type hemocyanins (like *Nautilus*, *Octopus*) is rather a more transient dissociation/ reassembly intermediate (Bonaventura *et al.*, 1981; van Holde & Miller, 1985).

The extracted cryoEM-structure and final molecular model of the subunit dimer is shown in Fig. 3 in [3].



**Fig. 53: Extracted subunit dimer of KLH1**

shown in side view (a), top view (b) and side view rotated by 180° as observed from the cylinder lumen (c). The different FU types are colored and labeled. The subunit dimer consists of subunit conformer 1 (a<sub>1</sub>-b<sub>1</sub>-c<sub>1</sub>-d<sub>1</sub>-e<sub>1</sub>-f<sub>1</sub>-g<sub>1</sub>-h<sub>1</sub>) and subunit conformer 2 (a<sub>2</sub>-b<sub>2</sub>-c<sub>2</sub>-d<sub>2</sub>-e<sub>2</sub>-f<sub>2</sub>-g<sub>2</sub>-h<sub>2</sub>). Each subunit dimer carries two separated arc FUs and one complete slab. The inset shows an extracted segment of the KLH1 decamer, as viewed from the side (upper image) and rotated by 180° (lower image). The wall segments of subunit dimers are interconnected *via* fine bridges along the major grooves (indicated by dashed lines). The subunit dimers are also interconnected *via* the arc-morphological unit interface g<sub>1</sub>-g<sub>2</sub> and the h<sub>2</sub>-h<sub>1</sub> interface between two adjacent slabs (indicated by yellow arrows).

The molecular model of the entire didecimer was assembled in CHIMERA. It reveals 12 types of inter-FU interfaces between wall-FUs, 2 types of arc-wall interface and the g<sub>1</sub>↔g<sub>2</sub> interface. These interfaces are also present in NpH and a comparative analysis allows now a detailed description of their similarities and differences at amino acid level.

Moreover, the molecular model of KLH1 reveals the molecular architecture of the additional features in KLH1: the contacts between the two decamers; the intriguing (slab) pentamer and its contacts to the cylinder-wall.

Furthermore, it allows detection of putative binding sites for divalent cations, required for assembly and also reveals the putative N-glycosylation pattern of the giant respiratory glycoprotein.

## 4.2 Molecular structure of the interfaces between functional units of the cylinder wall

For better orientation, throughout this section, the amino acid sequence of KLH-1 (Lieb & Markl, 2004) and a sketch of the tertiary structure of a single ~350k Da FU (OdH-g, Cuff *et al.*, 1998) are given in Fig. 4 in [3]. The participating amino acids in the interfaces between functional units are summarized in Table 3 in [3].

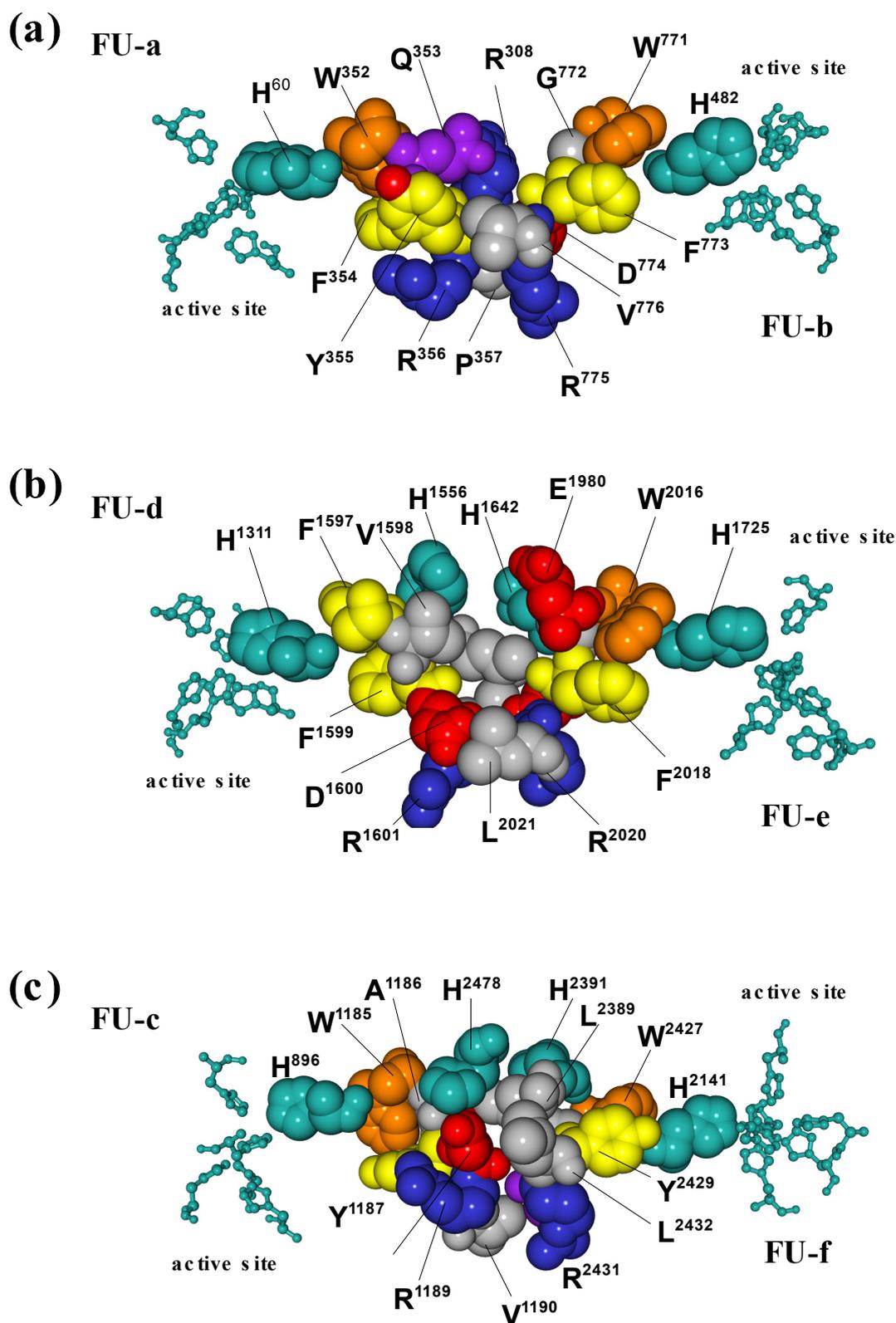
Moreover, we recommend consulting Fig. 6 in [2] and Fig. 46. Compared to NpH, the present molecular model revealed similar molecular interfaces at equivalent positions of the cylinder wall and their comprehensive description and comparison at amino acid level is now for the first time possible.

### 4.2.1 Morphological unit interfaces (a↔b, d↔e, c↔f)

As in NpH, the six different wall FUs (a, b, c, d, e, f) are arranged pair-wise as three morphological units and follow the architecture of the Octopus FU-g dimer crystal lattice, forming the interfaces a↔b, d↔e, c↔f. In each case, the two FU copies are associated at their  $\beta$ -sandwich domains in an anti-parallel manner. It should be noted, that the three morphological unit interfaces connect FUs from the same subunit. The basic features of this type of interface have already been described for the g↔g interface of Octopus hemocyanin in Cuff *et al.* (1998) and for the morphological unit interfaces of NpH in [2] (see also section C.3.3). The basic amino acid constellation is also true for the morphological unit interfaces of KLH1.

All in all, the three types of morphological unit interfaces of the hemocyanin cylinder wall appear as promising candidates for conformational signal transmission. In the center of these interfaces, two (W/F)X(F/Y)(D/E)R(L/V/P) (WX(F/Y)DRL in NpH) motives oppose each other, suggesting a direct allosteric signal transfer between the two neighbouring active sites, by mutually influencing an active site histidine.

In NpH, several histidine residues are also closely associated to these motives: two histidines at the a↔b, three at the d↔e and even four at the c↔f interface. A recent study from our group on the highly cooperative *Limulus* 8x6 arthropod hemocyanin (Martin *et al.*, 2005), revealed that the interfaces between hexamers are intriguingly histidine-rich.



**Fig. 54: Morphological unit interfaces localized in the cylinder wall of KLH1**

(a) morphological unit interface a-b. For comparison to NpH see Fig. 7b in [2]. (b) morphological unit interface d-e. For comparison to NpH see Fig. 7c in [2]. (c) morphological unit interface c-f. For comparison to NpH see Fig. 7d in [2].

Colour code of residues: red, acidic; blue, basic; purple, polar; light sea green, histidine; orange, tryptophan; grey, hydrophobic; yellow, phenylalanine; yellow/red, tyrosine; grey/yellow, cysteine and methionine. The active site histidine residues are shown in ball and stick representation, except for the one in focus which is shown in sphere mode.

The identified histidine-clusters were associated to transfer allosteric signal between the hexamers. Likewise, the histidine-rich clusters identified in the morphological unit interfaces of NpH, might reasonably be expected to mediate cooperativity between adjacent FUs. However, in contrast to NpH, the wall morphological unit interfaces of KLH1 are less histidine-rich.

In detail, the interface  $a \leftrightarrow b$  (Fig. 54 a) is mostly established by the opposing 352WQFYRP357 (FU-a) – 771WGFDRV776 (FU-b) motives. There is the possibility, that this motif influences an active site histidine (H60, H482) *via* the closely associated tryptophan residue (W352, W771) or vice versa. Such a signal could be transferred between the two anti-parallel FUs within the morphological unit *via* their contacting  $\beta$ -sandwich domains. However, histidine residues are not involved in this interface and in this case there are not further indications for such a mechanism.

The morphological unit interfaces  $d \leftrightarrow e$  and  $c \leftrightarrow f$  (Fig. 54 b, c) show a similar architecture to  $a \leftrightarrow b$ , but exhibit two central histidine residues respectively (H1556, H1642 in  $d \leftrightarrow e$ ; H2478, H2391 in  $c \leftrightarrow f$ ), which might contribute to allosteric interaction between adjacent FUs.

The structural differences between the three morphological unit interfaces of the cylinder wall suggest that they might play distinct roles in allosteric interaction.

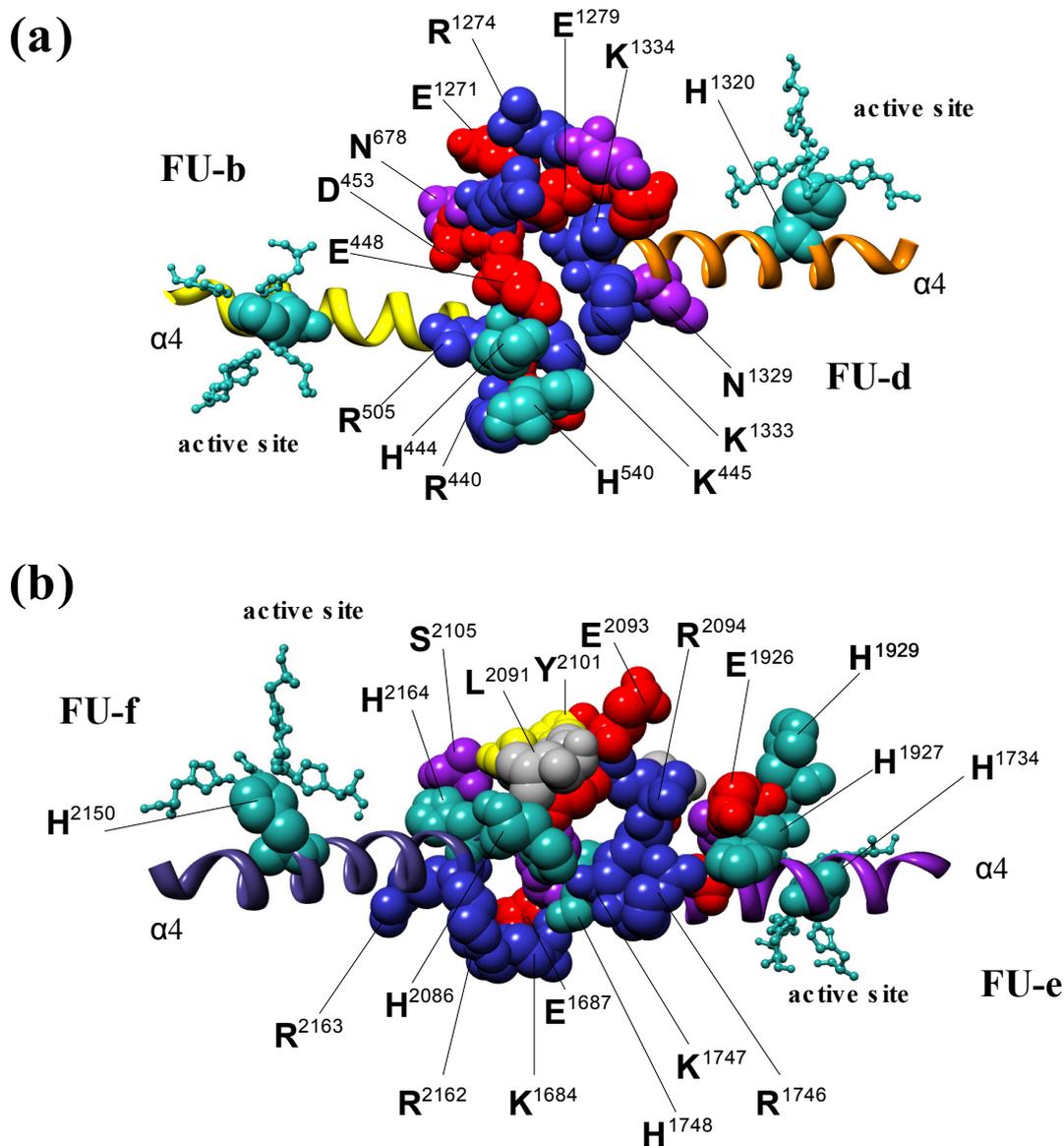
Although the wall morphological unit interfaces of KLH1 are less histidine-rich in comparison to those in NpH, they provide similar bonding and chemomechanical possibilities and still appear as firm candidates for centers of allostery, in terms of a direct site-to-site communication. Of course they are not identical, but this would be rather a surprise, since both molecules evolved at least 520 million years ago (Lieb *et al.*, 2001). Moreover, oxygen binding studies reported that cephalopod and gastropod hemocyanins exhibit different values of the Hill coefficient (e.g. van Holde *et al.*, 2000).

#### 4.2.2 The horizontal tier interfaces

The wall of the decamer has been described as superposition of three layers or tiers. Each tier consists of 10 morphological units (=20 FUs). The peripheral tiers contain 5 copies of the morphological units  $a$ - $b$  and  $c$ - $f$  each, whereas the central tier contains 10 copies of the morphological unit  $d$ - $e$ . The central tier is interconnected to the peripheral tiers by two types of horizontal tier interface:  $b \leftrightarrow d$  and  $e \leftrightarrow f$ . As in NpH, interface  $b \leftrightarrow d$  connects two FUs from the same subunit, whereas interface  $e \leftrightarrow f$  connects two FUs from different subunits within the same subunit dimer.

Interface  $b \leftrightarrow d$  (Fig. 55 a) exhibits a central cluster of charged amino acids, which shows a central cavity and several opportunities for strong electrostatic bonding. This cluster is in

direct contact with the near end of helices  $\alpha_4$ , which carry, at their far ends, the third active site histidine (H491, H1320). At the other end, associated to the cluster of charged amino acids, the  $\alpha_4$  helix of FU-b carries another histidine (H444) and an arginine residue (R505). This amino acid constellation, suggests that the active sites of FUs b and d might interact by dislocating the long helices  $\alpha_4$  that carry the third active site histidine.



**Fig. 55: Interfaces between the horizontal tiers in the cylinder wall of KLH1**

(a) Horizontal tier interface b-d; note the central cluster of charged residues, which provides several opportunities for electrostatic bonding; also note the possibility of interaction between the adjacent active sites by dislocating helix  $\alpha_4$  (for comparison to NpH, see Fig.8 a in [2]) (b) Horizontal tier interface e-f; note similarities and differences compared to (a) (for more details see main text)(for comparison to NpH, see Fig.8 b in [2]) Colour code as for Fig. 54.

Interface e-f (Fig. 55 b) shows a similar amino acid constellation; it exhibits a central cluster of charged residues associated to the near end of helices  $\alpha_4$ , which carries the third active site histidine (H1734, H2150), at their far end. Moreover, the near end of helix  $\alpha_4$  also carries a histidine residue (H2164, H1748). These histidine residues appear as candidates for mediating cooperativity between the two adjacent FUs.

However, in this case, the central cluster is dominated by basic residues, which form a central basic cavity that might bind an anion as allosteric regulator. In addition, three additional histidine residues are closely associated to the central cluster of charged aminoacids (H1929, H1927, H2086).

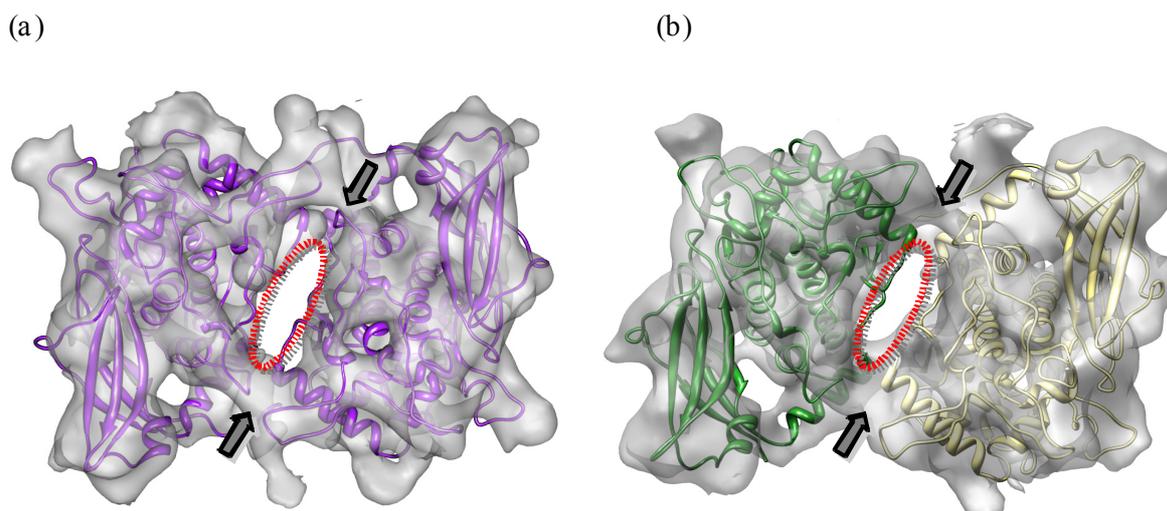
All in all, the horizontal tier interfaces appear as firm candidates for centers of allostery. Their differences indicate that they might contribute differently to the allostery of the whole molecule.

In *Nautilus* hemocyanin, the overall architecture of the horizontal tier interfaces is comparable to those in KLH1 and suggests the same mechanism for mediating cooperativity between the adjacent FUs. However, in NpH, a comparable central basic cavity is located in interface b $\leftrightarrow$ d and not in e $\leftrightarrow$ f. Thus, both molecules exhibit one potential binding site for an anion at their horizontal tier interfaces that might act as allosteric regulator. However, in NpH this site is located within a single subunit, whereas in KLH1 it is located at an interface between the two subunits of the subunit dimer.

### 4.2.3 Minor groove interfaces

The characteristic right handed oblique helical structure on the surface of molluscan hemocyanins, as observed in the first electron microscopical studies on hemocyanins (Mellema and Klug, 1972), is formed by the alternating major and minor grooves (indicated by diagonal lines in Fig. 40) and the elongated polypeptides. The minor groove is located within the subunit dimer whereas the major groove separates adjacent subunit dimers.

As in *Nautilus* hemocyanin, the minor groove is crossed by two interfaces: b $\leftrightarrow$ c, in the peripheral tier, and e $\leftrightarrow$ e, in the central tier, and by two well-defined bridges that correspond to the c $\rightarrow$ d and e $\rightarrow$ f linker, respectively. The b $\leftrightarrow$ c interface connects two adjacent morphological units (ab $\leftrightarrow$ cf), which stem from the same subunit. The e $\leftrightarrow$ e interface connects the two constituents of the subunit dimer. In both types of interface the two FUs are arranged at their core domains in an anti-parallel manner, showing a central window (red dashed circle in Fig. 56) and massive contacts in the periphery (arrows).



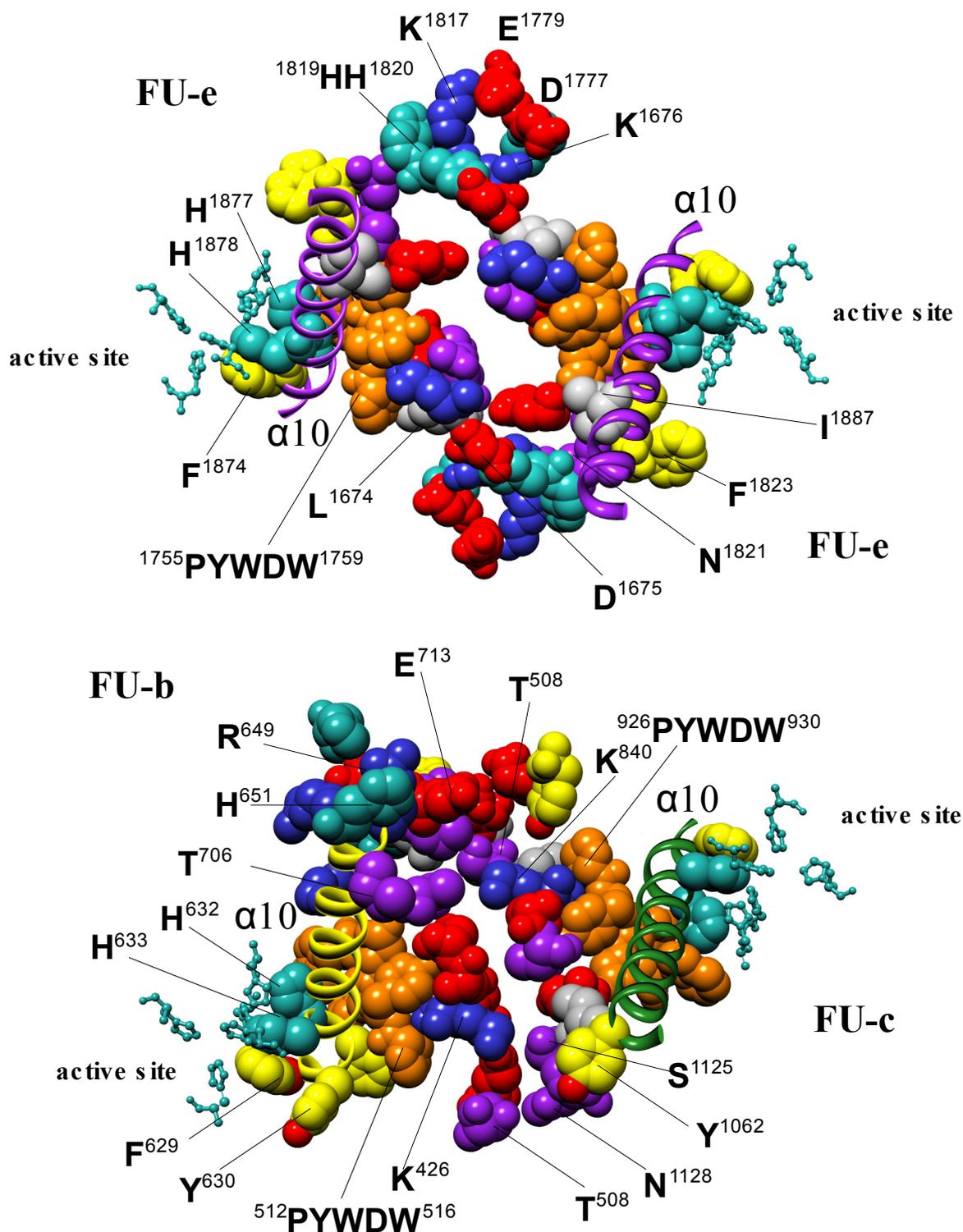
**Fig. 56: CryoEM structures and molecular models of the FUs involved in the minor groove interfaces**  
(For details see main text)

Both interfaces provide several opportunities for electrostatic bonding and allosteric interaction (see Fig. 57).

In detail, the massive contacts at the periphery of interface  $e \leftrightarrow e$  correspond to clusters of charged amino acids, which provide various possibilities for strong salt bridges (K1817-E1779, H1820-D1675, K1676-D1777). Moreover, further salt bridges might be formed between E1951 and K1669. Thus, the  $e \leftrightarrow e$  interface could provide a strong connection across the minor groove between the two constituents of the subunit dimer. In *Nautilus* hemocyanin, the equivalent interface shows in comparison, only two opportunities for salt bridges.

Thus, the present molecular model suggests that, due to the more “compact”  $e \leftrightarrow e$  interface and the additional reinforcement provided by the FU-h slab (see below), the KLH1 subunit dimer is expected to be more stable than the NpH subunit dimer. As already mentioned, this is also supported by previous studies on the isolation of stable subunit dimers in cephalopod, chiton and gastropod hemocyanins (Bonaventura *et al.*, 1981; van Holde *et al.*, 1985; Harris *et al.*, 1994).

Allosteric signal transfer between the two neighboring active sites appears to be possible by mutually dislocating helix  $\alpha 10$ . As in NpH, adjacent to the interface lies the invariant motif PYWDW that is believed to influence the oxygen binding behaviour (Cuff *et al.*, 1998; Lieb *et al.*, 2000). This motif is attached to helix  $\alpha 10$  that carries the sixth active site histidine (H1878).



**Fig. 57: Molecular interfaces at the minor groove in the cylinder wall of KLH1**

(a) Minor groove interface e-e; note the peripheral clusters of charged residues, which provides several opportunities for electrostatic bonding; also note the possibility of interaction between the adjacent active sites by dislocating helix  $\alpha 10$  via the invariant PYWDW motif (for comparison to NpH, see Fig.8 d in [2])  
 (b) Minor groove interface b-c; note similarities and differences compared to (a) (for more details see main text)(for comparison to NpH, see Fig.8 b in [2]) Colour code as for Fig. 52, with the exception that the PYWDW motif is coloured in orange.

There is the possibility that this motif influences the active site histidine of the opposing FU *via* the closely associated salt bridge (E1951-K1669) and I1886, which is attached to helix  $\alpha 10$ . Even a slight displacement of helix  $\alpha 10$  might substantially affect the oxygen binding properties. Moreover, communication between the PYWDW motives might also be possible *via* the peripheral clusters of charged residues, which also contain three histidines. As already mentioned, histidine-rich motives are candidates for mediating cooperativity between adjacent FUs. In NpH, the amino acid constellation in interface  $e \leftrightarrow e$  suggests a similar mechanism of interaction between the active sites of the anti-parallel FUs. However, the pathway of signal transfer between the invariant PYWDW motif and helix  $\alpha 10$  appears to be different (see section C.3.3.3).

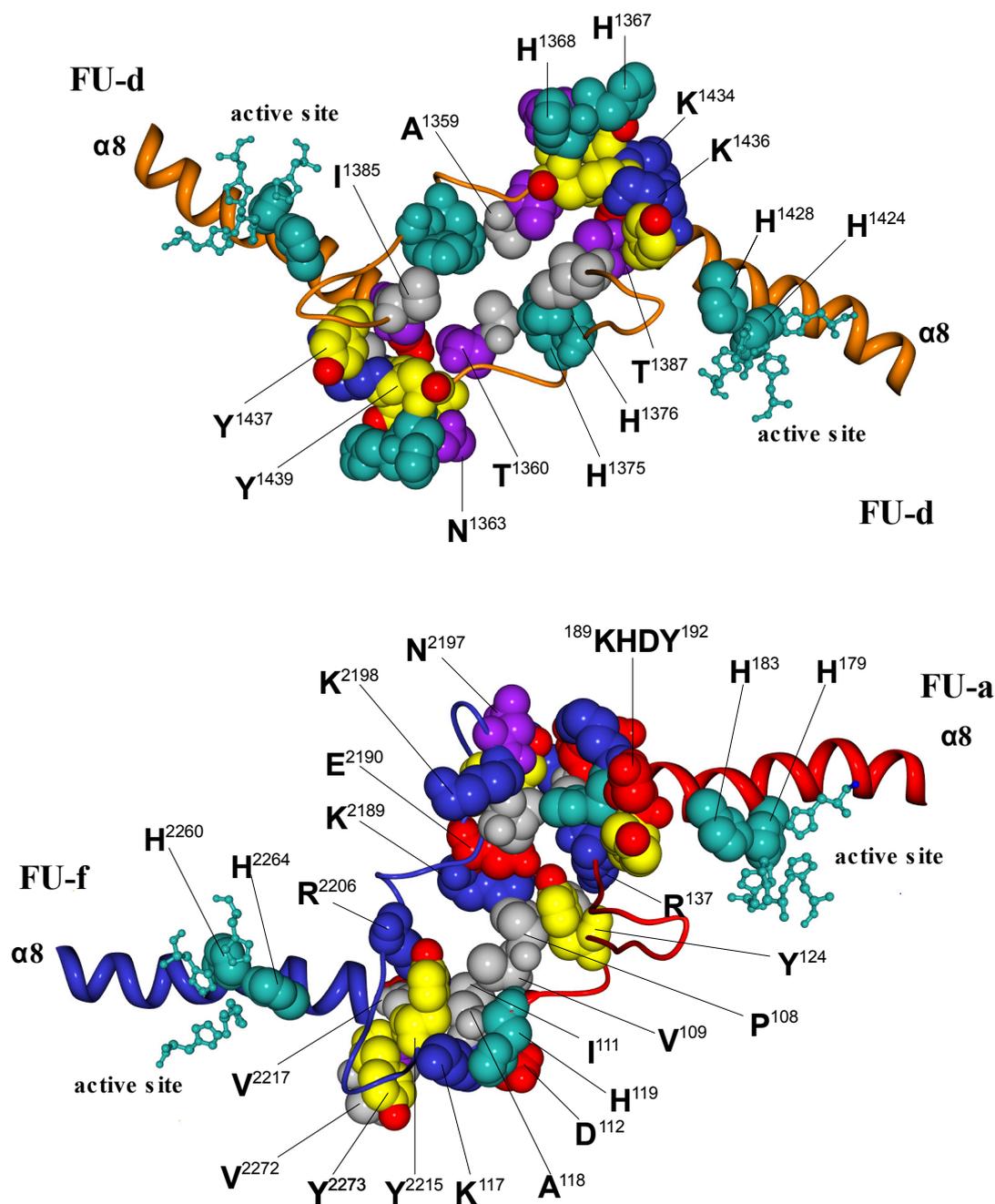
At first glance the situation in the  $b \leftrightarrow c$  minor groove interface appears to be different; this interface provides fewer opportunities for electrostatic bonding and the amino acid clusters in the periphery are less dominated by charged residues. Nevertheless, also in this case (and equivalently to NpH interface  $b \leftrightarrow c$ ), allosteric signal transfer between the neighbouring active sites appears to be possible by mutually dislocating helix  $\alpha 10$  *via* the invariant PYWDW motif. The near end of helix  $\alpha 10$  carries a tyrosine (FU-c, Y1062) or an arginine (FU-b, R649) which is surrounded by polar or charged residues respectively. These residue-clusters are closely associated to a lysine (K840, K426) which is in turn surrounded by residues of the invariant PYWDW motif, forming thereby an “arrow”-like structure. As already mentioned this motif is associated to helix  $\alpha 10$  and is believed to influence the oxygen binding behaviour (Cuff *et al.*, 1998; Lieb *et al.*, 2000). Thus, compared to the equivalent interface in NpH, the present analysis of the molecular interface  $b \leftrightarrow c$  suggests a similar mechanism for allosteric signal transfer between the two adjacent active sites.

#### 4.2.4 The major groove interfaces

As in NpH, the major groove in KLH1 is crossed by five different bridges which correspond to inter-FU-contacts ( $a \leftrightarrow f$ ,  $d \leftrightarrow d$ ,  $a \leftrightarrow e$ ,  $d \leftrightarrow b$  and  $a \leftrightarrow d$ ). All of them provide connection across the major groove between two adjacent subunit dimers.

Interface  $a \leftrightarrow d$  corresponds to the previously introduced structural feature of the molluscan hemocyanin wall, termed the anchor (see [1]; see [2]). As in NpH (see Fig. 9 d in [2]), also in the present structure (see Fig. 5 d in [3]), the anchor emerges as a connection formed between the  $\beta 8$ - $\beta 9$  loop of FU-a and FU-d. In both molecules, the  $\beta 8$ - $\beta 9$  loop is stabilized by a disulfide bridge, and in case of FU-a (but not in any other FU-type) is enlarged by aspartic acids (three aspartic acids in NpH, four in KLH1). This acidic cluster offers strong electrostatic bonding opportunities with polar and basic residues in the opposing  $\beta 8$ - $\beta 9$  loop of FU-d. Thus, the anchor provides a strong connection between the two adjacent

subunit dimers across the major groove in both hemocyanin molecules. Furthermore, this interface could also play a role in allosteric interaction. The striking similarity of the anchor interface in NpH and KLH1 suggests that this structural element might be of major importance for overall stability of the hemocyanin cylinder-wall.



**Fig. 58 : Molecular structure of the interfaces a↔f and d↔d at the major groove in the cylinder wall of KLH1**  
 (a) The a-f interface; note the opportunities for chemical bonding and for interaction between the two active sites. Also note the central and peripheral histidine residues. For comparison to NpH see Fig. 9 b in [2] (b) Interface d-f: note similarities and differences compared to a. For comparison to NpH see Fig. 9 a in [2]

The major groove interfaces  $a \leftrightarrow f$  and  $d \leftrightarrow d$  are structurally equivalent and connect the FUs at their core domains. Interface  $a \leftrightarrow f$  is localized in both peripheral wall tiers whereas interface  $d \leftrightarrow d$  is in the central wall tier. Both interfaces show a broad contact zone in the cryoEM-structure, but there is no covalent bonding *via* a linker peptide.

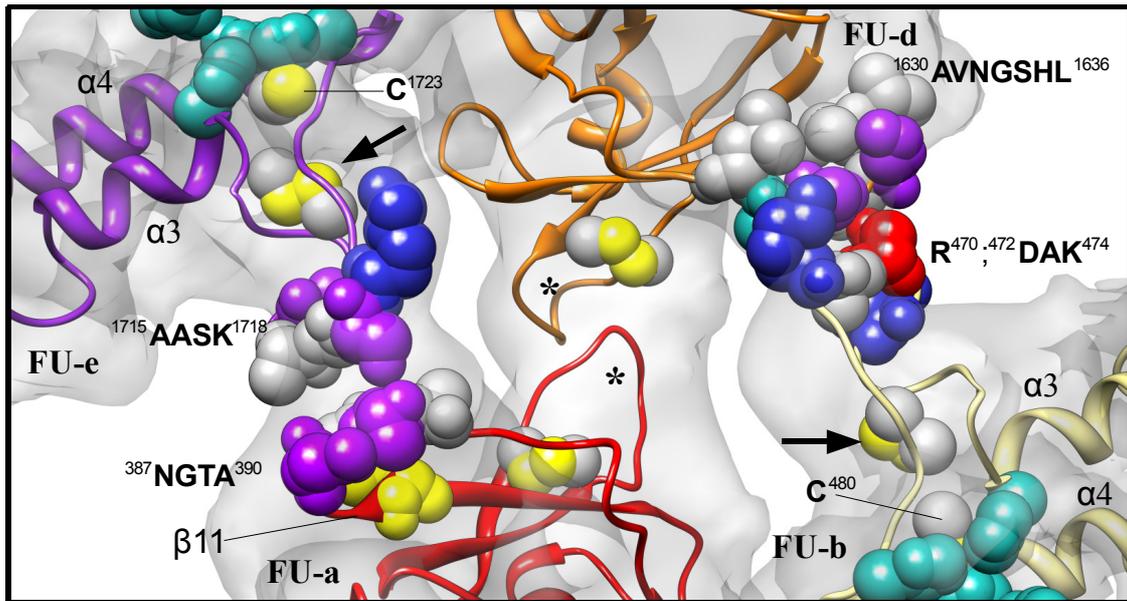
The general architecture of these interfaces is fully comparable to those in NpH (see Fig. 58). A major component of these interfaces is the long loop between  $\alpha 5$  and  $\alpha 6$  which provides several opportunities for electrostatic bonding and hydrophobic interaction; as in NpH, in case of interface  $d \leftrightarrow d$  four central histidines are present.

Furthermore, further contacts involve amino acids in the loop following helix  $\alpha 8$  (189KHDY192 in FU-a, 2270PQVY2273 in FU-f and 1434KGKY1437 in FU-d). Helix  $\alpha 8$  carries two active site histidine residues and therefore this amino acid constellation can not be excluded as possible candidate for allosteric signal transfer *via* the major groove.

Interestingly, compared to NpH, the distance between the two binding partners in interface  $a \leftrightarrow f$  is somewhat closer, although the association mode is identical. This small deviance could come from a deficiency in our molecular model and certainly, at 9 Å it is ambiguous as to whether such differences are real.

However, such deviances were not observed in the other 11 different types of wall interface. Moreover, it should be noted that in KLH1, the two FUs are involved in the slab-wall interfaces at the closed face and in the decamer/decamer interfaces at the open face of the decamer; these interfaces are absent in NpH. Thus, it might be possible, that a more compact  $a \leftrightarrow f$  interface is necessary for the formation of the additional interfaces in KLH1.

The major groove is further crossed by four smaller (but well defined) bridges which are constituted by two equivalent types of interface  $a \leftrightarrow e$  and  $d \leftrightarrow b$  (each with two copies). Both interfaces follow the architecture described for the equivalent interfaces in NpH and appear as promising candidates for conformational signal transmission in an analogous manner (see Fig. 59); moreover, in both hemocyanins these interfaces resemble the protomer-protomer contact in the crystal lattice of *Rapana* hemocyanin FU-e in that the tip of the  $\alpha 3$ - $\alpha 4$  loop touches the neighboring sandwich domain (Perbrandt *et al.*, 2003). However as already mentioned a major difference is that the rotation of the two Rth-e protomers against each other is rather different from our molecular models of KLH1 and NpH (40° deviation) and completely incompatible with both cryoEM-structures. This observation provides further evidence, that the basic association pattern of the protomers in the crystal structure of Rth is constrained by the crystal contacts; as already discussed, we assume that the crystallization process probably forced the assembling protomers into an artificial helical symmetry.



**Fig. 59: Major groove interfaces a↔f/ d↔b of KLH1.**

Allosteric interaction might be possible by movements of the two  $\beta$ -sandwich domains of FU-a and FU-d (red, orange) via the anchor interface (asterisks mark the  $\beta 8$ - $\beta 9$  loops forming the anchor interface; this interface is shown in Fig. 5d in [3]). Oxygenation-dependent movements of these domains might transmit forces to the  $\alpha 3$ - $\alpha 4$  loop of FU-d and FU-e respectively. These loops are stabilized by an invariant disulfide bridge (marked by arrows) and carry an additional strictly conserved cysteine (C1723, C480) which forms a thioether bridge with an active site histidine. This unusual thioether bridge probably contributes to the cooperative modulation of the oxygen affinity (Gielens *et al.*, 1992, Gielens *et al.*, 1997). For comparison to NpH see Fig. 9 c in [2].

#### 4.3 Putative binding sites for divalent cations within the decamer

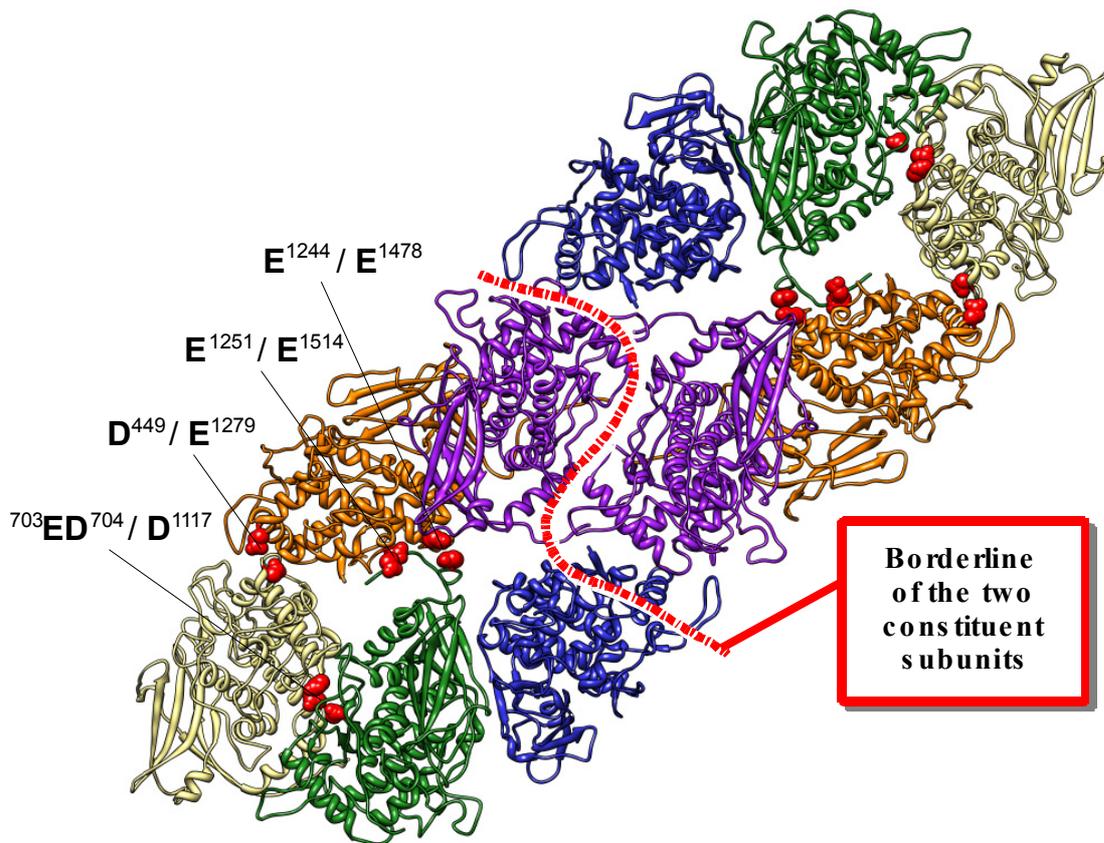
Molluscan hemocyanins require  $Mg^{2+}$  and/or  $Ca^{2+}$  for assembly from subunits and usually the quaternary structure is destabilized in the presence of chelating agents. In NpH we identified putative contacts between adjacent FUs, which involve a pair of carboxyl residues that are not compensated by available positively charged amino acids and therefore might bind a bridging divalent cation. In the cylinder wall of the KLH1 we also found a distribution of such putative structural reinforcements that are likely to stabilize the quaternary structure (see Fig. 60).

As in NpH, at the four morphological unit interfaces, no binding opportunities for bridging divalent cations can be detected. In contrast, two carboxyl pairs were detected in the horizontal tier interface b↔d (D449/E1279) and the minor groove interfaces b↔c (703ED704/D1117). The contact between the c→d linker and FU-d also shows two putative  $Ca^{2+}/Mg^{2+}$  binding sites (E1251/E1514; E1244/E1478). An additional site (E1246/E2492; not shown) bridges the two linker peptides c2→d2 and f2→g2 and probably serves as stabilizing

element of the stretched and exceptionally long f2→g2 linker. In case of the different folded f1→g1 linker such a contact is lacking.

The major groove interfaces lack such sites. Thus, linking of adjacent subunit dimers is not supported by divalent cations.

All five putative binding sites for a bridging divalent cation are localized within the same subunit, which is also true for the vast majority of such sites in NpH. However, in NpH an additional bonding opportunity was detected in the horizontal interface e↔f, bridging the two constituent subunits of the dimer.



**Fig. 60: Binding opportunities for bridging divalent cations.**

The subunit dimer (FU-a, FU-g and FU-h removed) with potential ligand carboxyl amino acids shown in red. For comparison to NpH see Fig- 10 d in [2])

At first glance, the absence of such bridging contacts in KLH1 between the two constituent subunits of the dimer seems to contradict observations that stabilizing the subunit dimer requires  $\text{Ca}^{2+}$  and/or  $\text{Mg}^{2+}$ .

However, it should be noted, that dissociation/reassociation experiments of *Helix pomatia* hemocyanin (HpH), reported a “loose” and “compact” conformation of the subunit dimer; the loose-compact transition was achieved by lowering the pH and/or by adding  $\text{Ca}^{2+}$  (Siezen and van Bruggen, 1974; Siezen and van Driel, 1974).

The compact conformation shows the typical “man-in boat” and parallelogram profiles (see Plate I b,c and Plate II b in Siezen and Van Bruggen, 1974) as seen in electron micrographs of purified subunit dimers of other 8FU-type hemocyanins (Harris *et al.*, 2004) and is fully comparable with the subunit dimer defined within the present study (see Fig. 53). Thus, without much doubt, the compact conformation of HpH corresponds to the stable subunit dimer required for nucleation of decamers.

These data suggest that for a stable subunit dimer, compaction of the subunit and consequently loose-compact transition is needed, which is most likely supported by divalent cations. This scenario is in full agreement with our present molecular model of KLH1, showing putative divalent cation-binding-sites exclusively between internal structures of the subunit.

#### 4.4 The decamer/decamer interfaces

The present study allows for the first time, description of the contact zones between the two halves of a molluscan hemocyanin didecamer at amino acid level.

The present cryoEM structure reveals 45 bridges representing four different types of interface, namely,  $b2 \leftrightarrow a2$ ,  $c2 \leftrightarrow a2$  (represented by two bridges in the density map),  $f2 \leftrightarrow c2$  (each of them are present in 10 copies) and  $f2 \leftrightarrow f2$  (present in five copies). The four types of interface are described in detail in relation to the molecular model in [3] (see Fig. 6 in [3]).

All in all, the interdecamer bridges show almost exclusively charged residues that might form contacts; they clearly appear as fixation points, providing strong attractive forces between the two decamers. This is in full agreement with earlier ultracentrifugation studies on a bivalve hemocyanin assuming rather few hydrophobic groups at the interdecamer interfaces (Herskovits *et al.*, 1991a).

Moreover, in interfaces  $f2 \leftrightarrow f2$  and  $b2 \leftrightarrow a2$ , we identified carboxyl residue pairs, at a distance of no more than 5 Å, that are not neutralized by adjacent positively charged residues; therefore they might bind a bridging divalent cation. Our data suggest that 15 bridging divalent cations reinforce the face-to-face assembly of two decamers.

This is in turn consistent with reassembly experiments in KLH1 and KLH2, showing that ion-stabilized decamers reassemble in the presence of divalent cations (Harris *et al.*, 1997). Thus, the hybrid model of KLH1 presented within this study provides for the first time the structural basis for these observations.

Moreover, as one might expect, cephalopod hemocyanins (such as *Octopus*, *Nautilus* and *Sepia*) which occur only as decamers, lack these sites; this underlines their likely importance for the formation of didecamers.

#### 4.5 Topology of the five arcs and the arc, arc/wall interfaces, hypothetical reassembly scenario

In NpH and KLH1 the wall of the molecule is symmetrical but the collar complex is asymmetric.

In NpH the collar complex consists of 5 discrete “arcs”, representing FU-g pairs (see [2]). The asymmetric arc derives from the fact, that the two subunits of the subunit dimer fold their tail into different conformations. We discovered that the two FU-g copies, united in one arc, differ in conformation with respect to their arc/wall interfaces; therefore we designated them as FU-g1 and g2.

Moreover, the five FU-g pairs in the collar of NpH follow the architecture of the Octopus FU-g dimer crystal lattice, forming thereby a  $g1 \leftrightarrow g2$  interface (see Fig. 40), which connects two adjacent subunit dimers. Thus, the pairing of the two subunits yields an asymmetric homodimer.

In order to explain this intriguing asymmetry of the collar, we proposed the following reassembly scenario: as soon as two dimers join each other, five FU-g1 copies shift into FU-g2 conformation, in order to form, together with the alternating five FU-g1 copies, the arc morphological units, connecting the two dimers *via* the major groove (see Fig. 10 c in [2]). The resulting  $g1 \leftrightarrow g2$  interface seems to be a well-conserved feature among molluscan hemocyanins, since all FUs of NpH and KLH1 (an exception is KLH1-h, see below) are arranged pair-wise as morphological units following this association mode.

Compared to NpH, the five arcs in KLH1 show at first glance a similar architecture; the main and obvious difference between the collar complexes of NpH and KLH1, is that the collar complex of KLH1 is enlarged by an outer annulus of five slabs, representing FU-pairs of the additional FU h.

Superposition and a more detailed comparison between the cryoEM structures of KLH1 and NpH showed that the five arcs are indeed topologically identical; the FU-g pairs show an identical OdH-g-like pairing mode. However, comparison of the molecular models revealed that the KLH1 arc is shifted ca 5 Å towards the open face of the decamer. Indeed, the f→g linker is one amino acid shorter than in NpH. Moreover, such a shift may be induced due to the presence of the additional 0.5 MDa outer slab pentamer and the additional arc-slab interfaces (see below). However, at 9 Å resolution, it is uncertain whether this shift is real.

The arc interface  $g1 \leftrightarrow g2$  is as already mentioned fully comparable to the interface of the Octopus FU-g dimer crystal lattice and the corresponding  $g1 \leftrightarrow g2$  interface in NpH (for comparison see Fig. 7b in [3] and Fig. 7a in [2]). As in NpH, the arc is attached to the cylinder wall *via* two interfaces ( $g1 \leftrightarrow d1$ ;  $g2 \leftrightarrow d2$ ). The architecture of these interfaces is fully comparable to those in NpH (for comparison see Fig. 7 c, d in [3] and Fig. 10 a, b in [2]).

The remarkable conservation of these structural features suggests that a similar reassembly scenario (as proposed for NpH) might explain the asymmetric collar in KLH1 and gastropod hemocyanins in general, but with a different starting point:

- during assembly, the first six FUs of each subunit might form - in an anti-parallel manner - the cylinder wall. This step probably takes place after compaction of the subunit, which is supported by divalent cations (see previous section)
- at the same time, 10 FU-h-copies, form homodimers (slabs) which might associate simultaneously in order to form the ring of the outer collar (it should be mentioned that such a decameric ring is believed to be the precursor of the extant decameric cylinder (Lieb *et al.*, 2001)). This annulus has opportunities to form strong salt bridges with the edge of the cylinder wall (see next section)
- the formation of this ring and its strong connections to the cylinder wall compels that the FU-g copies to be oriented towards the same cylinder opening. Such a movement of the arc FUs is allowed by their long linker peptides
- in order to stabilize the decameric structure, a further strong connection is needed between the adjacent subunit dimers (as already mentioned, each slab connects two subunits of the same subunit dimer). This strong connection is provided as soon 5 FU-g copies completely shift to FU-g2 conformation in order to form, together with the five alternating FU-g copies, the arcs which connect adjacent subunit dimers *via* the major grooves. This step probably completes and stabilizes the structure of the decamer.

In NpH, due to the absence of the outer collar, we suggest that arc formation probably takes place as soon as two dimers join each other; the arc serves thereby as a clamp between two adjacent subunit dimers and apparently the  $g1 \leftrightarrow g2$  interface is indispensable for the stabilization of the decamer.

The remarkable structural conservation of the arc structure and the arc-interface and also the fact that FU-g exhibits the most conservative sequence further support this hypothetical reassembly scenario and might explain why this behavior is somehow conserved in NpH, even in the absence of FU-h. At this point, it should be noted, that the seven-unit hemocyanins evolved from an eight-unit progenitor and not the other way around (Lieb *et al.*, 2001).

## 4.6 Interfaces within the Fu-h slab pentamer

The peripheral collar of KLH1 consists of five morphological units, termed the slabs, each comprising an FU-h pair (h1-h2) (see Fig. 61a). Functional units of type h are unique in many aspects. They have been detected only in gastropods and chitons and exhibit a unique tail extension of 95-100 amino acids, which is lacking in functional units a to g (e.g. Lieb *et al.*, 2000). The deviant primary structure of FU-h is not a surprise in view of its special contribution to the peripheral collar complex.

The present cryoEM structure reveals a thick bridge between the adjacent slabs and a broad contact zone between the two FU-h copies on the central region of the slab. Rigid body fitting of the X-ray-structure of the OdH-g monomer (Cuff *et al.*, 1998) and alternatively of the corresponding homology-modelled core/ $\beta$ -sandwich domain of KLH1-h was efficient, but left an empty space between the two anti-parallel FU-h copies within each slab (see Fig. 61 b). This region is apparently established by the 100-aa unique tail of FU-h.

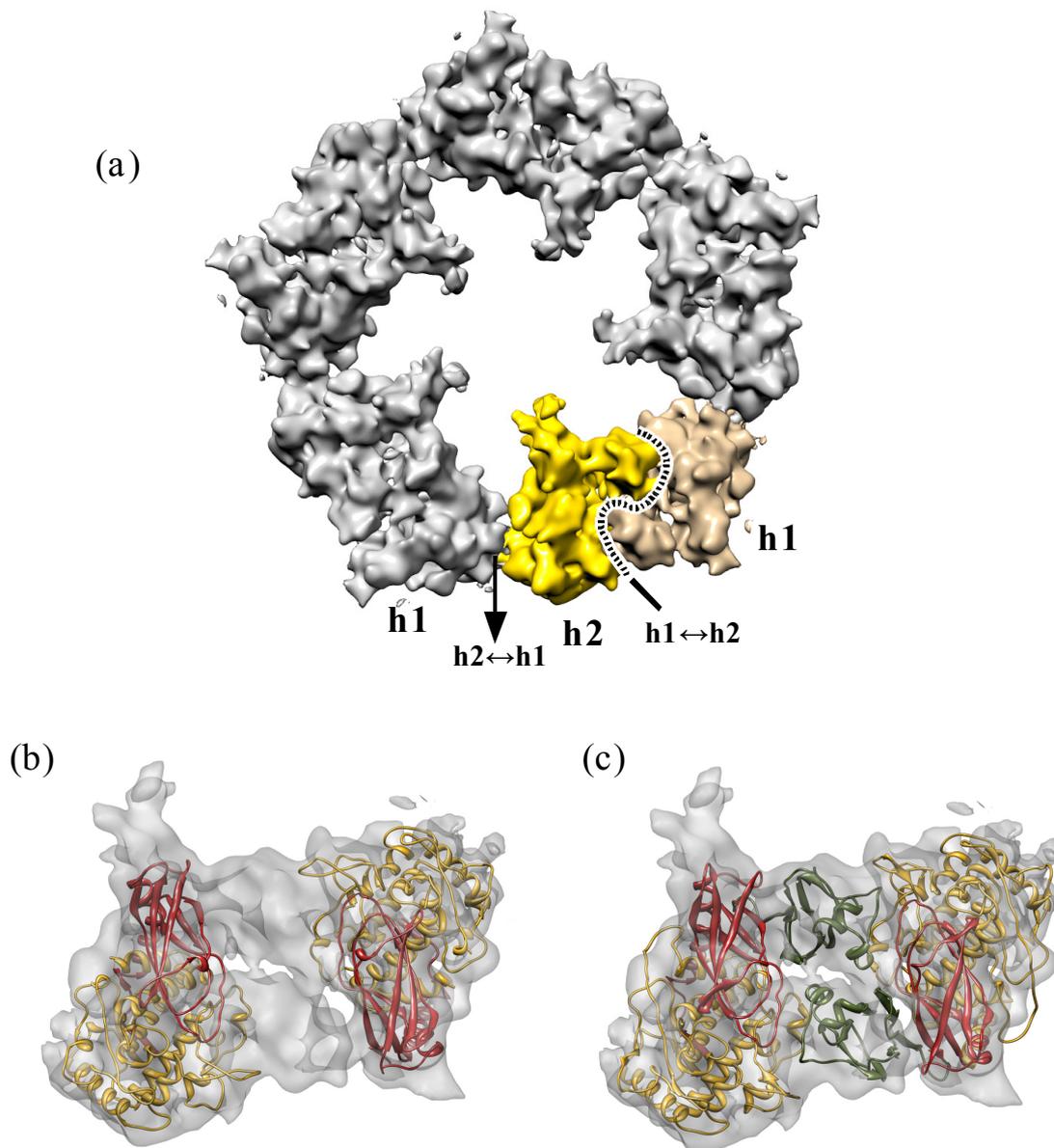
Indeed, the recent 4 Å X-ray structure of the KLH1-h dimer (Barends *et al.*, 2008) shows a C-terminal cupredoxin-like fold that exactly fills this space (see Fig. 61 c). The present cryoEM structure together with the preliminary fitting of the homology-modelled core/ $\beta$ -sandwich domain of KLH1-h revealed that the FU-h pairs within each slab strikingly resemble the crystallographic dimer of KLH1-h and confirmed that the assembly of the two protomers in the KLH1-h crystal followed a pattern that occurs in the native molecule.

The two FU-h copies within each slab are anti-parallel and connected *via* their cupredoxin domains that form a broad contact zone (h1 $\leftrightarrow$ h2 interface) with the opposing  $\alpha$ -helical (core) domains. However, due to the limited resolution of the crystal structure the orientation of side chains in the cupredoxin-like domains remained completely unsolved and homology modelling with comparable cupredoxins is critical because of low sequence identities. Therefore, molecular details of the h1 $\leftrightarrow$ h2 interface (see Fig. 8d in [3]) remain unclear. Moreover it is still unknown if the two FU-h copies are somehow allosterically coupled; at first glance, due to the large distance between the two active sites in the core domains, a direct site-to-site communication can be excluded. Moreover, the cupredoxin-like domain seems to lack a special active site.

It appears that the h1 $\leftrightarrow$ h2 interface might solely function as a stabilizer of the slab and therefore of the entire subunit dimer. As already mentioned, the slab morphological unit links *via* the h1 $\leftrightarrow$ h2 interface of two subunits from the same subunit dimer; this results in a more compact subunit dimer in comparison with the 7-FU-type cephalopod hemocyanins (see previous section).

In addition, the molecular model revealed that the thick bridges connecting adjacent slabs correspond to a broad contact zone between the core domains of two adjacent FU-h copies (interface h2 $\leftrightarrow$ h1), which stem from different subunit dimers. Therefore, this interface might

function as structural stabilizer of the peripheral collar ring. Indeed, it involves 20-30 polar or charged amino acids and provides several opportunities for electrostatic bonding. Thus, this interface might provide strong “glue” between two adjacent subunit dimers (similar to the arc morphological unit). Moreover, adjacent to the interface lies the invariant PYWDW motif which is believed to influence oxygen binding.



**Fig. 61: Structural organization of the peripheral collar**

(a) Extracted cryo EM structure of the slab pentamer, with one of the slabs highlighted in colour. The interface between the two FU-h copies within each slab ( $h1 \leftrightarrow h2$ ) and the slab-slab interface ( $h2 \leftrightarrow h1$ ) are also indicated. (b) Crystal structure of the OdH-g monomer (consisting of a core domain (shown in gold) and a  $\beta$ -sandwich domain (shown in red)) docked in the extracted cryoEM structure of the slab morphological unit. Note the empty spaces in the central region of the slab (c) fitting of the KLH1-h crystallographic dimer revealed that these density masses correspond to the tail -cupredoxin- like-extension of FU-h (shown in green)

#### 4.7 Interfaces between the peripheral and the internal collar

The present cryoEM structure reveals that each arc and slab, is interconnected *via* two clearly defined bridges, a narrow and a prominent one (see Fig. 8 b in [3]). The present molecular model revealed that the narrow bridge represents the two g→h (g1→h1, g2→h2) linker peptides, which run parallel and covalently link the arc with the two FU-h copies. As already mentioned, the arcs link adjacent subunit dimers and each subunit dimer is equipped with one complete slab. Consequently, the two FU-g copies within each arc are not covalently linked to the two FU-h copies of the same slab, but to two FU-h copies from different but adjacent slabs (see also Fig. 53 and Fig. 8 b in [3]).

The prominent slab-arc bridge represents a contact zone between KLH1-g1 and KLH1-h1. In contrast, an interface between KLH1-g2 and KLH1-h2 is lacking.

The g1↔h1 interface offers various possibilities for electrostatic bonding and hydrophobic interactions and therefore might provide a precise adjustment between arc and slab (see Fig. 9 (a) in [3]). The long g→h linker peptides appear to be rather a secondary link.

#### 4.8 Interfaces between peripheral collar and wall

Due to the respective local environment of the two FU-h (KLH-h1 and KLH1-h2) copies within each slab, their interfaces with neighboring wall FUs are different. According to the present molecular model and the 9 Å cryoEM structure of KLH1, KLH1-h1 shows contacts to KLH1-a1, -c1 and -f1 of the same subunit. This KLH1-f1 copy is also fixed to KLH1-h2. KLH1-h2 forms an additional bridge to the KLH1-a1 copy of the next subunit (see Fig. 9b in [3]).

Thus, the slab is attached to the cylinder wall exclusively *via* non-covalent contacts to FUs of the peripheral wall tier of subunit conformer 1. However, it should be noted, that a contact between slab and KLH1-b1 is absent.

The slab/wall interfaces are completely dominated by electrostatic bonding and probably the formation of strong salt bridges with the edge of the cylinder wall, forces the annulus into its asymmetric position. The slab/wall interfaces are described in detail in [3] (see the corresponding section in [3]; see also Fig. 9 in [3]).

#### 4.9 Glycosylation pattern

KLH is widely used as hapten carrier and immune stimulant and it is generally accepted that the sugar constituents of this glycoprotein, are likely to be implicated in the antigenicity and biomedical properties of KLH (see also section A.2.2.). The carbohydrate content of total KLH was calculated to be ca 4% of the molecular mass (van Kuik *et al.*, 1990).

The amino acid sequence, determined within our working group (Lieb & Markl, 2004) shows eight potential attachment sites (NXT/NXS) for N-linked glycans in KLH1 (see table 2). Only two FUs, KLH1-c and KLH1-e, lack such sites, whereas KLH1 -a, -b, -d, -f, -g and -h show either one or two.

Also, in other molluscan hemocyanins, N-glycans are lacking in case of FU-c, but FU-e is usually glycosylated (Lieb *et al.*, 2004). However, in isoform KLH2, the FU-c is decorated with a ~8 kDa O-glycosidically linked carbohydrate; this contrasts to all other KLH functional units (including KLH1-c), which lack O-linked glycosides (Stoeva *et al.*, 1999).

	<b><u>potential N-glycosylation sites</u></b>
KLH1-a	N387
KLH1-b	N529
KLH1-c	-
KLH1-d	N1632, N1264
KLH1-e	-
KLH1-f	N2129, N2468
KLH1-g	N2555
KLH1-h	N2362

**Table 2: Putative attachment sites of N-linked glycans in KLH1**

Analysis of the carbohydrates of KLH (isoforms 1 and 2) revealed a rich collection of N-linked glycans (Kurokawa *et al.*, 2002), suggesting that most of the potential attachment sites are used for this purpose.

Indeed, the present hybrid model of KLH1, reveals that the vast majority of the NXT/S type asparagines are associated with an obvious cryo-EM structure, that is not explained by the molecular model and is likely to represent a carbohydrate tree (see Fig. 10 in [3]). From the 8 different sites per subunit, only N2164 in KLH1-d and N2468 in KLH1-f are not associated with any conspicuous protrusion of the density map. Furthermore, these potential attachment

sites are buried deeply in the cryoEM-structure and probably are not accessible for sugar attachment.

Four of the carbohydrate side chains project from the cylinder wall (N387, N529, N1632, N1264). Another site (N2555) is part of FU-g and consequently, each arc (g1-g2) is tagged with a pair of sugar trees. Also, the peripheral collar components are tagged. The carbohydrate trees of KLH1-h2 and -h1 are associated with N3262. In the case of KLH1-h2, a carbohydrate side chain protrudes as five copies from each slab into the cylinder opening. The corresponding protrusion in the density map shows a characteristic diantennate shape and allowed docking of a sugar side chain extracted from the crystal structure of OdH-g, a complex oligosaccharide which is structurally related to several N-linked glycans found in KLH (Kurokawa *et al.*, 2002) (see Fig. 10 in [3]). In the alternating anti-parallel arranged KLH1-h1 copies, the corresponding protrusions contact the rim of the cylinder wall where it merges with KLH1-c1.

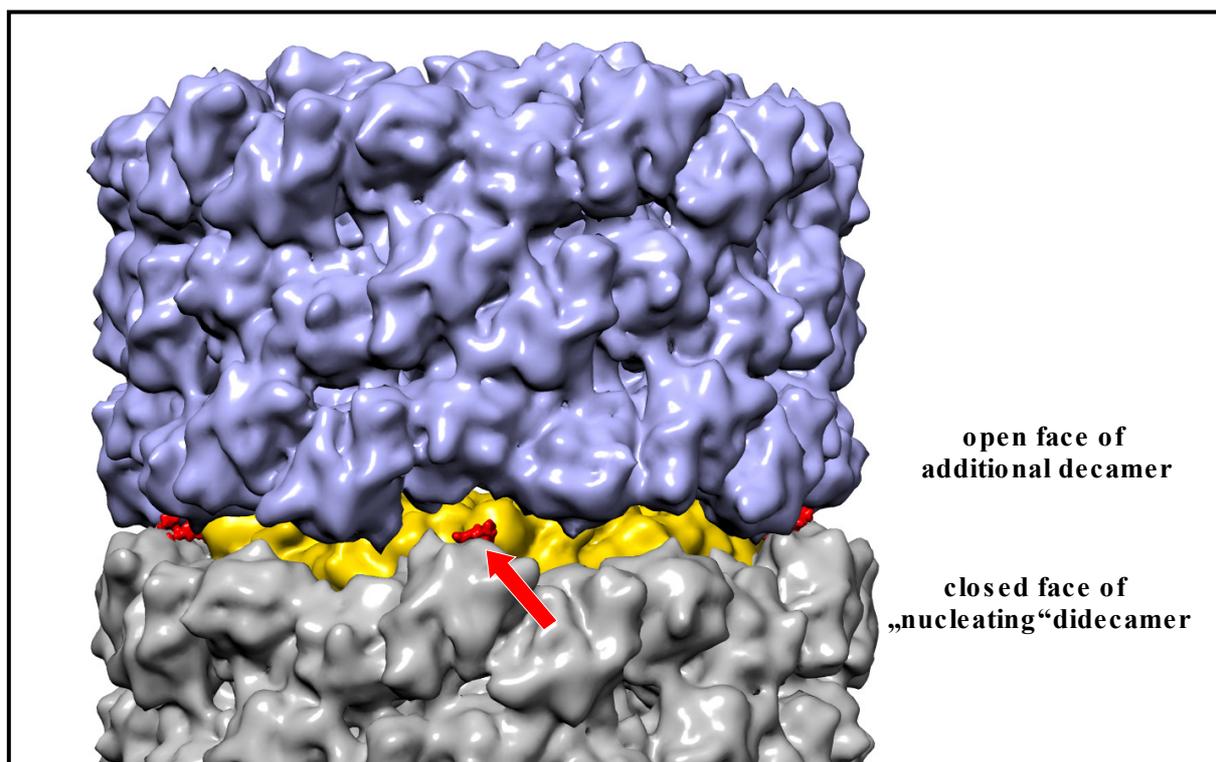
Thus, the present study reveals for the first time the principal localization and orientation of the glycan side chains in KLH isoform 1: thus, 120 N-linked evenly distributed carbohydrate chains protrude from this giant invertebrate protein.

#### 4.9.1 N-glycosylation sites and possible consequences for assembly

In *Megathura* and *Haliothis* hemocyanins a portion of native KLH2 and HtH2 forms assemblies larger than the didecamer. During multidecamer formation, additional decamers are attached, with their lower tier („open face”) to the upper tiers (“closed faces”) of a didecamer. This assembly mode can continue on both sides of the didecamer to form long tubes. In contrast, native KLH1 and HtH1 are fully restricted to a didecameric state (Keller *et al.*, 1999; Harris *et al.*, 1997; Harris *et al.*, 2000). However, during our experimental high Ca/Mg treatments (Harris *et al.*) we were able to produce multi-decamers from KLH1.

Lieb *et al.* (1999) proposed that the additional carbohydrate chains found in HtH1-f and HtH1-h could probably be involved in blocking multidecamer formation in HtH1, whereas absence of these side chains in HtH2-f and/or HtH2-h would allow such an oligomerisation in HtH2. The present hybrid model of KLH1, reveals that due to their topology, both FU-types could well interact with the lower tier (“open face”) of a neighbouring decamer. However, the corresponding additional N-glycosylation site in KLH1-f (N2470) appears not to be accessible for sugar attachment. In contrast, the specific attachment site of KLH1-h which is present in HtH1-h, but missing in KLH2-h and HtH2-h appears to be glycosylated. Moreover, these five copies of a large sugar moiety protruding from KLH1-h1 are deposited on the rim of the cylinder wall (closed face) and might well block multicamer formation (see Fig. 62). Thus, our data clearly support and provide the structural basis for the highly speculative (at that time)

concept of Lieb *et al.* (1999), but according to the reassociation data on KLH1 at high Ca/Mg concentrations (Harris *et al.*), the capability to form multidecamers, might be additionally supported by bridging divalent cations and consequently also depend on specific amino acids in the FU-repeat a-b-c-f.



**Fig. 62 : Hypothetical structural explanation for the inefficiency of KLH1 to form multidecamers**

Five sugar trees (one of them is indicated; red arrow) which protrude from the slab pentamer (shown in gold) are positioned on the rim of the wall at the closed face of the didecamer (cylinder wall shown in grey) and might sterically hinder the attachment of an additional decamer (shown in purple) and consequently multidecamer formation. The 9 Å volumes shown here were calculated from the molecular model of the KLH1-didecamer, tagged with carbohydrate trees at the h1 positions that correlate with empty masses in the cryoEM structure (red). The sugar side chains shown here were extracted from the crystal structure of OdH-g (Cuff *et al.*, 1997). Note, that according to Kurokawa *et al.*, (2002), the sugar trees of KLH1 are considerably larger.

#### 4.10 Putative model of the allosteric unit

In terms of mechanisms of oxygen binding, gastropod hemocyanins have not been examined in as much depth as have the cephalopod and the arthropod hemocyanins, and in only a few cases has an analysis of allosteric mechanisms been attempted.

Zolla *et al.* (1978) were able to fit cooperative binding data for the didecameric  $\beta$ -hemocyanin of *Helix*, using the classic MWC model and an allosteric unit size of 15 (possibly 16; Di Cera, 1990) below pH 8, and a size of 8 sites at higher pH.

The dimensions of a functional constellation (from 15, possibly 16, to 8 sites, depending on pH), corresponds very well with the number of FUs of a subunit and a subunit dimer. Moreover, the number of eight or 16 sites suggests that all eight types of FUs are involved with one or two copies ((2x) a-b-c-d-e-f-g-h) respectively. The authors explained this pH dependent change, in that some pair-wise coupling between monomers could relax at pH 8.0. On the other hand, isolated monomers do not bind oxygen cooperatively – to behave allosterically, the subunit must be in contact with other subunits, even though each subunit carries eight oxygen binding sites.

In native *Octopus* hemocyanin (decamer) the allosteric unit was found to encompass seven active sites and from their structural similarities the same is expected for *Nautilus* hemocyanin. The molecular model of NpH, indicated that the allosteric unit encompasses the densely packed series of six wall FUs that lie between the major and the minor groove (see Fig. 49) together with one FU-g copy (arc). These FUs stem from different subunits and this might explain why a single subunit is not cooperative. The remarkable conservation in KLH1 of the 15 molecular interfaces that also exist in NpH (see previous sections; [2]; [3]), suggests that basic functions are maintained similarly in NpH and the NpH-like core of KLH1. Thus, our proposed model of the allosteric unit can be adapted for the NpH-like core of KLH1. Nevertheless, in order to reach the experimentally determined number of 8 (vs 7 in NpH), we might also include one copy of the additional FU-h.

In addition, according to the present molecular model of KLH1, some of the contacts between decamers, suggest that, in relation to allostery, the decamers might not act as independent units. The inter-decamer interfaces might functionally link the 8-sites allosteric-units of the two adjacent decamers and this would result in an allosteric unit of 16 sites. According to this model, the allosteric unit of an isolated KLH1-decamer would encompass 8 sites, whereas the allosteric unit of the native molecule (the didecamer) would encompass 16 sites. This would explain the pH-dependent change in the dimensions of the allosteric unit in *Helix* described by Zolla *et al.* (1978). Indeed, the KLH1 didecamer (allosteric unit of 16 sites) can be rapidly split into an almost homogenous population of decamers (allosteric unit of 8 sites) by increasing the pH from 7.4 to 8.5 (Söhnngen *et al.*, 1997).

On the other hand, *Helix* hemocyanin shows a relative high degree of cooperativity ( $n_{50}=7$ ) (Zolla *et al.*, 1978), whereas KLH is apparently less cooperative ( $n_{50}=3$ ) in oxygen-binding (Senozan *et al.*, 1981). Moreover, a Hill Coefficient of less than one has been measured for *Archachalina* hemocyanin (Taiwo, 1992), which also exists as a didecamer. The cooperativity exhibited in the oxygen binding by these giant proteins is in general remarkably low. In addition, didecamers do not significantly differ in cooperativity from decamers (van Holde and

Miller, 1995). Therefore, we are still skeptical about the present indications for a 16-site-allosteric-unit and allosteric interaction between the two associated decamers.

However, it should be also mentioned, that most of the available oxygen-binding studies do not include careful characterization of the higher-order state of the respective hemocyanins during oxygen binding and were also performed under different pH-conditions (i.e. to mimic *in vivo*, a pH value of 7.2 was chosen for *Helix*, whereas a value of 7.9 was chosen for KLH).

Thus, this concept of a waggon-train-like allosteric unit of 16 sites is highly speculative and further cooperative binding data are needed to substantiate this idea. Moreover, considerations as to what extent the slab pentamer is involved in allosteric interaction require greater structural detail of the cupredoxin-like domain.

## 5. Conclusions and Outlook

In the present study, the quaternary structures of *Nautilus* and keyhole limpet hemocyanin were elucidated at 9 Å resolution by CryoEM single particle analysis. The presented reconstructions allowed for the first time construction of reliable molecular models, by advanced hology modelling and rigid body fitting, which in turn essentially solved the structure of these giant invertebrate respiratory proteins, after several decades of research in many laboratories.

In particular, the first hybrid pseudoatomic model of a molluscan hemocyanin decamer (NpH; published in 2007) was the first major step in order to describe the morphology of the molluscan hemocyanin decamer precisely, understand the assembly and reveal some fundamental features of this complex multi-subunit protein, such as the topology of the seven different functional units.

In short, NpH is a hollow cylindrical decamer of a 350 kDa subunit, which consists of seven different FU-types (termed FU-a to FU-g). In contrast to gastropod hemocyanins, a C'-terminal FU-h is missing. The first 6 FU-types form the cylinder wall, whereas five FU-g pairs (arcs) constitute the internal collar complex. The ten identical subunits are arranged in an anti-parallel manner as pairs (subunit dimers), which gives the wall the appearance of a right handed helix.

In our hybrid model, we identified the previously unknown subunit dimer and subunit pathway within the wall. Although several models of the molluscan hemocyanin quaternary structure based on low resolution cryoEM structures have been proposed, they are excluded for *Nautilus* hemocyanin.

All previously described but not well understood features of the cephalopod hemocyanin quaternary structure, such as the fenestrated wall, the major and minor grooves, the anchor structure and the internal collar, can now be explained at pseudo-atomic resolution.

The seven different FU types are arranged as pairs (a-b, d-e, c-f, g-g), which resemble in constellation the X-ray structure of the FU-dimer described for the *Octopus* FU-g.

Unexpectedly the arc pentamer was found to be shifted towards one cylinder opening, imposing an overall C5 symmetry on the decamer (rather than D5, which was used in all earlier 3D reconstructions of cephalopod hemocyanins). This arrangement corresponds to the arc described in gastropod hemocyanins and from these data it becomes clear that except for the absence of FU-h, the cephalopods fully conserved the original collar structure of their hemocyanin. This is in full agreement with molecular phylogeny data, showing that the seven-FU subunit is an apomorphy of the Cephalopoda. Moreover, in order to explain this local asymmetry of the homodecamer, which is somehow conserved even in the absence of FU-h, we propose a hypothetical reassembly scenario, suggesting that the asymmetric arc derives from the need to form the arc morphological unit; this appears to be a well conserved

feature, since all FUs of NpH are arranged pairwise as morphological units following an identical OdH-g like association mode.

Furthermore, the present molecular model of NpH, allowed identification of 15 types of interfaces between functional units, which can now be assigned to very few amino acids. Many of them serve as possible mediators of allosteric signals during oxygen binding and from their analysis, putative communication pathways between adjacent active sites were elucidated. Based on these results, we suggest a model for the allosteric unit, consisting of a straight row of 7 FUs (encompassing seven active sites), which are localised between minor and major groove and stem from two different subunits of the same subunit dimer. This 7-site structure fits the results of earlier oxygen-binding studies and we propose it as a working model for future research. Thus, the present all-atom model of NpH opens the door to a fundamental understanding of the function of this cooperative protein.

Moreover, the present hybrid model of NpH, allowed us for the first time to elucidate the glycosylation pattern of the decamer. Twelve protrusions from the cryoEM structure per subunit are associated with putative attachment sites for N-linked glycans, indicating a total of 120 sugar trees on the outer surface of this giant invertebrate glycoprotein.

The distribution of potential calcium/magnesium binding sites, which according to earlier studies are required for native hemocyanin assembly from subunits, has also emerged.

Comparable data on a gastropod hemocyanin didecamer (KLH1; published in 2009), essentially solved the structure of this promising tumour vaccine carrier and also confirmed and substantially broadened our structural analysis of NpH and clarified the difference between the cephalopod hemocyanin decamer and the more complex gastropod hemocyanin didecamer.

KLH1 is assembled from two asymmetric hollow cylindrical decamers at their open ends. Each decamer is based on a 400 kDa polypeptide, which is subdivided in eight functional units. The additional C'-terminal FU (FU-h) (which is absent in cephalopods) carries a unique C'-cupredoxin-like terminal extension of 100 amino acids which is lacking in functional units a to g. At the level of the peripheral tier 10 FU-h copies, arranged as pairs (termed slabs), form an outer collar ring.

The present hybrid model of KLH1 revealed that each KLH1-decamer consists of a NpH-like core (wall and arc) and a slab pentamer. In NpH and KLH1 the pathway and structure of the subunit segment a-b-c-d-e-f-g are highly conserved and in addition, we were also able to determine how the pathway continues from FU-g to FU-h.

All in all, the structure of NpH is very similar to wall-and-arc structure of the KLH1 decamer. Equivalent FUs occupy equivalent positions and are connected *via* similar bridges. The subunit dimer of KLH1, identified within this study, is fully comparable with the subunit dimer model that we proposed for NpH, but additionally equipped with a FU-h pair (slab). The slab

provides, apparently, an additional structural reinforcement (see below), suggesting greater stability of the KLH1-subunit dimer in comparison to NpH. Indeed, earlier studies demonstrated that in *Octopus* and *Nautilus* (7 FU type hemocyanins), the subunit dimer is a rather transient dissociation/reassembly intermediate, compared to other studied 8-FU-type hemocyanins. Our present data provide the structural basis for these observations.

Moreover, the NpH-like KLH1 core also shows the 15 types of inter-FU-interface that were detected in NpH. With respect to allostery, they provide similar bonding and chemomechanical possibilities. This suggests that basic functions and putative pathways for allosteric interaction between adjacent active sites are maintained similarly in NpH and KLH1, which is rather surprising for two hemocyanins that diverged 520 million years ago. Some of the observed amino acid exchanges might be responsible for the species-specific fine tuning of cooperativity during oxygen binding.

The major difference between the two hemocyanins concerns the additional features in KLH1: the contacts between the two halves of the didecamer and the slab pentamer.

The present cryoEM revealed that the FU-h pairs within each slab strikingly resemble the recently obtained crystallographic dimer of KLH1-h and confirmed that the assembly of the two protomers in the KLH1-h crystal followed a pattern that occurs in the native molecule. Moreover, our model revealed the various slab/slab, arc/slab and slab/wall interfaces. The two FU-h copies within each slab are anti-parallel and connected *via* their cupredoxin-like domains, which might function as a structural stabilizer of the slab. The slab pentamer is attached to one edge of the cylinder exclusively *via* interfaces dominated by electrostatic bonding.

The present model also fully explains the inter-decamer bridges, which have remained ambiguous, because of low resolution and unclear subunit topology. We detected a total of 45 bridges. Most contacts indicate strong electrostatic bonding and strong attractive forces between the two decamers. In some cases, we also detected binding sites for divalent cations, which explain the observations, that the decamer → didecamer transition can be controlled experimentally by the calcium/magnesium concentration level.

Furthermore, some of the contacts indicate that the two decamers might be allosterically coupled and give hints for a 16-site allosteric unit. However, our proposed model for the allosteric unit is highly speculative and certainly further oxygen-binding data are urgently needed, in order to understand the oxygen-binding properties of KLH1.

Our model also revealed the putative N-glycosylation pattern of KLH1. The sugar constituents of this glycoprotein are likely to be implicated in the antigenicity and biomedical properties of this protein. Six of the eight potential attachment sites per subunit appear to be accessible for sugar attachment. The KLH1 didecamer appears to be almost evenly tagged on its outer surface by a total of 120 sugar trees.

In conclusion, the present data provide fundamentally new insights into the architecture and function of molluscan hemocyanins.

However, it remains unknown as to why KLH1 forms exclusively didecamers, whereas the KLH2 didecamer is able to bind additional decamers to form tubular multidecamers. The present study revealed the protruding N3262 sugars (which are absent in KLH2) as potential blockers. However, the capability to form multidecamers, might also depend on specific amino acids in the FU-repeat a-b-c-f. Therefore, to complete the molluscan hemocyanin picture, further studies on the didecamer and tridecamer of KLH2 by cryoEM are currently in progress. A comparable molecular model of the KLH2-tridecamer should additionally reveal the didecamer-decamer interfaces and allow detection of residues exchanges, which might give KLH2 the ability to form multidecamers *in vivo*.

Another unexplored field constitutes oxygenation-dependent conformational changes in the quaternary structure. Currently, in our laboratory, highly resolved cryoEM structures of NpH in oxygenated and deoxygenated state have been produced (Moeller *et al.*; these results will be published elsewhere), which might assist the understanding as to how, during oxygen binding, allosteric signals are transferred between the functional units.

## D SUMMARY

This work presents the currently highest resolution cryoEM structures (9 Å) of a cephalopod hemocyanin decamer (*Nautilus pompilius* hemocyanin, NpH) and a gastropod hemocyanin didecamer (keyhole limpet hemocyanin isoform 1, KLH1) and their complete molecular models, obtained by advanced molecular modelling and rigid body fitting.

Hemocyanins are blue copper proteins that transport oxygen in the hemolymph of many arthropods and molluscs. Molluscan hemocyanins are found either as decamers (five subunit dimers assembled as a hollow cylinder), or didecamers (face-to-face assembly of two decamers), which sometimes bind more decamers to form tubular multi-decamers. In cephalopods, only the single decamer is present. The polypeptide subunit contains seven functional units (termed FU-a to g), each of which binds one oxygen molecule. Sixty FUs form the cylinder wall (FUs a to f, each of the 10 subunits) and the remaining 10 copies of FU-g fold in to form an internal collar complex.

The present molecular model of NpH fully explains the intricate quaternary structure of the decamer. It has allowed for the first time identification of the subunit dimer, the pathway of the subunit and 15 types of molecular inter-FU interfaces. Many of these interfaces have amino acid constellations that might transfer allosteric interaction between FUs and give hints for the allosteric unit. Moreover, the potential N-glycan and calcium/magnesium binding sites have emerged.

In contrast to NpH, gastropod hemocyanins (including KLH) are didecamers and the collar contains an additional FU-type, termed FU-h, which is enlarged by an extension of ~100 amino acids. KLH is obtained from the Californian keyhole limpet *Megathura crenulata* and intensively applied, in immuno-logical research and clinics, as an immunoactivator and tumour vaccine carrier; it occurs in two isoforms termed KLH1 and KLH2. The present hybrid model of KLH1 allows for the first time detailed insight into the quaternary structure of this gastropod hemocyanin at a pseudo-atomic level. We have found that the wall-arc structure of the KLH1 decamer is very similar to that of NpH. The pathway of the subunit segment a-b-c-d-e-f-g is conserved and with respect to allostery, the surprising conservation of the 15 molecular interfaces that also exist in NpH suggests that basic structures are maintained in both molecules. Moreover, we were able to trace how the pathway continues from FU-g to FU-h and completely solve the old mystery of the gastropod hemocyanin collar. In addition, the present model shows, for the first time, the contact zones between the two decamers. And ultimately, it reveals the potential attachment sites for N-linked glycans that might be primarily responsible for the observed immunological effects.

Thus, the present results essentially solve many long debated questions on the architecture and function of these giant respiratory invertebrate proteins.

## E ZUSAMMENFASSUNG

Diese Arbeit präsentiert die bislang höchst aufgelösten KryoEM-Strukturen für ein Cephalopoden hämocyandin Dekamer (*Nautilus pompilus* Hämocyandin, NpH) und ein Gastropoden Hämocyandin Didekamer (keyhole limpet hemocyandin isoform 1). Durch die Methoden des "molecular modeling" und "rigid-body-fitting" wurde auch eine detaillierte Beschreibung beider Strukturen auf atomarem Niveau erstmalig möglich.

Hämocyandine sind kupferhaltige Sauerstoff-Transport-proteine die frei gelöst in Blut zahlreicher Arthropoden und Mollusken vorkommen. Allgemein sind Molluskenhämocyandine als Dekamere (Hohlzylinder aus 5 Untereinheiten-dimere) oder Didekamere (Zusammenlagerung von zwei Dekameren) zu finden. Durch Anlagerung weiterer Dekamere bilden sich teilweise tubuläre Multidekamere. Hämocyandine der Cephalopoden bestehen ausschließlich aus solitären Dekameren. In *Octopus und Nautilus* bestehen die 10 Untereinheiten aus 7 funktionellen Einheiten (FU-a bis FU-g), wobei jede FU ein Sauerstoffmolekül binden kann. FUs a-f bilden die Wand des ringförmigen Moleküls und 10 Kopien der FU-g bilden einen sogenannten „inneren Kragenkomplex“. Das im Rahmen dieser Arbeit erstellte molekulare Modell von NpH klärt die Struktur des Dekamers vollständig auf. Wir waren zum ersten Mal in der Lage das Untereinheiten-dimer, den Verlauf der Polypeptidkette und 15 unterschiedliche Kontaktstellen zwischen FUs zu identifizieren. Viele der inter-FU-Kontakte weisen Aminosäurenkonstellationen auf, die die Basis für die Übertragung allosterischer Wechselwirkungen zwischen FUs darstellen könnten und Hinweise für den Aufbau der allosterischen Einheit geben. Potentielle Bindungsstellen für N-glykosidische Zucker und bivalente Kationen wurden auch identifiziert.

Im Gegensatz zu NpH, kommen Gastropoden Hämocyandine (inkl. KLH) hauptsächlich als Didekamere vor und der Kragenkomplex wird in diesem Fall aus 2 FUs gebildet (Fu-g und FU-h). Die zusätzliche C'-terminale FU-h zeichnet sich durch eine spezielle Verlängerung von ~ 100 Aminosäuren aus. KLH stammt aus der kalifornische Schnecke *Megathura crenulata* und kommt seit mehreren Jahrzehnten als Immunostimulator in der immunologischen Grundlagenforschung und klinischen Anwendung zum Einsatz. KLH weist zwei Isoformen auf, KLH1 und KLH2. Das vorliegende Modell von KLH1 erlaubt die komplexe Architektur dieses riesigen Proteins in allen Details zu verstehen, sowie einen Vergleich zum dem NpH Dekamer auf atomare Ebene. Es wurde gefunden, dass das Untereinheitensegment a-b-c-d-e-f-g, sowie die äquivalenten Kontaktstellen zwischen FUs stark konserviert sind. Dies deutet darauf hin, dass in Bezug auf die Übertragung allosterischer Signale zwischen benachbarten FUs, grundlegende Mechanismen in beiden Molekülen beibehalten wurden. Weiterhin, konnten die Verbindungen zwischen den zwei Dekameren erstmalig identifiziert werden. Schließlich, wurde die Topologie der N-glykosidischen Zucker, welche für die immunologischen Eigenschaften von KLH1 von großer Bedeutung sind, auch aufgeklärt.

Somit leistet die vorliegende Arbeit einen wesentlichen Schritt zum Verständnis der Quartärstruktur und Funktion der Molluskenhämocyandine.

## F APPENDIX

### 1. Abbreviations

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
Å	Angstrom
cDNA	Copy- or complement-DNA
CCF	Cross correlation function
cryoEM	Cryo-electron microscopy
CTF	Contrast transfer function
dpi	Dots per inch
ELISA	Enzyme-linked Immunosorbent Assay
FEG	Field-emission gun
Fig.	Figure
FSC (F)	Fourier Shell Correlation (function)
FT	Fourier transform
FU	Functional unit
g	Gravitational acceleration
GB	Gigabyte
GHz	Gigahertz
HAC	Hierarchical ascedant classification
kDa	Kilo Dalton
kV	Kilo Volt
min.	Minutes
mM	(mili) Molar
MRA	Multi-reference-alignment
MSA	Multivariate statistical analysis
MWC	Monod-Wyman-Changeaux
nm	Nanometer
OD	Optical density

PCIe	Peripheral Component InterconnectInterface express
pdb	Protein data bank
PSF	Point spread function
RAM	Random access memory
RCF	Rotational correlation function
RMSD	Root mean squate deviation
SCF	Sinogram correlation function
S/N	Signal to noise ratio
TEM	Transmission electron microscop, microscopy
TIFF	Tagged Image File Format
VRML	Virtual Reality Modeling Language

## 2. Abbreviations of hemocyanins

HtH	<i>Haliotis tuberculata</i> hemocyanin
KLH	<i>Megathura crenulata</i> hemocyanin
NpH	<i>Nautilus pompilius</i> hemocyanin
OdH	Octopus dofleini Hemocyanin
RtH	<i>Rapana thomasi</i> Hemocyanin
SoH	<i>Sepia officinalis</i> Hemocyanin

### 3. Abbreviation-code of amino acids

<b>A</b>	Ala	Alanine	<b>M</b>	Met	Methionine
<b>C</b>	Cys	Cysteine	<b>N</b>	Asn	Asparagine
<b>D</b>	Asp	Aspartic acid	<b>P</b>	Pro	Proline
<b>E</b>	Glu	Glutamic acid	<b>Q</b>	Gln	Glutamine
<b>F</b>	Phe	Phenylalanine	<b>R</b>	Arg	Arginine
<b>G</b>	Gly	Glycine	<b>S</b>	Ser	Serine
<b>H</b>	His	Histidine	<b>T</b>	Thr	Threonine
<b>I</b>	Ile	Isoleucine	<b>V</b>	Val	Valine
<b>K</b>	Lys	Lysine	<b>W</b>	Trp	Tryptophane
<b>L</b>	Leu	Leucine	<b>Y</b>	Tyr	Tyrosine

### 4. Database accession numbers

The present cryoEM density maps of NpH and KLH1 have been deposited in the EMD-database (EMBL-EBI) under the accession numbers EMD-1434 and EMD-1569.

(<http://www.ebi.ac.uk/msd-srv/emsearch/>).

The NpH sequence is available in the EMBL/ Genebank databases under the accession number CAF03590. The KLH1 sequence is available under the accession number CAG28309.

(<http://www.ebi.ac.uk/ebisearch/>).

The crystal structures of OdH-h, KLH1-h, and Rth-e possess the RSCB Protein Bank ID codes 1JS8, 3EU2 and 1LNL.

(<http://www.rcsb.org/pdb/home/home.do>).



## G BIBLIOGRAPHY

- 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.
- Altenhein B, Markl J, Lieb B. (2002):** Gene structure and hemocyanin isoform HtH2 from the mollusc *Haliotis tuberculata* indicate early and late intron hot spots. *Gene*, 301(1-2):53-60.
- Barends T, Büchler K, Gatsogiannis C, Schlichting I, Markl J, Decker H & Jaenicke E. (2008):** Hybrid methods in the structure determination of extremely large proteins. *Biophys. J.*, 94:1571.
- Bergmann S, Lieb B, Ruth P, Markl J. (2006):** The hemocyanin from a living fossil, the cephalopod *Nautilus pompilius*: protein structure, gene organization, and evolution. *J. Mol. Evol.*, 62(3):362-74
- Bergmann S, Markl J, Lieb B. (2007):** The first complete cDNA sequence of the hemocyanin from a bivalve, the protobranch *Nucula nucleus*. *J. Mol. Evol.*, 64(5):500-10.
- Boisset N, Mouche F. (2000):** *Sepia officinalis* hemocyanin: A refined 3D structure from field emission gun cryoelectron microscopy. *J.Mol. Biol.*, 296(2):459-72.
- Bonaventura C, Bonaventura J, Miller KI, Van Holde KE. (1981):** Hemocyanin of the chambered nautilus: structure-function relationships. *Arch. Biochem. Biophys.*, 211(2):589-98.
- Brouwer M, Ryan M, Bonaventura J, Bonaventura C. (1978):** Functional and structural properties of *Murex fulvescens* hemocyanin: isolation of two different subunits required for reassociation of a molluscan hemocyanin. *Biochemistry*, 17(14):2810-5.
- Carrera MR, Ashley JA, Parsons LH, Wirsching P, Koob GF, Janda KD. (1995):** Suppression of psychoactive effects of cocaine by active immunization. *Nature*, 378(6558):727-30.
- Carrera MR, Ashley JA, Hoffman TZ, Isomura S, Wirsching P, Koob GF, Janda KD. (2004):** Investigations using immunization to attenuate the psychoactive effects of nicotine. *Bioorg. Med. Chem.*, 12(3):563-70.
- Cheng K, Koeck PJ, Elmlund H, Idakieva K, Parvanova K, Schwarz H, Ternström T, Hebert H. (2006):** *Rapana thomasiana* hemocyanin (RtH): comparison of the two isoforms, RtH1 and RtH2, at 19A and 16A resolution. *Micron*, 37(6):566-76.
- Collman JP, Boulatov R, Sunderland CJ, Fu L. (2004):** Functional analogues of cytochrome c oxidase, myoglobin, and hemoglobin. *Chemical reviews*, 104(2):561-88.
- Connelly PR, Gill SJ, Miller KI, Zhou G, van Holde KE. (1989):** Identical linkage and cooperativity of oxygen and carbon monoxide binding to *Octopus dofleini* hemocyanin. *Biochemistry*, 28(4):1835-43.
- Cuff ME, Miller KI, van Holde KE, Hendrickson WA. (1998):** Crystal structure of a functional unit from *Octopus* hemocyanin. *J. Mol. Biol.*, 278(4):855-70.
- Curtis JE, Hersh EM, Harris JE, McBride C, Freireich EJ. (1970):** The human primary immune response to keyhole limpet haemocyanin: interrelationships of delayed hypersensitivity, antibody response and in vitro blast transformation. *Clin. Exp. Immunol.*, 6(4):473-91.
- Decker H, Sterner R. (1990):** Hierarchies in the structure and function of oxygen-binding proteins. *Naturwissenschaften*, 77(12):561-8.
- Decker H, Schweikardt T, Nillius D, Salzbrunn U, Jaenicke E, Tuczek F. (2007a):** Similar enzyme activation and catalysis in hemocyanins and tyrosinases. *Gene*, 398(1-2):183-91.
- Decker H, Hellmann N, Jaenicke E, Lieb B, Meissner U, Markl J (2007b):** Minireview: Recent progress in hemocyanin research. *Integrative and Comparative Biology*, 47(4):631-644.

- DeRosier, D.J., and Klug, A. (1968):** Reconstruction of three dimensional structures from electron micrographs. *Nature*, 217:130-134.
- Dewilde S, Van Hauwaert ML, Vinogradov S, Vierstraete A, Vanfleteren J, Moens L. (2001):** Protein and gene structure of a chlorocruorin chain of *Eudistylia vancouverii*. *Biochem. Biophys. Res. Commun.*, 281(1):18-24.
- Drexel R, Siegmund S, Schneider HJ, Linzen B, Gielens C, Preaux G, Lontie R, Kellermann J, Lottspeich F. (1987):** Complete amino-acid sequence of a functional unit from a molluscan hemocyanin (*Helix pomatia*). *Biol. Chem. Hoppe Seyler*, 368(6):617-35
- Di Cera E. (1990):** *Il Nuovo Cimento D*, 12:61-68.
- Dube P, Tavares P, Lurz R, van Heel M. (1993):** The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.*, 12(4):1303-9.
- Dunbrack RL Jr. (2002):** Rotamer libraries in the 21st century. *Curr. Opin. Struct. Biol.*, 12(4):431-40.
- 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.
- Flamm J, Bucher A, Höltl W, Albrecht W. (1990):** Recurrent superficial transitional cell carcinoma of the bladder: adjuvant topical chemotherapy versus immunotherapy. A prospective randomized trial. *J. Urol.*, 144(2 Pt 1):260-3.
- Foster D, Davie EW. (1984):** Characterization of a cDNA coding for human protein C. *Proc. Natl. Acad. Sci. USA.*, 81 (15): 4766-70.
- Gaykema WP, Hol, WG, Vereijken, JM, Soeter, NM., Bak, HJ, Beintema, JJ. (1984):** 3.2 Angstroms Structure of the Copper-Containing, Oxygen-Carrying Protein Panulirus Interruptus Haemocyanin. *Nature*, 309:23.
- Gebauer W, Harris JR, Heid H, Süling M, Hillenbrand R, Söhngen S, Wegener-Strake A, Markl J. (1994):** Quaternary structure, subunits and domain patterns of two discrete forms of keyhole limpet hemocyanin: KLH1 and KLH2. *Zoology*, 98, 51-68.
- Gebauer W, Stoeva S, Voelter W, Dainese E, Salvato B, Beltramini M, Markl J. (1999):** Hemocyanin subunit organization of the gastropod *Rapana thomasiana*. *Arch. Biochem. Biophys.*, 372(1):128-34.
- Gebauer W, Robin Harris J, Markl J. (2002):** Topology of the 10 subunits within the decamer of KLH, the hemocyanin of the marine gastropod *Megathura crenulata*. *J. Struct. Biol.*, 139(3):153-9.
- Genova-Kalou P, Dundarova D, Idakieva K, Mohammed A, Dundarov S, Argirova R. (2008):** Anti-herpes effect of hemocyanin derived from the mollusk *Rapana thomasiana*. *Z. Naturforsch. [C]*, 63(5-6):429-34.
- Gielens C, Bosman F, Préaux G, Lontie R. (1983):** Structural studies by limited proteolysis of the haemocyanin of *Sepia officinalis*. *Life Chem. Rep.*, Suppl. 1, 121-124.
- Gielens C, Idakieva K, De Maeyer M, Van den Bergh V, Siddiqui NI, Compennolle F. (2007):** Conformational stabilization at the active site of molluscan (*Rapana thomasiana*) hemocyanin by a cysteine-histidine thioether bridge A study by mass spectrometry and molecular modeling. *Peptides*, 28(4):790-7

- Gielens C, De Geest N, Xin XQ, Devreese B, Van Beeumen J, Préaux G. (1997):** Evidence for a cysteine-histidine thioether bridge in functional units of molluscan haemocyanins and location of the disulfide bridges in functional units d and g of the betaC-haemocyanin of *Helix pomatia*. *Eur. J. Biochem.*, 248(3):879-88.
- Goddard TD, Huang CC, Ferrin TE. (2007):** Visualizing density maps with UCSF Chimera. *J. Struct. Biol.*, 157(1):281-7.
- Grzych JM, Dissous C, Capron M, Torres S, Lambert PH, Capron A. (1987):** *Schistosoma mansoni* shares a protective carbohydrate epitope with keyhole limpet hemocyanin. *J. Exp. Med.*, 165(3):865-78.
- Guex N, Peitsch MC. (1997):** SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* 18:2714-2723.
- 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.
- Harris JR, Chen, S. & Markl, J. (1994):** Keyhole limpet hemocyanin (KLH): negative staining in the presence of trehalose and preliminary cryoelectron microscopy of unstained frozen-hydrated specimens. *Proc. Int. Conf. Electron. Mater.*, 13:557-558.
- Harris JR, Gebauer W, Söhngen S, Markl J. (1995):** Keyhole limpet hemocyanin (KLH): purification of intact KLH1 through selective dissociation of KLH2. *Micron*, 26:201-212.
- Harris JR, Gebauer W, Söhngen SM, Nermut MV, Markl J. (1997):** Keyhole limpet hemocyanin (KLH), II: Characteristic reassociation properties of purified KLH1 and KLH2. *Micron*, 28(1):43-56.
- Harris JR, Markl J. (1999):** Keyhole limpet hemocyanin (KLH): a biomedical review. *Micron*, 30(6):597-623.
- Harris JR, Scheffler D, Gebauer W, Lehnert R, Markl J. (2000):** *Haliotis tuberculata* hemocyanin (HtH): analysis of oligomeric stability of HtH1 and HtH2, and comparison with keyhole limpet hemocyanin KLH1 and KLH2. *Micron*, 31(6):613-22.
- Hartmann, H., Bongers, A. & Decker, H. (2004):** Small angle X-ray scattering-based three-dimensional reconstruction of the immunogen KLH1 reveals different oxygen-dependent conformations. *J. Biol. Chem.*, 279:2841-2845.
- Harris JR, Meissner U, Gebauer W, Markl J. (2004):** 3D reconstruction of the hemocyanin subunit dimer from the chiton *Acanthochiton fascicularis*. *Micron*, 35(1-2):23-6.
- Herscovitz HB, Harold WW, Stavitsky AB. (1972):** Immunochemical and immunogenic properties of a purified keyhole limpet haemocyanin. *Immunology*, 22:51-61.
- Herskovits TT. (1988):** Recent aspects of the subunit organization and dissociation of hemocyanins. *Comp. Biochem. Physiol. B*, 91(4):597-611
- Herskovits TT, Cousins CJ, Hamilton MG. (1991a):** The stabilizing influence of divalent ions and Na<sup>+</sup> on the di-decameric structure of *Yoldia limatula* hemocyanin. *Biochim. Biophys. Acta*, 1076(1):71-8.
- Herskovits TT, Hamilton MG. (1991b):** Higher order assemblies of molluscan hemocyanins. *Comp. Biochem. Physiol. B.*, 99(1):19-34.
- Heymann JB, Chagoyen M, Belnap DM. (2005):** Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *J. Struct. Biol.*, 151(2):196-207. Erratum in: *J. Struct. Biol.* 2006, 153(3):312.
- Higgins DG and Sharp PM. (1988):** CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73:237-244.

- Holmes MA, Le Trong I, Turley S, Sieker LC, Stenkamp RE. (1991):** Structures of deoxy and oxy hemerythrin at 2.0 Å resolution. *J. Mol. Biol.*, 218(3):583-93.
- Hubbard TJ, Blundell TL. (1987):** Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. *Protein Eng.*, 1(3):159-71.
- Idakieva K, Stoeva S, Voelter W, Gielens C. (2004):** Glycosylation of *Rapana thomasiana* hemocyanin. Comparison with other prosobranch (gastropod) hemocyanins. *Comp. Biochem. Physiol. B Biochem.*, 138(3):221-8.
- Jiang W, Baker ML, Jakana J, Weigele PR, King J, Chiu W. (2008):** Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature*, 451(7182):1130-4.
- Jurincic CD, Engelmann U, Gasch J, Klippel KF. (1988):** Immunotherapy in bladder cancer with keyhole-limpet hemocyanin: a randomized study. *J. Urol.*, 139(4):723-6.
- Kantelhardt SR, Wuhrer M, Dennis RD, Doenhoff MJ, Bickle Q, Geyer R. (2002):** Fuc(alpha1->3)GalNAc-: the major antigenic motif of *Schistosoma mansoni* glycolipids implicated in infection sera and keyhole-limpet haemocyanin cross-reactivity. *Biochem. J.*, 366(Pt 1):217-23.
- Keller H, Lieb B, Altenhein B, Gebauer D, Richter S, Stricker S, Markl J. (1999):** Abalone (*Haliotis tuberculata*) hemocyanin type 1 (HtH1). Organization of the approximately 400 kDa subunit, and amino acid sequence of its functional units f, g and h. *Eur. J. Biochem.*, 264(1):27-38.
- Klarman A, Shaklai N, Daniel E. (1975):** Oxygen binding by hemocyanin from *Levantina hierosolima*. I. Exclusion of subunit interactions as a basis for cooperativity. *Biochemistry*, 14(1):102-4.
- Klippel KF (1991):** Weniger Blasenkrebs-Rezidive durch Hämocyanin. *Immuntherapie des Krebses*, 43(7):29-31.
- Ko AI, Harn DA. (1987):** Characterization of protective and non-protective surface membrane carbohydrate epitopes of *Schistosoma mansoni*. *Mem. Inst. Oswaldo Cruz.*, 82 Suppl 4:115-9.
- Kurokawa T, Wuhrer M, Lochnit G, Geyer H, Markl J, Geyer R. (2002):** Hemocyanin from the keyhole limpet *Megathura crenulata* (KLH) carries a novel type of N-glycans with Gal(beta1-6)Man-motifs. *Eur. J. Biochem.*, 269(22):5459-73.
- Lambert O, Boisset N, Penczek P, Lamy J, Taveau JC, Frank J, Lamy JN. (1994a):** Quaternary structure of *Octopus vulgaris* hemocyanin. Three-dimensional reconstruction from frozen-hydrated specimens and intramolecular location of functional units Ove and Ovb. *J. Mol. Biol.*, 238(1):75-87.
- Lambert O, Boisset N, Taveau JC, Lamy JN. (1994b):** Three-dimensional reconstruction from a frozen-hydrated specimen of the chiton *Lepidochiton* sp. hemocyanin. *J. Mol. Biol.*, 244(5):640-7.
- Lambert O, Taveau JC, Boisset N, Lamy JN. (1995a):** Three-dimensional reconstruction of the hemocyanin of the protobranch bivalve mollusc *Nucula hanleyi* from frozen-hydrated specimens. *Arch. Biochem. Biophys.*, 319(1):231-43.
- Lambert O, Boisset N, Taveau JC, Préaux G, Lamy JN. (1995b):** Three-dimensional reconstruction of the alpha D and beta C-hemocyanins of *Helix pomatia* from frozen-hydrated specimens. *J. Mol. Biol.*, 248(2):431-48
- Lambert O, Boisset N, Taveau JC, Lamy JN. (1995c):** Three-dimensional reconstruction of *Sepia officinalis* hemocyanin from frozen-hydrated specimens. *Arch. Biochem. Biophys.*, 316(2):950-9.
- Lamy J, Gielens C, Lambert O, Taveau JC, Motta G, Loncke P, De Geest N, Préaux G, Lamy JN. (1993):** Further approaches to the quaternary structure of octopus hemocyanin: a model based on immunoelectron microscopy and image processing. *Arch. Biochem. Biophys.*, 305(1):17-29
- Lamy JN, Green BN, Toulmond A, Wall JS, Weber RE, Vinogradov SN. (1996):** Giant Hexagonal Bilayer Hemoglobins. *Chemical reviews*, 96(8):3113-3124.
- Lamy J, You V, Taveau JC, Boisset N, Lamy JN. (1998):** Intramolecular localization of the functional units of *Sepia officinalis* hemocyanin by immunoelectron microscopy. *J. Mol. Biol.*, 284(4):1051-74.

- Lang WH. (1988):** cDNA cloning of the Octopus dofleini hemocyanin: sequence of the carboxyl-terminal domain. *Biochemistry*, 27(19):7276-82.
- Lang WH, van Holde KE. (1991):** Cloning and sequencing of Octopus dofleini hemocyanin cDNA: derived sequences of functional units Ode and Odf. *Proc. Natl. Acad. Sci. USA*, 88(1):244-8.
- 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.
- Laskowski RA, Moss DS, Thornton JM. (1993):** Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.*, 231(4):1049-67.
- Launay G, Simonson T. (2008):** Homology modelling of protein-protein complexes: a simple method and its possibilities and limitations. *BMC Bioinformatics*, 9:427
- 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.
- Lee ST, Jiang YF, Park KU, Woo AF, Neelapu SS. (2007):** BiovaxID: a personalized therapeutic cancer vaccine for non-Hodgkin's lymphoma. *Expert Opin. Biol. Ther.*, 7(1):113-22.
- Lieb B, Altenhein B, Lehnert R, Gebauer W, Markl J. (1999):** Subunit organization of the abalone *Haliotis tuberculata* hemocyanin type 2 (HtH2), and the cDNA sequence encoding its functional units d, e, f, g and h.. *Eur. J. Biochem.*, 265(1):134-44.
- Lieb B, Altenhein B, Markl J. (2000):** The sequence of a gastropod hemocyanin (HtH1 from *Haliotis tuberculata*). *J. Biol. Chem.*, 275(8):5675-81.
- Lieb B, Altenhein B, Markl J, Vincent A, van Olden E, van Holde KE, Miller KI. (2001):** Structures of two molluscan hemocyanin genes: significance for gene evolution. *Proc. Natl. Acad. Sci. USA*, 98(8):4546-51.
- Lieb B, Markl J. (2004):** Evolution of molluscan hemocyanins as deduced from DNA sequencing. *Micron*, 35(1-2):117-9.
- Linzen B. (1989):** Blue blood: structure and evolution of hemocyanin. *Naturwissenschaften*, 76(5):206-11.
- Llorca O. (2005):** Introduction to 3D reconstruction of macromolecules using single particle electron microscopy. *Acta Pharmacol. Sin.*, 26(10):1153-64.
- Loehr JS, Lammers PJ, Brimhall B, Hermodson MA. (1978):** Amino acid sequence of hemerythrin from *Themiste dyscritum*. *J. Biol. Chem.*, 253(16):5726-31.
- Loewe, R. (1978):** Hemocyanin in spiders: V. Fluorimetric recording of oxygen binding curves, and its application to the analysis of allosteric interactions in *Eurypelma californicum* hemocyanin. *J. Comp. Physiol.*, 128B, 161-168.
- Loncke P, Vanderzande M, Gielens C, Préaux G. (1990):** Identification of the missing functional unit in *Octopus vulgaris* haemocyanin from a comparison with *Sepia officinalis* haemocyanin. *Invertebrate Dioxxygen Carriers* (G. Préaux and R. Lontie, Editors), Leuven University Press, Louvain, 295-298.
- Lovell SC, Word JM, Richardson JS, Richardson DC. (2000):** The penultimate rotamer library. *Proteins*, 40(3):389-408
- 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 single-particle reconstructions. *J. Struct. Biol.*, 128(1):82-97.

- Ludtke SJ, Baker ML, Chen DH, Song JL, Chuang DT, Chiu W. (2008):** De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure*, 16(3):441-8
- Markl J, Nour el Din M, Winter-Simanowski S, Simanowski UA. (1991):** Specific IgG activity of sera from Egyptian schistosomiasis patients to keyhole limpet hemocyanin (KLH). *Naturwissenschaften*, 78(1):30-1.
- Markl J, Decker H. (1992):** Molecular structure of the arthropod hemocyanins. *Adv. In Comp. & Envir. Physiol.*, 13:325-376.
- Markl J, Lieb B, Gebauer W, Altenhein B, Meissner U, Harris JR. (2001):** Marine tumor vaccine carriers: structure of the molluscan hemocyanins KLH and HtH. *J. Cancer Res. Clin. Oncol.*, 127 Suppl. 2:R3-9.
- Martin AG, Depoix F, Stohr M, Meissner U, Hagner-Holler S, Hammouti K, Burmester T, Heyd J, Wriggers W, Markl J. (2007):** Limulus polyphemus hemocyanin: 10 A cryo-EM structure, sequence analysis, molecular modelling and rigid-body fitting reveal the interfaces between the eight hexamers. *J. Mol. Biol.*, 366(4):1332-50.
- McFadden, DW, Riggs DR, Jackson, BJ, Vona-Davis L. (2003):** Keyhole limpet hemocyanin, a novel immune stimulant with promising anticancer activity in Barrett's esophageal adenocarcinoma. *Am. J. Surg.*, 186(5):552-5.
- Meissner U, Dube P, Harris JR, Stark H, Markl J. (2000):** Structure of a molluscan hemocyanin didecamer (HtH1 from *Haliotis tuberculata*) at 12 A resolution by cryoelectron microscopy. *J. Mol. Biol.*, 298(1):21-34.
- Mellema JE, Klug A.(1972):** Quaternary structure of gastropod haemocyanin. *Nature*, 239(5368):146-50.
- Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. (2006):** Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*, 7:339.
- Meyer D, Anderson DE, Gardner MB, Torres, J. (1998):** Hypervariable epitope constructs representing variability in envelope glycoprotein of SIV induce a broad humoral immune response in rabbits and rhesus macaques. *AIDS Res. Hum. Retroviruses*, 14:751-760.
- Miller KI. (1985):** Oxygen equilibria of Octopus dofleini hemocyanin. *Biochemistry*, 24(17):4582-6.
- Miller KI, Cuff ME, Lang WF, Varga-Weisz P, Field KG, van Holde KE. (1998):** Sequence of the Octopus dofleini hemocyanin subunit: structural and evolutionary implications. *J. Mol. Biol.*, 278(4):827-42.
- Mindell JA, Grigorieff N. (2003):** Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.*, 142(3):334-47.
- Moltedo B, Faunes F, Haussmann D, De Ioannes P, De Ioannes AE, Puente J, Becker MI. (2006):** Immunotherapeutic effect of Concholepas hemocyanin in the murine bladder cancer model: evidence for conserved antitumor properties among hemocyanins. *J. Urol.*, 176(6 Pt 1):2690-5.
- Mouche F, Boisset N, Lamy J, Zal F, Lamy JN. (1999):** Structural comparison of cephalopod hemocyanins: phylogenetic significance. *J. Struct. Biol.*, 127(3):199-212.
- Mouche F, Zhu Y, Pulokas J, Potter CS, Carragher B. (2003):** Automated three-dimensional reconstruction of keyhole limpet hemocyanin type 1. *J. Struct. Biol.*, 144(3):301-12.
- Naylor PH, Sztejn MB, Wada S, Maurer S, Holterman D, Kirkley JE, Naylor CW, Zook BC, Hitzelberg RA, Gibbs CJ Jr, et al. (1991):** Preclinical and clinical studies on immunogenicity and safety of the HIV-1 p17-based synthetic peptide AIDS vaccine—HGP-30-KLH. *Int. J. Immunopharmacol.*, 117-27.
- Nicholas KB, Nicholas HB Jr. and Deerfield DW II. (1997):** GeneDoc: Analysis and Visualization of Genetic Variation, *EMBNEW.NEWS* 4:14
- Oliva H, Moltedo B, De Ioannes P, Faunes F, De Ioannes AE, Becker MI. (2002):** Monoclonal antibodies to molluscan hemocyanin from *Concholepas concholepas* demonstrate common and specific epitopes among subunits. *Hybrid Hybridomics*, 21(5):365-74.

- Olsson CA, Chute R, Rao CN. (1974):** Immunologic reduction of bladder cancer recurrence rate. *J. Urol.*, 111(2):173-6
- Orlova EV, van Heel M. (1994):** Angular reconstitution of macromolecules with arbitrary point group symmetry. *Proceedings ICEM 13-Paris*, Vol. 1, 507-508.
- 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 Å resolution by electron cryomicroscopy and angular reconstitution. *J. Mol. Biol.*, 271(3):417-37
- Orlova EV, Saibil HR. (2004):** Structure determination of macromolecular assemblies by single-particle analysis of cryo-electron micrographs. *Curr. Opin. Struct. Biol.*, 14(5):584-90.
- 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. *Ultra-microscopy*, 53 (3): 251-70.
- Perbandt M, Guthöhrlein EW, Rypniewski W, Idakieva K, Stoeva S, Voelter W, Genov N, Betzel C. (2003):** The structure of a functional unit from the wall of a gastropod hemocyanin offers a possible mechanism for cooperativity. *Biochemistry*, 42(21):6341-6.
- Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC. (1960):** Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature*, 185: 416-22.
- 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.
- 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.
- Raghava GPS (2002):** APSSP2 : A combination method for protein secondary structure prediction based on neural network and example based learning. *CASP*, A-132.
- Riggs DR, Jackson B, Vona-Davis L, McFadden D. (2002):** In vitro anticancer effects of a novel immunostimulant: keyhole limpet hemocyanin. *J. Surg. Res.*, 108(2):279-84.
- Riggs DR, Jackson BJ, Vona-Davis L, Nigam A, McFadden DW. (2005):** In vitro effects of keyhole limpet hemocyanin in breast and pancreatic cancer in regards to cell growth, cytokine production, and apoptosis. *Am. J. Surg.*, 189(6):680-4.
- Rizvi I, Riggs DR, Jackson BJ, McFadden DW (2007):** Keyhole limpet hemocyanin: an effective adjunct against melanoma in vivo. *Am. J. Surg.*, 194(5):628-32.
- 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.
- Ruprecht J, Nield J. (2001):** Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction. *Prog. Biophys. Mol. Biol.*, 75(3):121-64.
- 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.
- Saxton WO, Frank J. (1977):** Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy*, 2(2-3):219-27.
- Saunders WB and Ward PD (1987):** Nautilus, the biology and paleobiology of a living fossil. *Topics in Geobiology*, 6:37-162.
- Schatz M. (1992):** Invariante Klassifizierung elektronenmikroskopischer Aufnahmen von eiseingebetteten biologischen Makromolekülen. Dissertation. Freie Universität, Berlin.
- Schatz M, Orlova EV, Dube P, Jager J, van Heel M. (1995):** Structure of Lumbricus terrestris hemoglobin at 30 Å resolution determined using angular reconstitution. *J. Struct. Biol.*, 114(1):28-40.
- Senozan NM, Landrum J, Bonaventura J and Bonaventura C. (1981):** Hemocyanin of the giant keyhole limpet *Megathura crenulata*. In: J. Lamy and J. Lamy, Editors, *Invertebrate Oxygen Binding Proteins*, Dekker, New York (1981), 703–717.
- Siezen RJ, van Bruggen EF. (1974):** Structure and properties of hemocyanins. XII. Electron microscopy of dissociation products of *Helix pomatia* alpha-hemocyanin: quaternary structure. *J. Mol. Biol.*, 90(1):77-89
- Siezen RJ, van Driel R. (1974):** Structure and properties of hemocyanins. XIII. Dissociation of *Helix pomatia* alpha-hemocyanin at alkaline pH. *J. Mol. Biol.*, 90(1):91-102
- Söhngen SM, Stahlmann A, Harris JR, Müller SA, Engel A, Markl J. (1997):** Mass determination, subunit organization and control of oligomerization states of keyhole limpet hemocyanin (KLH). *Eur. J. Biochem.*, 248(2):602-14.
- Somasundar P, Riggs DR, Jackson BJ, McFadden DW. (2005):** Inhibition of melanoma growth by hemocyanin occurs via early apoptotic pathways. *Am. J. Surg.*, 190(5):713-6.
- Streit K, Jackson D, Degnan BM, Lieb B. (2005):** Developmental expression of two *Haliothis asinina* hemocyanin isoforms. *Differentiation*, 73(7):341-9.
- Stoeva S, Schütz J, Gebauer W, Hundsdärfer T, Manz C, Markl J, Voelter W (1999):** Primary structure and unusual carbohydrate moiety of functional unit 2-c of keyhole limpet hemocyanin (KLH). *Biochim. Biophys. Acta*, 1435(1-2):94-109.
- Suhre K, Sanejouand YH. (2004):** ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.*, 32 (Web Server issue):W610-4.
- Swanson MA, Schwartz RS. (1967):** Immunosuppressive therapy. The relation between clinical response and immunologic competence. *N. Engl. J. Med.*, 277(4):163-70.
- Taiwo FA. (1992):** Oxygen-binding properties of hemocyanin from the giant African snail *Archachatina maginata*: pH and calcium ion effects. *Comp. Biochem. Physiol.*, 102A:225–227.
- 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 low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.*, 337(4):985-99.
- Tchorbanov A, Idakieva K, Mihaylova N, Doumanova L. (2008):** Modulation of the immune response using *Rapana thomasiana* hemocyanin. *Int. Immunopharmacol.*, 8(7):1033-8.
- Terwilliger NB. (1998):** Functional adaptations of oxygen-transport proteins. *J. Exp. Biol.*, 201(Pt 8):1085-98.

- 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.
- Thon F. (1966):** Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Z. Naturforschung*, 21a:476-478.
- Thuman-Commike PA (2001):** Single particle macromolecular structure determination via electron microscopy. *FEBS Lett.*, 505(2):199-205.
- Topham R, Tesh S, Westcott A, Cole G, Mercatante D, Kaufman G, Bonaventura C (1999):** Disulfide Bond Reduction: A Powerful, Chemical Probe for the study of Structure-Function Relationships in the Hemocyanins. *Arch. Biochem. Biophys.*, 369 (2):261-266
- Tzianabos, AO, (2000):** Polysaccharide immunomodulators as therapeutic agents: structural aspects and biological function. *Clin. Microbiol. Rev.*, 13(4):523-33.
- Yu X, Jin L, Zhou ZH.(2008):** 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, 453(7193):415-9.
- Vagin A, Teplyakov A. (1998):** A translation-function approach for heavy-atom location in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 54(Pt 3):400-2.
- Vagin A, Teplyakov A. (2000):** An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol. Crystallogr.*, 56(Pt 12):1622-4.
- Van Bruggen EF, Wiebenga EH., Gruber M. (1960):** Negative-staining electron microscopy of proteins at pH values below their isoelectric points. Its application to hemocyanin. *Biochim. Biophys. Acta*, 42: 171-172.
- 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, & Keegstra W. (1981):** IMAGIC: a fast, flexible and friendly image analysis software system. *Ultramicroscopy*, 7:113-130.
- 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 Heel M, Harauz G, Orlova, E.V., Schmidt R & Schatz M. (1996):** A new generation of the IMAGIC image processing system. *J. struct. Biol.*, 116,17-24.
- 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, Schatz M. (2005):** Fourier shell correlation threshold criteria. *J Struct Biol.*, 151(3):250-62.
- Van Holde KE. (1967):** Physical studies of hemocyanins. 3. Circular dichroism and absorption spectra. *Biochemistry*, 6(1):93-9.
- Van Holde KE, Miller KI. (1985):** Association-dissociation equilibria of Octopus hemocyanin. *Biochemistry*, 24(17):4577-82.
- Van Holde KE, Miller K, Schabtach E, Libertini L. (1991):** Assembly of Octopus dofleini hemocyanin. A study of the kinetics by sedimentation, light scattering and electron microscopy. *J. Mol. Biol.*, 217(2):307-21.
- Van Holde KE, Miller KI, Lang WH. (1992):** Molluscan hemocyanins: structure and function. *Adv. Comp. Environ. Physiol*, 13:257-300.

- Van Holde KE, Miller KI. (1995):** Hemocyanins. *Adv. Protein Chem.*, 47:1-81.
- Van Holde KE, Miller KI, van Olden E. (2000):** Allostery in very large molecular assemblies. *Biophys. Chem.*, 86(2-3):165-72.
- Van Kuik, JA, Kamerling JP, Vliegthardt, JFG. (1990):** Carbohydrate analysis of hemocyanins. Preaux, G., Lontie, R. (Eds.). *Invertebrate Oxygen Carriers*, Leuven University Press, Leuven, 157-163.
- Volbeda A, Hol WG. (1989):** Crystal structure of hexameric haemocyanin from *Panulirus interruptus* refined at 3.2 Å resolution. *J. Mol. Biol.*, 209(2):249-79.
- Vona-Davis L, Vincent T, Zulfiqar S, Jackson B, Riggs D, McFadden DW (2004):** Proteomic analysis of SEG-1 human Barrett's-associated esophageal adenocarcinoma cells treated with keyhole limpet hemocyanin. *J. Gastrointest. Surg.*, 8(8):1018-23.
- Wade RH, (1992):** A brief look at imaging and contrast transfer. *Ultramicroscopy*, 46:145–156.
- Ward, J.H. (1982)::** Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, 58:236–244.
- Weigle WO. (1964):** Immunochemical Properties of hemocyanin. *Immunochemistry*, 1:295-302.
- Westermann B, Beck-Schildwächter I, Beuerlein K, Kaleta EF, Schipp R. (2004):** Shell growth and chamber formation of aquarium-reared *Nautilus pompilius* (Mollusca, Cephalopoda) by X-ray analysis. *J. Exp. Zool. A Comp. Exp. Biol.*, 301(12):930-7.
- Wichertjes T, Gielens C, Loncke P., Préaux G., van Bruggen EF (1990):** Comparative electron microscopy of the haemocyanins of the cephalopods *Sepia officinalis* and *Octopus vulgaris*. *Invertebrate Dioxigen Carriers* (G. Préaux and R. Lontie, Editors), Leuven University Press, Louvain , 289–294.
- Wirguin I, Suturkova-Milosević L, Briani C, Latov N. (1995):** Keyhole limpet hemocyanin contains Gal(beta 1-3)-GalNAc determinants that are cross-reactive with the T antigen. *Cancer Immunol. Immunother.*, 40(5):307-10.
- Wriggers W, Milligan RA, McCammon JA. (1999):** Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.*, 125(2-3):185-95.
- Zhang X, Settembre E, Xu C, Dormitzer PR, Bellamy R, Harrison SC, Grigorieff N. (2008):** Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA.*, 105(6):1867-72.
- Zhou ZH. (2008):** Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr. Opin. Struct. Biol.*, 18(2):218-28.
- Zolla L, Kuiper HA, Vecchini P, Antonini E, Brunori M. (1978):** Dissociation and oxygen-binding behaviour of beta-hemocyanin from *Helix pomatia*. *Eur. J. Biochem.*, 87(3):467-73.

## H PUBLICATION LIST

Meissner U, Gatsogiannis C, Moeller A, Depoix F, Harris JR, Markl J. (2007): Comparative 11 Å structure of two molluscan hemocyanins from 3D cryo electron microscopy. *Micron*, 38 (7): 754-765

Gatsogiannis C, Moeller A, Depoix F, Meissner U, Markl J. (2007): Nautilus pompilius hemocyanin: 9 Å cryo-EM structure and molecular model reveal the subunit pathway and the interfaces between the 70 functional units. *J. Mol. Biol.*, 374(2): 465-486

Moeller A, Gatsogiannis C, Depoix F, Meissner U, Markl J. (2007): 8 Å Cryo-TEM and Single Particle Analysis of Nautilus Pompilius Hemocyanin under two states of oxygenation. *Microscopy and Microanalysis*, 13 (Suppl. 03) : 164-165

Gatsogiannis C, Büchler K, Depoix F, Markl J. (2008): 9 Å cryoEM-structure and molecular model of a gastropod hemocyanin didecamer (KLH1) reveals the architecture of the assymmetric collar. *EMC 2008*; Vol. 3: Life Science: 21-22.

Moeller A, Gatsogiannis C, Depoix F, Meissner U & Markl J (2008): Allosterism of *Nautilus pompilius* hemocyanin as deduced from 8 Å cryo-EM structures obtained under oxy and deoxy conditions. *EMC 2008*, Vol. 3: Life Science: 39-40.

Barends T, Büchler K, Gatsogiannis C, Schlichting I, Markl J, Decker H & Jaenicke E (2008): Hybrid methods in the structure determination of extremely large proteins. *Biophys. J.*, 94:1571.

Gatsogiannis C, Markl J. (2009): Keyhole limpet hemocyanin: 9-Å CryoEM structure and molecular model of the KLH1 didecamer reveal the interfaces and intricate topology of the 160 functional units. *J Mol Biol.*, 385(3): 963-83.

*I was also involved in one more project that was not presented above, but the preliminary results of this study can be found in:*

Kuehne C, Gatsogiannis C, Klamp T, Sahin U, Meissner U. (2007): 3D cryoelectron microscopy, molecular modelling and structural fitting with recombinant expressed virus like particles as part of drug design. *Microscopy and Microanalysis*, 13 (Suppl. 03): 162-163.

**Poster**

Gatsogiannis C, Meissner U, Depoix F, Moeller A, Markl J. (2005): 3D-reconstruction of Nautilus pompilius hemocyanin based on cryoelectron microscopy. 98 Jahresversammlung der Deutschen Zoologischen Gessellschaft.

Moeller A, Gatsogiannis C, Depoix F, Meissner U, Markl J. (2006): Cryoelectron microscopy and single particle analysis of Nautilus pompilius hemocyanin under different oxygenation states. Molluscan Forum (2006) of the malacological society of London.

Gatsogiannis C, Moeller A, Depoix F, Meissner U, Markl J (2006): 3D-structure of Nautilus pompilius hemocyanin by cryoTEM, Homology Modelling and Automated Fitting. Horizons in Molecular Biology, 3<sup>rd</sup> International PhD student symposium.

Meissner U, Depoix F, Martin AG, Stohr M, Gatsogiannis C, Moeller A and Jürgen Markl (2006): 3D-electron microscopy of arthropod and mollusc hemocyanins at 10 Å resolution. First International Congress of Respiratory Biology ICRB.

Gatsogiannis C, Moeller A, Depoix F, Meissner U, Markl J (2006): High resolution quaternary structure of Nautilus pomilius hemocyanin by cryoelectron microscopy, homology modelling and automated fitting. Deutsche Gessellschaft für Biophysik. Jahrestagung 2006.

Moeller A, Gatsogiannis C, Meissner U, Depoix F, Markl J (2006): Observing „Respiration” at the molecular level. Horizons in Molecular Biology. 3<sup>rd</sup> International PhD Student symposium.

Gatsogiannis C, Moeller A, Depoix F, Meissner U, Markl J (2007): 9 Å cryo-EM structures and molecular models of molluscan hemocyanin at different functional states. Horizons in Molecular Biology. 4<sup>th</sup> international PhD student symposium

Gatsogiannis C, Moeller A, Depoix F, Meissner U, Markl J (2007): Nautilus pompilius hemocyanin: cryoEM structure, molecular modelling and fitting of substructures. Deutsche Zoologische Gessellschaft. 100<sup>th</sup> Annual Meeting.

Gatsogiannis C, Büchler K, Depoix F & Markl J. (2008): 9 Å cryo-EM structure and molecular model of a gastropod hemocyanin didecamer (KLH1) reveals the architecture of the asymmetric collar. Gordon Conference: Three Dimensional electron microscopy. Technical advances and recent breakthroughs in 3DEM.

# I Danksagung – Acknowledgement - Ευχαριστία

Ich möchte besonders Herrn [REDACTED] [REDACTED] für die Möglichkeit, diese Arbeit durchzuführen, sowie für eine optimale Betreuung und ständige, aktive Unterstützung bei der Arbeit danken.

Vielen Dank an Herrn [REDACTED] für die Zweitkorrektur meiner Arbeit.

[REDACTED] gebührt ein herzlicher Dank für die Bereitstellung zahlreicher Cryo-EM-Bilder.

Herrn [REDACTED] und Herrn [REDACTED] möchte ich für die Betreuung im Rahmen des Graduiertenkollegs „Antigenspezifische Immuntherapie“ danken.

[REDACTED] danke ich für die Bereitstellung des aufgereinigtes Hämocyanins der Schlüssellochschncke *Megathura crenulata* (KLH1).

[REDACTED], [REDACTED], [REDACTED] danke ich für die Bereitstellung des Hämocyanins von *Nautilus pompilius* am Beginn meiner Doktorarbeit

Ein herzlicher Dank geht an [REDACTED] für die hervorragende Zusammenarbeit.

[REDACTED] danke ich für für wertvolle Beiträge zur Interpretation der Cupredoxin-Domäne von KLH1-h.

I would like to thank [REDACTED] for critical reading of the manuscripts and my thesis and language corrections.

[REDACTED] danke ich für die ständige Hilfsbereitschaft bei allen schwierigen Computerproblemen und zusammen mit [REDACTED] auch für das mir entgegengebrachtes Vertrauen bei der Betreuung zahlreicher Studenten.

Ein besonderer Dank gilt den ehemaligen und aktuellen Zimmerkollegen, [REDACTED] [REDACTED] für die zahlreichen und sehr unterhaltsamen Gespräche der letzten Jahre.

Bedanken möchte ich mich auch bei der [REDACTED] für das hervorragende Arbeitsklima.

Ιδιαίτερα ευχαριστώ τους γονείς μου και τον αδερφό μου, που όλα αυτά τα χρόνια με στηρίζουνε και με βοηθούνε ενεργά να πετύχω τους στόχους μου.

Τη σύντροφο της ζωής μου ████████ ευχαριστώ που είναι κάθε στιγμή δίπλα μου....για όλα αυτά που ζήσαμε μαζί ως τώρα και για τις συγκινήσεις που μας περιμένουνε μαζί με την κορούλα μας ████████, ένα ευχαριστώ θα ήταν πολύ λίγο....

