This is a postprint of an article published in
Kayser, H., Mann, K., Machaidze, G., Nimtz, M., Ringler, P., Müller,
> S.A., Aebi, U.
> Isolation, Characterisation and Molecular Imaging of a
> High-Molecular-Weight Insect Biliprotein, a Member of the Hexameric
> Arylphorin Protein Family
> (2009) Journal of Molecular Biology, 389 (1), pp. 74-89.

Isolation, characterization and molecular imaging of a high molecular-weight insect biliprotein, a member of the hexameric arylphorin protein family.

Hartmut Kayser^{1*}, Karlheinz Mann², Gia Machaidze³, Manfred Nimtz⁴, Philippe Ringler³, Shirley A. Müller³ and Ueli Aepli³

¹ Institut für Allgemeine Zoologie und Endokrinologie, Universität Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany,

² Max-Planck-Institut für Biochemie, Abteilung Proteomics und Signaltransduktion, Am Klopferspitz 18, 82152 Martinsried, Germany,

³ Maurice E. Müller Institut für Strukturbiologie, Biozentrum, Universität Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland

⁴ Helmholtz-Zentrum für Infektionsforschung GmbH, Strukturbiologie, Inhoffenstrasse 7, 38124 Braunschweig, Germany,

* corresponding author

Email address of the corresponding author: hartmut.kayser@uni-ulm.de

Abbreviation: CBP, *Cerura* biliprotein

Running title: Hexameric arylphorin-type insect biliprotein

Summary

The abundant blue hemolymph protein of last instar larvae of the moth *Cerura vinula* was purified and characterized by protein-analytical, spectroscopic and electron microscopic methods. Amino acid sequences obtained from a large number of cleavage peptides revealed high similarity of the blue protein with arylphorins from a number of other moth species. In particular there is a high abundance of the aromatic amino acids tyrosine and phenylalanine amounting to about 19% of total amino acids, and a low content of methionine (0.8%) in the *Cerura* protein. The mass of the native protein complex was studied by size-exclusion chromatography, analytical ultracentrifugation, dynamic light scattering and scanning transmission electron microscopy and found to be around 500 kDa. Denaturing gel electrophoresis and mass spectrometry suggested the presence of two proteins with masses of about 85 kDa. The native *Cerura* protein is, therefore, a hexameric complex of two different subunits of similar size, as is known for arylphorins. The protein was further characterized as a weakly acidic ($p_I \sim 5.5$) glycoprotein containing mannose, glucose and N-acetylglucosamine in an approximate ratio of 10:1:1. The structure proposed for the most abundant oligosaccharide of the *Cerura* arylphorin was the same as already identified in arylphorins from other moths. The intense blue color of the *Cerura* protein is due to non-covalent association with a bilin of novel structure at an estimated protein subunit-to-ligand ratio of 3:1. Transmission electron microscopy of the biliprotein showed single particles of cylindrical shape measuring about 13 nm in diameter and 9 nm in height. A small fraction of particles of the same diameter but half the height was likely a trimeric arylphorin dissociation intermediate. Preliminary 3D reconstruction based on averaged TEM projections of the individual particles revealed a double-trimeric structure for the hexameric *Cerura* biliprotein complex suggesting it to be a dimer of trimers.

(word count: 300 words)

Keywords:

Hexamerin, arylphorin, biliprotein, sequence similarity, glycoprotein, particle reconstruction, STEM mass measurement, insects.

Introduction

Insect hemolymph contains several proteins which, although involved in different functions, probably have a common phylogenetic origin in arthropod evolution represented by a protein with six highly conserved histidine residues providing two copper-binding sites.¹⁻³ Divergent evolution involving gene duplications gave rise to the prophenoloxidasases and hemocyanins, both of which retained the two copper sites involved in oxygen binding. Prophenoloxidasases, in their processed active form, are involved in the oxidation of tyrosine for melanin formation and cuticle sclerotization, while hemocyanins transport oxygen in respiration in spiders, higher crustaceans and some insects.

During evolution of the ancestral protein some proteins lost the histidines essential for binding copper, and consequently oxygen, while retaining sequence similarity to prophenoloxidasases and hemocyanins. The evolution of these copper-less proteins followed two major, divergent routes in insects that resulted in two groups of proteins with strikingly different amino acid profiles. One group, the arylphorins,⁴ features an unusually high proportion of the aromatic amino acids tyrosine and phenylalanine, giving them their name and typically accounting for about 20% of the total number of amino acids. The copper-less homologs forming the other group are rich in methionine (up to ~10 %), but low in aromatic amino acids.^{5,6} Proteins in both groups have the same quaternary structure as the hemocyanins, forming hexameric complexes of ~80 kDa subunits. These complexes with sizes of ~500 kDa are collectively called hexamerins and are usually hetero-oligomers of subunits with very similar amino acid sequences.^{2,7,8} Recent studies of hexamerins from stoneflies, an ancient group of insects, suggested that the typical arylphorins were the first to diverge from hemocyanins, and that hexamerins low in aromatic amino acid content appeared later and adopted other functions.⁹ While the crystal structures of a few arthropod hemocyanins have been obtained and found to be highly conserved, comparable data for an arylphorin or methionine-rich hexamerin are still missing.^{10,11}

Besides the well-characterized arylphorins and methionine-rich hexamerins widespread in lepidopteran insects, other types of hexamerins have been found that do not fit into the two major groups and may

hence have more specific functions. Examples of these are the arylphorins from dipterans, hexamerins binding juvenile hormone, riboflavin, or biliverdin, hexamerins the expression of which is repressed or enhanced by juvenile hormone, and hexamerins rich in glutamine.¹²⁻¹⁷ In many cases, the expression of hexamerins is under developmental and tissue-specific control. Arylphorins, for example, are typically produced in the larval fat body, accumulate in the hemolymph before larval-pupal transformation and are thought to deliver amino acids required for the formation of the adult insect and of the sclerotized cuticle that consumes large quantities of aromatic amino acids. Therefore, these proteins are generally described as storage proteins, though details of their dynamics and fate are rare.^{7,8,18,19}

According to recent studies besides serving as mass resources for metamorphosis, cuticle formation and egg development, hexamerins may also function in cellular control. In termites, two different hexamerins rich in aromatic amino acids play a major role in caste determination in cooperation with juvenile hormone.^{20,21} An arylphorin subunit, acting as a monomer, has been shown to stimulate cell proliferation and differentiation in several species of moth both *in vitro* and *in vivo*.²²⁻²⁴ In the wax moth, the haemolytic and cytotoxic effects of hemolymph have been attributed to a typical hexameric arylphorin, which may thus contribute to the insect's immune response.^{25,26}

The impressive accumulation of storage proteins just before metamorphosis is readily observed when they are associated with coloured ligands such as the blue bilins that originate from heme cleavage. A number of such insect biliproteins have been characterized and many of them are lipocalins as judged from their monomer sizes of ~20 kDa.²⁷ However, definite structural assignment has only been possible for the two biliproteins for which the crystal structures have been solved.^{28,29} A few biliproteins can be tentatively classified as hexamerins based on subunit size, but none of these represents a true arylphorin with a high (~20%) content of aromatic amino acids. The four cyanoproteins of the bug, *Riptortus clavatus*, are true biliproteins. They are present in eggs and hemolymph as complexes formed from two different subunits.^{15,30} However, the proportion of aromatic amino acids present is only about 15%, significantly lower than that of typical arylphorins possibly indicating a different role of these cyanoproteins the expression of which is activated by juvenile hormone.

We have isolated and characterized a hexamerin-type storage protein from the larval hemolymph of a moth, *Cerura vinula*. The protein is a typical arylphorin with high sequence similarity to a wide range of arylphorins, a high content of aromatic amino acids and a low content of methionine. It occurs as a hexameric complex formed from two different subunits with the typical quaternary structure of hexamerins. Surprisingly, this arylphorin is also a biliprotein with a non-covalently bound biliverdin-based chromophore conferring an intense blue colour. In addition to the physicochemical characterization of this *Cerura* biliprotein (CBP) we also present electron microscopic images and a preliminary three-dimensional (3D) structural reconstruction of the protein particles. To the best of our knowledge, this is the first molecular study of a true arylphorin as a biliprotein.

Results

Purification of CBP

Hemolymph of *C. vinula* late last instar larvae contains the blue CBP as a major protein constituent. Together with yellow carotenoproteins also present in the hemolymph, CBP produces a dark green colouration. Based on the spectral data of samples of the isolated blue protein (see below), the purities of CBP preparations were >60% after ammonium sulfate precipitation and >95% after chromatography on DEAE-Sephacel. Size exclusion chromatography provided pure CBP, according to SDS-PAGE analysis.

A solution of purified blue CBP showed absorbance maxima at 277 nm, 373 nm and 666 nm and absorbance ratios of ~ 1.5 for $A_{666\text{nm}}/A_{373\text{nm}}$ and of ~ 0.2 for $A_{666\text{nm}}/A_{277\text{nm}}$ (Figure 1). The maxima at 373 nm and 666 nm are due to the association of CBP with a bilin. This was easily liberated from the dissolved holo-protein by the addition of methanol at $\geq 50\%$ (v/v) final concentration. The blue pigment was isolated from the supernatant by solid or liquid phase extraction. Results so far suggest a novel biliverdin-based structure for the CBP ligand (Kayser, unpublished data). Since this chromophore is different from those of the known biliverdin isomers its true extinction coefficient is unknown. However,

as the UV absorption (Soret band), contrary to that in the visible range, was only slightly different in the bound and free CBP ligand (Kayser, unpublished data) a preliminary estimate for the bilin-to-protein ratio has been obtained as reported below.

Chromatofocusing showed CBP to be a weakly acidic protein exhibiting isoelectric heterogeneity. The majority of the protein eluted at a pH of ~5.5, and the rest at a pH around 5 as a broad tailing peak only partially separated from the predominant form. All fractions showed identical visible absorption spectra characteristic of the native protein-bound bilin ligand.

The size of the native CBP complex was studied by several methods. Size-exclusion chromatography (TSK HW 55-S) performed in 50 mM Tris/HCl (pH 7.4), consistently provided a value of ~430 kDa for several fresh protein isolates (Figure 2). The same result was obtained using 200 mM potassium phosphate (pH 7.00) as eluant. A single batch of CBP prepared in the same way provided a value of ~230 kDa (i.e., it behaved like catalase) in repeated experiments calibrated with seven reference proteins. The reason for this exceptional result is unknown. Exposure of CBP to high pH (pH 9.6; 100 mM Gly/NaOH), even over night, did not result in dissociation of the complex into its subunits as shown by size-exclusion chromatography performed in the same buffer.

On analytical ultracentrifugation of CBP, sedimentation equilibrium runs, performed at a protein concentration of 1.7 mg/mL, showed one major fraction with a molecular mass ~500 kDa. A sedimentation velocity run, performed at 0.42 mg/mL, provided a sedimentation coefficient $S_{20,w}$ of 17.8 and a frictional ratio f/f_0 of 1.25. The latter is close to the value of 1.28 determined for bovine serum albumin,³¹ suggesting that CBP has a spherical shape. Dynamic Light Scattering, as a supplementary method, provided an estimate of ~570 kDa for the molecular mass of CBP.

CBP was subjected to two modes of mass spectrometric analysis. Whereas MALDI/TOF-MS produced a relatively broad molecular ion signal in the range of 84-87 kDa (data not shown), after deconvolution ESI-MS yielded a cluster of clearly separated molecular ion signals in the range of 85.2-85.9 kDa

confirming the oligomeric structure of CBP and unequivocally demonstrating the heterogeneity of its subunits. The most intense molecular ion signals were detected at 85,246 Da, 85,408 Da and 85,569 Da (Figure 3). The mass differences of approximately 162 Da suggest the presence of glycoforms differing by hexose residues as is typical for high-mannose type structures. Weaker signals are compatible with the presence of a second, by approximately 90-kDa larger protein bearing similar modifications (see below). Additional minor signals were present at in the range of 42.6–43.2 kDa; these might represent doubly charged molecular ion signals.

Under the denaturing conditions of SDS-PAGE freshly prepared samples of CBP behaved as a single band with a M_r of ~72,000. Occasionally, small amounts of a doublet protein band with a M_r $\geq$ 40,000 and a band with a M_r of ~150,000 were also present (Figure 4). CBP pre-treated with DTT at 70°C provided the same pattern of bands. Prolonged storage of CBP at ~5 °C (for several weeks or months) produced more of the proteins with M_r of $\geq$ 40,000 and additional bands in the range of M_r of 20,000–25,000 (not present in Figure 4). The visible absorption spectra of such ‘aged’ samples were identical to those of fully native ones. Both samples also behaved identically on size-exclusion chromatography. The mechanism underlying this change of band pattern is unknown.

Taken together, the results from the various approaches to mass determination convincingly suggested that native CBP is a hexameric ~500 kDa-complex most likely composed of non-identical subunits with sizes in the 85-kDa range.

Amino acid composition and ligand-to-protein ratio

A significant feature of the amino acid composition of CBP was the relatively high content of aromatic amino acids with Tyr (9.3%) and Phe (9.6%) making up 18.9% of the total (Table 1). As further significant features, CBP was also rich in Asx (11.9%) and Glx (12.7%) but low in Met (0.8%). Calculated for an 85-kDa subunit, the number of Tyr residues obtained by acid hydrolysis (58.1 residues) was in good agreement with results from spectroscopy (57.1 residues); the latter method also provided data on the Trp content (Table 1). For the bilin-binding protein from *P. brassicae*, studied as a reference

by the same spectroscopic method, the residue numbers for Tyr and Trp were 16.0 and 5.1, respectively, in perfect agreement with sequence data.³²

An extinction coefficient of $167 \text{ mM}^{-1} \text{ cm}^{-1}$ for 294.5 nm in 0.1 N NaOH (spectral intersection of Tyr and Trp absorption)³³ was obtained for an 85-kDa subunit of apo-CBP by calculation. This compares well with the experimentally determined value of $155 \text{ mM}^{-1} \text{ cm}^{-1}$ ($n = 4$). Given the molar ratio of Tyr and Trp in CBP and the change of Tyr absorption with pH, an extinction coefficient for neutral media of $120 \text{ mM}^{-1} \text{ cm}^{-1}$ was estimated for the apo-protein absorption maximum around 277 nm. Based on this value and the molar Soret absorption of the biliverdin isomers the molar bilin-to-protein subunit ratio was estimated to be 1:3. A similar value was obtained from the determination of the holoprotein concentration by the Bradford reagent and, after precipitation of the apo-protein with methanol, the concentration of the free bilin using both absorption peaks. This estimate of stoichiometry suggests that the hexameric CBP complex contains two molecules of the bilin ligand.

Peptide sequences and sequence comparisons

CBP was further characterized by Edman sequencing of the intact protein and of isolated internal peptides after cleavage by CNBr and LysC. Mass spectrometric *de novo* sequencing of peptide mixtures obtained by tryptic and GluC digestion of the isolated protein preparation and of protein bands partially separated by SDS-PAGE was performed as a complementary technique (Table 2). Two N-terminal sequences of approximately equal concentration were detected by Edman sequencing of intact CBP: G T V P T K F N V K and T T I K H N L K T K E V D A E F V A R Q. No significant similarity to other proteins was found for the first sequence, probably due to the relatively low number of identified amino acids, whereas the second sequence was found to be very similar to the N-terminus of subunits from insect arylphorins. A large number of internal peptides were identified by Edman sequencing and/or *de novo* MS sequencing after proteolytic cleavage (Table 2). It should be noted that the mass spectrometric technique applied is not capable of distinguishing between L (Leu) and I (Ileu) residues. Therefore, in MS-derived sequences the amino acid L stands for either L or I. The amino acid sequences obtained by the different techniques are in excellent agreement. Database searches using FASTA or MPsrch programmes provided by EBI

(www.ebi.ac.uk) revealed that almost all peptides showed high similarity to arylphorins from various insects, mostly from moth species. The identification of two different N-terminal sequences by Edman sequencing suggest the presence of at least two different subunits in the hexameric CBP complex. This was confirmed by the detection of very similar, but clearly different internal peptides by Edman as well as MS sequencing. Finally, it should be noted that no sequences incompatible with arylphorins were detected, excluding the presence of a single unrelated contaminating protein (Table 2).

Carbohydrate composition

The glycoprotein nature of CBP was further demonstrated by its adsorption to ConA-sepharose. Complete desorption required the use of both α -methyl-glucopyranose and α -methyl-mannopyranose in the eluent allowing 33% and 67%, respectively, to be recovered by successive elution. When α -methyl-mannopyranose was used first, CBP was even recovered to 94% in the eluate. These results indicate that mannose is the major sugar moiety of CBP.

The presence of carbohydrates was confirmed by a semi-quantitative analysis of the carbohydrate components applying GC/MS after methanolysis and trimethylsilylation of the holoprotein. N-acetylglucosamine, mannose and glucose were detected in a ratio of approximately 1:10:1 suggesting the presence of high-mannose-type glycans. The total amount of carbohydrate was estimated to be approximately 5 % of the holo-protein mass.

Glycosylation sites

Four glycopeptides were identified by mass spectrometric analysis of the proteolytic digests of the protein (Table 2). A series of molecular ions differing by 162 mass units with characteristic fragmentation behaviour upon collision-induced dissociation (CID) were detected that corresponded to glycopeptides with an identical peptide backbone but different numbers of hexose residues. The very high abundance of carbohydrate specific fragments allowed the clear identification of the carbohydrate part as high mannose-type glycans with 8-10 hexose residues as the dominant structures. From the abundant fragment generated by cleavage of the chitobiose linkage, the mass of the peptide plus one remaining N-

acetylglucosamine residue could be readily determined. However, amino acid sequence specific fragments were of relatively low abundance or completely missing. Therefore, MS could not provide peptide sequences of the intact glycopeptides. Four additional peptides were successfully obtained after enzymatic removal of the glycans from the peptides by PNGase F, resulting in deamination of Asn, the specific site for N-glycosylation in the tripeptide consensus sequence Asn X Ser/Thr (N X S/T), to Asp (D). One of these was also identified by Edman sequencing (see Table 2). All peptides containing the modified consensus sequence D X S/T showed strong similarity to insect arylphorins.

Carbohydrate structure

To obtain more information on the oligosaccharide moieties, the intact glycoprotein was enzymatically deglycosylated and the resulting free N-glycans were isolated and analysed by MALDI-MS and ESI-MS/MS before and after reduction and permethylation (data not shown). The MALDI mass spectrum of the native N-glycans agreed very well with the data obtained from analysis of the glycopeptides. Three glycans with the composition Hex₈₋₁₀HexNAc₂ at a ratio of approximately 1:10:5 were observed, suggesting the presence of high mannose type N-glycans with 8 or 9 mannose residues and, as a relatively unusual structure, a glycan with nine mannose residues, one glucose and one chitobiose unit. This structural assignment was compatible with the carbohydrate compositional analysis of the intact protein, which showed the presence of small amounts of glucose in addition to mannose and N-acetylglucosamine. Interestingly, the proposed structure has already been described as the dominant glycan of the arylphorin of the Chinese oak silkworm.³⁴ To further demonstrate the identity of the isolated N-glycans with the structures described in the literature, the glycans were reduced, permethylated and analysed by ESI-MS/MS. The masses and fragmentation pattern were in complete agreement with the assumed structures (Figure 5). In the case of the structure containing one glucose residue, a fragment generated by the elimination of four hexose units was detected, which can only be explained by the linkage of the additional hexose to the branch linked α 1-3 to the inner common core mannose of the standard Man₉ high mannose-type structure, as shown for the largest N-glycan described by others.³⁴

Electron microscopy and 3D analysis

Transmission electron microscopy (TEM) of negatively stained CBP at different concentrations (range of 0.014 to 0.5 mg/mL) revealed a fairly homogeneous fraction of single particles. According to the various top- and side-view projections these were 13 nm in diameter and about 9 nm in height (Figure 6). A low number (estimated <10%) of smaller particles with the same diameter but about half the height were present in some preparations (Figure 6). Particles associated with others were also occasionally observed.

Images of unstained CBP particles recorded by scanning transmission electron microscopy (STEM) for mass measurement, did not show the smaller class of particles, but some association was also observed (Figure 7a). Analysis of the images provided a mass of 471 ± 47 kDa ($n = 1541$; Figure 7a) for the predominant particle fraction, in accordance with the results from the other methods detailed above and confirming the hexameric nature of these complexes. STEM imaging of similarly prepared but negatively stained grids revealed structural details of the complexes present without the application of averaging procedures. Six domains can be distinguished on the roughly circular top-views. These form the corners of two, centrally aligned, overlapping triangles giving the structure a 3-fold symmetry. There are also signs of a central ring-like domain. In addition, some side-views have a clear 2-fold symmetry (Figure 7b and c).

Particle projections selected from TEM images of negatively stained samples were used for preliminary 3D reconstruction. Two of the 2D averages initially generated are shown in Figure 8a. As indicated by the STEM single-shot images, the top-view average has a triangular-to-circular cross-section and six distinct domains organised in two sets of three. These features and the possible ring-like structure towards the centre of the average became more pronounced on the application of D3 symmetry (data not shown). The various side-views, one of which is shown in Figure 8a, were rectangular in shape, often with indications of a 2-fold symmetry axis. D3 symmetry was assumed to generate the final low resolution 3D model (Figure 8c), which has a resolution of 25 Å. Re-projections were calculated and compared to the initial averages as a control (Figure 8b); the two views shown are positioned immediately adjacent to the initial averages to facilitate comparison.

Discussion

We have identified the predominant larval blue hemolymph protein from the moth *C. vinula*, the *Cerura* biliprotein (CBP), as a typical member of the arylphorin group of hexameric hemocyanin-related protein complexes. This assignment is based on the physicochemical characterization of the native holoprotein and its subunits, on the high content of aromatic amino acids and on the high similarity of the peptide sequences to arylphorins from insects, almost all from other moth species.

CBP occurs as an oligomeric complex with a mass of ~500 kDa, as determined by several methods, and is composed of six ~85 kDa subunits. This is in accordance with other hexamerins that have been characterized from a diversity of insects.² Data from mass spectrometry and from sequence analyses of cleavage peptides suggest the presence of two non-identical subunits of very similar size and sequence, which is common in this group of proteins. Based on the monomer size(s) obtained by mass spectrometry, around 85 kDa, the expected mass of the hexamer is 510 kDa, a value also supported by analytical ultracentrifugation (indicated mass ~500 kDa), and STEM mass measurements of individual complexes (indicated mass 471 ± 47 kDa). A somewhat higher estimate, of 570 ± 30 kDa, was obtained by dynamic light scattering. The remarkable underestimation of the CBP mass as a 430-kDa complex by gel permeation chromatography may result from interaction of the N-glycans attached to the protein with the Fractogel matrix. Interestingly, SDS PAGE also gave an underestimate of the molecular mass indicating a M_r of ~72,000 for the subunits, corresponding to a M_r of ~430,000 for the hexameric CBP complex.

Some arylphorins, such as calliphorin from blowfly, *Calliphora vicina*, have been observed to reversibly dissociate into subunits at alkaline pH, eventually producing trimers as intermediates.^{18,35} The CBP complex did not dissociate under these conditions, like the arylphorin (manducin) of tobacco hornworm, *Manduca sexta*.³⁶

The name-giving physicochemical feature of the arylphorin class of insect hexamerins is their high content of the aromatic amino acids Tyr and Phe, typically amounting to ~20% of the total amino acids, and a low content of Met. Other predominant amino acids are Asx and Glx.^{18,37} This feature was clearly reflected by the amino acid profile of CBP. The other major class of hexamerins are the Met-rich storage proteins with Met amounting up to 10%.^{6,18} Assignment of CBP to the arylphorin type of hexamerins was confirmed by the high similarities of a fairly large number of CBP cleavage peptides, many of them comprising more than 20 residues, to sequences from other arylphorins, almost all from other moth species. This information is currently being used in a cloning approach to obtain the complete sequences of the two CBP subunits.

Arylphorins are known to be glycosylated.^{4,34,38} The glycoprotein nature of CBP was initially indicated by its affinity to ConA-Sepharose and was subsequently firmly established by mass spectrometry. The sugar components were identified as high-mannose-type glycans that are well-known in insects. N-Glycans from mammals, in contrast, are rich in galactose.³⁹ Fucose, which is frequently found in proteins produced by insect cell lines,³⁹ was absent in CBP. Interestingly, however, a low amount of fucose is present in the arylphorin from the silkworm, *Antheraea pernyi*.³⁴ In CBP, the glycans are composed of mannose, N-acetylglucosamine and glucose with an approximate ratio of 10:1:1. The proposed structure for the Glc₁Man₉GlcNAc₂ moiety is identical to that presented by Kim et al. for the dominant glycan of the Chinese oak silkworm on the basis of mass and NMR spectra.³⁴ Moreover, the same oligosaccharide structure (lacking the terminal glucose, however) was previously identified by NMR spectroscopy in a glycopeptide obtained from the arylphorin from another moth, the tobacco hornworm, *M. sexta*.³⁸ In the three moth species, the carbohydrate content is in the range of 2-5% of the total protein mass.

Several glycopeptides identified in CBP showed the conserved tripeptide consensus sequence AsnXSer/Thr with Asn as the site of glycosylation. While two to three potential Asn-linked glycosylation sites are present in the arylphorin gene sequences from several insects,⁴⁰⁻⁴⁴ glycopeptides from only two sites were found by mass spectrometry in the protein from *A. pernyi*, identified as Asn 196 and Asn 344.⁴⁴ To locate these sites in the hexamer complex, these authors constructed a three-dimensional model of

arylphorin based on crystal structure data for the hemocyanins from spiny lobster, *Panulirus interruptus*, and horseshoe crab, *Limulus polyphemus*. According to this model, Asn 196, which carries the Glc₁Man₉GlcNAc₂ glycan, is buried inside the hexamer and thus protected from solvent. This glycosylation site at a position around 200 seems to be conserved among all hexamerins,¹⁵ and probably corresponds to the glycosylation site contained in peptides T11 and T13 from CBP. The second site, at Asn 344 in the *A. pernyi* protein, linked to Man₅₋₆GlcNAc₂, is located at the hexamer surface in this model. Hence, it may represent a processed glycan. As this site is apparently not present in any of the identified peptide sequences from CBP the actual number of sugar moieties in this aarylphorin remains open at present.

Except for a few hemocyanins, no crystal structure from any other hexamerin is available at present. However, the conserved physicochemical features and sequence similarities among hexamerins, including the hemocyanins, suggest that the overall structures of the subunits as well as their hexameric complexes are also conserved.

On this conceptual basis, a preliminary model of the aarylphorin hexamer from blowfly has been deduced and supplemented by an electron microscopic study of the morphology of negatively stained protein particles.³⁵ These particles exhibited a triangular-to-circular profile in top views with a diameter of 12 nm and a rectangular shape in side views, indicating a height of 10 nm. The aarylphorin CBP particles are of comparable shape in the two views and of similar size, having a diameter of 13 nm and a height of 9 nm (Figures 6 and 7). Moreover, the presence of six distinct domains and the symmetry of the CBP complex revealed by STEM single-shot images (Figure 7) and TEM single particle analysis and 3D reconstruction (Figure 8) are compatible with structural information available for other hexamerins, such as the aarylphorin from the blowfly *Calliphora vicina* and the hemocyanins from the lobsters

Panulirus interruptus and *Palinurus elephas*.^{34,45,46}

Gelöscht:

In addition to this major population of particles, the preparation of CBP also contained a small number of particles with the same diameter but about half the height. As these must also have half the mass they are

probably trimers formed by the partial dissociation of the native hexamer. If this interpretation is correct, the CBP complex must be a dimer of trimers. Markl et al. arrived at the same conclusion in their study of the blowfly arylphorin.³⁵ In contrast to the blowfly protein, however, specific higher aggregates of CBP were neither detected by electron microscopy nor by gel permeation chromatography.

CBP is an interesting variant of the arylphorin-type of hexamerins as it is associated with a tetrapyrrol ligand of unknown structure and function. CBP shares the latter aspect with other insect biliproteins that are synthesized according to developmental and tissue-specific programs.^{27,47} As the ligand is clearly not present at an equimolar ratio to the hexamer subunits, but rather at a ratio of 1:3, the location and mode of binding within the presumed trimeric subcomplex is rather speculative at present, calling for crystal structure analysis of this arylphorin. The novel biliverdin-based chromophore, though non-covalently associated with the protein, is obviously bound in a most specific way that induces high optical activity in the ligand, which is optically inactive in its free state.⁴⁸

Materials and Methods

Protein purification

CBP protein was isolated from the dark green hemolymph of late last instar larvae of the moth *Cerura vinula* (Notodontidae) entering the wandering stage. Hemolymph was obtained by bleeding from an abdominal leg and mixed with in the same volume of cold 50 mM Tris/HCl buffer (pH 7.4) supplemented with a few grains of N-phenyl thiourea to prevent blackening by phenoloxidase activity. After clearing the buffer-diluted hemolymph by centrifugation at 10,000g and 5 °C for 10 min (Sorvall RC-5C with rotor SS-34), CBP was obtained as a blue or blue-green precipitate by the addition of solid ammonium sulphate in the saturation range of 50-60 % (at 0 °C). The precipitate was collected by centrifugation at 20,000g for 30 min and then re-dissolved in Tris buffer. The bulk of yellow carotenoprotein(s) remained in the supernatant, which was discarded. The dissolved pellet containing the blue CBP and some yellow protein(s) were concentrated and intensively dialysed against the 50 mM Tris/HCl buffer containing 30

mM NaCl, pH 7.4, using Centriprep 50 centrifugation units (Amicon-Millipore). CBP was further purified by chromatography on a column of DEAE Sephacel (2.5 x 34 cm) applying a linear gradient of NaCl (30-400 mM) in 50 mM Tris/HCl buffer, pH 7.4, at a flow rate of 1.8 mL/min and monitoring at 280 nm. Fractions were collected and spectrally characterized by the absorbance ratios of $A_{666\text{nm}}/A_{373\text{nm}}$ and $A_{666\text{nm}}/A_{278\text{nm}}$, respectively. Spectra were recorded in the range of 250 to 750 nm using a Perkin Elmer Lambda 15 UV/VIS spectrophotometer and quartz cuvettes of 10 mm or 2 mm light path. Fractions were pooled according to purity and concentrated for either re-chromatography on DEAE Sephacel (eventually of non-peak fractions) or for size-exclusion chromatography on a Fractogel TSK HW 55-S (Merck) column (2.5 x 36 cm) run with 0.05 M Tris/HCl buffer, pH 7.4, at a rate of 0.8 mL/min. The fractions were spectrally examined for CBP, eventually pooled and concentrated as above. Protein was quantified using the Bradford reagent (Bio-Rad) and bovine serum albumin fraction V (Serva) as a standard.

Chromatofocusing

A column (1 x 50 cm) of Polybuffer exchanger PBE 94 (Pharmacia) was equilibrated with 25 mM histidine x HCl, pH 6.1. Elution was performed with Polybuffer 74 x HCl of pH 4.0 at a flow rate of 0.8 mL/min. Approximately 5 mL of elution buffer was applied to the column prior to the sample, as recommended by the manufacturer. Elution was monitored at 280 nm. The collected fractions were characterized by their pH values and full range-absorption spectra.

ConA-Sepharose chromatography

A small volume (approximately 1.5 mL) of ConA-Sepharose (Pharmacia) equilibrated with 10 mM Tris/HCl containing 500 mM NaCl, 1 mM MnCl_2 and 1 mM CaCl_2 , pH 7.5, was incubated with approximately 0.3 mL of CBP in Tris/HCl buffer, pH 7.4, overnight in a 3-mL filtration column (Isolute). Elution was sequentially performed with 0.5 M α -methyl-glucopyranose and 0.5 M α -methyl-mannopyranose, both dissolved in equilibration buffer, and monitored at 280 nm to assure complete elution by each displacing sugar. The relative concentrations of CBP in the eluates were calculated on the basis of the absorbances at 280 nm.

Analytical size exclusion chromatography

Size exclusion chromatography on a TSK HW 55-S column (2.5 x 34 cm) was also used to estimate the molecular mass of native CBP. The standard running buffer was 0.05 M Tris/HCl, pH 7.4. In some experiments 200 mM potassium phosphate, pH 7.0, and 100 mM Gly/NaOH, pH 9.6, respectively, were used. The reference proteins (Pharmacia, Bio-Rad) were thyroglobulin (669 kDa), ferritin (440 kDa), catalase (232 kDa), bovine γ -globulin (158 kDa), bovine serum albumin (67 kDa) and chicken ovalbumin (44 kDa), used in various combinations. In addition, the *Pieris* bilin-binding protein (19.8 kDa) from earlier work was used.²⁸

Dynamic light scattering

A DynaPro Plate Reader (Wyatt Technology Corp., Santa Barbara, USA) was used for mass determination of CBP in 0.05 M Tris/Cl-buffer, pH 7.4, at several dilutions.

Analytical ultracentrifugation

Analytical ultracentrifugation was performed in a Beckman XL-A analytical ultracentrifuge at 20 °C. CBP was dissolved in 25 mM Tris/HCl buffer pH 7.4 containing 150 mM NaCl. Sedimentation velocity runs were performed in a 12 mm double-sector cell at a protein concentration of 0.42 mg/mL and at 50,000 rpm; the scanning wavelength was 280 nm. Sedimentation equilibrium runs were performed for different protein concentrations at 4400 rpm and 4500 rpm in double-sector cells of different thickness (3 mm, 4 mm and 12 mm). The data were analysed using the SedFit software⁵⁷ for sedimentation velocity runs and the SEGAL program⁵⁸ for sedimentation equilibrium runs. The partial specific volume of CBP was 0.73 cm³/g and the solution density was 1.005 g/cm³. A solution viscosity of 1.03 centipoises was taken for the conversion of *S* to *S*_{20,w}. The total conversion factor including the solution density was 1.05.

Polyacrylamide gel electrophoresis (PAGE)

PAGE in the presence of SDS was performed with pre-cast NuPAGE Novex Bis-Tris 10% and 4-12% gels run with MES or MOPS buffers. In some experiments the protein samples were heated to 70 °C for 5 min in the presence of 2 mM DTT before loading. The SeeBlue Plus2 mixture was used as molecular

weight markers. Proteins were stained with SimplyBlue™ SafeStain. Operation conditions were as recommended by the manufacturer (Invitrogen).

Preparation of apo-CBP and isolation of the bilin ligand

Methanol was added to a buffer solution of native holo-CBP to a final concentration of $\geq 50\%$ (v/v) to achieve immediate and complete precipitation of the apo-protein and extraction of the bilin. After storage in the cold for several hours and centrifugation, the bilin ligand was isolated by passing the blue supernatant, after dilution with water to about 20% methanol, through a Sep-Pak C18 cartridge (Waters), washing with 1 mL 20% (v/v) methanol followed by elution of the blue pigment with 100% methanol. The bilin concentration was estimated by assuming a millimolar extinction coefficient of $50 \text{ mM}^{-1} \text{ cm}^{-1}$ for the Soret absorption maximum around 373 nm of both the protein-bound and the free pigment and a value of $15 \text{ mM}^{-1} \text{ cm}^{-1}$ for the absorption maximum around 630 nm for the free bilin.⁴⁹ Alternatively, the bilin was extracted from the supernatant with chloroform after dilution with water and acidification to pH < 3 with citric acid. The apo-CBP pellet was lyophilised and re-dissolved as required for further studies (see below).

Amino acid composition

The amino acid composition of (holo-)CBP was analysed in duplicate on a Biotronic LC 3000 analyser after hydrolysis in 6 M HCl at 110 °C for 24 h.⁵⁰ The content of Tyr was also determined by pH-difference spectroscopy of apo-CBP from the extinction difference at 295 nm of the solution in 0.1 N NaOH/4 M urea *minus* that in 0.1 N HCl using an extinction coefficient of $2.33 \text{ mM}^{-1} \text{ cm}^{-1}$.⁵¹ From these results, the content of Trp relative to that of Tyr was spectroscopically obtained in 0.1 N NaOH.⁵² As a reference, the aromatic amino acids of the *Pieris* bilin-binding protein³² were quantified with the same spectroscopic methods. The extinction coefficient of apo-CBP in 0.1 N NaOH/4 M urea was calculated according to Goodwin and Morton.³³

Protein cleavage, peptide isolation and sequence analysis

The proteins were reduced and pyridylethylated as described.⁵⁰ CNBr cleavage was performed in 70 % (v/v) formic acid by addition of a small crystal of CNBr and incubation at 23 °C for 24 h in the dark. The fragments were suspended in 0.1 % trifluoroacetic acid containing 25 % (v/v) acetonitrile. The insoluble material was removed by centrifugation and the soluble fragments were separated using a Superdex Peptide column HR 10/30 (Amersham Biosciences) in the same solvent at a flow rate of 0.3 mL/min and by C18 reversed phase HPLC.⁵⁰ The pellet of fragments insoluble in trifluoroacetic acid/acetonitrile was dissolved in 0.1 M ammonium acetate, pH 5, containing 4 M guanidinium hydrochloride, and separated using a Superdex 75 (10/300; Amersham Biosciences) column in the same solvent at a flow rate of 0.3 mL/min. The fractions were acidified to pH ~1 with trifluoroacetic acid and desalted using a short reversed phase column (HD-Gel-RP-7s-300, 4 x 40 mm; Grom, Rottensburg, Germany) with a gradient of 15-65 % (v/v) acetonitrile in 0.1 % trifluoroacetic acid over 20 min at a flow rate of 0.5 mL/min. Cleavage with *Achromobacter* lysyl endopeptidase (WAKO Chemicals, Neuss, Germany) was done in 0.1 M ammonium hydrogen carbonate containing 4 M urea at an enzyme to substrate ratio of 1:100 at 23 °C for 16 h. The peptides were separated using a Superdex Peptide column and reversed phase HPLC as above. N-terminal amino acid sequence analysis and sequencing of peptides was performed using the Procise 492 sequencer (Applied Biosystems).

Carbohydrate compositional analysis

Monosaccharides linked to CBP were analyzed on a Finnigan GCQ GC/MS system after methanolysis, N-reacetylation and trimethylsilylation, as the corresponding methyl glycosides on a 30 m DB5 capillary column.⁵³

Enzymatic liberation of N-glycans

For the liberation of N-glycans, a 100–300 µg sample of holo-CBP was incubated with 25 mU of recombinant PNGase F (Boehringer Mannheim) over night at 37 °C. Complete digestion was monitored by the specific mobility shift of the protein on SDS-PAGE. The released N-glycans were recovered in the supernatant after precipitation with four volumes of 100% ethanol at –20 °C and centrifugation at 18,000g

for 10 min at 4 °C. The resulting pellet was washed three times with 1 mL of 80% (v/v) ethanol. The pooled supernatants containing the oligosaccharides were dried in a vacuum centrifuge (Speed Vac concentrator, Savant Instruments Inc., USA). The glycan samples were desalted using Hypercarb cartridges (100 mg). Prior to use, the columns were washed three times with 500 µL 80% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid followed by three washes with 500 µL water. The samples were diluted with water to a final volume of 300 µL before loading onto the cartridges. The loaded cartridges were washed rigorously with water. Oligosaccharides were eluted with 1.2 mL 25% (v/v) acetonitrile containing 0.1% (v/v) trifluoroacetic acid. The eluate was neutralized with 2 M NH₄OH and dried in a Speed Vac concentrator.

Reduction and permethylation of the liberated N-glycans

N-glycans were reduced using NaBH₄ and permethylated.⁵⁴ The permethylated sample was used for MALDI/TOF-MS and ESI-MS/MS analysis.

In-gel tryptic digestion

Samples (ca. 20 µg) of CBP holoprotein were separated on 15% polyacrylamide gels and stained by Coomassie Brilliant Blue. Selected gel areas corresponding to the protein bands were excised and washed twice with 500 µL 50 mM NH₄HCO₃, and twice with 500 µL 50% (v/v) acetonitrile containing 0.1% (v/v) formic acid. Gel pieces were dried in a Speed Vac concentrator and re-hydrated after adding 100 µL of a trypsin solution (2 µg/mL in 50 mM NH₄HCO₃) followed by a second addition of 100 µL of the same trypsin solution after 2 h. After incubation for 7 h at 37 °C, the supernatants containing the tryptic peptides were transferred to a centrifuge tube. Gel pieces were washed twice with 500 µL 50 mM NH₄HCO₃ and twice with 500 µL 50% acetonitrile/0.1% formic acid, as above. Washes were pooled and dried in vacuum. Peptide mixtures were re-dissolved in 50 µL water and samples were divided into two equal parts. One part was subjected to PNGase F digestion as described above; the other part served as control. Trifluoroacetic acid at 0.1% (v/v) was added to the samples and tryptic peptides with and without PNGase F treatment were purified using ZipTip pipette tips (Millipore) containing C18 reversed-phase

material, according to the manufacturer's instructions. After washing with 0.1% (v/v) formic acid, elution of peptides was performed with 10 μ L 0.1% (v/v) formic acid in 60% (v/v) acetonitrile.

Enzymatic deglycosylation of glycopeptides for MS peptide sequencing

Approximately 1 nmol of the tryptic digest was dissolved in 50 μ L of 50 mM sodium phosphate buffer, pH 7.5, containing 10 mM EDTA and 0.02% NaN_3 . Two units of recombinant PNGase F (Roche Diagnostics GmbH, Mannheim, Germany) were added and the mixture was incubated for 12 h at 37 °C. After desalting, the deglycosylated peptides were analyzed as described below. All glycopeptides contained the glycosylation sequence Asn X Ser/Thr with Asn being converted to Asp by the action of PNGase F.

MALDI-TOF MS

Enzymatically liberated oligosaccharides were analyzed in the positive ion mode on a Bruker Ultraflex time-of-flight (TOF/TOF) instrument using the reflectron with 2,5-dihydroxybenzoic acid containing approximately 5 μ M NaCl as matrix. Reduced and permethylated glycans were similarly studied using a matrix of α -cyano-4-hydroxy-cinnamic acid. Samples of 1 μ L at a concentration of 1–10 pmol/ μ L were mixed with equal amounts of matrix solution. This mixture was spotted onto a stainless steel target and dried at room temperature before analysis.

Electrospray MS/MS

The (glyco)-peptides were analysed by ESI-MS/MS as described.⁵⁵ Briefly, the sample (~3 μ L dissolved in methanol/water 2:1 (v/v) with 1% formic) was placed in a gold-coated nanospray glass capillary (Protana, Odense, Denmark) and the tip of the capillary was placed orthogonally in front of the entrance hole of a quadrupole time-of-flight (Q-TOF 2) mass spectrometer (Micromass, Manchester, UK) equipped with a nanospray ion source. A voltage of ~1000 V was applied. The intact protein was studied in a similar way, dissolved in 2:1 (v/v) water/methanol containing 1% formic acid. The resulting spectrum, showing multiple charged protein ions, was deconvoluted using the Max.Ent.1 algorithm (Micromass). For collision induced dissociation (CID) experiments, parent ions were selectively

transmitted from the quadrupole mass analyser into the collision cell. Argon was used as the collision gas and the kinetic energy was set to 20-40 eV for fragmentation of the protonated (glyco)-peptides, whereas for the reduced and permethylated N-glycans detected as sodium adducts, a significant higher voltage of 60-90 eV was necessary to induce dissociation. The resulting daughter ions were then separated by an orthogonal time-of-flight (TOF) mass analyser. The acquired MS/MS spectra of proteolytic peptides were enhanced (Max. Ent. 3; Micromass) and used for manual *de novo* sequencing. Database searches⁵⁶ for similar proteins were performed using the fasta33 algorithm (<http://www.ebi.ac.uk/fasta33/>).

Transmission electron microscopy

Five μL of CBP solution in 50 mM Tris/HCl buffer, pH 7.4, was adsorbed onto the fresh glow-discharged copper grid (mesh 200) coated with parlodium/carbon film for 30 s, washed three times with quartz double-distilled water and once with 1% freshly diluted uranyl acetate; the uranyl acetate solution was filtered through a 0.2 μm filter before use. Samples were stained in 1% uranyl acetate for 10 sec. Excess stain on the grids was blotted and subsequently air-dried. The grids were examined in a transmission electron microscope Hitachi 7000 operated at 100 kV. Images were recorded at nominal magnifications of 50,000x to 100,000x on Kodak SO163 photographic film. The negatives were scanned into a computer using a Primescan D 7100 (Heidelberger Druckmaschinen AG, Heidelberg, Germany) for further processing.

Scanning transmission electron microscopy (STEM) and mass measurements

A HB-5 scanning transmission electron microscope (Vacuum Generators, East Grinstead, UK) interfaced to a modular computer system (Tietz Video and Image Processing Systems, Ganting, Germany) was used for imaging. Samples were prepared on 200-mesh-per-inch gold-plated copper grids, as previously.⁵⁹ Grids were washed on 4 droplets of quartz double-distilled water and either negatively stained (2% uranyl acetate) for structural examination or plunge frozen and freeze-dried for mass measurement. Grids were imaged at recording doses of 395 to 510 electrons/ nm^2 for the mass measurements (pixel size 0.908 nm) and doses of 7810 to 10200 electrons/ nm^2 for the negative stain microscopy (pixel size 0.326 nm). Beam-induced mass-loss was experimentally determined by repeatedly scanning of selected grid regions and

corrections. Tobacco mosaic virus (kindly supplied by R. Diaz-Avalos, Institute of Molecular Biophysics at Florida State University) was used for absolute mass calibration. Mass analysis was carried out with the IMPSYS program package.⁵⁹ In the final step, the mass values were corrected for beam-induced mass-loss, scaled as indicated by the mass standard, binned into a histogram and fitted with a single Gaussian curve. The illustrations included in Figure 7 b (insets) and c, were generated by imposing a Gaussian blur on the images and adjusting the contrast.

Single particle analysis and 3D reconstruction

A total of 680 variously oriented, negatively stained CBP particles were manually selected from the digitised TEM micrographs using the Boxer software of the EMAN program package.⁶⁰ Selected particles were aligned, classified by multi-variate statistical analysis and averaged (onto which no symmetry was imposed) for subsequent single particle reconstruction and 3D analysis. Determination of the symmetry of the CBP particle, initial model generation and the subsequent refinement of 3D model were performed using the EMAN software package. For 3D reconstruction several initial 3D models – 3-fold symmetry (C3), 6-fold symmetry (C6), dihedral (double 3-fold) symmetry (D3) and without any symmetry (non-symmetry) were generated. For all 3D models several cycles of iterative refinement were performed. Dihedral (D3) rotational symmetry provided the best convergence and was therefore used to obtain a starting model for the refined final 3D model. The 2D projection images generated on re-projection of this model served as reference images for particle classification. After classification the particles were manually aligned and averaged to generate the refined 3D model. This procedure was iteratively repeated 8 times. The final 3D model of CBP was based on 80 class averages. 3D rendering and manipulation of the model were performed using the Chimera software.⁶¹

Acknowledgements

We thank Ariel Lustig for studying CBP by analytical ultracentrifugation and Lars Prade for data from dynamic light scattering. The STEM microscopy was supported by the Maurice E. Müller Institute of Switzerland and Swiss National Foundation grant number SNF 501 221 to Andreas Engel who we thank for aiding this project.

References

- 1 Hughes, A. L. (1999). Evolution of the arthropod prophenoloxidase/hexamerin protein family. *Immunogenetics* 49, 106-114.
- 2 Burmester, T., Massey, Jr., H. C., Zakharkin S. O. & Benes, H. (1998). The evolution of hexamerins and the phylogeny of insects. *J. Mol. Evol.* 47, 93-108.
- 3 Burmester, T. & Hankeln, T. (2007). The respiratory proteins of insects. *J. Insect Physiol.* 53, 285-294.
- 4 Telfer, W. H., Keim, P. S. & Law, J. H. (1983). Arylphorin, a new protein from *Hyalophora cecropia*: comparison with calliphorin and manducin. *Insect Biochem.* 13, 601-613.
- 5 Tojo, S., Nagata, M. & Kobayashi, M. (1980). Storage proteins in the silkworm, *Bombyx mori*. *Insect Biochem.* 10, 289-303.
- 6 Ryan, R.O., Keim, P.S., Wells, M.A. & Law, J.H. (1985). Purification and properties of a predominantly female-specific protein from the hemolymph of the larva of the tobacco hornworm, *Manduca sexta*. *J. Biol. Chem.* 260, 782-787.
- 7 Telfer, W. H. & Kunkel, J. G. (1991). The function and evolution of insect storage hexamers. *Annu. Rev. Entomol.* 36, 205-228.
- 8 Beintema, J. J., Stam, W. T., Hazes, B. & Smidt, M. P. (1994). Evolution of arthropod hemocyanins and insect storage proteins (hexamerins). *Mol. Biol. Evol.* 11, 493-503.
- 9 Hagner-Holler, S., Pick, Ch., Girgenrath, S., Marden, J. H. & Burmester, T. (2007). Diversity of stonefly hexamerins and implication for the evolution of insect storage proteins. *Insect Biochem. Mol. Biol.* 37, 1064-1074.
- 10 Volbeda, A. & Hol, W.G.J. (1989). Crystal structure of hexameric haemocyanin from *Panulirus interruptus* refined at 3.2 Å resolution. *J. Mol. Biol.* 209, 249-279.
- 11 Magnus, K. A., Lattman, E. E., Volbeda, A. & Hol, W. G. J. (1991). Hexamers of subunit II from *Limulus* hemocyanin (a 48-mer) have the same quaternary structure as whole *Panulirus* hemocyanin molecules. *Proteins: Struct. Funct. Genet.* 9, 240-247.

- 12 Munn, E. A., Feinstein, A. & Greville, G. D. (1971). The isolation and properties of the protein calliphorin. *Biochem. J.* 124, 367-374.
- 13 Braun, R. P. & Wyatt, G.R. (1996). Sequence of the hexameric juvenile hormone-binding protein from the hemolymph of *Locusta migratoria*. *J. Biol. Chem.* 271, 31756-31762.
- 14 Magee, J., Kraynack, N., Massey, H.C. Jr. & Telfer, W.H. (1994). Properties and significance of a riboflavin-binding hexamerin in the hemolymph of *Hyalophora cecropia*. *Arch. Insect Biochem. Physiol.* 25, 137-157.
- 15 Miura, K., Shinoda, T., Yura, M., Nomura, S., Kamiya, K., Yuda, M. & Chinzei, Y. (1998). Two hexameric cyanoprotein subunits from an insect, *Riptortus clavatus*. Sequence, phylogeny and developmental and juvenile hormone regulation. *Eur. J. Biochem.* 258, 929-940.
- 16 Jones, G., Brown, N., Manczak, M., Hiremath, S. & Kafatos, F. C. (1990). Molecular cloning, regulation, and complete sequence of a hemocyanin-related, juvenile-hormone-suppressible protein from insect hemolymph. *J. Biol. Chem.* 265, 8596-8602.
- 17 Bitonidi, M. M. G., Nascimento, A. M., Cunha, A. D., Guidugli, K. R., Nunes, F. M. F. & Simões, Z. L. P. (2006). Characterization and expression of the Hex 110 gene encoding a glutamine-rich hexamerin in the honey bee, *Apis mellifera*. *Arch. Insect Biochem. Physiol.* 63, 57-72.
- 18 Levenbook, L. (1985). Insect storage proteins. In: Kerkut, G.A., Gilbert, L.I. (Eds.), *Comprehensive Insect Biochemistry, Physiology and Pharmacology*, vol. 10. Pergamon Press, Oxford, pp. 307-446.
- 19 Burmester, T. (1999). Evolution and function of the insect hexamerins. *Eur. J. Entomol.* 96, 312-225.
- 20 Zhou, X., Oi, F. M. & Scharf, M. E. (2006). Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4499-4504.
- 21 Zhou, X., Tarver, M. R., Bennett, G. W., Oi, F. M. & Scharf, M. E. (2006). Two hexamerin genes from the termite *Reticulitermes flavipes*: sequence, expression, and proposed functions in caste regulation. *Gene* 376, 47-58.
- 22 Blackburn, M. B., Loeb, M. J., Clark, E. & Jaffe, H. (2004). Stimulation of midgut stem cell proliferation by *Manduca sexta* alpha-arylphorin. *Arch. Insect Biochem. Physiol.* 55, 26-32.

- 23 Smagghe, G., Vanhassel, W., Moeremans, C., De Wilde, D., Goto, S., Loeb, M. J., Blackburn, M. B. & Hakim, R. S. (2005). Stimulation of midgut stem cell proliferation and differentiation by insect hormones and peptides. *Ann. N. Y. Acad. Sci.* 1040, 472-475.
- 24 Hakim, R. S., Blackburn, M. B., Corti, P., Gelman, D. B., Goodman, C., Elsen, K., Loeb, M. J., Lynn, D., Soin, T. & Smagghe G. (2007). Growth and mitogenic effects of arylphorin *in vivo* and *in vitro*. *Arch. Insect Biochem. Physiol.* 64, 63-73.
- 25 Memmel, N. A., Trewitt, P. M., Silhacek, D. L. & Kumaran, A. K. (1992). Nucleotide sequence and structure of the arylphorin gene from *Galleria mellonella*. *Insect Biochem. Molec. Biol.* 22, 333-342.
- 26 Beresford, P. J., Basinski-Gray, J. M., Chiu, J. K. C., Chadwick, J. S. & Aston, W. P. (1997). Characterization of hemolytic and cytotoxic gallysins; a relationship with arylphorins. *Dev. Comp. Immunol.* 21, 253-266.
- 27 Kayser, H. (2005). Lipocalins and structurally related ligand-binding proteins. In: Gilbert, L.I., Iatrou, K., Gill, S. (Eds.), *Comprehensive Molecular Insect Science*, vol. 4. Elsevier, Oxford, pp. 267-306.
- 28 Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. & Kayser, H. (1987). Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J. Mol. Biol.* 198, 499-513.
- 29 Holden, H. M., Rypniewski, W. R., Law, J. H. & Rayment, I. (1987). The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 Å resolution. *EMBO J.* 6, 1565-1570.
- 30 Chinzei, Y., Haruna, T., Miura, K., Numata, H. & Nakayama, S. (1990). Purification and characterization of biliverdin-associated cyanoprotein from eggs and hemolymph of the bean bug, *Riptortus clavatus* (Heteroptera: Alydidae). *Insect Biochem.* 20, 545-555.
- 31 Edsall, J. (1953). The size, shape and hydration of protein molecules. In: Neurath, H., Bailey, K. (Eds.), *The Proteins: Chemistry, Biological Activity, and Methods*. Academic Press, New York, pp. 549-726.

- 32 Suter, F., Kayser, H. & Zuber, H. (1988). The complete amino-acid sequence of the bilin-binding protein from *Pieris brassicae* and its similarity to a family of serum transport proteins like the retinol-binding protein. *Biol. Chem. Hoppe-Seyler* 369, 497-505.
- 33 Goodwin, T. W. & Morton, R. A. (1946). The spectrophotometric determination of tyrosine and tryptophan in proteins. *Biochem. J.* 40, 628-632.
- 34 Kim, S., Hwang, S.K., Dwek, R.A., Rudd, P.M., Ahn, Y.H., Kim, E.-H., Cheong, C., Kim, S.I., Park, N.S. & Lee, S.M. (2003). Structural determination of the N-glycans of a lepidopteran arylphorin reveals the presence of a monoglucosylated oligosaccharide in the storage protein. *Glycobiology* 13, 147-157.
- 35 Markl, J., Burmester, T., Decker, H., Savel-Niemann, A., Harris, J. R., Söling, M., Naumann, U. & Scheller, K. (1992). Quaternary and subunit structure of *Calliphora* arylphorin as deduced from electron microscopy, electrophoresis, and sequence similarities with arthropod hemocyanin. *J. Comp. Physiol. B*, 162, 665-680.
- 36 Kramer, S.J., Mundall, E.C. & Law, J.H. (1980). Purification and properties of manducin, an amino acid storage protein of the haemolymph of larval and pupal *Manduca sexta*. *Insect Biochem.* 10, 279-288.
- 37 Haunerland, N.H. & Bowers, W.S. (1986). Arylphorin from the corn earworm, *Heliothis zea*. *Insect Biochem.* 16, 617-625.
- 38 Ryan, R.O., Anderson, D.R., Grimes, W.J. & Law, J.H. (1985). Arylphorin from *Manduca sexta*: carbohydrate structure and immunological studies. *Arch. Biochem. Biophys.* 243, 115-124.
- 39 Tomiya, N, Narang, S., Lee, Y. C. & Betenbaugh, M. J. (2004). Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. *Glycoconjugate* 21, 343-360.
- 40 Fujii, T., Sakurai, H., Izumi, S. & Tomino, S. (1989). Structure of the gene for the arylphorin-type storage protein SP 2 of *Bombyx mori*. *J. Biol. Chem.* 264, 11020-11025.
- 41 Jamroz, R.C., Beintema, J.J., Stam, W.T. & Bradfield, J.Y. (1996). Aromatic hexamerin subunit from adult female cockroaches (*Blaberus discoidalis*): molecular cloning, suppression by juvenile hormone, and evolutionary perspectives. *J. Insect Physiol.* 42, 115-124.

- 42 Zheng, Y., Yoshiga, T. & Tojo, S. (2000). cDNA cloning and deduced amino acid sequences of three storage proteins in the common cutworm, *Spodoptera litura*. *Appl. Entomol. Zool.* 35, 31-39.
- 43 Ashfaq, M., Sonoda, S. & Tsumuki, H. (2007). cDNA characterization and expression analysis of two arylphorin-like hexameric protein genes from the diamondback moth, *Plutella xylostella* (L.). *Arch. Insect Biochem. Physiol.* 64, 175-185.
- 44 Jung, H., Kim, Y.H. & Kim, S. (2005). Structural basis for the presence of monoglucosylated oligosaccharide in mature glycoproteins. *Biochem. Biophys. Res. Commun.* 331, 100-106.
- 45 Wichertjes, T., Keegstra, W., Neuteboom, B., Hazes, B., Beintema, J.J. & van Bruggen, E.F. (1989). Crystallization properties and structure of *Panulirus interruptus* haemocyanin. *Eur J Biochem.* 184, 287-96.
- 46 Meissner, U., Stohr, M., Kusche, K., Burmester, T., Stark, H., Harris, J.R., Orlova, E.V & Markl, J. (2003). Quaternary structure of the European spiny lobster (*Palinurus elephas*) 1x6-mer hemocyanin from cryoEM and amino acid sequence data. *J Mol Biol.* 325, 99-109.
- 47 Kayser, H. (1985). Pigments. In: Kerkut, G.A., Gilbert, L.I. (Eds.), *Comprehensive Insect Biochemistry, Physiology and Pharmacology*, vol. 10. Pergamon Press, Oxford, pp. 367-415.
- 48 Scheer, H. & Kayser, H. (1988). Conformational studies of biliproteins from the insects *Pieris brassicae* and *Cerura vinula*. *Z. Naturforsch.* 43c, 84-90.
- 49 McDonagh, A. F. (1979). Bile pigments: bilatrienes and 5,15-biladienes. In: *The Porphyrins* (Dolphin, D., editor), Academic Press, New York, USA, 6, 293-491.
- 50 Mann, K. & Siedler, F. (2004). Ostrich (*Struthio camelus*) eggshell matrix contains two different C-type lectin-like proteins. Isolation, amino acid sequence, and posttranslational modifications. *Biochem. Biophys. Acta* 1696, 41-50.
- 51 Donovan, J. W. (1964). The spectrophotometric titration of the sulfhydryl and phenolic groups of aldolase. *Biochemistry* 3, 67-74.
- 52 Bencze, W. L. & Schmid, K. (1957). Determination of tyrosine and tryptophan in proteins. *Analyt. Chem.* 29, 1193-1196.
- 53 Chaplin, M. F. (1982). A rapid and sensitive method for the analysis of carbohydrate components in glycoproteins using gas-liquid chromatography. *Anal. Biochem.* 123, 336- 341.

- 54 Anumula, K. R. & Taylor, P. B. (1992). A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates. *Anal. Biochem.* 203, 101–108.
- 55 Nimtz, M., Conradt, H. S. & Mann, K. (2004). LacdiNAc (GalNAc β 1-4GlcNAc) is a major motif in N-glycan structures of the chicken eggshell protein ovocleidin-116. *Biochim. Biophys. Acta* 1675, 71-80.
- 56 Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85, 2444- 2448.
- 57 Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J.* 78, 1606-1619.
- 58 Machaidze, G. & Lustig, A. (2006). SEGAL, a semi-automatic program for fitting sedimentation equilibrium patterns from analytical ultracentrifugation. *J. Biol. Phys. Chem.* 6, 91–102.
- 59 Müller, S. A., Goldie, K. N., Bürki, R., Häring, R. & Engel, A. (1992). Factors influencing the precision of quantitative scanning transmission electron microscopy. *Ultramicroscopy* 46, 317-334.
- 60 Ludtke, S., Baldwin, P. & Chiu, W. (1999). EMAN: a semi-automated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 128, 82-97.
- 61 Pettersen, E., Goddard, D., Huang, C., Couch, G., Greenblatt, D., Meng, C. & Ferrin, E. (2004) UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. <http://www.cgl.ucsf.edu/chimera/>.

Figure legends

Figure 1. UV/VIS absorption spectrum of CBP. The absorbance maxima at 373 nm and 666 nm are due to association of the protein with a biliverdin-based novel ligand.

Figure 2. Size-exclusion chromatography of CBP on TSK HW 55-S with 50 mM Tris/HCl (pH 7.4). Open symbols represent protein standards: 1, thyroglobulin; 2, ferritin; 3, catalase; 4, aldolase; the filled symbol marks CBP, which behaved as a 430-kDa protein.

Figure 3. Deconvoluted ESI mass spectrum of CBP. The heterogeneity of the protein is clearly visible. The mass increments of approximately 162 Da between the dominant peaks suggest the presence of glycoforms.

Figure 4. Analysis of CBP by SDS-PAGE using a 10 % gel and MES buffer. Lanes 1-4: samples from different fractions from chromatofocusing followed by gel permeation chromatography. The arrow indicates the position of the CBP band. Sample loads ranged between 13 and 18 µg. Lane 5: protein standards.

Figure 5. MALDI mass spectrum of the enzymatically liberated native glycans of CBP. The molecular ions (sodium adducts) are compatible with monosaccharide compositions of Hex₈₋₁₀HexNAc₂ strongly suggesting the presence of high-mannose type glycans. The most likely structures of individual glycans, confirmed by carbohydrate compositional analysis and MS/MS analysis after reduction and permethylation, are presented. Red squares, N-acetylglucosamine; green circles, mannose; blue rhombs, glucose.

Figure 6. TEM images of negatively stained CBP at concentrations of 0.04 mg/ml (a) and 0.014 mg/ml (b). The round and rectangular projections arising from the main particle population correspond to top and side views of the complex, respectively, indicating that it is about 13 nm in diameter (top views and

length of the rectangles) and 9 nm high (width of the rectangle). A fraction of smaller projections document the presence of particles with the same diameter but only half the height (probably trimeric half-CBP). (c) Enlarged views showing some structural details and one of the smaller projections (bottom). The scale bar corresponds to 100 nm in (b), and to 10 nm in (c).

Figure 7. STEM analysis of CBP. (a) STEM dark-field image of the unstained protein particles and histogram displaying the mass distribution obtained from the sample. The Gaussian indicates a mass of 471 ± 47 kDa ($n = 1541$) for the main population of complexes. (b) Single-shot, STEM dark-field image of negatively stained CBP showing both rectangular side-views and more circular top-views of the particles. Insets: the adjacent particle displayed to enhance the main features. (c) Four side- and one top-view image of the CBP complex recorded at a magnification of 1000,000x. Six domains arranged in two triangles of equivalent regions can be distinguished in the latter. The view lower right, prepared as the insets in (b), aids their identification. The scale bar corresponds to 50 nm in (a), 20 nm in (b) and the boxes are 20 nm wide in (c).

Figure 8. 3D reconstruction of the hexameric CBP complex. Projections from 680 manually collected single complexes were aligned, classified, averaged and used to generate a 3D model. (a) Top-view (left panel) and side-view (right panel) from class averages. (b) Corresponding re-projections of a top and a side view of the final model for comparison. (c) 3D model generated from the average projections. The scale bar corresponds to 10 nm.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8

Table 1. Amino acid composition of CBP. Data of acid hydrolysates represent mean values from duplicate analyses. Trp was determined by spectrophotometry (n = 4). Numbers of residues were calculated for a subunit mass of 85 kDa, as obtained by mass spectrometry.

Amino acid	Mol%	Mol / subunit (nearest integer)
<hr/>		
ASX	11.9	74
THR	5.0	31
SER	4.3	27
GLX	12.7	79
GLY	4.2	26
ALA	4.7	29
VAL	5.8	36
MET	0.8	5
ILE	3.7	23
LEU	6.4	40
TYR	9.3	58
PHE	9.6	60
HIS	2.4	15
LYS	7.2	45
ARG	4.6	29
PRO	5.2	32
TRP	2.0	13
<hr/>		
Total	99.9	622