MOLECULAR STUDIES OF HEMOCYANIN EXPRESSION IN THE DUNGENESS CRAB

by

GREGOR DURSTEWITZ

A DISSERTATION

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 1996

"Molecular Studies of Hemocyanin Expression in the Dungeness Crab," a dissertation prepared by Gregor Durstewitz in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology. This dissertation has been approved and accepted by:

_____

Dr. Nora B. Terwilliger, Chair of the Examining Committee

_____

Date

Committee in charge:

Dr. Nora B. Terwilliger, Chair
Dr. Roderick Capaldi
Dr. Ry Meeks-Wagner
Eric Schabtach
Dr. Kensal van Holde
Dr. Tom Stevens

Accepted by:

_____

Vice Provost and Dean of the Graduate School

An Abstract of the Dissertation of

Gregor Durstewitz for the degree of Doctor of Philosophy in the Department of Biology to be taken June 1996

Title: MOLECULAR STUDIES OF HEMOCYANIN EXPRESSION IN THE DUNGENESS CRAB

Approved: _____
Dr. Nora B. Terwilliger

This study investigates developmentally regulated changes in the expression of the copper based respiratory protein hemocyanin (Hc) in the Dungeness crab (*Cancer magister*). Hc gene expression was studied by Northern blot analysis. All six protein subunits were purified and their amino-terminal sequences determined. Subunit-specific oligonucleotide primers were designed based on these amino-terminal sequences and on a conserved region near the active site. These primers were then used to PCR-amplify stretches of cDNA coding for developmentally regulated Hc subunit 6 to be used as subunit-specific probes in Northern blots. Animals were raised under controlled conditions, and total RNA was isolated from 13 developmental stages and 6 tissue types, run on formaldehyde agarose gels, blotted onto nylon membranes and probed with radioactive ^{32}P-labeled adult Hc-

specific cDNA probes. Results indicate that adult Hc biosynthesis occurs in hepatopancreas tissue only and is initiated during the 6th juvenile instar stage, as indicated by the appearance of subunit 6 mRNA. A model is proposed to explain the observed changes in subunit stoichiometries between juvenile and adult Hc.

cDNA coding for developmentally regulated Hc subunit 6 and another putative Hc subunit obtained from a cDNA library screen were sequenced with the dideoxy method. The complete cDNA sequence of subunit 6 showed an open reading frame of 650 amino acids homologous in sequence to other arthropodan Hcs. Functional domains within the proteins were identified, and both were aligned with proteins displaying apparent sequence similarities. A comparison of structural parameters (predicted hydrophilicities, surface probabilities and regional backbone flexibilities) provided evidence for a remarkable degree of structural conservation among crustacean Hcs, chelicerate Hcs, insect hexamerins and arthropodan prophenoloxidases.

Parsimony analysis of the aligned sequences allowed a phylogenetic reconstruction of their evolutionary history. Confidence limits were established with the bootstrap approach. The most parsimonious phylogenetic tree consistent with the dataset identified crustacean Hcs, insect hexamerins, chelicerate Hcs and prophenoloxidases as a monophyletic group relative to molluscan Hcs and non-

arthropodan tyrosinases. Results for individual clades were evaluated and discussed in the light of the evolutionary history of the Hc gene family.

CURRICULUM VITAE

NAME OF AUTHOR: Gregor Durstewitz

PLACE OF BIRTH: Frankfurt am Main, Germany

DATE OF BIRTH: March 30, 1960

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon
Freie Universität Berlin
Universität Tübingen

DEGREES AWARDED:

Doctor of Philosophy in Biology, 1996,
University of Oregon
Diplom (Master of Science) in Biochemistry, 1987,
Freie Universität Berlin
Vordiplom (Bachelor of Science) in Biochemistry, 1983,
Universität Tübingen

AREAS OF SPECIAL INTEREST:

Zoology
Marine Biology
Natural History

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, Oregon Institute of Marine Biology and Department of Biology, University of Oregon, Eugene, 1989-1996

Research Assistant, Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany, 1986-1987

Teaching Assistant, University of Tübingen Medical School, Tübingen, Germany, 1985

Military Service, 1978-1980

PUBLICATIONS:

Brink, L. and Durstewitz, G. (1996) Field guide to marine birds and mammals of Oregon's Bay Area. Oregon Marine Studies Association, Charleston.

Durstewitz, G. and Terwilliger, N.B. (1995) Developmental changes in hemocyanin expression in the Dungeness crab: Northern blots and cDNA sequence of a developmentally regulated subunit. Am. Zool. 35, 65A.

Durstewitz, G. and Terwilliger, N.B. (1995) Northern blot analysis of the differential expression of hemocyanin subunits in various tissues and developmental stages of the Dungeness crab *(Cancer magister)*. Physiol. Zool. 68, 82.

Durstewitz, G., Joslyn, A., Otoshi, C. and Torchin, M. (1993) Guide to intertidal invertebrates. An introduction to tidepool life on the Oregon coast. Oregon Marine Studies Association, Charleston.

Durstewitz, G., O'Brien, K. and Terwilliger, N.B. (1992) Specific DNA probes for the analysis of hemocyanin subunit expression in *Cancer magister*. Am. Zool. 32, 34A.

Durstewitz, G. and Tesche, B. (1987) Adsorption of macromolecules onto support films for electron microscopy. Eur. J. Cell Biol. 44, suppl. 19, 15.

Terwilliger, N.B. and Durstewitz, G. (1996) Molecular studies of the sequential expression of a respiratory protein during crustacean development, in: *Molecular Zoology: Advances, Strategies and Protocols*, eds: Ferraris, J.D. and Palumbi, S.R., Wiley-Liss, 353-368.

ACKNOWLEDGEMENTS

This dissertation is dedicated to my parents Hildegard and Dr. Josef Durstewitz who--sometimes to my surprise--have whole-heartedly supported my every move. Many thanks go to my boss Dr. Nora Terwilliger for continuous support throughout the project and for her tolerance in view of my annual northward migrations. I salute my teachers Dr. Christian Bardele and Dr. Wilhelm Harder for inspiring an interest in Zoology and Marine Biology. I am deeply grateful to Dr. Yi-Lin Yan for taking such an interest in my study; without her support and on the spot technical advice in all aspects of molecular biology this study would not have been possible. Bob "Harley" Hanner pointed out the pitfalls of phylogenetic reconstruction. Kristin O'Brien and Dr. Margaret Ryan chipped in with animal husbandry and lab chores. Finally, and maybe most importantly, I thank Clete Otoshi, Kraig Slack and Nicole Apelian for putting up with me and being such good friends through all these years.

This study was supported by National Science Foundation grants DCB 8908362 and IBN 9217530 to NBT.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

Chapter I

Chapter II

Chapter III

Page

Chapter IV

CHAPTER I

INTRODUCTION

The process of respiration is essential for life, both at the cellular and the organismal level. In fact, gas exchange is so fundamental that it was used as one of the criteria in the search for organic life on Mars in the experiments of NASA's VIKING mission in 1976. The term *respiration* generally implies the use of oxygen (O_2) in cellular oxidation. Yet, for most of our own planet's history, O_2 was not readily available.

On prehistoric earth, about 2 billion years ago, O_2 levels in the essentially anaerobic atmosphere slowly began to rise; cyanobacteria had developed O_2 producing water-based photosynthesis. This novel process, however, produced as byproducts extremely reactive oxygen radicals (O^-) that proved to be potent cellular toxins. Procaryotic life responded by first developing an antidote to oxygen poisoning (superoxide dismutase, SOD), and then employing molecular oxygen (O_2) in a new metabolic process, biological oxidation or "respiration". The use of O_2 was a major step in the evolution of life on earth because the respiratory chain with O_2 as the final electron acceptor extracts 18

times as much energy from glucose as does anaerobic fermentation. Today, most organisms use O_2 as the final electron acceptor in the respiratory chain.

Organisms with a low metabolic rate or a high surface to volume ratio may rely exclusively on the process of diffusion of O_2 across their body surface to provide their cells with O_2. Most higher animals, however, have developed specific respiratory organs to allow more efficient exchange of respiratory gases. These organs, gills, lungs and tracheal systems, are poised between the internal milieu and the environment.

Most aquatic organisms rely on gills for gas exchange between their body and the aquatic medium. Gills are thin-walled evaginations or appendages of the body surface, often perfused by capillary nets, that allow diffusion of O_2 from the water to the animal's circulatory system. Air breathing creatures employ lungs. Those are invaginations of the body that - in the case of vertebrates - are evolutionarily derived from the anterior part of the digestive tract. To increase respiratory surfaces, the airways usually branch out and end in multiple small thin-walled chambers, the alveoles, where gas exchange between the air and the circulatory system takes place. The third major type of respiratory organ is the tracheal system used by many terrestrial arthropods, especially insects. The tracheal tubes are invaginations of the chitinized body surface that

branch out and permeate the whole animal, allowing air direct access to the respiring tissues.

In order to increase transport efficiency for the respiratory gases, O_2 and CO_2, beyond the limitations of mere diffusion, many animals use circulatory systems for active gas transport between the respiratory organs and the tissues. To overcome the intrinsically low solubility of O_2 in aqueous media, the blood often contains O_2 transporting respiratory proteins. These proteins reversibly bind O_2 at the respiratory surfaces and transport it via the circulatory system to the respiring cells within where O_2 is released. Respiratory proteins can occur extracellularly, i.e., freely suspended in the blood or hemolymph, or they can be contained within specific circulating blood cells. Frequently these respiratory proteins are large multisubunit proteins that show pronounced cooperativity and allosteric regulation. The assembly of respiratory proteins into high molecular weight aggregates or their packaging into blood cells both allow for substantial O_2 transport capacities while maintaining moderate osmotic values within the circulatory system.

Three main classes of respiratory proteins have been described in animals: The hemoglobins, the hemerythrins and the hemocyanins. Hemoglobin is a ubiquitous molecule. Along with its monomeric cousin myoglobin, it occurs in virtually all animal phyla as well as in plants, protists and

bacteria. A notable exception are the archaebacteria (Wittenberg, 1992). In hemoglobin, O_2 is reversibly bound to an Fe^{2+} contained within a porphyrin ring system called a heme. This heme disc is embedded in a precise location within a single polypeptide chain. The Fe^{2+} is complexed by four nitrogen atoms in the center of the heme disc, as well as by a specific histidine residue of the surrounding polypeptide chain. The remaining 6^{th} coordination site reversibly binds O_2. The heme group is responsible for O_2 binding and the characteristic absorption spectrum that makes hemoglobin containing solutions like blood appear red, while the polypeptide portion of the molecule, the globin, provides substrate specificity and allosteric regulation. In vertebrates, hemoglobin is contained within red blood cells and each hemoglobin molecule is composed of four such subunits, each with the potential of binding one molecule of O_2. The large multisubunit extracellular hemoglobins of many invertebrates are sometimes called erythrocruorins. In some annelids, the peripheral vinyl group in the heme disc is replaced by a formyl group. These pigments appear green in solution and hence are called chlorocruorins.

Hemerythrins occur, contained within pink blood cells, in brachiopods, sipunculids, priapulids and one family of annelids. Here Fe^{2+} is not complexed within a heme disc, but instead directly bound to amino acid side chains. In each subunit, two Fe^{2+} cooperate to bind one molecule O_2.

Hemerythrins appear purple when oxygenated and can form multisubunit complexes (often octamers; Kurtz, 1986).

Hemocyanin (Hc), the "blue protein" (Fredericq, 1878), is the respiratory protein responsible for O_2 transport in many arthropods and molluscs. It occurs freely dissolved in the hemolymph of those phyla. The metal responsible for reversible O_2 binding is not iron but copper. Two Cu^+ ions are bound by amino acid side chains (the distal nitrogens of three histidine residues, see below) and cooperate in the binding of one O_2. During oxygenation, a peroxide (O_2^{2-}) bridge is formed between the Cu^+ ions, oxidizing them to Cu^{2+}. Crystallographic analysis of oxygenated *Limulus* Hc (Magnus et al., 1994) has recently provided direct evidence that O_2 is bound between the two Cu^{2+} ions in a transverse orientation ($\eta^2:\eta^2$, the two oxygen atoms lying in a plane perpendicular to the Cu^{2+}-Cu^{2+} axis), forming a Kitajima complex (Kitajima et al., 1992):

In addition to the absorbance at 280 nm due to the aromatic amino acids of the peptide portion of the molecule, the absorption spectrum of oxygenated Hc displays two additional peaks at 345 and 560-600 nm. The latter one explains the blue color of oxy-Hc in solution. Upon deoxygenation, Hc-containing solutions become colorless.

While mollusc and arthropod Hcs share certain features, like the structure of their active site and a similar absorption spectrum, in many ways they are quite different. Molluscan Hcs are large, cylindrical molecules composed of 10 subunits or multiples thereof. Each subunit is a polypeptide chain with a molecular weight of about 400 kDa. It contains 7 (*Octopus, Nautilus*) or 8 functional units or domains. Each domain has one O_2 binding site (Lontie et al., 1973; van Holde and Miller, 1995).

Arthropodan Hcs, on the other hand, are composed of heterogeneous subunits with molecular weights of about 75 kDa each. A single subunit contains two Cu binding sites, CuA and CuB. At each site, the Cu atom is coordinated with three histidine ligands (Volbeda and Hol, 1989a). Both sites participate in the binding of one molecule O_2. X-ray crystallography of Hc from *Panulirus* and *Limulus* (Volbeda and Hol, 1989b; Hazes et al., 1993) has shown that an arthropod Hc subunit consists of 3 domains (Fig.1), the quite variable and mainly α-helical domain 1, the highly conserved domain 2 containing the active site, and domain 3, the ß-barrel structure.

Arthropodan subunits self-assemble into hexamers or multiples thereof. In the hemolymph of the adult Dungeness crab, *Cancer magister,* we find 2-hexamer 25S Hc as well as 1-hexamer 16S Hc (Ellerton et al., 1970) as shown in Fig.2.

Fig.1: Structure of *Limulus* hemocyanin according to X-ray crystallography at 2.18 Å resolution (Hazes, 1993, reprinted with the permission of Cambridge University Press). A, whole subunit. B, domain 1. C, domain 2. D, domain 3. Solid black spheres in center: Cu^{+} ions. Black sphere to the left: Cl^{-} ion. Shaded sphere: Ca^{2+} ion.

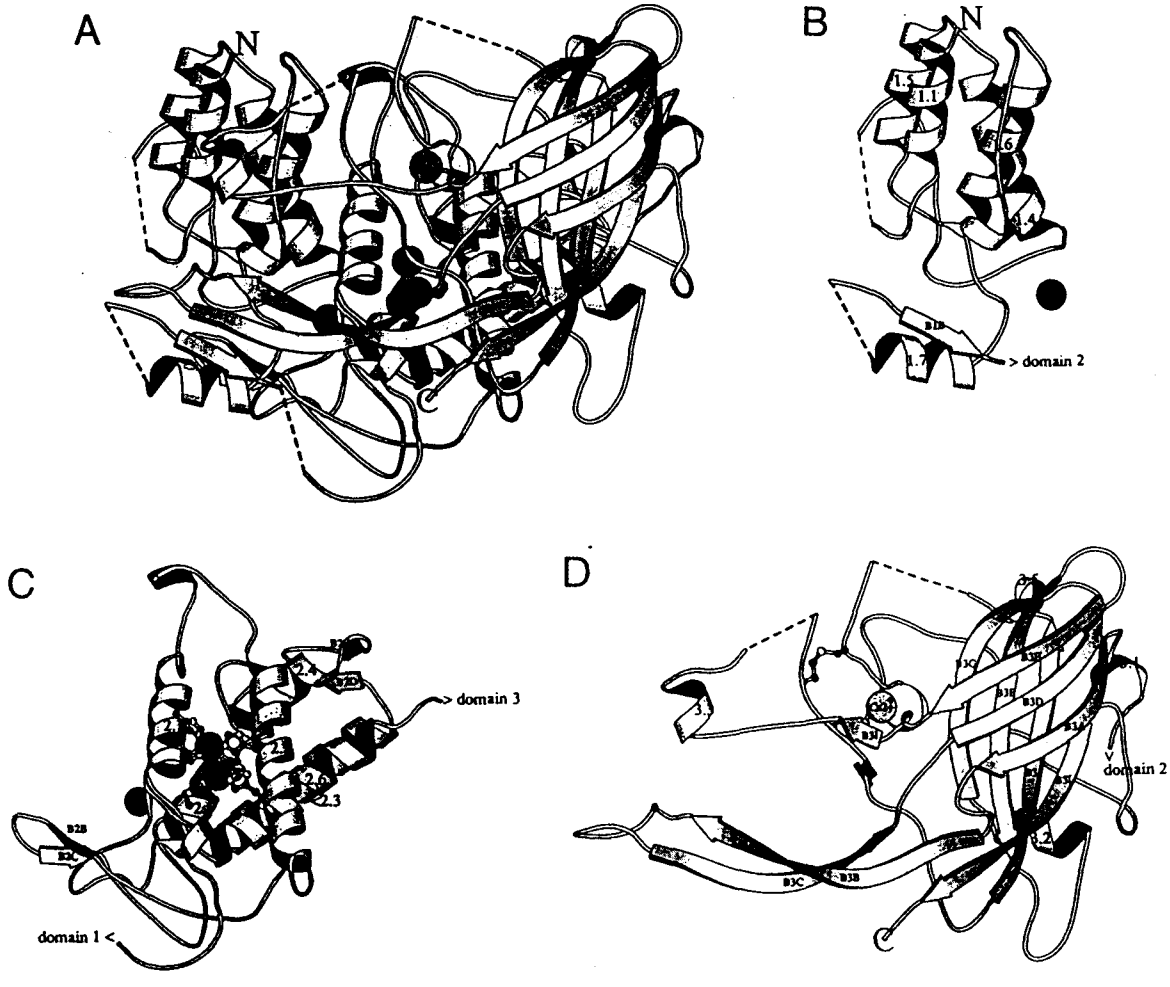
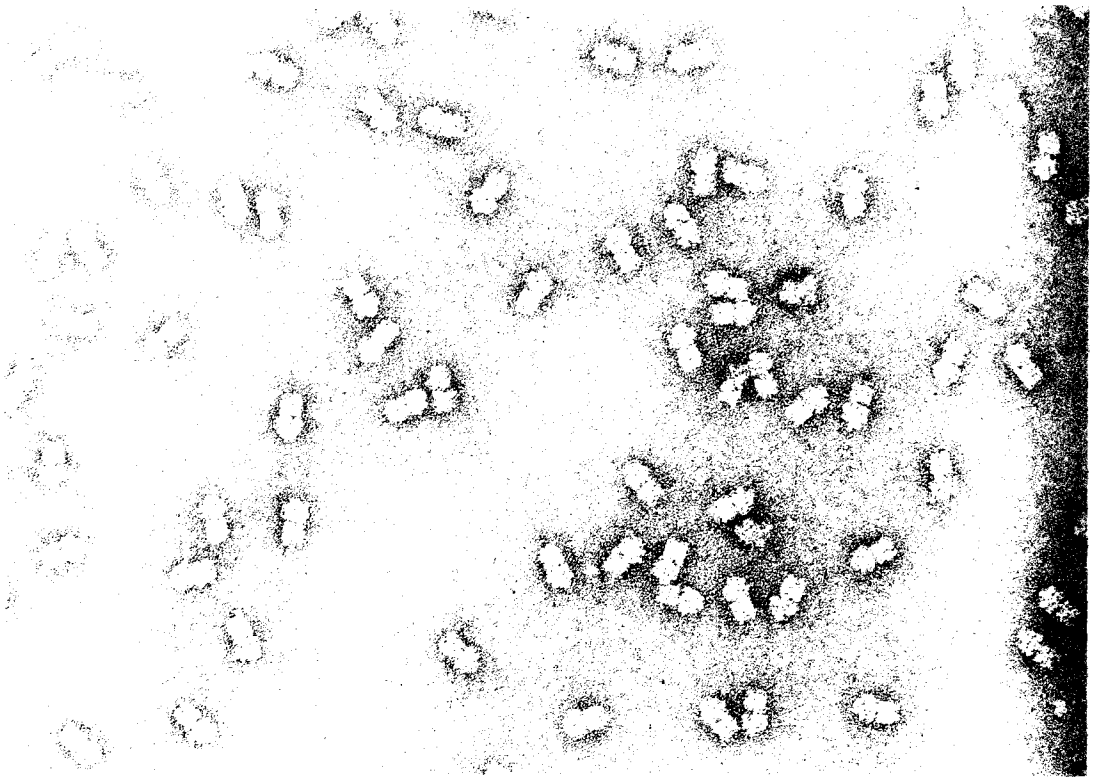
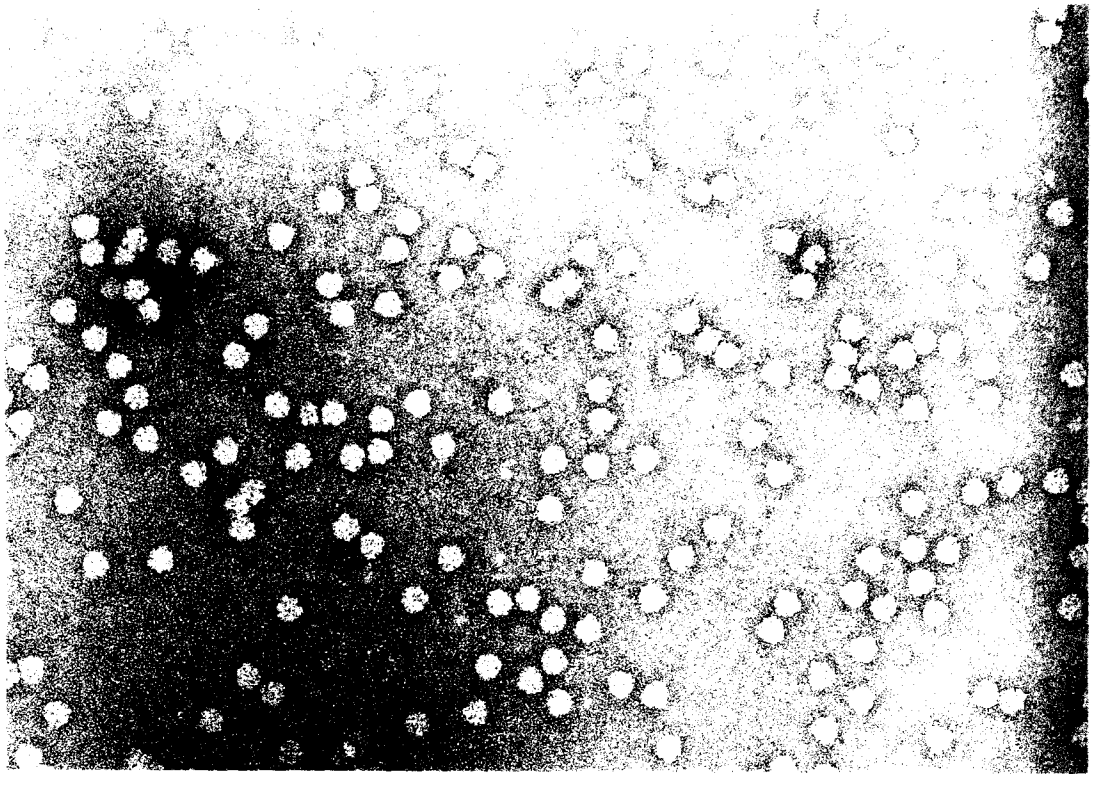

A
N
B
N
> domain 2
C
> domain 3
domain 1 <
D
domain 2

Fig.2: Electron micrograph of 25S (top) and 16S (bottom) hemocyanin from the Dungeness crab. Size of hexamer: 15 nm. Negative stain. Electron micrograph by Eric Schabtach.

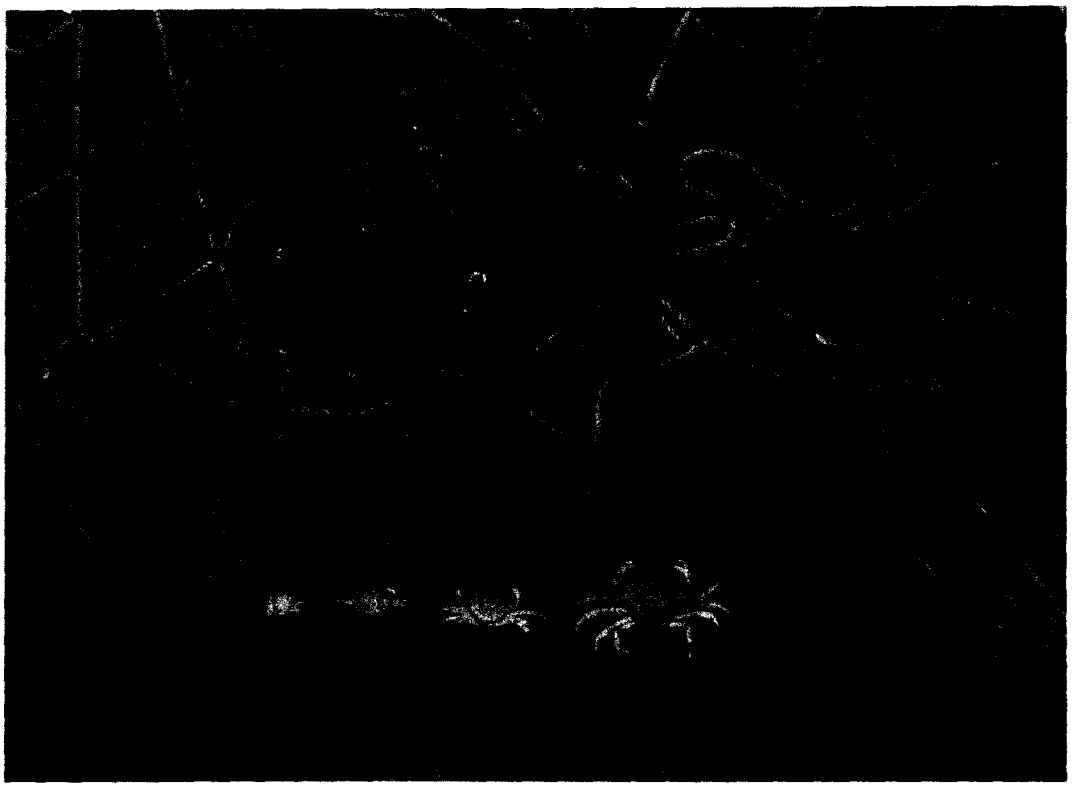




The life cycle of the Dungeness crab *Cancer magister* includes planktonic oceanic larval stages and benthic estuarine juveniles and adults (Fig.3). In the coastal waters of the Pacific Northwest, Dungeness crab embryos hatch in winter from eggmasses attached to the female's pleopods. After five zoeal stages in the offshore plankton, the final larval stage, the megalopa, enters Oregon coastal and estuarine waters in late April. The transport mechanism from oceanic to estuarine waters is not entirely clear but may involve diurnal vertical migrations in and out of the Ekman layer as well as onshore transport in surface slicks generated by internal waves on the thermocline (Shanks, 1983 and 1986; Hobbs et al., 1992). Megalopas are very active swimmers. Once in the estuary, they soon metamorphose into benthic juvenile instars. Growing rather rapidly, they molt periodically and reach maturity after about two years (MacKay, 1942).

Benthic juveniles and adults are exposed to much greater variations in salinity, temperature and O_2 partial pressure than their planktonic larval stages. In summer, rapid solar heating of exposed tidal flats during low tides can dramatically increase temperature and, through evaporation, salinity. In winter, freshets caused by the infamous Oregon rains can significantly reduce salinities of estuarine surface waters. On estuarine mudflats, the substrate turns anaerobic almost immediately below the

Fig.3: Adult Dungeness crab (*Cancer magister*) with juvenile instars. Photograph by Dr. R.C. Terwilliger.



surface. These challenges are particularly serious for juvenile crabs; they live right on the mudflats, while the adults generally prefer the estuary's deep subtidal channels where conditions are more uniform.

During its life cycle, the crab adapts to various environments and lifestyles by initiating behavioral, morphological, physiological and biochemical changes. One of these adaptations is an apparent change in both structure and function of its respiratory protein Hc during the juvenile-adult transition (Terwilliger and Terwilliger, 1982).

Adult *C. magister* 25S Hc is composed of six different types of subunits as shown by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), 2 stage PAGE and limited proteolysis (Larson et al., 1981). One of these, subunit 6, is absent in both 25S and 16S megalopa Hc (Terwilliger and Terwilliger, 1982) and typically doesn't appear in hemolymph Hc until sometime during the 6th juvenile instar stage. In addition, the stoichiometry of two other Hc subunits, 4 and 5, changes during the transition from larval to adult crab.

The developmental shift in subunit composition results in a new population of adult Hc molecules that have a higher affinity for oxygen than does juvenile Hc (Terwilliger et al., 1986, Terwilliger and Brown, 1993). This larval-adult shift in Hc is analogous, then, to the fetal-adult shift in

hemoglobins seen in humans and other mammals (Ingermann, 1993) and to that seen in some invertebrate extracellular hemoglobins (Schin et al., 1979; Heip et al., 1980). We hypothesized that the differences between juvenile and adult Hc are due to an ontogenetically regulated change in Hc gene expression in an effort to adapt to changing ecological conditions during the crab's life cycle (from a freeswimming planktonic larva to a benthic adult crab) and to accommodate a parallel ontogeny of ionic regulatory capabilities (Brown and Terwilliger, 1992, and Terwilliger and Brown, 1993).

We view Hc as a model system, illustrating how structural changes (in subunit composition) affect functional characteristics (like O_2 affinity) of the whole molecule. Several attributes make the Dungeness crab an ideal beast for our study: (1), both larval and adult stages of the animal can be collected easily and in sufficient numbers (zoeas from the eggmasses attached to the pleopods of pregnant females, megalopas with plankton nets from estuarine surface waters and adults of both sexes with baited crabrings or by SCUBA diving) (2), the megalopa stage of this species is markedly larger than any megalopa of the other Pacific Coast crustaceans and thus can be identified at a glance (3), the Dungeness crab can be raised through the different larval and juvenile stages under controlled conditions in a running seawater system at ambient temperature and salinity (4), their respiratory protein Hc

is easily obtained by bleeding a crab at one of its legjoints, and changes in an individual crab's hemolymph proteins can be analyzed over time by repeated sampling without killing the animal (Otoshi, 1994; Terwilliger and Otoshi, 1994). Thus, the Dungeness crab appears to fit the August Krogh Principle: "For a large number of problems there will be some animal of choice on which it can be most conveniently studied" (Krogh, 1929). This study is divided into three parts:

Chapter II describes how conserved functional domains in the respiratory protein Hc have been used to develop Hc-specific primers and probes as tools for the study of gene expression in the Dungeness crab.

Chapter III is a study of developmental changes in the expression of Hc during the life cycle of the Dungeness crab, using the subunit-specific probes developed in chapter II to investigate those changes at the molecular level. This is the first described case of an ontogenetic change in a copper protein, and a model is presented to explain the observed subunit stoichiometries in juvenile and adult Hc.

Chapter IV presents the complete cDNA- and protein sequence of the developmentally regulated Hc subunit investigated in chapter III as well as the sequence of another putative Hc subunit. Both sequences are aligned with proteins showing apparent sequence similarities. Functional domains are identified, and sequence-based predictions of

hydrophilicities, surface probabilities and regional backbone flexibilities are screened for signs of structural conservation. The alignment is evaluated by parsimony analysis, and the resulting most parsimonious phylogenetic tree is discussed in the light of the evolutionary history of the Hc family of proteins.

CHAPTER II

DEVELOPMENT OF PRIMERS AND PROBES FOR ANALYSIS OF *CANCER MAGISTER* HEMOCYANIN EXPRESSION

Previously Published in:

MOLECULAR ZOOLOGY: Advances, Strategies, and Protocols.

eds. Joan D. Ferraris and Stephen R. Palumbi,

Wiley-Liss, 1996, 353-368.

under the title:

MOLECULAR STUDIES OF THE SEQUENTIAL EXPRESSION OF A RESPIRATORY PROTEIN DURING CRUSTACEAN DEVELOPMENT

Nora Barclay Terwilliger

and

Gregor Durstewitz

Oregon Institute of Marine Biology and Department of Biology, University of Oregon, Eugene, OR 97403

ABSTRACT

Oxygenation properties of hemocyanin, the copper-containing oxygen transport protein found in arthropod hemolymph, are responsive to both the external milieu and the internal metabolism of the organism. During development from a swimming megalopa to a crawling crab, hemocyanin subunit composition and oxygen affinity change in the *Cancer magister*, the Dungeness crab. To understand the molecular mechanisms responsible for these ontogenetic changes, we are investigating patterns of tissue-specific and developmental stage-specific hemocyanin expression. To identify site of biosynthesis and onset of adult hemocyanin biosynthesis, we used PCR and a combination of both specific and universal degenerate primers to develop hemocyanin-specific cDNA probes. These probes were then used for both Northern blot analysis of mRNA transcripts and for cDNA library screening. A developmentally regulated hemocyanin cDNA has been cloned and is being sequenced.

INTRODUCTION

Hemocyanin, like the other oxygen-transporting molecules, hemoglobin and hemerythrin, combines reversibly with oxygen at the respiratory surface of the organism and carries oxygen via the circulatory system to cells and tissues far from the animal's surface. Thus the hemocyanin molecule is poised between the external environment and the internal milieu. Its functional properties, including oxygen affinity and cooperativity, respond to changes in both external and internal parameters such as temperature, salinity, pH and metabolic compounds (van Holde and Miller, 1982, for review).

Hemocyanins occur in only two phyla, the Mollusca and the Arthropoda. We now know that these hemocyanins are two very different proteins, even though they share the functional property of combining reversibly with oxygen, and they each contain two copper atoms at their active sites. Indeed, one of the copper binding sites of molluscan hemocyanin, CuB, shows clear sequence homology to the CuB region of arthropodan hemocyanins (Drexel et al., 1987). There is no significant homology between the molluscan and arthropodan CuA sites, however, nor between the rest of the amino acid sequences as far as is known (Volbeda and Hol, 1989b, Lang and van Holde, 1991). Not surprisingly,

molluscan and arthropodan hemocyanin molecules show marked differences in quaternary structure and subunit size (for reviews see Markl and Decker, 1992; van Holde et al., 1992).

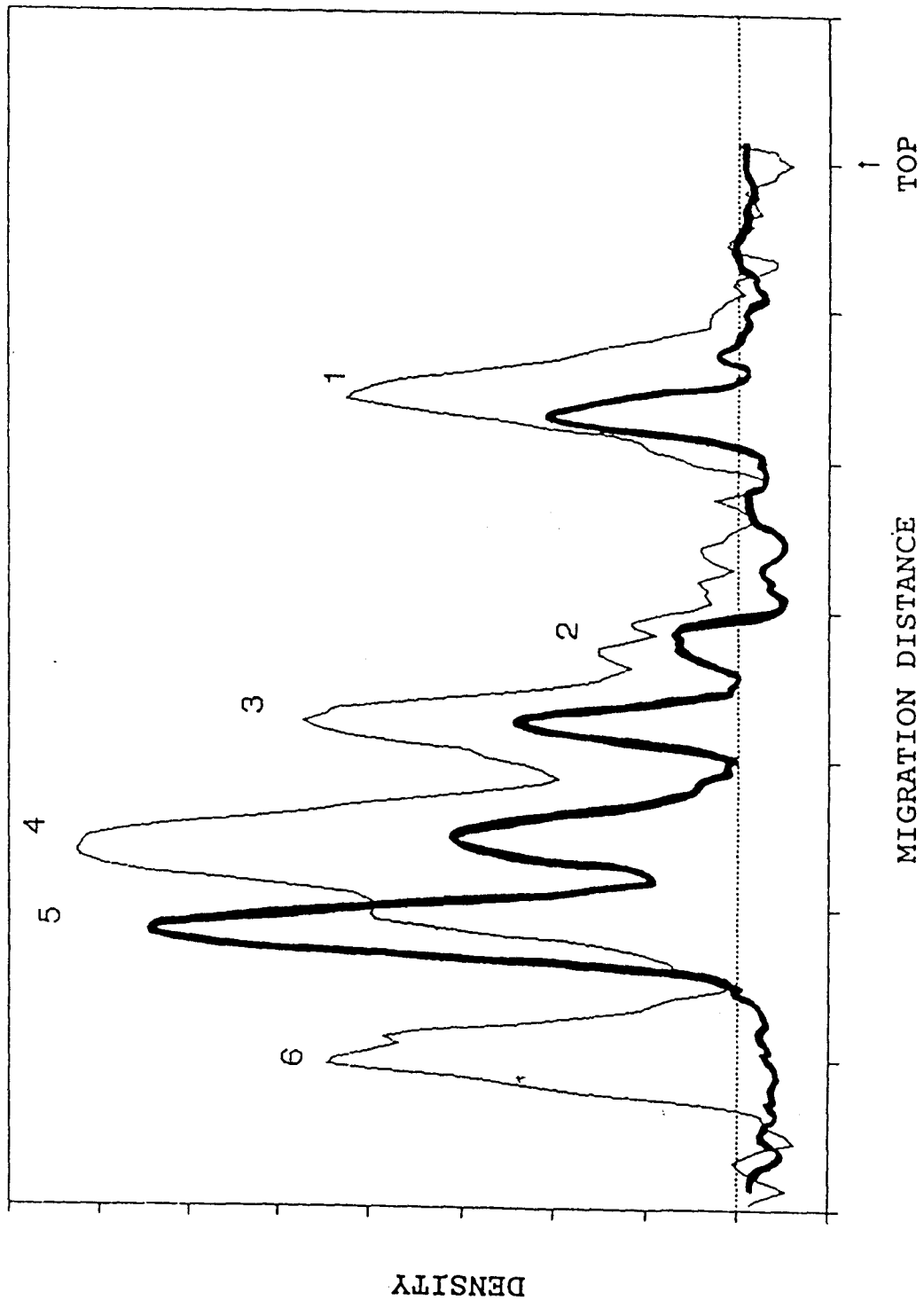
Arthropodan hemocyanins, found in chelicerates, crustaceans, and one myriapod, are made up of individual polypeptide chains or subunits of about 75 kDa. Based on the two arthropodan hemocyanins whose crystal structures are known, *Panulirus interruptus* a and *Limulus polyphemus* II (Gaykema et al., 1984; Volbeda and Hol, 1989a; Hazes et al., 1993), each polypeptide chain is bean shaped and composed of three structural regions or domains. Domain 1 is made up of seven α-helices while domain 2, containing the two copper-binding sites, CuA and CuB, is composed of two pairs of anti-parallel α-helices. The third domain is a seven-stranded ß-barrel with two long loops that wrap around domain 2 to interact with domain 1. Domain 2, containing the functional copper sites where oxygen is bound, is the most conserved region of the subunit, a feature that has been helpful in the present study. The subunits self-assemble to form extracellular hexameric and multi-hexameric oligomers that circulate in the hemolymph. Subunit heterogeneity in the oligomers varies among the arthropodan hemocyanins, ranging from 2 to as many as 12 different polypeptide chains, depending on the species (van Holde and Miller, 1982, for review). Functional studies on hemocyanins from a number of different arthropods indicate that subunit

composition affects the oxygen-binding properties of the oligomer (Sullivan et al., 1974; Truchot, 1992, for review).

The hemocyanin of the Dungeness crab, *Cancer magister*, is an especially intriguing hemocyanin to study because it changes in both structure and function during development of the crab from megalopa through the early juvenile instars to the adult (Terwilliger and Terwilliger, 1982). Megalopa and early juvenile crab hemocyanins are composed of five different subunits, numbered in order of increasing mobility in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as shown in Figure 1. Adult hemocyanin contains the same five subunits plus another, subunit 6, that is not present in the megalopa and young juvenile stages. Furthermore, the relative amounts of two other subunits, 4 and 5, switch during the change from juvenile to adult. These latter three subunits - 4, 5 and 6 - appear to be developmentally regulated, in contrast to subunits 1, 2 and 3 whose stoichiometries are constant during development.

As the hemocyanin subunit composition changes from juvenile to adult pattern, so too does the oxygen affinity of the hemocyanin (Terwilliger et al., 1985; Terwilliger and Brown, 1993). Megalopa and juvenile hemocyanins have an intrinsically lower oxygen affinity than does adult hemocyanin when measured at the same pH and ionic composition. As subunit 6 appears and subunits 4 and 5 reverse their relative concentrations, the oxygen affinity

Fig.1: Gel scan comparing five subunits of megalopa hemocyanin (broad tracing) and six subunits of adult hemocyanin (thin tracing) from *Cancer magister* separated by electrophoresis on 7.5% SDS PAGE.


DENSITY
MIGRATION DISTANCE
TOP
1
2
3
4
5
6

of the hemocyanin increases to adult levels. In addition, the changes in hemocyanin structure and functional properties during development are integrated with the ontogeny of hemolymph ion regulation (Brown and Terwilliger, 1992; Terwilliger and Brown, 1993).

There are many examples of ontogenetic changes in protein expression from a variety of phyla. The fetal-maternal shift in mammalian hemoglobin is one of the classic examples (Bunn et al., 1977). This ontogenetic shift in Dungeness crab hemocyanin is the first documented change in a copper protein whose biochemical and physiological roles in both the adult and the juvenile organism have been well studied (McMahon et al., 1979; Graham et al., 1983; Morris and McMahon, 1989; Brown, 1991).

Cancer magister has a number of attributes that make it a particularly suitable organism for these biochemical, physiological and molecular studies. First, wild megalopas can be collected easily in sufficient numbers when they return to Oregon coastal waters and estuaries in the spring after several months of oceanic larval life. Second, the megalopa stage of this species is markedly larger than any of the other Pacific Coast megalopas and thus can be identified at a glance rather than having to be laboriously sorted out from a swirl of similar beasts varying only in length of rostral spine or patterns of hairs on hairy little legs. Third, the megalopas can be raised through the

different instars in a running seawater system at ambient temperature and salinity. In addition, hemocyanin protein is easily obtained by bleeding a crab, and changes in an individual crab's hemolymph proteins can be analyzed over time by repeated sampling (Otoshi, 1994; Terwilliger and Otoshi, 1994). Thus the crab fits the August Krogh principle, "For a large number of problems there will be some animal of choice on which it can be most conveniently studied" (Krogh, 1929).

As more information has been obtained on developmental changes in hemocyanin structure and function, more questions have arisen, ones that could best be approached using molecular techniques. This chapter asks the following questions and describes the molecular strategies we are using to answer them.

1. Where is *C. magister* hemocyanin synthesized? Hemocyanin circulates as an extracellular protein in the hemolymph, but where are the cells located that synthesize the hemocyanin? The hepatopancreas has been implicated as the site of synthesis in several studies of crustacean hemocyanin biosynthesis (Senkbeil and Wriston, 1981; Préaux et al., 1986; Hennecke et al., 1990). Hemocyanin synthesis has also been described in horseshoe crab blood cells within sinuses around the eye (Fahrenbach, 1970; Wood and Bonaventura, 1981), scorpion endocuticle (Alliel et al., 1983), tarantula heart (Kempter, 1986; Markl et al., 1990),

and crab reticular connective tissue around several organs (Ghiretti-Magaldi et al., 1973, 1977). Is the site of synthesis species specific, or are there multiple sites within an animal? Our approach to determine where *C. magister* hemocyanin is synthesized was to look for hemocyanin mRNA in various tissues.

2. When does synthesis of hemocyanin subunit 6 begin? We knew when hemocyanin containing subunit 6, "adult hemocyanin," was first detectable in the hemolymph, but we wished to know when synthesis of subunit 6 first occurred. Such information might provide clues about the mode of assembly of the multihexameric molecules as well as insights into the regulation of biosynthesis. Ultimately, this approach may reveal whether the onset of adult hemocyanin biosynthesis is solely regulated by an internal developmental program or is correlated to extrinsic environmental cues as well. To begin to answer this question, we need to investigate hemocyanin mRNA expression in specific developmental stages of *C. magister*.

3. What is the evolutionary relationship of crustacean hemocyanin to other arthropod hemolymph proteins, including chelicerate hemocyanin, arthropod cryptocyanin (Terwilliger and Otoshi, 1994) and insect hemolymph proteins? To answer this question, we decided to construct a *C. magister* cDNA library and determine hemocyanin cDNA sequences.

WHERE IS *C. MAGISTER* HEMOCYANIN SYNTHESIZED ?

We chose a combination of approaches to address this question. They included designing a hemocyanin-specific degenerate primer (a short DNA sequence complementary to hemocyanin mRNA) and also preparing *C. magister* cDNA from crab hepatopancreas. With the hemocyanin-specific primer and a commercially available oligo-dT primer plus the crab cDNA, we would try to amplify by the polymerase chain reaction (PCR) any crab cDNA complementary to hemocyanin mRNA. The amplified hemocyanin cDNA fragment could then be used as a probe to assay hemocyanin mRNA expression in different tissues of the crab using Northern blots.

Design of Hemocyanin-Specific Primers

Since the CuA-binding site in domain 2 is highly conserved in all crustacean and chelicerate hemocyanin subunits thus far sequenced (Beintema et al., 1994), we expected it to be a conserved feature in *C. magister* hemocyanin subunits as well. Therefore a 32 bp oligonucleotide primer (CuA primer I) was designed based on a conserved sequence of 10 amino acids within the CuA site of another crustacean hemocyanin polypeptide, subunit a of the spiny lobster, *Panulirus interruptus* (Bak and Beintema, 1987). Due to the degeneracy of the genetic code, the "primer" actually consisted of a family of oligonucleotides

that represented all possible ways of coding for the short sequence of amino acids. The primer (degeneracy = 16384) was synthesized in a Model 380B automated DNA synthesizer (Applied Biosystems, Inc.) at the University of Oregon Biotechnology Laboratory. Its sequence was:

```
                                                        AA sequence of
                                                        subunit a from
    E   L   F   F   W   V   H   H   Q   L   T           P. interruptus

5' GAA-TTT-TTT-TTT-TGG-GTT-CAT-CAT-CAA-TTT-AC  3'  CuA primer I
   G C C   C   C       C   C   C   G C C
       A               A               A
       G               G               G
```

Another primer was designed to specifically amplify mRNA coding for hemocyanin subunit 6, the subunit present only in adult *C. magister*. We needed a short amino acid sequence from subunit 6 hemocyanin to develop this primer. All 6 subunits of adult *C. magister* hemocyanin were purified and the N-terminal amino acid sequence of each was determined (unpublished data). This 5′ subunit 6 primer, a degenerate 26 bp oligonucleotide (degeneracy = 65536, see below), was based on the unique N-terminal amino acid sequence of subunit 6:

```
                                                        N-terminal AA
                                                        sequence of sub 6
    S   A   G   G   A   F   D   A   Q                   from C. magister

5' TCT-GCT-GGT-GGT-GCT-TTT-GAT-GCT-CA   3'  5' sub 6 primer
   AGC   C   C   C   C   C   C   C
     A   A   A   A   A           A
     G   G   G   G   G           G
```

Isolation of Total RNA

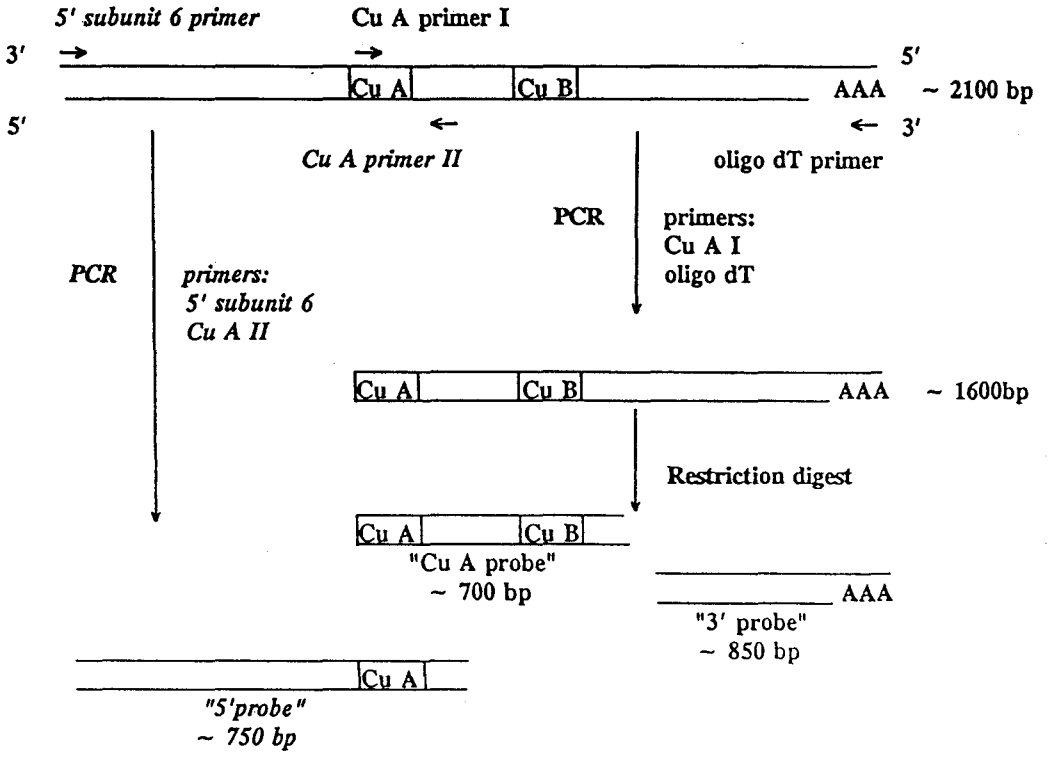
Adult crab hepatopancreas was chosen for the initial source of RNA for PCR, based on crab mRNA literature and the abundance of hepatopancreas tissue. Adult male Dungeness crabs were collected from Coos Bay and quickly killed. Tissue samples (1 g) were immediately dissected, thoroughly rinsed with *C. magister* saline buffer (Brown and Terwilliger, 1992), placed in liquid nitrogen, and ground to a fine powder with mortar and pestle. Total RNA was isolated with the guanidinium isothiocyanate method using the *Rapid* Total RNA Isolation Kit (5 Prime → 3 Prime, Inc.). This standard procedure quickly inactivates cellular RNases that were likely to be present in high concentrations in actively metabolizing hepatopancreas tissue.

Development of Probes: Reverse Transcription of mRNA, PCR Amplification, Cloning and Sequencing of Hemocyanin cDNA.

First strand cDNA was prepared from hepatopancreas total RNA using AMV reverse transcriptase to synthesize complementary DNA from the RNA template (see appendix A). Using the CuA primer I plus an oligo-dT primer directed against the poly (A)+ tail of mRNA, the PCR reaction was carried out with hepatopancreas cDNA as template (Fig. 2 and appendix A). PCR products were size analyzed on 1.2% agarose Tris-acetic acid-EDTA (TAE) minigels; several major fragments between 1200 and 2000 bp had been generated as

Fig.2: Schematic PCR amplification of *Cancer magister* cDNA using two different combinations of primers (CuA I and oligo dT, *5′ subunit 6* and *CuA II*) and an EcoRV restriction digest to obtain three hemocyanin-specific cDNA probes (CuA probe, 3′ probe and *5′ probe*).

Adult *Cancer magister* hepatopancreas cDNA



seen in Figure 3. These fall within the expected size range for a hemocyanin cDNA amplified from the CuA site to the 3′ end, based on known hemocyanin sequences (Linzen et al., 1985). The PCR products of interest were cloned (see appendix B) into a Bluescript II SK+ phagemid vector (Stratagene, Inc.). A positive clone containing a 1600 bp insert was selected, and sequencing grade DNA was prepared from the clone with a QIAGEN maxiprep kit according to manufacturer's instructions. The 1600 bp insert was then partially sequenced with a Sequenase kit (Version 2.0, US Biochemical, Inc.) using T3 and T7 sequencing primers and the dideoxy sequencing method. The bestfit alignment (FASTA algorithm, Devereux et al., 1984) of the derived amino acid sequence of the 5′ end of the 1600 bp PCR fragment and the corresponding CuA-coding portion of *Panulirus interruptus* subunit a is shown in Figure 4. With 81% amino acid similarity (chemical similarity) and 76% amino acid identity between the two sequences, the PCR fragment is clearly a hemocyanin cDNA. Choosing the conserved portion, the CuA site, of the crustacean hemocyanin subunit as the basis for our primer was a successful strategy for obtaining hemocyanin cDNA. Which of the 6 possible hemocyanin subunits of *C. magister* this cDNA represented could not be identified at this point. The cDNA insert was mapped by restriction analysis and the 5′ and 3′ ends were sequenced with the dideoxy technique. Digestion of the 1600 bp cDNA fragment by

Fig.3: PCR products on 1.2% agarose TAE minigel. Target sequence: adult *Cancer magister* hepatopancreas cDNA; primers: CuA I and oligo dT. From left to right, products of PCR reactions run at 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM and 8 mM Mg^{2+}; right lane, 1 kb ladder (Gibco BRL). Box outlines 1200 - 2000 bp PCR fragments excised from gel and cloned.

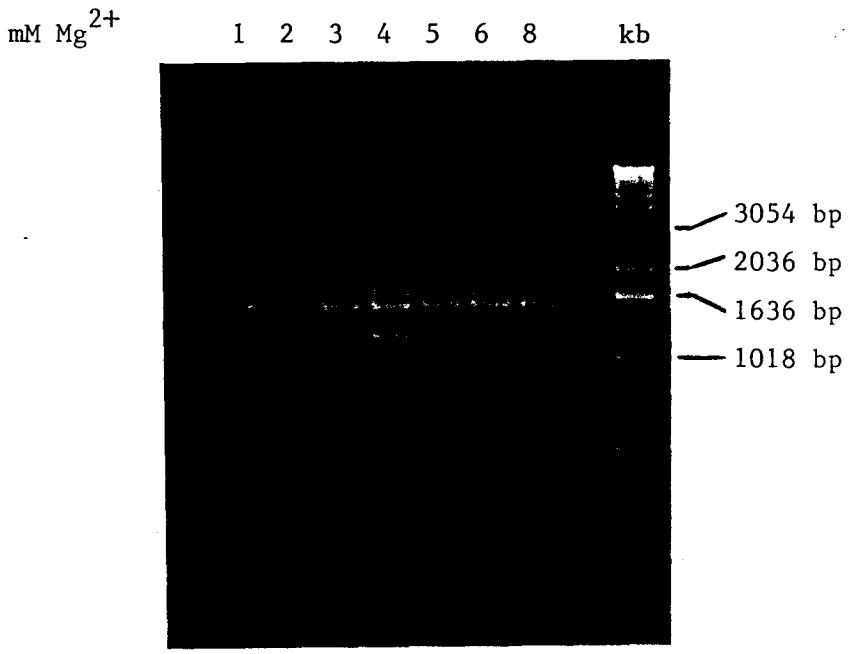

mM Mg^{2+}
1 2 3 4 5 6 8
kb
3054 bp
2036 bp
1636 bp
1018 bp

Fig.4: Amino acid sequence alignment of 5′ end of 1600 bp PCR fragment from *Cancer magister* (1603.T) with *Panulirus interruptus* hemocyanin subunit a (Pinta). CuA portion of Pinta sequence shown is numbered after Linzen et al. (1985).

```
                                                    10        20        30
1603.T                                             ELFFWVHHQLTVRFDAERLSNHLPDVDELH
                                                   ||||||||||||:||| ||||| |  |||||
Pinta. IGMNIHHVTWHMDFPFWWEDSYGYHLDRKGELFFWVHHQLTARFDFERLSNWLDPVDELH
       190        200        210        220        230        240

                40        50        60        70
1603.T WDDVIHEGFDPQAVYKYGGYFPSRPDNIHFEDVDGVADVRDM
       || :|:|||:| : ||||| || |||||||||||||||:|:|:
Pinta. WDRIIREGFAPLTSYKYGGEFPVRPDNIHFEDVDGVAHVHDLEITESRIHEAIDHGYITD
       250        260        270        280        290        300
```

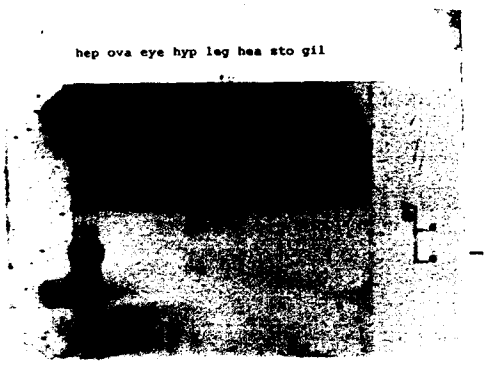
the restriction enzyme EcoRV yielded two fragments, a 700 bp "CuA probe" and an 850 bp "3′ probe" (Fig. 2).

Using the 5′ subunit 6 primer in combination with a different CuA primer, CuA II, whose sequence was derived from the 1600 bp clone, we were able to amplify, clone and sequence a 750 bp fragment corresponding to the 5′ end of the hemocyanin subunit 6 cDNA. Its 3′ end overlapped with the CuA site of the 1600 bp PCR fragment (Fig. 2). Thus with PCR and different combinations of primers, we were able to amplify different regions of hemocyanin subunit 6 cDNA. These regions could now be used as probes to identify transcripts of hemocyanin subunits in various tissues.

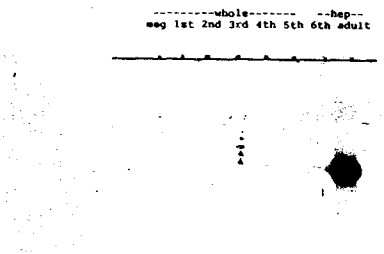
Northern Blots: Screening Tissues for Hemocyanin mRNA

Total RNA was prepared as described above from eight different tissues of adult *C. magister*, including hepatopancreas, ovary, eyestalks, hypodermis, leg muscle, heart, stomach and gill in order to see where hemocyanin mRNA was located. Equal amounts of RNA from each tissue were subjected to denaturing gel electrophoresis and the RNA was transferred onto a nitrocellulose membrane as described in appendix C. The membrane was then hybridized to the 5′ probe that had been ^{32}P-random prime labeled ("Probe-Eze" Random-Prime-Labeling Kit, 5′--->3′ Inc.). After hybridization, the blot was evaluated by autoradiography (Fig. 5A). A 2600 bp

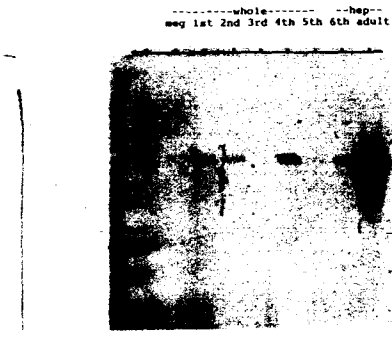
Fig.5: Northern blots of *Cancer magister* RNA hybridized with ^{32}P random prime labeled probes. (A) Total RNA from different tissues of adult *C. magister*. Probe, 750 bp 5′ probe. Hep, hepatopancreas; ova, ovary; eye, eyestalk; hyp, hypodermis; leg, leg muscle; hea, heart; sto, stomach; gil, gill. (B) mRNA from different developmental stages of *C. magister*. Probe, 750 bp 5′ probe. Meg, megalopa; 1st, 1st instar; 2nd, 2nd instar; 3rd, 3rd instar; 4th, 4th instar; 5th, 5th instar; 6th hep, 6th instar hepaptopancreas; adult hep, adult hepatopancreas. (C) Same as (6B), but hybridized with 700 bp CuA probe.



- 2.6 kb



-2.6 kb



- 2.6 kb

transcript, about the size expected for a full-length hemocyanin mRNA, was present only in the hepatopancreas RNA sample. Thus the hepatopancreas appears to be the site of hemocyanin synthesis in the adult Dungeness crab.

WHEN DOES SYNTHESIS OF ADULT HEMOCYANIN BEGIN ?

Northern Blots: Screening for Stage Specific Hemocyanin mRNA

Messenger RNA was prepared from different developmental stages to determine the stage at which adult hemocyanin synthesis begins. Megalopa larvae collected in May from Coos Bay, Oregon, were maintained in running seawater aquaria at ambient temperature and salinity at the Oregon Institute of Marine Biology. The megalopas quickly molted into first instar juvenile crabs, and the juveniles, fed a diet of mussel, fish and squid, continued molting through the instar stages in synchrony with field-caught juveniles. As each developmental stage reached intermolt, aliquots of megalopa through 5th instar juvenile were harvested by quick-freezing batches of whole animals in liquid nitrogen. The aliquots were stored at -80°C until the older instars had developed, at which point 1 g samples of frozen whole animals were ground to a fine powder with mortar and pestle in liquid nitrogen, and RNA was isolated. The larger stages, 6th instar and adult, were dissected as described above; 1 g samples of hepatopancreas were quickly rinsed and frozen.

Total RNA was isolated with the guanidinium isothiocyanate method using a *RAPID* Total RNA Isolation Kit, and poly (A)+ mRNA was prepared with oligo-dT spin columns (5 Prime → 3 Prime, Inc.). The mRNA yield was quantified by measuring absorbance at 260 nm.

Eight developmental stages were assayed by Northern blots for the presence of hemocyanin subunit 6 mRNA. Equal amounts (0.3 μg) of mRNA from each stage were separated according to size by electrophoresis and blotted as described in appendix C. Results of hybridizing the blot to the 750 bp 5′ probe are shown in Figure 5B. Hemocyanin mRNA was detected only in the adult stage, where a 2.6 kb transcript hybridized strongly. When a duplicate blot was incubated with the 700 bp CuA probe (Fig. 5C), the adult sample again gave a strong signal. Some degree of hybridization occurred in the earlier stages as well. These results suggest that synthesis of hemocyanin subunit 6 occurs after the 6th instar stage. The low levels of hybridization of the earlier stages with the CuA probe probably indicate cross-reactions with mRNA transcripts coding for one or more of the other hemocyanin subunits, 1 - 5, that are present in megalopa and early instars or low levels of subunit 6 mRNA. Experiments in progress to further pinpoint the onset of adult hemocyanin synthesis will include Northern blots of mRNA from instars 7, 8 and 9 as the juveniles develop.

SEQUENCING HEMOCYANIN SUBUNITS

Construction of a *Cancer magister* cDNA Library

The third question we wished to investigate, the cDNA sequence of the developmentally regulated hemocyanin subunit 6, led us to construct a *C. magister* cDNA library. The cDNA library would eventually give us the complete sequences of all hemocyanin subunits including the 5′ non-coding regions. Adult crab hepatopancreas mRNA was used to create the library in a lambda phage vector (Lambda-ZAP cDNA Synthesis Kit, Stratagene). In this procedure, double stranded cDNA was synthesized from mRNA; each cDNA was then inserted into the multiple cloning site of a Bluescript SK^- vector within a Lambda ZAP II phage. The library was amplified once on *Escherichia coli* XL-1 blue MRF′ cells to a final titer of 1.3×10^7 plaque forming units/μl. An aliquot was plated out and screened with filter lifts using the random prime labeled 5′ probe described above, according to standard procedure (Sambrook et al., 1989). Positive clones were excision rescued according to manufacturer's instructions. Plasmid DNA was isolated by alkaline lysis miniprep (Sambrook et al., 1989) and subjected to restriction enzyme analysis. The longest clones that had hybridized to the 5′ probe, including an 1800 bp cDNA, were selected for sequencing.

Creating Nested Deletions for DNA Sequencing

Initial sequencing of the 5′ and 3′ ends of the 1800 bp cDNA showed that it was a hemocyanin cDNA. To sequence it entirely, overlapping nested deletions were created by cleavage with DNase I in the presence of Mn^{2+} (Lin et al., 1985; C. Thisse and B. Thisse, personal communication) as described in Figure 6 and appendix D. The resulting nested deletions derived from the 1800 bp cDNA clone provided a family of clones with overlapping staggered deletions extending from various sites in the sequence to the 5′ end. Fifteen clones, ranging in size from 150 to 1800 bp (Fig. 7), were selected for sequencing by the dideoxy method, using a Sequenase kit as described above. To date, 1200 bp at the 5′ end of the cDNA have been sequenced, along with 200 bp at the 3′ end, and the full sequence is expected shortly.

CONCLUSIONS

We view hemocyanin as a model system for a developmentally regulated multisubunit protein. Several molecular approaches to the study of hemocyanin ontogeny have been presented here. Most have led to more questions or opened up other avenues of research. How is hemocyanin gene

Fig.6: Formation of DNaseI nested deletions for DNA sequencing. Bluescript SK^- carrying the 1800 bp hemocyanin cDNA insert (stippled) is randomly linearized by digestion with DNaseI in the presence of Mn^{2+}. The linearized fraction is cut in two by digestion with SmaI (E1), polyethylene glycol precipitated, and repaired with Klenow fragment. Both segments are recircularized with T4 DNA ligase. The construct carrying the 5′ end is recut with EcoRI (E2) and then the mixture is used to transform competent *E. coli* XL-1 Blue cells. Clones where DNase I had cut outside the insert and clones containing 5′ portions of the hemocyanin cDNA insert are rendered incapable of transformation because they are linearized by E2. The remaining "useful" clones (those carrying cDNA inserts with 5′ end deletions of various lengths) have no E2 site, remain circularized, and can transform. Insert size is determined by restriction analysis with XbaI and XhoI (E3 and E4). Clones of appropriate length are then selected for dideoxy sequencing.

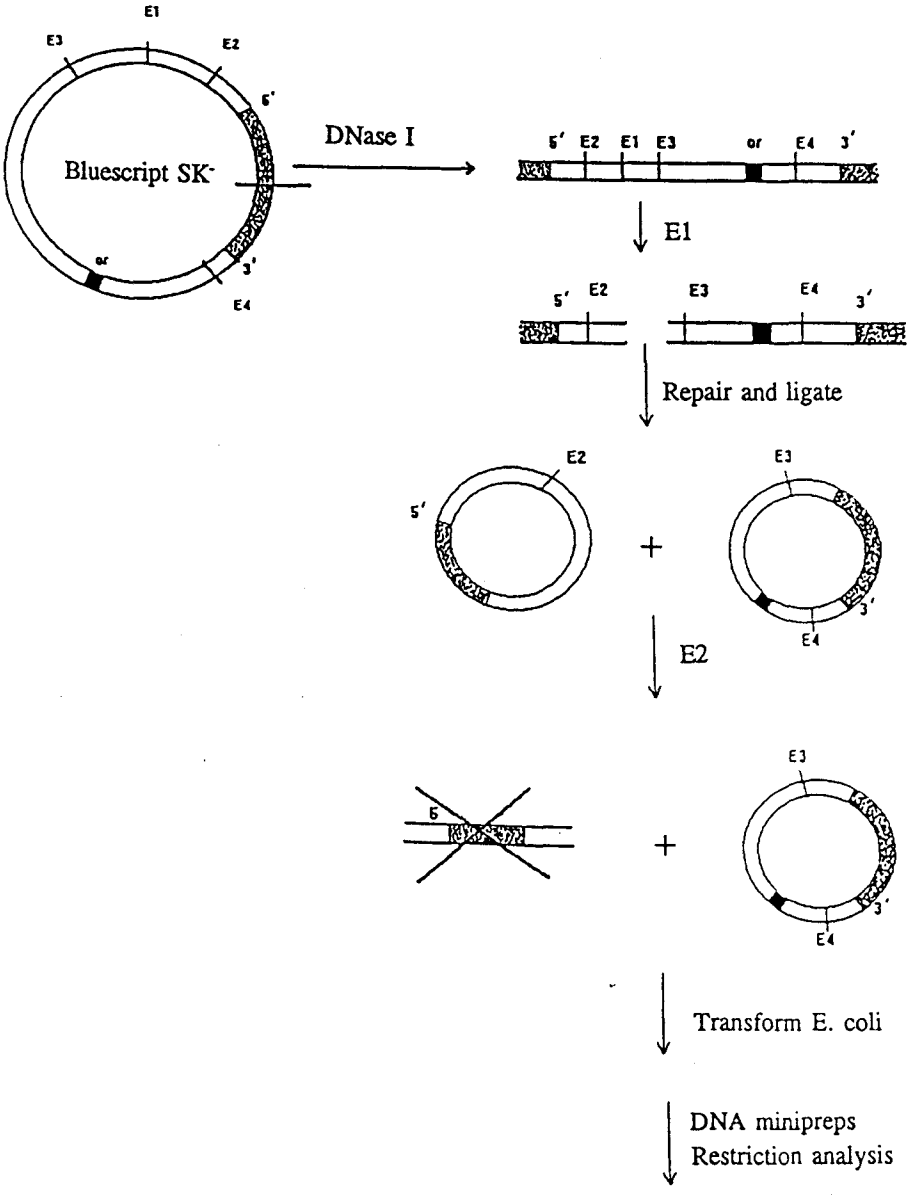
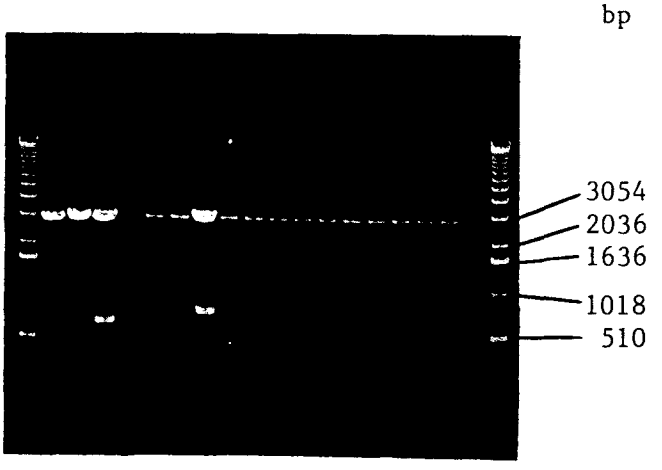

E1
E3
E2
5'
Bluescript SK-
or
3'
E4
DNase I
5' E2 E1 E3 or E4 3'
E1
5' E2
E3 E4 3'
Repair and ligate
E2
5'
E3
3'
E4
E2
E3
5'
3'
E4
Transform E. coli
DNA minipreps
Restriction analysis

Fig.7: Restriction analysis of DNase I nested deletions of 1800 bp hemocyanin cDNA. Miniprep DNA (0.1 μg) from each clone was digested with 10 units restriction enzymes (XbaI/XhoI) in 10 μl total volume for 2 hours at 37°C and analyzed on 1.2% TAE minigel. Clones shown in lanes 3 to 19 were selected for dideoxy sequencing. Insert size increases from ≈150 bp (left) to ≈1700 bp (right). Lanes 1 and 20, 1 kb ladder (Gibco BRL); lane 2, Bluescript vector (2.9 kb) linearized with XbaI/XhoI.


bp
3054
2036
1636
1018
510

expression controlled? How do individual subunits contribute to the functional properties of the whole protein? Does the cDNA library we constructed contain blueprints of other exciting crab hemolymph proteins? What insights into the evolution of arthropod hemolymph proteins and other oxygen-binding proteins can be derived from the cDNA sequence of this arthropodan hemocyanin? A better understanding of this hemolymph protein will lead to further understanding of how animals cope with the challenge of oxygen transport during development.

Chapter II has shown how conserved sequences within a functional domain of the protein Hc can be used to develop Hc subunit-specific cDNA probes. These probes will be used in chapter III to investigate ontogenetic changes and tissue-specific differences in Hc expression during the life cycle of the Dungeness crab.

CHAPTER III

NORTHERN BLOT ANALYSIS OF HEMOCYANIN EXPRESSION IN THE DUNGENESS CRAB - DEVELOPMENTAL CHANGES AND TISSUE-SPECIFIC DIFFERENCES

Gregor Durstewitz

and

Nora Barclay Terwilliger

Oregon Institute of Marine Biology and Department of Biology, University of Oregon, Eugene, OR 97403

ABSTRACT

The copper based respiratory protein hemocyanin (Hc) undergoes a developmental shift in subunit composition analogous to that seen in mammalian hemoglobin. We studied Hc gene expression in the Dungeness crab, *Cancer magister*, by Northern blot analysis. Animals were raised under controlled conditions, and total RNA was isolated from 13 developmental stages as well as from 6 tissue types in the adult animal. RNA was run on formaldehyde agarose gels, blotted onto nylon membranes and probed with ^{32}P-labeled adult Hc-specific cDNA probes. Results indicate that adult Hc biosynthesis occurs in hepatopancreas tissue only. Analysis of various developmental stages shows that expression of adult-type Hc, as indicated by the appearance of Hc subunit 6 mRNA, begins during the 6th juvenile instar. A model is proposed to explain the observed subunit stoichiometries in juvenile and adult Hc.

INTRODUCTION

Respiratory proteins function to combine reversibly with oxygen at the respiratory surfaces of an animal and to carry oxygen via the circulatory system to the tissues inside. In many arthropods and molluscs, these oxygen transport proteins are hemocyanins (Hc), large multisubunit molecules that occur extracellularly in the hemolymph (van Holde and Miller, 1982; van Holde and Miller, 1995). Arthropodan Hcs are composed of heterogeneous subunits with molecular weights of about 75 kDa. A single subunit contains two copper (Cu) binding sites, CuA and CuB, each site complexing one Cu atom by three histidine ligands (Volbeda and Hol, 1989). Both sites participate in the binding of one oxygen molecule. In deoxygenated Hc, the Cu ions are in the Cu^{+} state. During oxygenation, a peroxide (O_2^{2-}) bridge is formed between the Cu^{+} ions, oxidizing them to Cu^{2+} (Freedman et al., 1976). The 75 kDa subunits self assemble into hexamers or multiples of hexamers. In the hemolymph of the adult Dungeness crab, *Cancer magister,* we find 2-hexamer 25S Hc as well as 1-hexamer 16S Hc (Ellerton et al., 1970).

The Hc of *C. magister* is particularly interesting because it changes in both subunit composition and function during development of the crab from megalopa and early juvenile instar stages to adult (Terwilliger and

Terwilliger, 1982). Adult *C. magister* 25S Hc is composed of six different types of subunits as shown by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), 2 stage PAGE and limited proteolysis (Larson et al., 1981). One of these, subunit 6, is absent in both 25S and 16S megalopa Hc (Terwilliger and Terwilliger, 1982) and typically does not appear in hemolymph Hc until sometime during the 6th juvenile instar. In addition, the stoichiometry of two other Hc subunits, 4 and 5, changes during the transition from larval to adult crab.

The developmental shift in subunit composition results in a new population of adult Hc molecules that have a higher affinity for oxygen than juvenile Hc (Terwilliger et al., 1986; Terwilliger and Brown, 1993). The larval-adult shift in Hc is analogous, then, to the fetal-adult shift in hemoglobins seen in humans and other mammals (Ingermann, 1992) and to the developmental change seen in some invertebrate extracellular hemoglobins (Schin et al., 1979; Heip et al., 1980). We hypothesized that the differences between juvenile and adult Hc are due to an ontogenetically regulated change in Hc gene expression. To test this hypothesis, we determined the site of Hc synthesis in adult *C. magister* and then investigated when during development adult-specific subunit 6 mRNA is first expressed.

Several different tissues have been proposed as possible sites of Hc synthesis in arthropods. Among

chelicerates, these include cyanocytes (cells containing crystals made of Hc-shaped particles) in sinuses of the compound eye (Fahrenbach, 1970; Wood and Bonaventura, 1981), endocuticle tissue (Alliel et al., 1983) and cells in the inner heart wall (Kempter, 1983; Markl et al., 1990). Tissues implicated in crustacean Hc synthesis include hepatopancreas (Senkbeil and Wriston, 1981; Preaux et al., 1986; Hennecke et al., 1990; Rainer and Brower, 1993), eyestalk cyanocytes (Schönenberger et al., 1980) and reticular connective tissue around pyloric stomach, opthalmic artery and hepatopancreas (Preaux et al., 1986; Ghiretti-Magaldi et al., 1973; Ghiretti-Magaldi et al., 1977). These results indicated that *C. magister* Hc could potentially be made in several locations within the crab.

We developed a Hc subunit 6 specific cDNA probe that allowed us to identify adult Hc mRNA. Using this probe, we could detect potential sites of active Hc synthesis and exclude the possibility that these were merely sites of Hc storage or degradation. To investigate when the shift in synthesis from juvenile to adult Hc occurs, Northern blots of RNA from different developmental stages were probed with the subunit 6 specific cDNA. In this paper we present the results of our studies on Hc ontogeny, determining when adult Hc, identified by the presence of subunit 6 mRNA, is first expressed during development. The data allow us to present a model for the shift in subunit stoichiometries

between juvenile and adult Hcs.

MATERIALS AND METHODS

Sample Collection and Total RNA Isolation. Adult male *Cancer magister* were caught in the Coos Bay estuary (Oregon) by scuba diving or using baited crabrings. Crabs were quickly killed, and tissue samples (100 mg) were dissected, thoroughly rinsed with *C. magister* hemolymph buffer (50 mM Tris-HCl, 454 mM NaCl, 11.5 mM KCl, 13.5 mM $CaCl_2$, 18 mM $MgCl_2$, 23.5 mM Na_2SO_4, pH 7.6) and frozen in liquid nitrogen. Total RNA was isolated from tissues by the guanidinium isothiocyanate method using a RAPID Total RNA Isolation Kit (5 Prime -> 3 Prime, Inc.). Total RNA yield was quantified by measuring absorbance at 260 nm.

Ovigerous female *C. magister* were collected from the Pacific Ocean near Coos Bay; zoea larvae that hatched from the fertilized egg masses on the females' pleopods were harvested and frozen in liquid nitrogen. Megalopas were collected from surface waters of Coos Bay in late spring and maintained in running seawater aquaria at the Oregon Institute of Marine Biology as described (Brown and Terwilliger, 1992). As the young crabs grew and molted, samples of each developmental stage from megalopa to 5th instar were obtained by quick-freezing batches of whole animals in liquid nitrogen. For RNA isolation of these early

developmental stages, 100 mg aliquots of frozen whole animals were dropped into liquid nitrogen and ground to a fine powder. For the larger developmental stages, 6th instar to adult, 100 mg samples of hepatopancreas tissue were dissected from freshly killed animals, rinsed with hemolymph buffer and quick-frozen in liquid nitrogen. All samples were stored at -80°C. Total RNA was isolated using the guanidinium isothiocyanate method as described above.

Purification and N-Terminal Sequence Analysis of *Cancer magister* Hc Subunits. Fresh hemolymph was obtained from the sinus at the base of a walking leg of adult male crabs using a 22 gauge needle and syringe. Hemolymph was allowed to agglutinate on ice for 30 min and then centrifuged at 12000 g for 10 min in a Sorvall RC2-B refrigerated centrifuge at 4°C. The supernatant was applied to a BioGel A-5m chromatography column (1.8 x 115 cm) equilibrated in column buffer (0.05 M Tris-HCl, pH 7.5, 0.1 M NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$) in order to separate the 25S 2-hexamer Hc from the 16S 1-hexamer Hc. Individual subunits were purified by subjecting the 25S Hc fraction to a 2 step combination of alkaline PAGE (pH 8.9) and SDS-PAGE (Terwilliger and Terwilliger, 1982). The SDS gel was electroblotted onto an Immobilon-P PVDF membrane (Millipore, Inc.) in a Western blot procedure, stained with 0.1% Coomassie Blue, 50% methanol, 10% acetic acid (Cleveland et al., 1977) for 1 min

and destained 20 min in 50% methanol, 10% acetic acid. Bands representing each of six individual Hc subunits (about 100 pmole protein per subunit) were excised from the PVDF membrane and used to obtain N-terminal sequences in an automated protein sequencer in the University of Oregon Biotechnology Laboratory.

Hc-Specific Primers and Probes. Three oligonucleotide primers, CuA I, 5′ subunit 6 and CuA II, were designed as indicated in Results and synthesized in an Applied Biosystems Model 380B DNA synthesizer at the University of Oregon Biotechnology Laboratory. Reverse transcription (RT) and polymerase chain reaction (PCR) amplification of mRNA from adult *C. magister* hepatopancreas, using these primers, were carried out as described (Terwilliger and Durstewitz, 1996) in order to develop Hc-specific probes.

Northern Blot Analysis. Aliquots (5 μg) of total RNA prepared from each tissue type and developmental stage were denatured and run on a 1.2% agarose formaldehyde gel. RNA was pressure blotted onto a Hybond-N nylon membrane (Amersham), UV-crosslinked with 120 mJoule in a UV-Stratalinker (Stratagene) and baked for 2 h at 80°C.

The blots were then prehybridized under agitation for 2 h at 42°C in a solution of 50% formamide (freshly deionized), 5x SSPE (saline sodium phosphate EDTA buffer, 3M

NaCl, 0.2M $NaH_2PO_4*H_2O$, 20mM EDTA), 2x Denhardt's reagent and 0.1% SDS (Sambrook et al., 1989). Hybridizations were performed in glass bottles in a hybridization oven (Hybaid, Inc.) overnight at 42°C with a ^{32}P-random-prime-labeled probe specific for *C. magister* Hc mRNA (see Results). Blots were washed 4x for 20 min at 45°C in 2x SSC (saline sodium citrate buffer, 3M NaCl, 0.3M sodium citrate), 0.1% SDS and evaluated by autoradiography.

Following autoradiography, blots were stained for total RNA. They were immersed in staining solution (0.03% methylene blue, 0.3 M sodium acetate, pH 5.2) for 30 sec, then washed several times under agitation in distilled water until background disappeared (Wilkinson et al., 1990).

RESULTS

Designing Hc-Specific Primers for PCR. Our first goals were to develop primers and to amplify Hc coding sequences by PCR. The amplified cDNAs could then be used as probes in Northern blots. Because the CuA binding site in arthropod Hc domain 2 is highly conserved in all crustacean and chelicerate Hc subunits thus far sequenced (Beintema et al., 1994), we predicted it would be a conserved feature in all six *C. magister* Hc subunits as well. Accordingly, primer CuA I (5'- GAA-TTT-TTT-TTT-TGG-GTT-CAT-CAT-CAA-TTT-AC-3'), a 32

bp degenerate oligonucleotide, was designed based on the amino acid sequence NH_2-E L F F W V H H Q L T-COOH within the CuA site of Hc subunit a of the spiny lobster, *Panulirus interruptus* (Bak and Beintema, 1987). Reverse translation of part of the unique N-terminal amino acid sequence of Hc subunit 6 (Fig. 1) allowed the synthesis of a second degenerate primer, 5′-TCT-GCT-GGT-GGT-GCT-TTT-GAT-GCT-CA-3′, specific for the 5′ end of Hc subunit 6 cDNA. The third primer, CuA II, was an anti-sense primer based on the sequence of our 1914 bp PCR product (see below). This primer, 5′-CAC-TGC-CTG-GGG-ATC-GAA-GCC-CTC-ATG-3′, was designed to be specific for a region just downstream of the CuA site in *C. magister* Hc subunit 6.

PCR Amplification of Hc Subunit 6 cDNA. The locations of the primers relative to Hc subunit 6 cDNA are shown in Fig. 2. The first PCR experiment, using the CuA I primer plus a universal oligo-dT primer directed against the poly-A^+ tail of mRNA (gift from Dr. Ry Meeks-Wagner), amplified several major fragments between 1200 and 2000 bp in size. These fall within the expected size range for a Hc cDNA fragment extending from the CuA site to the 3′ end, based on known Hc sequences (Linzen et al., 1985). A 1914 bp PCR product was cloned into a Bluescript II SK^+ phagemid vector (Stratagene) as described before (Terwilliger and Durstewitz, 1996). Second, using the 5′ subunit 6 primer in combination with

Fig.1: N-terminal amino acid sequences of *Cancer magister* hemocyanin subunits 1 - 6. Sequence alignment manual. Alignment gap, (...); ambiguous identity, (?).

```
1   A D . . . . L A H ? . Q . Q . A . V N ? L L R K I R S P
2   A C . . . . L A A R . . . Q ? A . V N ? L L Y K I Y
3   . D N F ? S L A ? . . . K Q ? E D V ? H
4   . D S A G . . A P D . . . Q
5   A . S P G ? . A S D V Q K ? ? ? . V
6   . D S A G G . A F D A Q K Q ? . D V N S A L D K ? . S
```

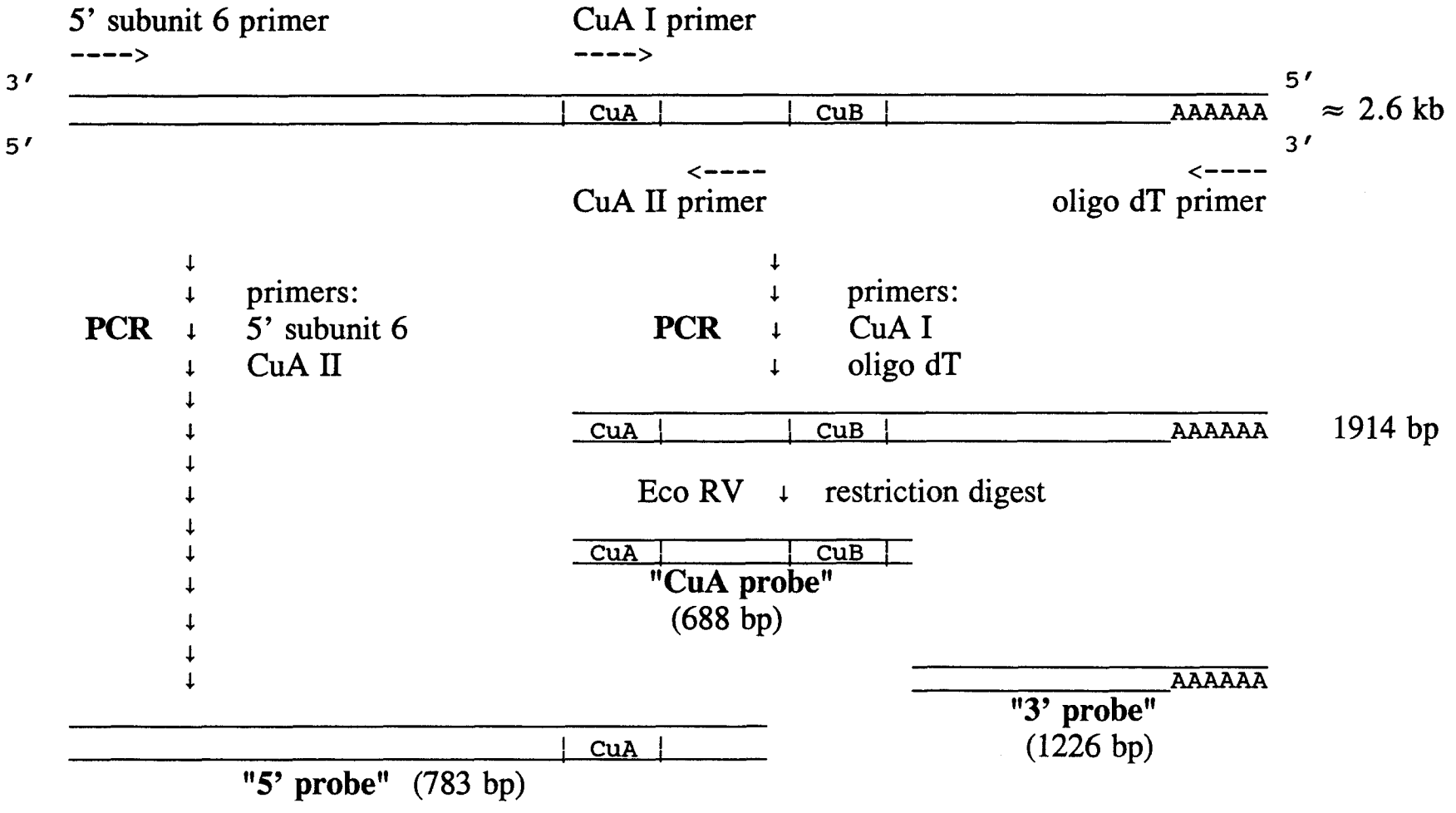
primer CuA II, we were able to amplify and clone a 783 bp fragment corresponding to the 5′ end of Hc subunit 6 cDNA. We refer to this fragment as "5′ probe."

Dideoxy sequencing of both clones confirmed they were Hc coding sequences (Genbank accession number: U48881, manuscript in preparation). The 5′ end of the 1914 bp PCR product displayed 76% amino acid identity and 81% similarity with Hc subunit a from *Panulirus interruptus*, indicating that it was a Hc cDNA. The 3′ end of the 783 bp 5′ probe overlapped by 133 bp and was identical in sequence with the CuA site within the 1914 bp cDNA (see Fig. 2). Digestion of the 1914 bp fragment with restriction enzyme Eco RV yielded two fragments, a 688 bp "CuA probe" and a 1226 bp "3′ probe" (Fig. 2).

Thus, using different combinations of primers, we were able to amplify three distinct regions of the Hc subunit 6 cDNA, the 5′ end, the CuA region and the 3′ end. These cDNAs could now be used as probes to identify subunit 6 transcripts from various tissues and developmental stages.

Probe Specificity. The three probes obtained by PCR could be expected to hybridize equally well with Hc subunit 6 mRNA. They might differ in the extent of crossreactivity with mRNAs corresponding to the other five Hc subunits (and, possibly, with other related proteins) because some regions of the Hc protein, and hence the cDNA, show a higher degree

Fig.2: PCR amplification of various regions of Hc subunit 6 cDNA (modified from Terwilliger and Durstewitz, 1996; reprinted by permission of John Wiley and Sons, Inc.).


5' subunit 6 primer
---->
CuA I primer
---->
3'
5'
CuA
CuB
AAAAAA
≈ 2.6 kb
5'
3'
<----
CuA II primer
<----
oligo dT primer
PCR
primers:
5' subunit 6
CuA II
PCR
primers:
CuA I
oligo dT
CuA
CuB
AAAAAA
1914 bp
Eco RV
restriction digest
CuA
CuB
"CuA probe"
(688 bp)
AAAAAA
"3' probe"
(1226 bp)
CuA
"5' probe" (783 bp)

of sequence conservation than others.

In order to assay crossreactivity of the probes with other Hc subunits, three identical Northern blots were prepared using equal amounts of RNA from adult crab and several early juvenile stages. Autoradiograms of all three blots showed a 2.6 kb transcript in RNA from adult animals (Fig. 3). This is about the size expected for a full length Hc mRNA. The CuA and 3′ probes, but not the 5′ probe, also hybridized to a lesser extent to juvenile transcripts of approximately the same size. We interpret these juvenile bands as crossreactivity of the probes with mRNAs of Hc subunits other than subunit 6, since we know juvenile stage Hcs do not contain subunit 6 but are composed of subunits 1 to 5 (Terwilliger and Terwilliger, 1982). The 3′ probe showed the greatest crossreactivity, probably due to its long poly-A^+ tail. The 5′ probe displayed the least degree of subunit crossreactivity, hybridizing only with adult mRNA, and was therefore chosen for the following experiments.

Northern Blot Analysis of Tissue-Specific Hc Expression. Northern blots of total RNA from six different tissues of the Dungeness crab, heart, leg muscle, hypodermis, stomach, gill and hepatopancreas, were probed with the 5′ probe (Fig. 4). Equal amounts of RNA were loaded as indicated by methylene blue total RNA stain. The

Fig.3: Northern blots of RNA to test probe specificity using (left) 5′, (middle) CuA and (right) 3′ probes for Hc subunit 6. M, megalopa; 1-6, 1st to 6th juvenile instars; A, adult crab.

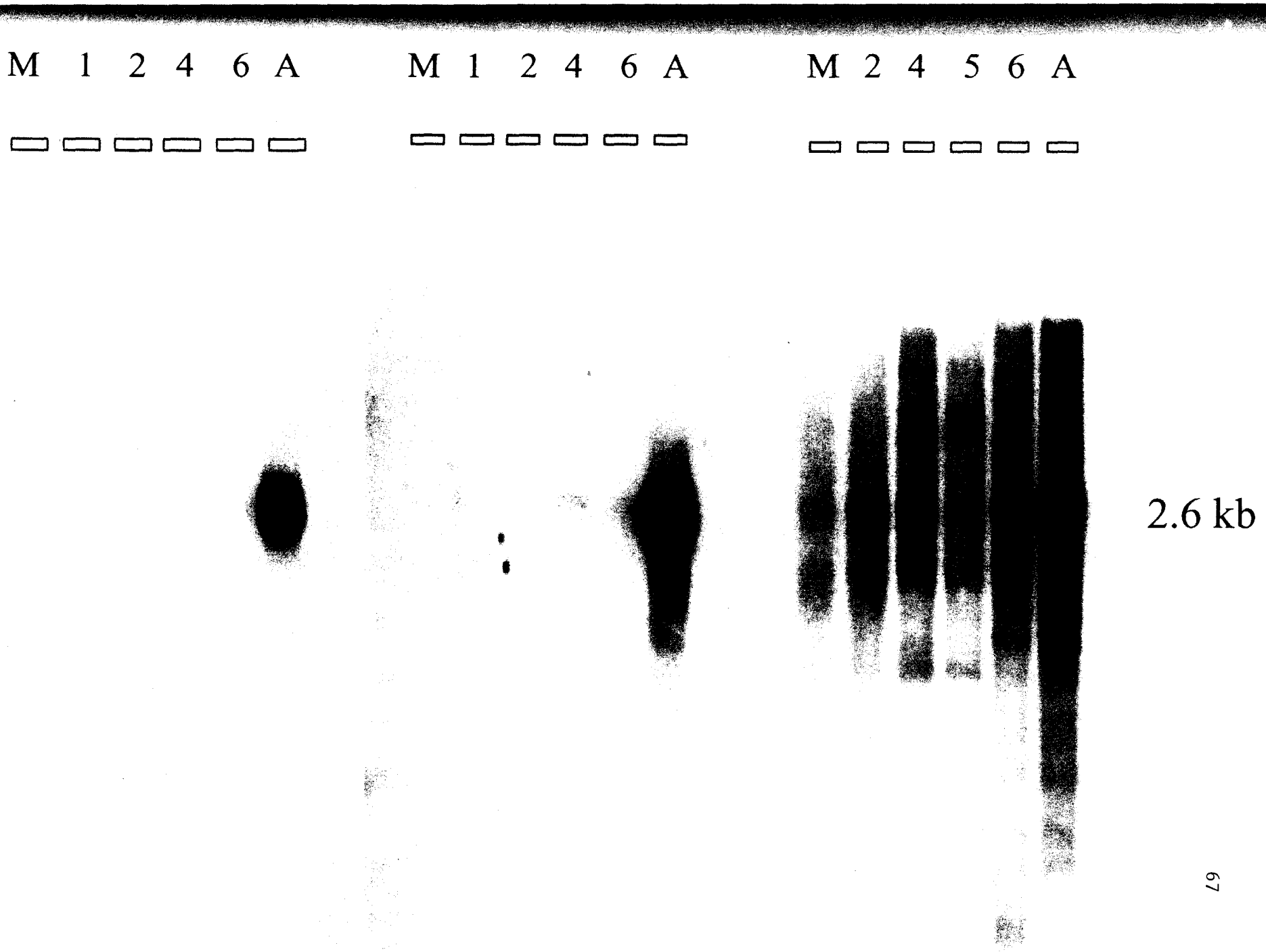
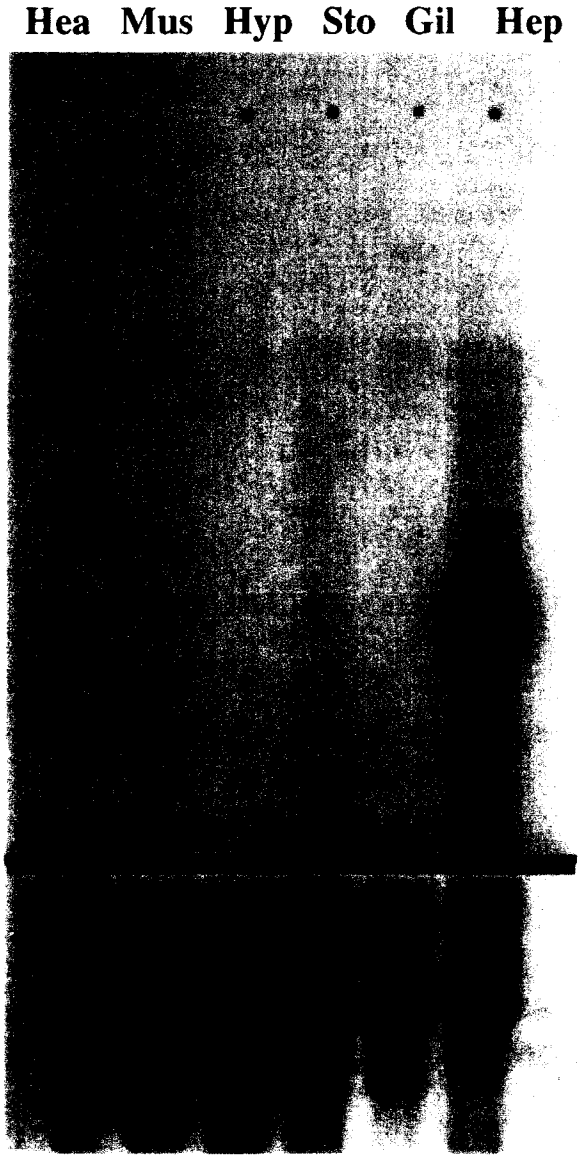

M 1 2 4 6 A
M 1 2 4 6 A
M 2 4 5 6 A
2.6 kb

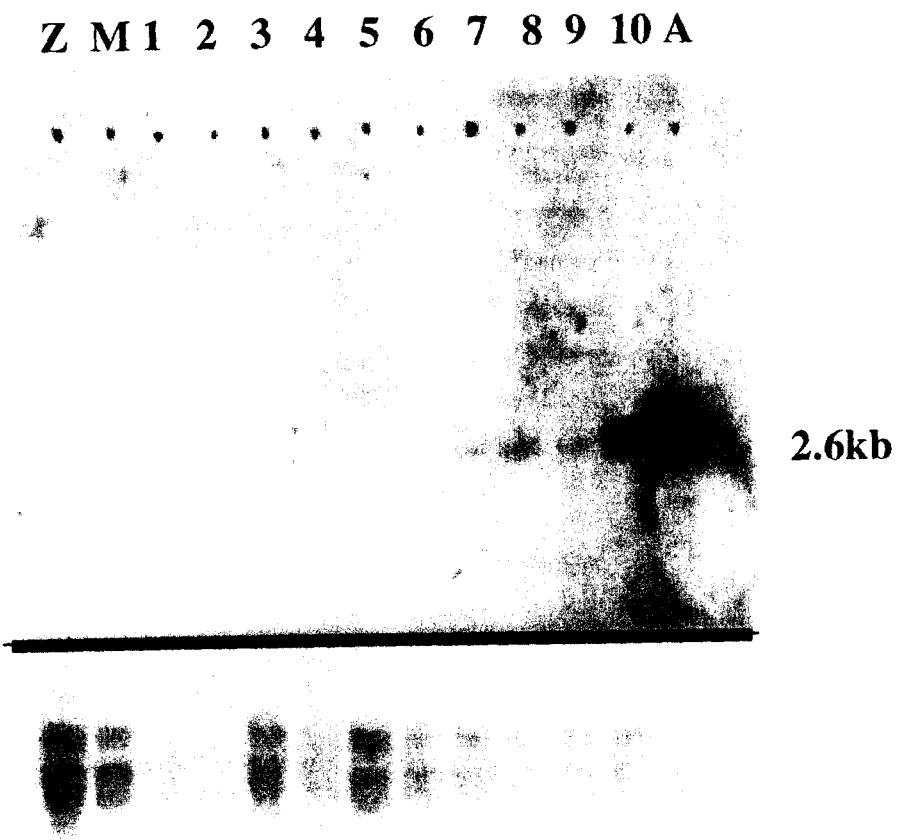
Fig.4: Northern blot of RNA from various tissues of adult *Cancer magister*. Upper: autoradiogram using 5′ probe, random-prime-labeled with ^{32}P. Lower: total RNA stain. Hea, heart; Mus, leg muscle; Hyp, hypodermis; Sto, stomach; Gil, gill; Hep, hepatopancreas.


Hea Mus Hyp Sto Gil Hep
2.6 kb

autoradiogram showed a 2.6 kb transcript only in the RNA sample from hepatopancreas tissue. We conclude that hepatopancreas is the site of Hc synthesis in the adult Dungeness crab.

Northern Blot Analysis of Developmental Changes in Hc Expression. No mRNA transcripts were detected in zoea, megalopa or 1st through 5th juvenile instar when using the 5′ probe (Fig. 5). A 2.6 kb transcript was present in the 6th instar and older stages. All lanes contained equal amounts of total RNA as indicated by the methylene blue stain. The steady state level of transcript increases with age of crab. These results indicate that Hc subunit 6 biosynthesis begins during the 6^{th} juvenile instar, and the steady state level of mRNA continues to increase as the crab approaches maturity.

Fig.5: Northern blot of RNA from different developmental stages of *Cancer magister*. Upper: autoradiogram using 5′ probe, random-prime-labeled with ^{32}P. Lower: total RNA stain. Z, zoea; M, megalopa; 1-10, 1st - 10th juvenile instars; A, adult.


Z M 1 2 3 4 5 6 7 8 9 10 A
2.6kb

DISCUSSION

Site of Hc Synthesis. The presence of Hc mRNA in hepatopancreas of *C. magister* is consistent with other studies on crustacean Hc synthesis. Active Hc biosynthesis has been demonstrated in hepatopancreas of lobster, crayfish and 2 species of crab by either mRNA *in vitro* translation or radioisotope incorporation studies (Senkbeil and Wriston, 1981; Preaux et al., 1986; Hennecke et al., 1990; Rainer and Brouwer, 1993). Two other sites also have been implicated in Hc biosynthesis in crustaceans, the wall of the pyloric stomach in *Carcinus maenas* (Ghiretti-Magaldi et al., 1977) and cells in the eyestalk of *Squilla mantis* (Schönenberger et al., 1980), but the evidence appears more indirect. While the presence of Hc was convincingly shown by immunofluorescence, the studies left open the possibility that these were sites of Hc storage or degradation. Alternatively, these tissues could contain Hc synthesizing cells that had been transported by the hemolymph from the hematopoetic tissue to their present location as suggested for *Limulus* cyanocytes (Fahrenbach, 1970). Hemocyanin synthesis by hemocytes or cyanocytes has been ruled out in several crustaceans (Senkbeil and Wriston, 1981; Hennecke et al, 1990), however, and may be a phenomenon restricted to the chelicerates

(Fahrenbach, 1970; Kempter, 1983; Markl et al, 1990; Voit and Schneider, 1986). Another potential site of Hc synthesis in crustaceans is the reticular connective tissue, based on morphology, immunohistochemistry and mRNA *in vitro* translation (Preaux et al., 1986; Ghiretti-Magaldi et al., 1973; Ghiretti-Magaldi et al., 1977). Preliminary studies in our laboratory identify reticular connective tissue in *C. magister* as a site of synthesis not for Hc but for a closely related non-respiratory protein, termed cryptocyanin (unpublished data, NBT; Terwilliger and Bremiller, 1995). The results presented here, in which Hc mRNA was present only in hepatopancreas and not in the other five tissues examined, identifies the hepatopancreas as the sole source of Hc in the crab, *C. magister*.

Ontogeny of Hc. The fetal-maternal shift in mammalian hemoglobin expression has been studied extensively. In invertebrates, Heip et al. (1980) described ontogenetic changes in the composition of hemoglobin in the brine shrimp *Artemia salina*. Our study now presents a case of developmentally regulated expression of the Cu-based oxygen transport protein Hc. Results show that expression of adult-type Hc as indicated by the appearance of subunit 6 mRNA in hepatopancreas tissue begins during the 6th juvenile instar. Onset and continuation of subunit 6 synthesis is indicated by increasingly strong probe hybridization to a transcript

of about 2.6 kb starting during the 6th instar (Fig. 5). Use of either the CuA or the 3′ probe results in crossreactions with other mRNAs, presumably those corresponding to Hc subunits 1 to 5 that are expressed during all developmental stages. This is evident from the less intense 2.6 kb signal that appears throughout the early stages (Fig. 3). An alternative explanation for the presence of subunit 6 on SDS-PAGE gels of adult Hc is that this new subunit could be the product of posttranslational modification or proteolysis of an already existing polypeptide rather than marking the onset of expression of a different gene. The Northern blot procedure we used in this study allows us to exclude this possibility for three reasons. First, any of our three Hc-specific cDNA probes should have hybridized strongly to a potential "pre-subunit 6" transcript being expressed in the early juvenile stages. No such signal was found. Furthermore, the 5′ probe we used in these blots is clearly subunit 6 - specific because its 5′ end was designed to match that subunit's unique N-terminal sequence. Finally, the appearance of a new species of mRNA that hybridizes to our subunit 6 - specific probe in the hepatopancreas of 6th instar crabs coincides with the appearance of a new Hc subunit in the animal's hemolymph as shown by SDS-PAGE. These data also exclude the possibility that subunit 6 might be stored intracellularly until its eventual release into the hemolymph during the 6th instar. The observed structural

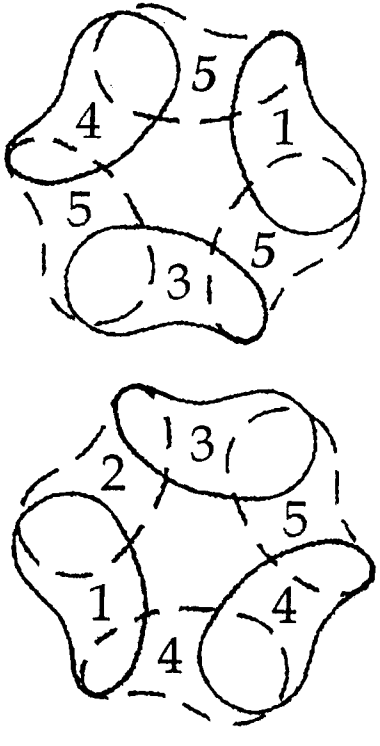
shift from juvenile to adult Hc is accompanied by functional changes, most notably an increase in oxygen affinity (Terwilliger et al., 1986). This may be an adaptation to changing ecological conditions during ontogenesis from a freeswimming planktonic larva to a benthic adult crab. The shift may also be part of a developmental pattern in which the changes in Hc function counterbalance a parallel ontogeny of ionic regulatory capabilities in *C. magister* (Terwilliger and Brown, 1993; Brown and Terwilliger, 1992).

The observed developmental differences in subunit stoichiometries require the initiation of expression of subunit 6, downregulation of subunit 5 production and an increase in subunit 4 synthesis (Fig. 6). The proportions of each subunit in megalopa and early juvenile Hc are constant as are those in adult Hc, and, consistent with other studies on adult crustacean Hc (Markl and Decker, 1992), there appears to be only one type of 25S molecule in these age groups. Intermediate stage juveniles having Hcs with subunit stoichiometries and oxygen affinities approaching those of the adult (Terwilliger and Brown, 1993) probably have a mixture of both types of 25S Hc.

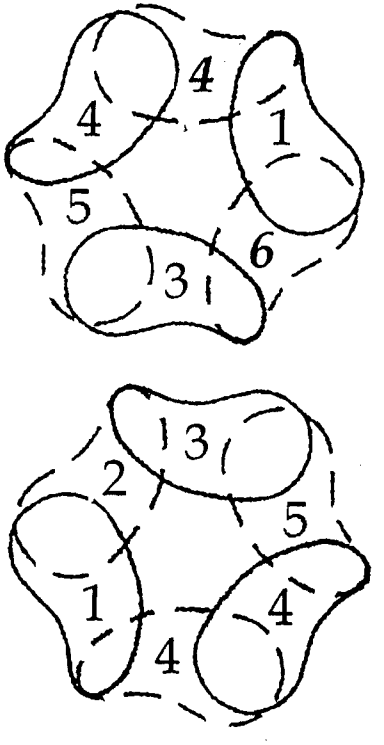
Our model, adapted from X-ray crystallography of one-hexamer *Panulirus interruptus* Hc (Volbeda and Hol, 1989) and image analysis of two-hexamer *Cancer pagurus* Hc (de Haas et al., 1991), both crustacean Hcs, requires a minimal number of changes to account for the observed developmental changes

Fig.6: Subunit composition of juvenile and adult two-hexamer Hc. Stoichiometries determined from SDS-PAGE scans, using JAVA Peakfit (Jandel Scientific) to correlate relative areas under the curves. In native protein, upper hexamer would be perpendicular to lower one (rotation around the long axis of the two-hexamer) (36), and both intra- and interhexameric distances would be shorter. Developmentally regulated subunits are shaded.

juvenile



adult



subunit	1	2	3	4	5	6		1	2	3	4	5	6
%composition	18.8	9.5	17	22.6	32.1	0		16.9	7.9	19.8	30.5	12.9	12.4
copies/25S	2	1	2	3	4	0		2	1	2	4	2	1

in subunit expression. In the native protein, the upper hexamer would be perpendicular to the lower one (90° rotation along the long hexamer-hexamer axis, see Fig.2 in chapter I and de Haas et al., 1991). In Fig.6 we present both hexamers parallel to each other and have expanded the distance between subunits for clarity.

Based on the interhexameric contacts described by de Haas et al. (1991), subunit 3 has been positioned as the primary linker subunit. Subunit 3 is invariably present in *Cancer magister* 25S two-hexamer Hc and is always absent in the 16S one-hexamer Hc fraction (Terwilliger and Terwilliger, 1982). We know the 3-3 linkage in two-hexamer 25S Hc is neither a disulfide-linked dimer as has been shown for the Hc of the spider *Cupiennius* (Markl et al., 1976) nor one dependent on divalent cations at neutral pH as in the Hc of the lobster *Homarus* (Pickett et al. 1966). Instead, the 3-3 interhexamer linkage is probably due to a combination of weak non-covalent forces, since high alkalinity plus absence of divalent cations cause the two-hexamer Hc of *Cancer magister* to dissociate (Ellerton et al., 1970).

The subunits in arthropod Hc hexamers are arranged either as a dimer of trimers or a trimer of dimers; crustacean Hc hexamers appear to function as a dimer of trimers in that the upper and lower trimers of the hexamer are in relatively loose contact with each other (see Markl and Decker, 1992, for review). While there are many possible

positions for the remaining subunits, our model presents the most parsimonious arrangement, the one implying the fewest subunit substitutions consistent with our stoichiometric data. One hexamer of each two-hexamer oligomer, the one composed of (1-3-4) and (2-5-4), is identical in both juvenile and adult Hc. In addition, the other hexamer in both juvenile and adult 25S Hc also contains a (1-3-4) trimer. Only one trimer varies between juvenile and adult Hc, its composition being either (5-5-5) or (5-4-6). Furthermore, contacts between the two hexamers of both juvenile and adult Hc are invariant as well, with subunits 3-3 the primary and subunits 2-5 the secondary contact. The subunits that change during development (5-5 to 4-6) are thus on the periphery of the molecule, minimizing any changes in cooperativity between the two hexamers. This peripheral location is consistent with our observation that the primary functional difference between juvenile and adult Hc is in oxygen affinity, not in cooperativity (Terwilliger et al., 1986; Terwilliger and Brown, 1993). Our model, then, accounts for the developmental changes in Hc subunit composition by substitutions in only two of the twelve subunit positions.

The data presented here show that first, hemocyanin is synthesized in one tissue, the hepatopancreas, and second, expression of adult Hc subunit 6 begins during 6th instar in the crab, *C. magister* . The model we have presented suggests

how initiation of synthesis of one subunit and changes in rate of synthesis of two other subunits can alter the respiratory physiology of the crab. These results lead to further questions about stage specific regulation of synthesis of six different polypeptide chains and their coordinated assembly into developmentally appropriate multisubunit molecules. Future studies will hopefully enhance understanding of the molecular mechanisms controlling these ontogenetic changes.

In chapter III we investigated developmental changes in the expression of the respiratory protein Hc at the molecular level and described the first documented case of ontogenetic change in a copper-based respiratory protein. In chapter IV we will present the complete cDNA- and protein sequence of this developmentally regulated subunit, along with the sequence of another putative Hc subunit from *Cancer magister*. We will then align both *Cancer magister* sequences with proteins that display apparent sequence similarities. Computer-assisted predictions of hydrophilicity, surface probability and regional backbone flexibility will be compared among the taxa in order to detect a possible conservation of structural features. The sequence alignment will be evaluated by parsimony analysis, and the resulting most parsimonious phylogenetic tree will be discussed in the light of the evolutionary history of the Hc gene family.

CHAPTER IV

cDNA SEQUENCE OF A DEVELOPMENTALLY REGULATED HEMOCYANIN SUBUNIT IN *CANCER MAGISTER*: IMPLICATIONS FOR THE PHYLOGENY OF THE HEMOCYANIN GENE FAMILY

Gregor Durstewitz

and

Nora Barclay Terwilliger

Oregon Institute of Marine Biology and Department of Biology, University of Oregon, Eugene, OR 97403.

ABSTRACT

Complete cDNA sequences coding for one and possibly two hemocyanin (Hc) subunits in the Dungeness crab (*Cancer magister*) are presented. One, the developmentally regulated Hc subunit 6, was amplified in two fragments from hepatopancreas tissue of the crab using the polymerase chain reaction (PCR). The amplified fragments were cloned into a Bluescript II SK^+ vector and sequenced with the dideoxy-method. The sequence shows an open reading frame of 650 amino acids. The other one was obtained by a hepatopancreas tissue cDNA library screen and contains a 191 amino acid deletion between residues 410 and 601. Both sequences are aligned with 17 other proteins displaying apparent sequence similarities. Functional domains are identified, and a comparison of predicted hydrophilicities, surface probabilities and regional backbone flexibilities provides evidence for a remarkable degree of structural conservation among the proteins surveyed. Parsimony analysis of these sequences allows phylogenetic reconstruction of their evolutionary history. Confidence limits were established with the bootstrap approach. The most parsimonious phylogenetic tree consistent with the dataset identifies 4 monophyletic groups on the arthropod branch: crustacean Hcs, insect hexamerins, chelicerate Hcs and arthropod prophenoloxidases. They form a monophyletic group relative

to molluscan Hcs and non-arthropod tyrosinases. Results for individual clades are evaluated and discussed in the light of the evolutionary history of the Hc gene family.

INTRODUCTION

Hemocyanins and related copper proteins are ancient molecules. They probably arose about 1.6 billion years ago when the earth's atmosphere changed from a reducing to an oxidizing environment. Prior to that time, most of the earth's available copper (Cu) was precipitated in insoluble sulfides, CuS and Cu_2S (Ochiai, 1983), and was therefore virtually inaccessible to living organisms. Due to the rise of oxygen-producing photosynthesis about 2 bya, significant amounts of Cu were being oxidized to Cu^{2+}. In this form it is readily dissolved in aquatic systems, distributed in the biosphere and therefore available to living organisms.

Hemocyanin (Hc), the oxygen transport protein of many arthropods and molluscs, was named for the blue color it displays in the oxygenated state (Fredericq, 1878). They occur freely dissolved in the hemolymph. Though similar in function to molluscan Hc, arthropodan Hc is radically different in molecular architecture. Arthropodan Hc is composed of heterogeneous subunits with molecular weights of about 75 kDa. These subunits self-assemble into hexamers or

multiples thereof. Each subunit contains two Cu binding sites, CuA and CuB, that cooperate to reversibly bind one molecule of oxygen.

The evolutionary relationship between the Hcs has long been the topic of speculation: are arthropodan and molluscan Hcs homologous gene products or the result of convergent evolution ? They are certainly very different in sequence as well as in subunit structure and composition (van Holde and Miller, 1995), but there is good evidence for a common origin of at least part of their active site (Drexel et al., 1987). On the basis of sequence comparisons, other members of a putative Hc gene family have recently been identified. They include the tyrosinases (Lerch et al., 1986), prophenoloxidases (Aspan et al., 1995) and insect storage proteins or hexamerins (Munn and Greville, 1969; Telfer and Massey, 1987). The latter do not have Cu binding sites. Other Cu proteins like the plastocyanins, Cu-dependent cytochrome c oxidase, ceruloplasmin, azurins, laccase, ascorbate oxidase or Cu-dependent superoxide dismutase show no structural or sequence similarity with the Hc family (Markl and Decker, 1992).

In this paper we present the complete protein sequences of 2 Hc subunits, Cmag6 and CmagX, from a brachyuran crustacean, the Dungeness crab *Cancer magister*. In brachyurans, Hc occurs in the hemolymph predominantly as a 2-hexamer molecule, and subunit sequences of multihexameric

crustacean Hcs have been unavailable up to now. Cmag6 is the first Cu-based respiratory protein whose expression appears to be developmentally regulated (Terwilliger and Terwilliger, 1982; Durstewitz and Terwilliger, 1996). Unlike the molt cycle related changes in hemolymph concentration of insect storage proteins (Telfer and Kunkel, 1991), the changes in Hc expression during the transition from juvenile to adult Hc in the Dungeness crab are of a more permanent nature; they are initiated at a certain developmental stage and persist for the rest of the crab's life.

We use these sequences, as well as those of potentially related proteins, to address the following questions. (1), is there evidence for homology and a "Hc gene family" based on conserved structural features among these proteins ?, and (2), what relationship between Hc and other potentially related proteins is suggested by parsimony analysis of protein sequence alignments ?

The results of both sequence and structural comparisons will be used to shed light on the evolutionary relationships among Hc-type proteins.

MATERIALS AND METHODS:

Total RNA Isolation

Fresh tissue samples (100 mg) from adult male *C. magister* hepatopancreas tissue were rinsed with *C. magister*

hemolymph buffer (50 mM Tris-HCl, 454 mM NaCl, 11.5 mM KCl, 13.5 mM $CaCl_2$, 18 mM $MgCl_2$, 23.5 mM Na_2SO_4, pH 7.6; Terwilliger and Brown, 1993), frozen in liquid nitrogen and ground to a fine powder with mortar and pestle. Total RNA was isolated with the guanidinium isothiocyanate method using a RAPID Total RNA Isolation Kit (5 Prime -> 3 Prime, Inc.). Total RNA yield was quantified by measuring absorbance at 260 nm.

Reverse Transcription and Polymerase Chain Reaction (PCR) Amplification of Hemocyanin Coding Sequences

In an eppendorf tube, 1 µl total RNA (1 µg/µl) was diluted with 10.65 µl autoclaved water, and 0.75 µl oligo dT-primer (0.27 µg/µl) was added. The mixture incubated for 3 min at 65°C and then was allowed to cool down to room temperature. Next were added in this order: 4 µl 5x reverse transcription buffer (250 mM Tris-HCl pH 8.5, 200 mM KCl, 30 mM $MgCl_2$), 1 µl 20 mM dithiothreiotol (DTT), 1 µl 25 mM dNTPs, 1 µl RNAsin (10 u/µl). The reaction was incubated at 42°C for 90 min and then diluted to a total volume of 500 µl. A 10 µl aliquot of this reverse transcription reaction was added to 18.5 µl water, 5 µl 10x PCR-buffer (670 mM Tris-HCl), 4 µl 2.5 mM dNTPs, 5 µl 10x bovine serum albumin (1 µg/µl), 1 µl of each primer (0.2 µg/µl), 0.5 µl Taq polymerase (5 u/µl) and 5 µl 40 mM $MgCl_2$. PCR was carried out using the following protocol: (denature: 94°C for 40

sec, anneal: 55°C for 40 sec, polymerize: 72°C for 1 min) * repeat for 35 cycles, then 5 min at 72°C and hold at 4°C. 10 µl aliquots of each reaction were analyzed on 1.2% agarose Tris-acetate/EDTA (TAE) minigels.

Cloning and Sequencing of PCR-Amplified Hemocyanin cDNA

Unless indicated otherwise, the following procedures were performed in accordance with Sambrook et al. (1989). PCR products (40 µl total) were separated by size through electrophoresis on 1.2% agarose TAE maxigels. Bands of interest were excised under UV-light and purified in a glassmilk procedure (GENECLEAN II kit, Bio 101, Inc.). Ends were repaired with Klenow polymerase and 5′ ends phosphorylated with T4 polynucleotide kinase. A Bluescript II SK^+ vector (Stratagene) was cut with restriction endonuclease SmaI and dephosphorylated using calf intestinal phosphatase (CIP). In a total volume of 20 µl, the inserts were blunt-end-ligated into 50 ng vector DNA in a molar ratio insert/vector of ≈ 3:1, using 1 Weiss unit T4 DNA ligase. Ligation occurred overnight at 16°C. Competent Escherichia coli XL-1 Blue cells were transformed with 50 ng ligated DNA and plated out on LB-Amp plates. Positive clones were selected and DNA was isolated in alkaline lysis minipreps. When desired, inserts were excised and analyzed using restriction enzymes EcoRV and XbaI. cDNA inserts were

sequenced using the dideoxy method with a SEQUENASE 2.0 kit (US Biochemical) using radioactive ^{35}S labeled nucleotides (NEN-Du Pont). T3 and T7 were used as initial sequencing primers. Further primers were designed as 17mers based on stretches of cDNA sequence located approximately at the -50 bp position relative to the end of the known region.

Screening a cDNA Library of *Cancer magister* Hepatopancreas Tissue

A cDNA library of adult *C. magister* hepatopancreas tissue was created as described before (Terwilliger and Durstewitz, 1996) and screened with a ^{32}P random-prime-labeled 783 bp *C. magister* Hc specific probe (Durstewitz and Terwilliger, 1996). Positive clones were analyzed for insert size. An 1800 bp insert (CmagX) was sequenced by creating overlapping nested deletions: The clone containing the CmagX cDNA fragment was digested with DNase I in the presence of Mn^{2+} (Lin et al., 1985; Terwilliger and Durstewitz, 1996). The resulting population of plasmids contained overlapping nested deletions and was sequenced with the dideoxy method.

RESULTS

The complete cDNA sequence coding for *C. magister* Hc

subunit 6 was amplified in 2 overlapping fragments by PCR. The template was 1st strand cDNA derived from hepatopancreas total RNA. The 4 primers used to amplify Hc coding sequences were (1) a degenerate primer based on the unique N-terminal amino acid sequence of *C. magister* Hc subunit 6 (5′ TCT-GCT-GGT-GGT-GCT-TTT-GAT-GCT-CA 3′, "5′ sub 6 primer", Durstewitz and Terwilliger, 1996), (2) an antisense primer based on the 3′ PCR product (5′ CAC-TGC-CTG-GGG-ATC-GAA-GCC-CTC-ATG 3′, "CuA II primer", see Fig.1), (3) a degenerate primer based on a conserved sequence within the copper A site (CuA) of arthropod Hc (5′ GAA-TTT-TTT-TTT-TGG-GTT-CAT-CAT-CAA-TTT-AC 3′, "CuA I primer") and (4) a universal oligo-dT primer. Using the primer combinations (1 and 2) plus (3 and 4), the reaction generated 2 overlapping cDNA fragments (Fig.1). Each fragment was blunt-end-cloned into a Bluescript II SK^+ vector (Fig.2) and sequenced. Fragment pSK 6 coded for the 5′ end, the other, pSK 1602, for the 3′ end of Hc subunit 6. The identical overlap between both clones was 133 bp. The complete cDNA sequence of *C. magister* Hc subunit 6 (GenBank accession # U48881) with the correct protein reading frame (Cmag6) is shown in Fig.3.

The subunit is composed of 650 amino acid residues. The six histidines marked by an asterisk have been implicated in Cu binding and are highly conserved among other arthropod Hc subunits (Linzen et al., 1985; Beintema et al., 1994). The molecular weight of subunit 6, calculated from the amino

Fig.1: PCR-amplification of various regions of *Cancer magister* Hc subunit 6 cDNA.

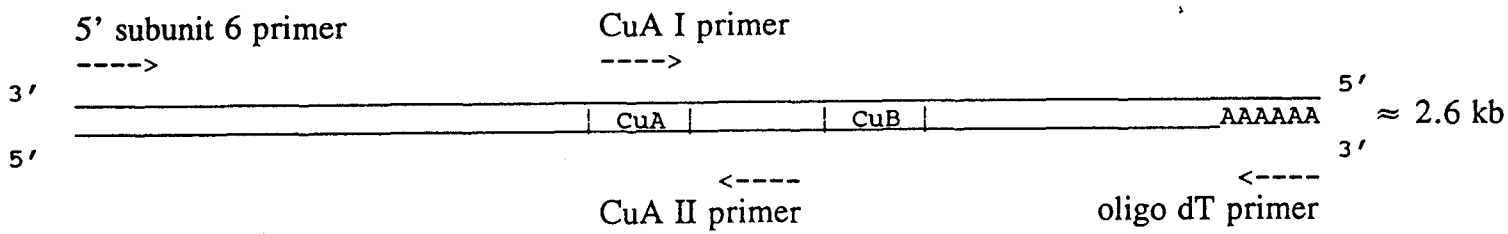
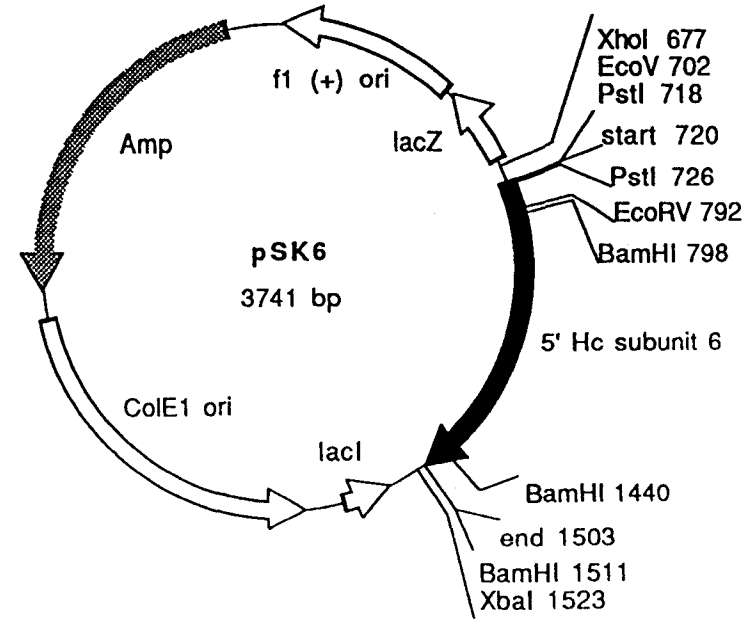

5' subunit 6 primer
---->
CuA I primer
---->
3'
5'
CuA
CuB
AAAAAA
5'
3'
≈ 2.6 kb
<----
CuA II primer
<----
oligo dT primer

Fig.2: cDNA clones of *Cancer magister* Hc subunit 6. Top, clone pSK 6 with 5′ fragment. Bottom, clone pSK 1602 with 3′ fragment.


XhoI 677
EcoV 702
PstI 718
start 720
PstI 726
EcoRV 792
BamHI 798
f1 (+) ori
Amp
lacZ
pSK6
3741 bp
5' Hc subunit 6
ColE1 ori
lacI
BamHI 1440
end 1503
BamHI 1511
XbaI 1523

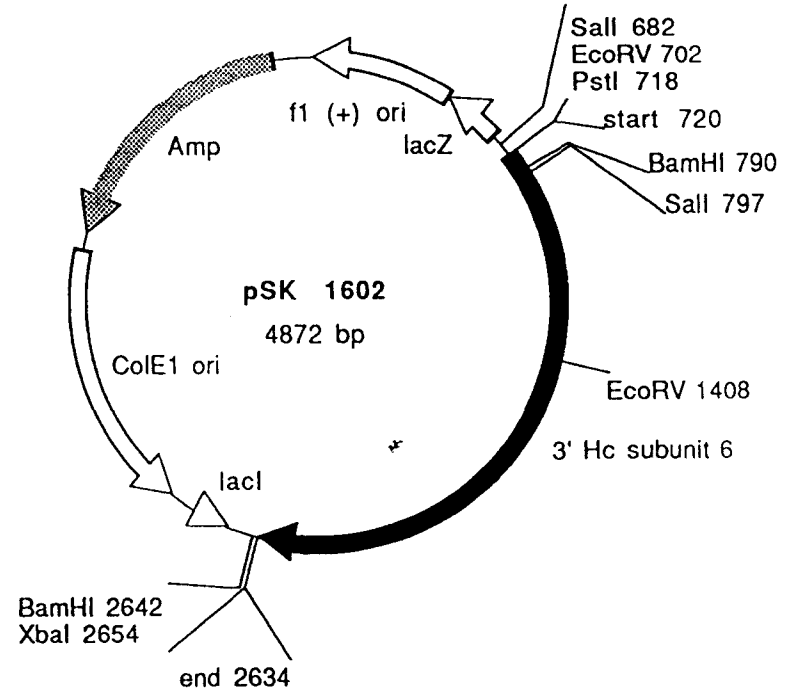

SalI 682
EcoRV 702
PstI 718
start 720
BamHI 790
SalI 797
f1 (+) ori
Amp
lacZ
pSK 1602
4872 bp
ColE1 ori
EcoRV 1408
3' Hc subunit 6
lacI
BamHI 2642
XbaI 2654
end 2634

Fig.3: *Cancer magister* Hc subunit 6. cDNA sequence and correct protein reading frame. *, conserved histidine, presumably acting as Cu ligand.

```
   ACTGCAGGCGGAGCGTTCGACGCGCAGAAGCAGCACGATGTCAACAGCGCTCTGTGGAAG
 1 ---------+---------+---------+---------+---------+---------+ 60
   TGACGTCCGCCTCGCAAGCTGCGCGTCTTCGTCGTGCTACAGTTGTCGCGAGACACCTTC

   T  A  G  G  A  F  D  A  Q  K  Q  H  D  V  N  S  A  L  W  K

   GTCTACGAGGATATCCAGGATCCCCACCTAATACAACTTTCCCAGAACTTCGACCCGCTC
61 ---------+---------+---------+---------+---------+---------+ 120
   CAGATGCTCCTATAGGTCCTAGGGGTGGATTATGTTGAAAGGGTCTTGAAGCTGGGCGAG

   V  Y  E  D  I  Q  D  P  H  L  I  Q  L  S  Q  N  F  D  P  L

   TCCGGCCACTATGACGACGATGGTGTCGCCGCCAAGCGCCTCATGAAGGAGCTCAACGAA
121 ---------+---------+---------+---------+---------+---------+ 180
   AGGCCGGTGATACTGCTGCTACCACAGCGGCGGTTCGCGGAGTACTTCCTCGAGTTGCTT

   S  G  H  Y  D  D  D  G  V  A  A  K  R  L  M  K  E  L  N  E

   AACCGCTTGCTGAAGCAGAACCACTGGTTCTCACTGTTCAACACCCGCCAGCGCGAGGAG
181 ---------+---------+---------+---------+---------+---------+ 240
   TTGGCGAACGACTTCGTCTTGGTGACCAAGAGTGACAAGTTGTGGGCGGTCGCGCTCCTC

   N  R  L  L  K  Q  N  H  W  F  S  L  F  N  T  R  Q  R  E  E

   GCTCTCATGCTCTACGACGTCCTCGAACACTCCACAGACTGGAGCACCTTCGCCGGCAAC
241 ---------+---------+---------+---------+---------+---------+ 300
   CGAGAGTACGAGATGCTGCAGGAGCTTGTGAGGTGTCTGACCTCGTGGAAGCGGCCGTTG

   A  L  M  L  Y  D  V  L  E  H  S  T  D  W  S  T  F  A  G  N

   GCTGCCTTCTTCCGCGTTAGCATGAACGAGGGCGAGTTCGTTTACGCACTGTACGCTGCC
301 ---------+---------+---------+---------+---------+---------+ 360
   CGACGGAAGAAGGCGCAATCGTACTTGCTCCCGCTCAAGCAAATGCGTGACATGCGACGG

   A  A  F  F  R  V  S  M  N  E  G  E  F  V  Y  A  L  Y  A  A

   GTTATCCACTCTGAGCTGACACAACACGTGGTGCTACCACCCCTCTACGAGGTCACTCCT
361 ---------+---------+---------+---------+---------+---------+ 420
   CAATAGGTGAGACTCGACTGTGTTGTGCACCACGATGGTGGGGAGATGCTCCAGTGAGGA

   V  I  H  S  E  L  T  Q  H  V  V  L  P  P  L  Y  E  V  T  P

   CACCTCTTCACCAACAGCGAGGTGATCCAAGAAGCCTACAAAGCCAAGATGACCCAGACT
421 ---------+---------+---------+---------+---------+---------+ 480
   GTGGAGAAGTGGTTGTCGCTCCACTAGGTTCTTCGGATGTTTCGGTTCTACTGGGTCTGA

   H  L  F  T  N  S  E  V  I  Q  E  A  Y  K  A  K  M  T  Q  T
```

```
    GCCGCCAAGATTGAGTCCCACTTCACCGGCAGCAAGAGTAACCCGGAACAGCGTGTGGCC
481 ---------+---------+---------+---------+---------+---------+ 540
    CGGCGGTTCTAACTCAGGGTGAAGTGGCCGTCGTTCTCATTGGGCCTTGTCGCACACCGG

    A  A  K  I  E  S  H  F  T  G  S  K  S  N  P  E  Q  R  V  A

                                      *           *
    TACTTCGGCGAGGACATCGGCATGAATACCCATCACGTCACCTGGCATTTGGAGTTCCCC
541 ---------+---------+---------+---------+---------+---------+ 600
    ATGAAGCCGCTCCTGTAGCCGTACTTATGGGTAGTGCAGTGGACCGTAAACCTCAAGGGG

    Y  F  G  E  D  I  G  M  N  T  H  H  V  T  W  H  L  E  F  P

    TTCTGGTGGGACGACGCCCATGAGAACCACCACATCGAGCGCAAGGGCGAGAGCTGTTCT
601 ---------+---------+---------+---------+---------+---------+ 660
    AAGACCACCCTGCTGCGGGTACTCTTGGTGGTGTAGCTCGCGTTCCCGCTCTCGACAAGA

    F  W  W  D  D  A  H  E  N  H  H  I  E  R  K  G  E  S  C  S
              *
    TCTTGGGTCCACCACCAGCTCACTGTCCGCTTCGACGCCGAGCGTCTGTCTAACTACTTG
661 ---------+---------+---------+---------+---------+---------+ 720
    AGAACCCAGGTGGTGGTCGAGTGACAGGCGAAGCTGCGGCTCGCAGACAGATTGATGAAC

    S  W  V  H  H  Q  L  T  V  R  F  D  A  E  R  L  S  N  Y  L

    GATCCCGTCGACGAACTCCACTGGGACGATGTCATCCATGAGGGCTTCGATCCCCAGGCA
721 ---------+---------+---------+---------+---------+---------+ 780
    CTAGGGCAGCTGCTTGAGGTGACCCTGCTACAGTAGGTACTCCCGAAGCTAGGGGTCCGT

    D  P  V  D  E  L  H  W  D  D  V  I  H  E  G  F  D  P  Q  A

    GTGTATAAGTACGGCGGATATTTCCCCTCCCGCCCTGACAATATCCACTTTGAAGATGTG
781 ---------+---------+---------+---------+---------+---------+ 840
    CACATATTCATGCCGCCTATAAAGGGGAGGGCGGGACTGTTATAGGTGAAACTTCTACAC

    V  Y  K  Y  G  G  Y  F  P  S  R  P  D  N  I  H  F  E  D  V

    GATGGTGTTGCTGATGTTCGTGACATGCTTTTGTATGAAGAACGTATTCTTGACGCTACT
841 ---------+---------+---------+---------+---------+---------+ 900
    CTACCACAACGACTACAAGCACTGTACGAAAACATACTTCTTGCATAAGAACTGCGATGA

    D  G  V  A  D  V  R  D  M  L  L  Y  E  E  R  I  L  D  A  T

    GCTCATGGCTACGTGCGGATCAACGGTCAGATCGTTGACCTGAGAAACAATGATGGCATC
901 ---------+---------+---------+---------+---------+---------+ 960
    CGAGTACCGATGCACGCCTAGTTGCCAGTCTAGCAACTGGACTCTTTGTTACTACCGTAG

    A  H  G  Y  V  R  I  N  G  Q  I  V  D  L  R  N  N  D  G  I
```

```
     GATCTCCTTGGAGACGTGATTGAATCTTCCTTATACAGCCCCAATCCTCAGTACTACGGC
961 ---------+---------+---------+---------+---------+---------+ 1020
     CTAGAGGAACCTCTGCACTAACTTAGAAGGAATATGTCGGGGTTAGGAGTCATGATGCCG

     D  L  L  G  D  V  I  E  S  S  L  Y  S  P  N  P  Q  Y  Y  G
           *              *
     GCCCTGCACAACACAGCTCATATGATGCTTGGCCGCCAGGGTGACCCTCATGGAAAGTTC
1021 ---------+---------+---------+---------+---------+---------+ 1080
     CGGGACGTGTTGTGTCGAGTATACTACGAACCGGCGGTCCCACTGGGAGTACCTTTCAAG

     A  L  H  N  T  A  H  M  M  L  G  R  Q  G  D  P  H  G  K  F

     GACCTTCCTCCCGGTGTTCTGGAGCACTTCGAGACCGCAACACGTGATCCCGCTTTCTTC
1081 ---------+---------+---------+---------+---------+---------+ 1140
     CTGGAAGGAGGGCCACAAGACCTCGTGAAGCTCTGGCGTTGTGCACTAGGGCGAAAGAAG

     D  L  P  P  G  V  L  E  H  F  E  T  A  T  R  D  P  A  F  F
           *
     CGTCTACACAAGTACATGGATAACATCTTCAGAAAACACAAGGACAGCCTGCCACCCTAC
1141 ---------+---------+---------+---------+---------+---------+ 1200
     GCAGATGTGTTCATGTACCTATTGTAGAAGTCTTTTGTGTTCCTGTCGGACGGTGGGATG

     R  L  H  K  Y  M  D  N  I  F  R  K  H  K  D  S  L  P  P  Y

     ACTAAGGAAGAGCTTAACTTTGAGGGTGTTAACATCGATAACTTCTACATTAAGGGAAAT
1201 ---------+---------+---------+---------+---------+---------+ 1260
     TGATTCCTTCTCGAATTGAAACTCCCACAATTGTAGCTATTGAAGATGTAATTCCCTTTA

     T  K  E  E  L  N  F  E  G  V  N  I  D  N  F  Y  I  K  G  N

     TTGGAAACCTATTTTGAGACCTTCGAGTACAGTCTTGTGAATGCTGTTGACGACACAGAA
1261 ---------+---------+---------+---------+---------+---------+ 1320
     AACCTTTGGATAAAACTCTGGAAGCTCATGTCAGAACACTTACGACAACTGCTGTGTCTT

     L  E  T  Y  F  E  T  F  E  Y  S  L  V  N  A  V  D  D  T  E

     GATGTCGATGACGTGGATATCTTCACGTATATTTCACGCTTGAATCATAAGGAATTTTCA
1321 ---------+---------+---------+---------+---------+---------+ 1380
     CTACAGCTACTGCACCTATAGAAGTGCATATAAAGTGCGAACTTAGTATTCCTTAAAAGT

     D  V  D  D  V  D  I  F  T  Y  I  S  R  L  N  H  K  E  F  S

     TTTGTTGGTGATGTCACCAATGAACTTGATCATGATGTACTAGCCACTGTGCGCATCTTT
1381 ---------+---------+---------+---------+---------+---------+ 1440
     AAACAACCACTACAGTGGTTACTTGAACTAGTACTACATGATCGGTGACACGCGTAGAAA

     F  V  G  D  V  T  N  E  L  D  H  D  V  L  A  T  V  R  I  F
```

```
     GCCTGGCCGCACGAGGACAACAATGGAGTGGCGTTCAGCTTCAACGATGGTCGCTGGAAC
1441 ---------+---------+---------+---------+---------+---------+ 1500
     CGGACCGGCGTGCTCCTGTTGTTACCTCACCGCAAGTCGAAGTTGCTACCAGCGACCTTG

     A  W  P  H  E  D  N  N  G  V  A  F  S  F  N  D  G  R  W  N

     GCCATCGAAATGGACAAGTTCTGGGTTATGTTGCATCCCGGCCACAACCACATCGAGCGA
1501 ---------+---------+---------+---------+---------+---------+ 1560
     CGGTAGCTTTACCTGTTCAAGACCCAATACAACGTAGGGCCGGTGTTGGTGTAGCTCGCT

     A  I  E  M  D  K  F  W  V  M  L  H  P  G  H  N  H  I  E  R

     TCGTCTCATGACTCCTCCGCGACCGTTCCTGATATACCCAGCTTCCAAATCATTAAGGAC
1561 ---------+---------+---------+---------+---------+---------+ 1620
     AGCAGAGTACTGAGGAGGCGCTGGCAAGGACTATATGGGTCGAAGGTTTAGTAATTCCTG

     S  S  H  D  S  S  A  T  V  P  D  I  P  S  F  Q  I  I  K  D

     AGGACCAATGAAGCGATAGCTCAGAACAAGGAACTCCATATTGAAGAATTTGAAAGCGGT
1621 ---------+---------+---------+---------+---------+---------+ 1680
     TCCTGGTTACTTCGCTATCGAGTCTTGTTCCTTGAGGTATAACTTCTTAAACTTTCGCCA

     R  T  N  E  A  I  A  Q  N  K  E  L  H  I  E  E  F  E  S  G

     CTTGGCCTGCCAAACAGGTTCCTCATTCCCAAGGGCAATGTGAAGGGCCTTGACATGGAT
1681 ---------+---------+---------+---------+---------+---------+ 1740
     GAACCGGACGGTTTGTCCAAGGAGTAAGGGTTCCCGTTACACTTCCCGGAACTGTACCTA

     L  G  L  P  N  R  F  L  I  P  K  G  N  V  K  G  L  D  M  D

     GTAATGGTGGCCATCACGAGCGGAGAGGCGGATGCTGCCGTTGAAGGGTTGCACGAAAAC
1741 ---------+---------+---------+---------+---------+---------+ 1800
     CATTACCACCGGTAGTGCTCGCCTCTCCGCCTACGACGGCAACTTCCCAACGTGCTTTTG

     V  M  V  A  I  T  S  G  E  A  D  A  A  V  E  G  L  H  E  N

     ACTTCCTTCAACCACTACGGCTGTCCTGACGGCACCTACCCAGACAAGAGGCCCCACGGT
1801 ---------+---------+---------+---------+---------+---------+ 1860
     TGAAGGAAGTTGGTGATGCCGACAGGACTGCCGTGGATGGGTCTGTTCTCCGGGGTGCCA

     T  S  F  N  H  Y  G  C  P  D  G  T  Y  P  D  K  R  P  H  G

     TACCCACTGGACCGCCACGTCGACGATGAGCGCATCATCAATGACTTGCACAACTTCAAG
1861 ---------+---------+---------+---------+---------+---------+ 1920
     ATGGGTGACCTGGCGGTGCAGCTGCTACTCGCGTAGTAGTTACTGAACGTGTTGAAGTTC

     Y  P  L  D  R  H  V  D  D  E  R  I  I  N  D  L  H  N  F  K

     CACATTCAGGTCAAGGTGTTCCATCATGCG
1921 ---------+---------+---------+
     GTGTAAGTCCAGTTCCACAAGGTAGTACGC
     H  I  Q  V  K  V  F  H  H  A
```

acid residues, was determined to be 74903 Da, as opposed to an estimate of 67300 Da, based on its mobility on SDS-PAGE gels, by Larson et al. (1981). The subunit is quite acidic (isoelectric point pI = 5.02). A glycosylation site (Asn-Thr-Ser) occurs at residue 600.

The sequence of another putative Hc subunit, CmagX, was obtained from a *C. magister* hepatopancreas cDNA library. We call it a "putative Hc subunit" for three reasons: (1), it was obtained through a cDNA library screen with a Hc-specific probe, (2), it shows an extremely high degree of sequence similarity with Cmag6 (85% sequence identity, Fig.9), and (3), all of its potential Cu ligands are conserved. It is unknown, however, which Hc subunit, if any, this clone represents, whether it might even code for a crustacean storage protein or prophenoloxidase or whether it reflects an error in reverse transcription or 2nd strand cDNA synthesis. Its 484 amino acid open reading frame sports a 191 residue deletion between Cmag6 residues 410 and 601. This extensive deletion extends from the C-terminal part of domain 2, just beyond the second Cu binding site (CuB), well into domain 3. However, all putative Cu binding histidine residues are preserved. In addition to that, it shows a typical signal peptide of 21 hydrophobic residues at the N-terminal end, indicating the gene product is targeted for secretion. Whether other arthropod Hc subunits (including Cmag6) contain a signal peptide is not known. The upstream

primer used to amplify Cmag6 by PCR is identical to the N-terminal sequence of the mature subunit, and any sequence information upstream of that primer would be lost.

A protein sequence alignment of both *C. magister* Hc subunits, Cmag6 and CmagX, with other members of the Hc family is shown on Fig.4. Alignment was done by hand. It was our goal to include in our analysis representatives of all major groups within the Hc family of proteins. Among these, the O_2 transporting Hcs of arthropods and molluscs are respiratory proteins. Tyrosinases and prophenoloxidases (Lerch et al., 1986), both binuclear copper proteins, are enzymes involved in dopa and melanin biosynthesis, catalyzing the hydroxylation of mono- and the oxidation of diphenols. Recent studies (Aspan et al., 1995) assign prophenoloxidases a key role in the arthropod immune system. Another group of proteins, the hexamerins, is also found in insect hemolymph. One of several functions assigned to these hexamerins is that of storage proteins during insect metamorphosis (Telfer and Kunkel, 1991). Hexamerins include the arylphorins, proteins rich in aromatic amino acids, and the methionin-rich storage proteins. Although structurally similar to arthropod Hcs, hexamerins contain no copper.

Figures 5-7 compare predictions of structural features of the arthropod proteins aligned in Fig.4 as predicted by the PEPTIDESTRUCTURE program (GCG Sequence analysis software package, Devereux et al., 1984). Hydrophilicity (window size

Fig.4: Sequence alignment of Hc-type proteins. Residue numbers refer to Cmag6. Other numbers indicate protein domain. *, conserved histidine, presumably acting as Cu ligand.
Cmag6 = *C. magister* Hc subunit 6; CmagX = Possible *C. magister* Hc subunit with deletion between residues 410 and 601; Pintc = *Panulirus interruptus* Hc subunit c; Pinta = *Panulirus interruptus* Hc subunit a; Penv1 = *Penaeus vannamei* Hc subunit 1; LimII = *Limulus polyphemus* Hc subunit II; Euryd = *Eurypelma californica* Hc subunit d; Eurye = *Eurypelma californica* Hc subunit e; Anda6 = *Androctonus australis* Hc subunit 6; BombA = *Bombyx mori* storage protein 2; MsexA = *Manduca sexta* arylphorin subunit alpha; Tni_M = *Trichoplusia ni* basic juvenile hormone sensitive hemolymph protein 1; BombM = *Bombyx mori* sex-specific storage protein 1; PapPO = *Pacifastacus leniusculus* prophenoloxidase; DrpPO = *Drosophila melanogaster* prophenoloxidase; Octoe = *Octopus dofleini* Hc domain e; Hpomd = *Helix pomatia* Hc domain d; NeuTy = *Neurospora crassa* tyrosinase; HumTy = Human tyrosinase.

```
          11111111111111111111111111111111111111111111111111

Cmag6     ...............................TAGGAFDAQKQHDVNSALWK 20
CmagX     MKLLVLFA.LVAAAVAWPSFGM......MADSAGAPDAHKQHDVNSVLWK
Pintc     ........................ADCQAGDSADKLLAQKQHDVNYLVYK
Pinta     ...............................DALGTGNAQKQQDINHLLDK
Penv1     MRVLVVL.GLVAA..............AAFQVASADVQQQKDVLYLLNK
LimII     ....................................TLHDKQIRVCHLFEQ
Euryd     ....................................TIADHQARILPLFKK
Eurye     ....................................PDKQKQLRVISLFEH
Anda6     ....................................TVADKQARLMPLFKH
BombA     MKSVLILAGLVAVALSSAVPKP...STIKSKNVDAVFVEKQKKILSFFQD
MsexA     MKTVVILAGLVALALSSAVPPPKYQHHYKTSPVDAIFVEKQKKVFSLFKN
Tni_M     MRVLVLVASLGLR..GSVVKDDTTVVIGKDNMVTMDIKMKELCILKLLNH
BombM     MRVLVLLACLAAASASAISGGYGTMVFTKEPMVNLDMKMKELCIMKLLDH
PapPO     MQVTQKLLRRDTE.....................MADAQKQL..LYLFER
DrpPO     MTNTDLKALELMFQRPLEPAFT............TRDSGKTVLELPDSFY
Octoe     ..................................................
Hpomd     ..................................................
NeuTy     ..................................................
HumTy     ....................................MLLAVLYCLLWS

          11111111111111111111111111111111111111111111111111

Cmag6     VYEDIQDPHLIQLSQNFDPLSG..HYDDDGVAAKRLMKELNENRLLKQNH 68
CmagX     VYEDIQDPHLIQLSQNFDPLSG..HYDDDGVAAKRLMKELNENRLLKQNH
Pintc     LYGDIRDDHLKELGETFNPQGDLLLYHDNGASVNTLMADFKDGRLLQKKH
Pinta     IYEPTKYPDLKEIAENFNPLGDTSIYNDHGAAVETLMKELNDHRLLEQRH
Penv1     IYGDIQDGDLLATANSFDPVGNLGSYSDGGAAVQKLVQDLNDGKLLEQKH
LimII     LSSATVIGD..........GD.........KHKHSDRLKNVGKLQPGA
Euryd     LTSLSP..........DPLP..........EAERDPRLKGVGFLPRGT
Eurye     MTSIN...........TPLP..........RDQIDARLHHLGRLPQGE
Anda6     LTALTR..........EKLP..........LDQRDERLKGVGILPRGT
BombA     VSQLNTDDEYYKIGKDYDIEMN.MDNYTNKKAVEEFLKMYRTG.FMPKNL
MsexA     VNQLDYEAEYYKIGKDYDVEAN.IDNYSNKKVVEDFLLLYRTG.FMPKGF
Tni_M     ILQPTMYDDIREVAREWVIEEN.MDKYLKTDVVKKFIDTFKMG.MLPRGE
BombM     ILQPTMFEDIKEIAKEYNIEKS.CDKYMNVDVVKQFMEMYKMG.MLPRGE
PapPO     PYDPINAPRADGSFLYAVAGAXTVATRFGVAPTSTVTVPARPDADRRLLG
DrpPO     TDRYRNDTEEVGNRFSKDVDLKIPIQELSNVPSLEFTKKIGLKNQFSLFN
Octoe     ..................................................
Hpomd     ..................................................
NeuTy     ..................................................
HumTy     FQTSAGHFPRACVSSKNLMEKECCPPWSGDRSPCGQLSGRGSCQNILLSN
```

```
          11111111111111111111111111111111111111111111111111

Cmag6     WFSLFNTRQREEALMLYDVLEHSTDWSTFAGNAAFFRVSM..........  108
CmagX     WFSLFNTRQRKEALMLYDVLEHSTDWSTFAGNAAFFRVHM..........
Pintc     WFSLFNTRQREEALMMHRVLMNCKNWHAFVSNAAYFRTNM..........
Pinta     WYSLFNTRQRKEALMLFAVLNQCKEWYCFRSNAAYFRERM..........
Penv1     WFSLFNTRHRNEALMLFDVLIHCKDWASFVGNAAYFRQKM..........
LimII     IFSCFHPDHLEEARHLYEVFWEAGDFNDFIEIAKEARTFV..........
Euryd     LFGSFHEEHLAEAIVFIEIIHDAKNFDDFLALATNARAVV..........
Eurye     LFSCFHEEDLEEATELYKILYTAKDFDEVINLAKQSRTFV..........
Anda6     LFSCFHARHLAEATELYVALYGAKDFNDFIHLCEQARQIV..........
BombA     EFSVFYDKMRDEAIALLDLFYYAKDFETFYKSACFARVHL..........
MsexA     EFSIFYERMREEAIALFELFYYAKDFETFYKTASFARVHV..........
Tni_M     VFVHTNELHLEQAVKVFKIMYSAKDFDVFIRTACWLRERI..........
BombM     TFVHTNELQMEEAVKVFRVLYYAKDFDVFMRTACWMRERI..........
PapPO     RAPSVPRGAVFSFFIRSHREAARDLCDVLMKTQNSTDLMQLAASVRRHV.
DrpPO     NRHREIASELITLFMSAPNLRQFVSLSVYTKDRVNPVL............
Octoe     ..................................................
Hpomd     ..................................................
NeuTy     ..................................................
HumTy     APLGPQFPFTGVDDRESWPSVFYNRTCQCSGNFMGFNCGNCKFGFWGPNC


          11111111111111111111111111111111111111111111111111

Cmag6     NEGEFVYALYAAVIHSELTQHVVLPPLYEVTPHLFTNSEVI....QEAY.  153
CmagX     NEGEFVYALYAAVIHSELTQHVVLPPLYEVTPHLFTNSEVI....QEAY.
Pintc     NEGEYLYALYVSLIHSGLGEGVVLPPLYEVTPHMFTNSEVI....HEAY.
Pinta     NEGEFVYALYVSVIHSKLGDGIVLPPLYQITPHMFTNSEVI....DKAY.
Penv1     NEGEFVYALYVAVIHSSLAEQVVLPPLYEVTPHLFTNSEVI....EEAY.
LimII     NEGLFAFAAEVAVLHRDDCKGLYVPPVQEIFPDKFIPSAAI....NEAF.
Euryd     NEGLYAFAMSVALLSRDDCNGVVIPPIQEVFPDRFVPAETI....NRAL.
Eurye     NEGLFVYAVSVALLHRDDCKGIVVPAIQEIFPDRFVPTETI....NLAV.
Anda6     NEGMFVYAVSVAVLHREDCKGITVPPIQEVFPDRFVPAETI....NRAN.
BombA     NQGQFLYAFYIAVIQRPDCHGFVVPAPYEVYPKMFMNMEVL....QKIY.
MsexA     NEGMFLYAYYIAVIQRMDTNGLVLPAPYEVYPQYFTNMEVL....FKVD.
Tni_M     NGGMFVYALTACVFHRTDCRGITLPAPYEIYPYVFVDSHII....NKAF.
BombM     NGGMFVYAFTAACFHRTDCKGLYLPAPYEIYPYFFVDSHVI....SKAF.
PapPO     NENLFIYALSFTILRKQELRGVRLPPILEVFPHKFIPMEDLTSMQVEVNR
DrpPO     ....FQYAYAVAVAHRPDTREVPITNISQIFPSNFVEPSAFRDARQEASV
Octoe     ........................EGNEYLVRKNVERLSLSEMNSLIHAFR
Hpomd     ......................DAVTVASHVRKDLDTLTAGEIESLRSAFL
NeuTy     .STDIKFAITGVPTTPSSNGAVP.LRRELRDLQQNYPEQFNLYLLGLRDF
HumTy     TERRLLVRRNIFDLSAPEKDKFFAYLTLAKHTISSDYVIPIGTYGQMKNG
```

```
          1111111111111111111111111111111111111111222222222

Cmag6     .........KAKMTQTAAK...........IESHFTGSKSNPEQRVAYFG 183
CmagX     .........KAKMTQTAAK...........IESHFTGSKSNPEQRVAYFG
Pintc     .........KAQMTNTPSK...........FESHFTGSKKNPEQHVAYFG
Pinta     .........SAKMTQKQGT...........FNVSFTGTKKNREQRVAYFG
Penv1     .........RAKQKQTPGK...........FKSSFTGTKKNPEQRVAYFG
LimII     .........KKAHVRPEFDESP......ILVDVQDTGNILDPEYRLAYYR
Euryd     .........KVDKISDPNKD........TVVPIQKTGNIRDPEYNVAYFR
Eurye     .........KEAANHPDQD.........ISVHVVETGNILDEEYKLAYFK
Anda6     .........KEASNHPDQQS........IVVEAEETGNILDPEYKLSYFR
BombA     VTKMQHGLINPEAAAKYGIHK.ENDYFVYKANYSNAVLYNNEEQRLTYFT
MsexA     RIKMQDGFLNKDLAAYYGMYH.ENDNYVFYANYSNSLSYPNEEERIAYFY
Tni_M     MMKMTKAARDPVMLDYYGIKVTDKNLVVIDWRKGVRRTLT.EHDRISYFT
BombM     MMKMTKAAKDPVLWKYYGITVTDDNLVVIDWRKGVRRSLSQN.DVMSYFM
PapPO     ...................TPPTATTPLVIEYGPEFANTNQKAEHRVSYWR
DrpPO     ...................IGESGARVHVDIPQNYTASDREDEQRLAYFR
Octoe     RMQKDKSSDGFEAIASFHALPPLCPSPTAKHRHAC..........CLHGM
Hpomd     DIQQDHT...YENIASFHGKPGLCQH..EGHKVAC..........CVHGM
NeuTy     QGLDEAKLDSYYQVAGIHGMPFKPWAGVPSDTDWSQPGSSGFGGYCTHSS
HumTy     STPMFNDINIYDLFVWMHYYVSMDALLGGSEIWRDID........FAHEA


          2222222222222222222222222222222222222222222222222
            |<-helix 2.1->|                |<----helix 2.2----
                  *   *                          *
Cmag6     EDIGMNTHHVTWHLEFPFWWDDAHENHHIERKGES...CSSWVHHQLTVR 230
CmagX     EDIGMNTHHVTWHLEFPFWWDDAHENHHIERKGES....FFWVHHQLTVR
Pintc     EDVGMNTHHVLWHMEFPFWWEDS.SGRHLDRKGES....FFWVHHQLTVR
Pinta     EDIGMNIHHVTWHMDFPFWWEDS.YGYHLDRKGE....LFFWVHHQLTAR
Penv1     EDIGLNTHHVTWHMEFPFWWNDA.YGHHLDRKGE....NFFWIHHQLTVR
LimII     EDVGINAHHWHWHLVYPSTWNPKYFGKKKDRKGE....LFYYMHQQMCAR
Euryd     EDIGINSHHWHWHLVYPAFYDADFFGKIKDRKGE....LFYYMHQQMCAR
Eurye     EDVGTNAHHWHWHIVYPATWDPAFMGRMKDRKGE....LFYYMHQQMCAR
Anda6     EDIGINAHHWHWHIVYPATWNPTVMGKEKDRKGE....LFFYMHQQMCAR
BombA     EDIGMNAYYYYFHSHLPFWWTSEKYGALKERRGE....VYFYFYQQLLAR
MsexA     EDIGLNSYYYYFHMHLPFWWNSEKYGPFKERRGE....IYYYFYQQLIAR
Tni_M     EDIDLNTYMYYLHMSYPFWMTDDMYTVNKERRGE.....IMGTYTQLLAR
BombM     EDVDLNTYMYYLHMNYPFWMTDDAYGINKERRGE....IMMYANQQLLAR
PapPO     EDFGINSHHWHWHLVYPIEMN.....VNRDRKGE....LFYYMHQQMVAR
DrpPO     EDIGVNSHHWHWHLVYPTTGPTEV..VNKDRRGE....LFYYMHHQILAR
Octoe     ......ATFPHWHRLYVVQFEQALHRHGATVG..................
Hpomd     ......PTFPSWHRLYVEQVEEALLDHGSSVA..................
NeuTy     ......ILFITWHRPYLALYEQALYASVQAVAQKFPVEGGLRAKYVAAAK
HumTy     ......PAFLPWHRLFLLRWEQEIQKLTGDEN..................
```

```
          2222222222222222222222222222222222222222222222222
          ------->|
Cmag6     FDAERLSNYLDPVDELHW.DDVIHEGFDPQAVYK.YGGYFPS.RPDNIHF 277
CmagX     FDAERLSNYLDPVDELHW.DDVIHEGFAPHTMYK.YGGYFPS.RPDNVHF
Pintc     YDAERLSNHLDPVEELSW.NKAIDEGFAPHTAYK.YGGYFPS.RPDNVHF
Pinta     FDFERLSNWLDPVDELHW.DRIIREGFAPLTSYK.YGGEFPV.RPDNIHF
Penv1     FDAERLSNYLDPVGELQW.NKPIVDGFAPHTTYK.YGGQFPA.RPDNVKF
LimII     YDCERLSNGMHRMLPFNN.FDEPLAGYAPHLTHV.ASGKYYSPRPDGLKL
Euryd     YDCERLSVGLQRMLPFQN.IDDELEGYSPHLSSL.VSGLSYGSRPAGMHL
Eurye     YDCERLSNGMRRMIPFSN.FDEKLEGYSAHLTSL.VSGLPYAFRPDGLCL
Anda6     YDSERLSNGLQRMIPFHN.FDEPLEGYAPHLTSL.VSGLQYASRPEGYSI
BombA     YYFERLTNGLGKIPEFSW.YSPIKTGYYPLMLTK..FTPFAQ.RPDYYNL
MsexA     YYLERLTNGLGEIPEFSW.YSPVKTGYYP.MLYG.SYYPFAQ.RPNYYDI
Tni_M     LRLERLSHEMCDIKSIMW.NEPLKTGYWPKIRLH.TGDEMPV.RSNNKII
BombM     MRLERLSHKMCDVKPMMW.NEPLETGYWPKIRLP.SGDEMPV.RQNNMVV
PapPO     YDWERLSVNLNRVEKLENWRVPIPDGYFSKLTANNSGRPWGT.RQDNTFI
DrpPO     YNVERFCNNLKKVQPLNNLRVEVPEGYFPKILSSTNNRTYPA.RVTNQKL
Octoe     ...VPYWDWTRPISKIPDFIASEKYSDPFTKIEVYNPFNHGHISFISEDT
Hpomd     ...VPYFDWISPIQKLPDLISKATYYNSREQRFDPNPFFSGKVA..GEDA
NeuTy     DFRAPYFDWASQPPKGTLAFPESLSSRTIQVVDVDGKTKSINNPLHRFTF
HumTy     .FTIPYWDWRDAEKCDICTDEYMGGQHPTNPNLLSPASFFSSWQIVCSRL


          2222222222222222222222222222222222222222222222222

Cmag6     EDVDGVADVRDMLLYEERILDATAHGYVR.IN....GQIVDLRNND.... 318
CmagX     EDVDGVARVRDMLILESRIRDAIAHGYVTGRT....GSIISISDSH....
Pintc     SDVDGVARVRDMSMTEDRIRDAIAHGYIDALD....GSHIDIMNSH....
Pinta     EDVDGVAHVHDLEITESRIHEAIDHGYITDSD....GHTIDIRQPK....
Penv1     EDVDDVARIRDMVIVESRIRDAIAHGYIVDSE....GKHIDISNEK....
LimII     RD.LGDIEISEMVRMRERILDSIHLGYVISED....GSHKTLDELH....
Euryd     RD.INDCSVQ.MERWRERILDAIHTGLVTDSH....GKEIKITEEN....
Eurye     HD.LKDIDLKEMFRWRERILDAIDSGYYIDNE....GHQVKLDIVD....
Anda6     HD.LSDVDVQDMVRWRERILDAINMHYIVDKD....NNKIPLDIEH....
BombA     HTEENYERVRFLDTYEKTFVQFLQKDHFEAF.....GQKIDFHDPK....
MsexA     HNDKNYEQIRFLDMFEMTFLQYLQKGHFKAF.....DKEINFHDVK....
Tni_M     VTKENVKVKRMLDDVERMLRDGILTGKIERRD....GTIINLKKAE....
BombM     ATKDNLKMKQMMDDVEMMIREGILTGKIERRD....GTVISLKKSE....
PapPO     KDFRRNDAGLDFIDISDMEIWRSRLMDAIHQGYMLNRNGERVPLSDNVTT
DrpPO     RDVDRHDGRVE...ISDVERWRDRVLAAIDQGYVEDSSGNRIPL.DEV..
Octoe     TTKREVSEYLFEHPVLGKQTWLFDNIAL.ALEQTDYCDF..........
Hpomd     VTTRDPQPELFNNN.......YFYEQALYALEQDNFDDF..........
NeuTy     HPVNPSPGDFSAAWSRYPSTVRYPNRLTGASRDERIAPILANELASLRNN
HumTy     EEYNSHQSLCNGTPEGPLRRNPGNHDKSRTPRLPSSADVEFCLSLTQYES
```

```
          2222222222222222222222222222222222222222222222222
                                                |<---helix 2.5-->|
                                                   *        *
Cmag6     ...GIDLLGDVIESSLYSP.N.....PQYYGA.LHN....TAHMMLGRQG 354
CmagX     ...GIDVLGDVIESSLYSP.N.....PEYYGA.LHN....TAHMMLGRQG
Pintc     ...GIEFLGDIIESSGYSA.N.....PGFYGS.LHN....TAHIMLGRQG
Pinta     ...GIELLGDIIESSKYSS.N.....VQYYGS.LHN....TAHVMLGRQG
Penv1     ...GIDILGDIIESSLYSP.N.....VQYYGA.LHN....TAHIVLGRQG
LimII     ...GTDILGALVESSYESV.N.....HEYYGN.LHN....WGHVTMARIH
Euryd     ...GINVIGALIESSHDSV.N.....KPYYGT.LHN....WGHVMIARIH
Eurye     ...GINVLGALIESSFETK.N.....KLYYGS.LHN....WGHVMMARLQ
Anda6     ...GTDILGDIIESSDESK.N.....VEYYGS.LHN....WGHVMMANIT
BombA     ...AINFVGNYWQDNADLY.G.....EEVTKD.YQRSYEVFARRVLGAAP
MsexA     ...AVNFVGNYWQANADLY.N.....EEVTKL.YQRSYEINARHVLGAAP
Tni_M     ...DVEHLARLLLGGMGLV.G.....DDAKFMHMMH....LMKRLLSYNV
BombM     ...DIENLARLVLGGLEIV.G.....DDAKVIHLTN....LMKKMLSYGQ
PapPO     GKRGIDILGDAFEADAQLSPN.....YLFYGD.LHN....TGHVLLAFCH
DrpPO     ..RGIDILGNMIEASPVLSIN.....YNFYGN.LHN....EGHNIISFAH
Octoe     ............................EIQLEIVHN....AIHSWIGGKE
Hpomd     ............................EIQFEVLHN....ALHSWLGGHA
NeuTy     VSLLLLSYKDFDAFSYNRWDPNTNPGDFGSLEDVHN....EIHDRTGGNG
HumTy     GSMDKAANFSFRNTLEGFASPLTGIADA.SQSSMHN....ALHIYM..NG


          2222222222222222222222222222222222222222222222222
                                              |<---helix 2.6---->|
                                                 *
Cmag6     DPHGKFDLPPGVLEHFET.ATR.....DPAFFRLHKYMDNIFRKHKD.SL 397
CmagX     DPHGKFDLPPGVLEHFET.ATR.....DPAFFRLHKYMD..........
Pintc     DPTGKFDLPPGVLEHFET.STR.....DPSFFRLHKYMDNIFREHKD.SL
Pinta     DPHGKFNLPPGVMEHFET.ATR.....DPSFFRLHKYMDNIFKKHTD.SF
Penv1     DPHGKFDLPPGVLEHFET.ATR.....DPSFFRLHKYMDNIFKEHKD.NL
LimII     DPDGRFHEEPGVMSDTST.SLR.....DPIFYNWHRFIDNIFHEYKN.TL
Euryd     DADGRYRTNPGVMDDTST.SLR.....DPIFYRYHRWMDNIFQEYKH.RL
Eurye     DPDHRFNENPGVMSDTST.SLR.....DPIFYRYHRFIDNIFQKYIA.TL
Anda6     DPDHRFQENPGVMSDTST.SLR.....DPIFYRWHRFIDNIFQEHKK.SF
BombA     MPFDKYTFMPSAMDFYQT.SLR.....DPAFYQLYNRIVEYIVEFKQ.YL
MsexA     KPFNKYSFIPSALDFYQT.SLR.....DPVFYQLYDRIINYINEFKQ.YL
Tni_M     YNFDKYTYVPTALDLYST.CLR.....DPVFWRLMKRVTDTFFLFKK.ML
BombM     YNMDKYTYVPTSLDMYTT.CLR.....DPVFWMIMKRVCNIFTVFKN.ML
PapPO     DNDNSHREEIGVMGDSAT.ALR.....DPVFYRWHKFVDDIFQEYKL.TQ
DrpPO     DPDYRHLEDFGVMGDVTT.AMR.....DPIFYRWHGFIDTVFNKFKT.RL
Octoe     ..............EHSLNHLHYAAYDPIFYLHHSNVDRLWVIWQ....
Hpomd     ..............KYSFSSLDYTAFDPVFFLHHANTDRLWAIWQ....
NeuTy     ................HMSSLEVSAFDPLFWLHHVNVDRLWSIWQDLNP
HumTy     ................TMSQVQGSANDPIFLLHHAFVDSIFEQWLRRHR
```

```
         22333333333333333333333333333333333333333333333333

Cmag6    PPYTKEELNFEGVNIDNFYIKGNLETYFETFEYSLVNAVDDTED.VDD.. 444
CmagX    ..................................................
Pintc    TPYTRDELEFNGVSIDSIAIEGTLETFFENFEYSLLNAVDDTVD.IAD..
Pinta    PPYTHDNLEFSGMVVNGVAIDGELITFFDEFQYSLINAVDSGEN.IED..
Penv1    PPYTKADLEFSGVSVTELAVVGELETYFEDFEYSLINAVDDAEG.IPD..
LimII    KPYDHDVLNFPDIQVQDVTLHARVDNVVHFTMREQELELKHGINPGNA..
Euryd    PSYTHQQLDFPGVRISRVTVRSKVPNILHTYSKDSLLELSHGLNIKGH..
Eurye    PHYTPEDLTCPGVHVVNVTVNAKVPNVVTTFMKEAELELSYGIDFGSD..
Anda6    HPYTKEELSFPGVEVVGVSINSKTANVITTLIKESLLELSHGINFGTD..
BombA    KPYTQDKLYFDGVKITDVKVD.KLTTFFENFEFDASNSVYFSKEEIKN..
MsexA    QPYNQNDLHFVGVKISDVKVD.KLATYFEYYDFDVSNSVFVSKKDIKN..
Tni_M    PKYTREDFDFPGVKIEKFTTD.KLTTFIDEYDMDITNAMFLDDVEMKKKR
BombM    PKYTREQFSFPGVKVEKITTD.ELVTFVDEYDMDISNAMYLDATEMQNKT
PapPO    PPYTMEDLSLPGVVLDKVGVVRNDQLNTLTTGWSVREFEASRGLDFNSPN
DrpPO    NPYNAGELNFDGITVDYIEAKIGKSNTKANTLLTYWQKSSADLAAGLDFG
Octoe    ..............................................ELQ
Hpomd    ..............................................ELQ
NeuTy    NSFMTPRPAPYSTFVAQ..............................E
HumTy    PLQEVYPEANAPIGHNR..............................E


         33333333333333333333333333333333333333333333333333

Cmag6    ...VDIFTYISRLNHKEFSFVGDVTNELDHDVLATVRIFAWPHEDNNGVA 491
CmagX    ..................................................
Pintc    ...VEILTYIERLNHKKFSFLILVTNNNNTEVLATVRIFAWPLRDNNGIE
Pinta    ...VEINARVHRLNHKEFTYKITMSNNNDGERLATFRIFLCPIEDNNGIT
Penv1    ...VEISTYVPRLNHKEFTFRIDVENGGA.ERLATVRIFAWPHKDNNGIE
LimII    ...RSIKARYYHLDHEPFSYAVNVQNNSASDKHATVRIFLAPKYDELGNE
Euryd    ...IQVKYNYEHLDHEPYNYEIEVDNRTGEARETCVRIFLAPKYDELGNR
Eurye    ...HSVKVLYRHLDHEPFTYNISVENSSGGAKDVTMRIFLGPKYDELGNR
Anda6    ...QSVKVKYHHLDHEPFTYNIVVENNSGAEKHSTVRIFLAPKYDELNNK
BombA    .NHVHELRCATRLNHSPFNVNIEVD..SNVASDAVVKMLLAPKYDDNGIP
MsexA    .FPYGYKVRQPRLNHKPFSVSIGVK..SDVAVDAVFKIFLGPKYDSNGFP
Tni_M    .SDMTMVARMARLNHHPFKVTVDVT..SDKTVDCVVRIFIGPKYDCLGRL
BombM    .SDMTFMARMRRLNHHPFQVSIDVM..SDKTVDAVVRIFLGPKYDCMGRL
PapPO    PVTAHYPSRPCTLHLPSPDNKQHRKPKS.....VTVRIYMAPKHNERGLE
DrpPO    PTTDRNIFASFTHLQNAPFTYTFNVTNNGARRTGTCRIFICPKVDERNQA
Octoe    KLRGLNAYESHCALELMKVPLKPFSFGAPYNLNDLTTKLSKPEDMFRYKD
Hpomd    RYRGLPYNEADCAINLMRKPLQPFQDKKL.NPRNITNIYSRPADTFDYRN
NeuTy    GESQSKSTPLEPFWDKSAANFWTSEQVKDSITFGYAYPETQKWKYSSVKE
HumTy    SYMVPFI.PLYRNGDFFISSKDLGYDYSYLQDSDPDSFQDYIKSYLEQAS
```

```
          333333333333333333333333333333333333333333333333

Cmag6     FSFNDGRWNAIEMDKFWVMLHPGHNHIERSSHDSSATVPDIPSFQIIKDR 541
CmagX     ..................................................
Pintc     YSFNEGRWRALELDRFWVKVKHGHHQITRQSTESSVTVPDVPSLQTLIDR
Pinta     LTLDEARWFCIELDKFFQKVPKGPETIERSSKDSSVTVPDMPSFQSLKEQ
Penv1     YTFDEGRWNAIELDKFWVSLKGGKTSIERKSTESSVTVPDVPSIHDLFAE
LimII     IKADELRRTAIELDKFKTDLHPGKNTVVRHSLDSSVTLSHQPTFEDLLHG
Euryd     LLLEEQRRLYIELDKFHRRLEPGKNVLVRASGDSSVTLSKVPTFEELESG
Eurye     LQPEQQRTLNIELDKFKATLDPGKNVVTRDHRNSTVTVEQSVPVKKLREE
Anda6     LEPDEQRRLFIELDKFFYTLTPGKNTIVRNHQDSSVTISKVRTFDQLGAG
BombA     LTLEDNWMKFFELDWFTTKLTAGQNKIIRNSNEFVIFKED....SVPMTE
MsexA     IPLAKNWNKFYELDWFVHKVMPGQNHIVRQSSDFLFFKED....SLPMSE
Tni_M     MSVNDKRMDMIEMDTFLYKLETGKNTIVRNSLEMHGVIEQRPWTRRILNN
BombM     MSVNDKRLDMFELDSFMYKLVNGKNTIVRSSMDMQGFIPEYLSTRRVMES
PapPO     MGFMEQRLLWAEMDKFTQDLKPGQNQIVRASNLSSITNPSGYTFRSLEAV
DrpPO     LNLEEQRLLAIEMDKFTVDLVPGENTIRRQSTESSVAIPFERSFRPVGAD
Octoe     NFHYEYDILDINSMSINQIESSYIRHQKDHDRVFAGFLLSGFGSSAYATF
Hpomd     HFHYEYDTLELNHQTVPQLENLLKRRQ.EYGRVFAGFLIHNNGLSADVTV
NeuTy     YQAAIRKSVTALYGSNVFANFVENVADRTPALKKPQATGEESKSTVSAAA
HumTy     RIWSWLLGAAMVGAVLTALLAGLVSLLCRHKRKQLPEEKQPLLMEKEDYH


          333333333333333333333333333333333333333333333333

Cmag6     TNEAIAQ...NKELHIEEFESGLG.LPNRFLIPKGNVKGLDMDVMVAITS 587
CmagX     ..........................................FHLVVFVSD
Pintc     ADAAISS...GCALHLEDYESALG.LPNRFLLPKGQAQGMEFNLVVAVTD
Pinta     ADNAVNG...GHDLDLSAYERSCG.IPDRMLLPKSKPEGMEFNLYVAVTD
Penv1     AEAGGAG.........LAKFESATG.LPNRFLLPKGNDRGLEFDLVVAVTD
LimII     VGLNEH..........KSEYCSCG.WPSHLLVPKGNIKGMEYHLFVMLTD
Euryd     NANVNPN..........EYCSDG.KPEHMLVPRGKERGMDFYLFVMLTD
Eurye     GGVAG............EYCSCG.WPEHMLIPKGNHRGMDFELFVIVTD
Anda6     EGVSEDST..........EYCSCG.WPEHMLIPRGSHKGMEFELFVMLTD
BombA     IMKMLD....EGKVPFDMSEEFCY.MPKRLMLPRGTEGGFPFQLFVFVYP
MsexA     IYKLLD....EGKIPSDMSNSSDT.LPQRLMLPRGTKDGYPFQLFVFVYP
Tni_M     MIGTVGTISKTVDVESWWYK.RHR.LPHRMLLPLGRRGGMPMQMFVIVTP
BombM     EMMPSG..DGQTMVKDWWCKSRNG.FPQRLMLPLGTIGGLEMQMYVIVSP
PapPO     NPANPGPPANAETNFCGCG......WPEHLLLPRGKPEGMTYQLFFMLTD
DrpPO     YQPKAADELARFK.FCGCG......WPQHLLLPKGNAQGMLFDLFVMISD
Octoe     EICIEGGEC......HEGSHFAVLGGSTEMPWAFDRLYKIEITDVLSDMH
Hpomd     YVCVPSGPKGKNDCNHKAGVFSVLGGELEMPFTFDRLYKLQITDTIKQLG
NeuTy     AHAVELSGAKKVAEKVHNVFQHAEEKAQKPVVPVKDTKAESSTAAGMMIG
HumTy     SLYQSHL.....................................
```

```
                                                                    110

         333333333333333333333333333333333333333333333333333

Cmag6    GEADAAV.EGLHENTSFNHYGCP...DGT..YPDKRRHGYPLDRHVDDER 631
CmagX    GAKDAAI.DGLLENTSFNHYG...AHSGK..YPDKQPHPYPLDRRVDDKR
Pintc    GRTDAAL.DDLHENTKFIHYGY....DRQ..YPDKRPHGYPLDRRVDDER
Pinta    GDKDTEG.HNGGHDYGGTHAQCGV.HGEA..YPDNRPLGYPLERRIPDER
Penv1    GDADSAV.PNLHENTEYNHYG...SHGV...YPDKRPHGYPLDRKVPDER
LimII    WDKDKVD.GSESVACVDAVSYCGA.RDHK..YPDKKPMGFPFDRPIHTEH
Euryd    YEEDSVQGAGEQTIDQDAVSYCGA.KDQK..YPDKKAMGYPFDRPIQVRT
Eurye    YAQDAVNGHGENAECVDAVSYCGA.KDQK..YPDKKPMGFPFDRVIEGLT
Anda6    HDEDTVAGLSENAVCSDAVSYCGA.RDDR..YPDKKAMGFPFDRKIEART
BombA    ..FD...NKG........KD.LAP.FESF..VLDNNLLASLWIAPLLMHY
MsexA    YQ......AVP.......KE.MEP.FKSI..VPDSKPFGYPFDRPVHPEY
Tni_M    VK...TNLLLPNLDMNIMKERKTC.AGAS..VSTRCRSGFPFDRKIDMTH
BombM    VR...TGMLLPTLDMTMMKDRCAC.RWSS..CISTMPLGYPFDRPIDMAS
PapPO    LEKDQVD.QPAGPRR..CANAVSFCGILDSKFPDKRPMGFPFDRRPPPRL
DrpPO    YSQDSVE.QPKTPNDACSTAYSFCGLKDKLYPDRRTMGYPFDRRLPNANL
Octoe    LAFDSA..FTIKTKIVAQNGTELPASILPEATVIRIPPSKQDA.......
Hpomd    LKVNNAASYQLKVEIKAVPGTLLDPHILPDPSIIFEPGTKER........
NeuTy    LSIKRPSKLTASPGPIPESLKYLAPDGKYTDWIVNVRAQKHGLGQSFRVI
HumTy    ..................................................


         333333333333333333333333333333333333333333333333333

Cmag6    IINDLH.NFKHIQVKVFHHA............................... 650
CmagX    IITGVT.NFKGMDVKVYHVEEQ.............................
Pintc    IFEALP.NFKQRTVKLYSHEGVDGG..........................
Pinta    VIDGVS.NIKHVVVKIVHHLEHHD...........................
Penv1    VFEDLP.NFKHIQVKVFNHGEHIH...........................
LimII    ISDFLTNNMFIKDIKIKFHE................................
Euryd    PSQFKTPNMAFQEIIIQYEGHKH............................
Eurye    LEEFLTPSMSCTDVRIKYTDIK..............................
Anda6    AAEFLTPNMGLTDIKIKFHG................................
BombA    SR.FLTCISRIFSFTT.RVNGSLTNSIFLRMTHMIMLFQKIKF.......
MsexA    FKQPNMHFEDVHVYHEGEQFPYKFNVPFYVPQKVEV...............
Tni_M    FFTRNMKFTDVMIFRKDLSLSNTIKDVDMSDMMMKKDDLTYLDSDMLVRW
BombM    FFTSNMKFADVMIYRKDLGMSNTSKTTSEMVMM..KDDLTYLDSDMLVKR
PapPO    QDAEVTSVADYARLSNMTVQDITITFLTTASRSRHDGPI............
DrpPO    TELVGAFGNMAKTDLRIVFNDRVIDKA.........................
Octoe    ..................................................
Hpomd    ..................................................
NeuTy    VFLGEFNPDPETWDDEFNCVGRVSVLGRSAETQCGKCRKDNANGLIVSGT
HumTy    ..................................................
```

```
          333333333333333333333333333333333333333333333333
Cmag6     ..................................................
CmagX     ..................................................
Pintc     ..................................................
Pinta     ..................................................
Penv1     ..................................................
LimII     ..................................................
Euryd     ..................................................
Eurye     ..................................................
Anda6     ..................................................
BombA     ..................................................
MsexA     ..................................................
Tni_M     SYKAVMMMSKDDMMRM..................................
BombM     TYKDVMMMSSMMN.....................................
PapPO     ..................................................
DrpPO     ..................................................
Octoe     ..................................................
Hpomd     ..................................................
NeuTy     VPLTSLCCRILWAASSRASSLRMSSRICAPT...................
HumTy     ..................................................
```

Fig.5: Hydrophilicity plots of proteins aligned in Fig.4. Abbreviations as above. Residue numbers apply to Cmag6.

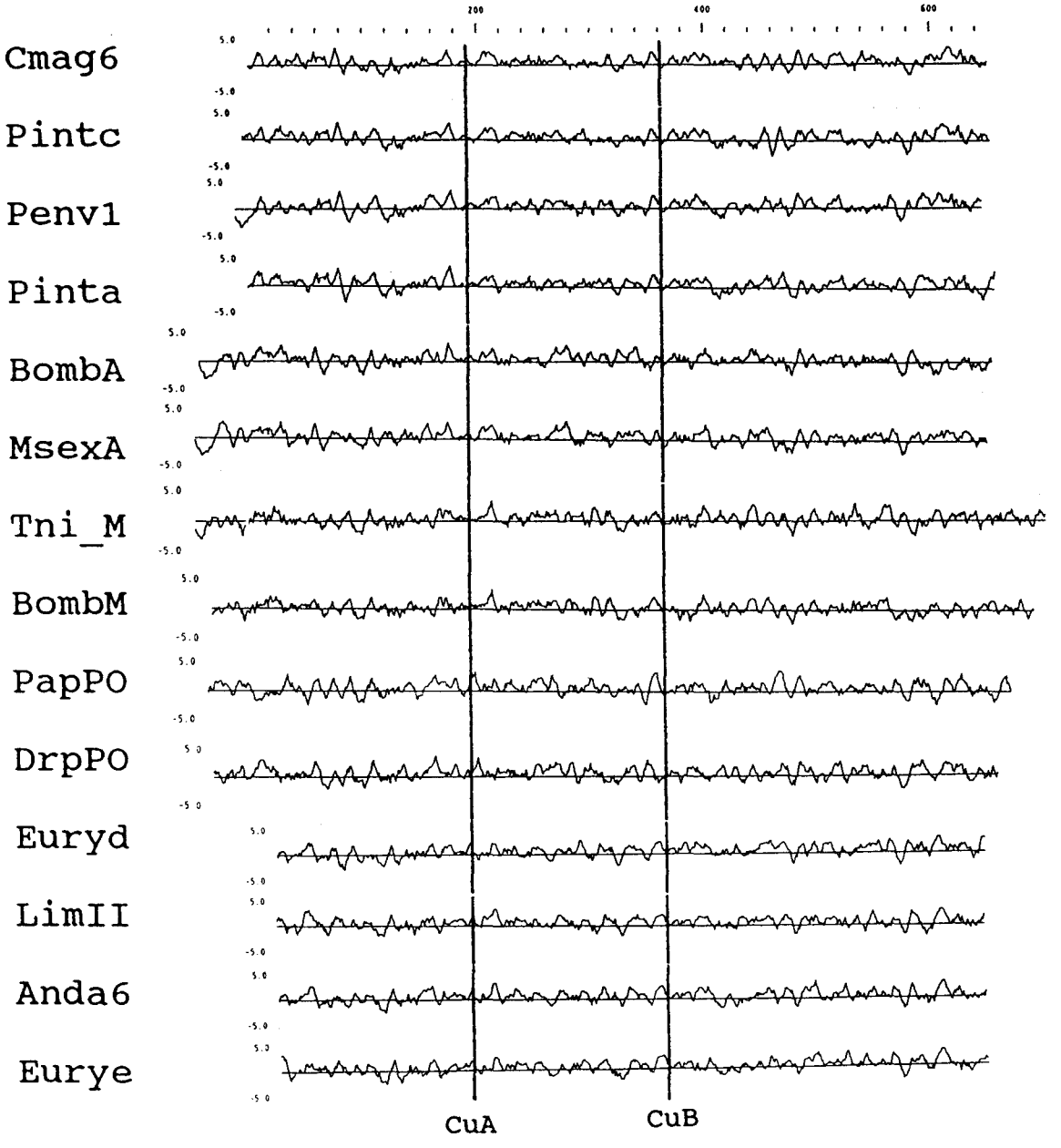

200
400
600
Cmag6
Pintc
Penv1
Pinta
BombA
MsexA
Tni_M
BombM
PapPO
DrpPO
Euryd
LimII
Anda6
Eurye
5.0
-5.0
CuA
CuB

Fig.6: Surface probability plots of proteins aligned in Fig.4. Abbreviations as above. Residue numbers apply to Cmag6.

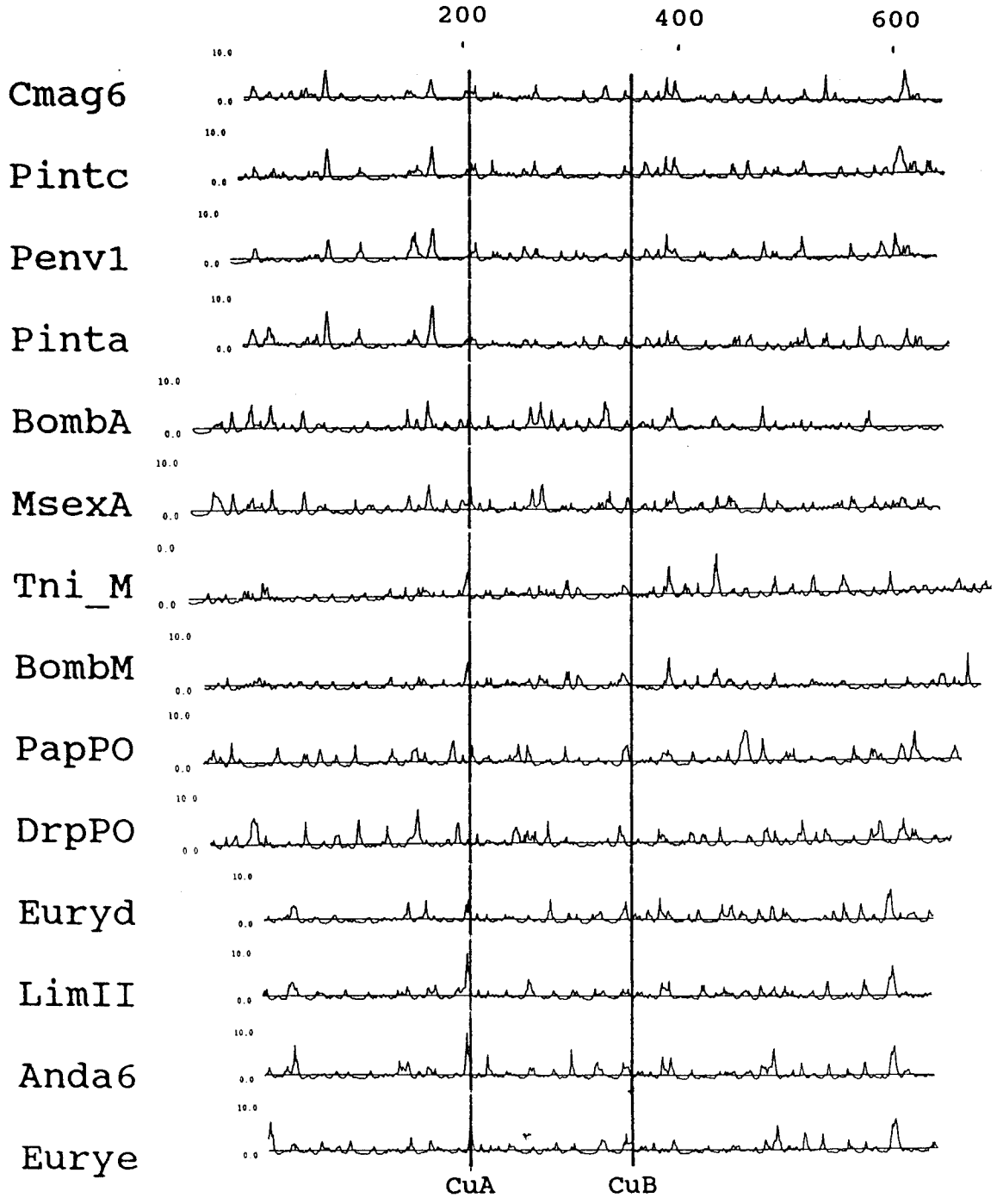
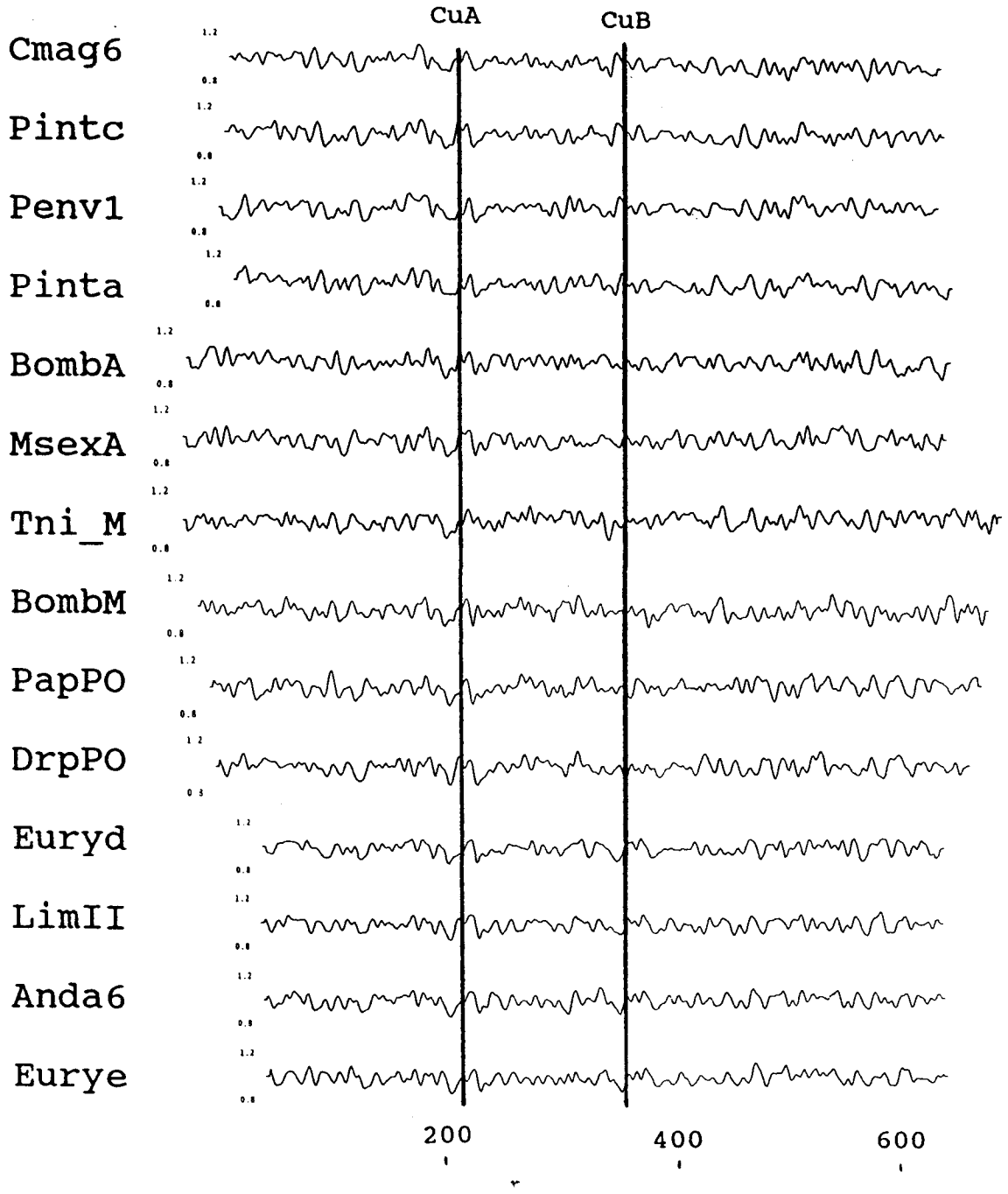

200
400
600
Cmag6
Pintc
Penv1
Pinta
BombA
MsexA
Tni_M
BombM
PapPO
DrpPO
Euryd
LimII
Anda6
Eurye
CuA
CuB

Fig.7: Regional backbone flexibility plots of proteins aligned in Fig.4. Abbreviations as above. Residue numbers apply to Cmag6.


CuA
CuB
Cmag6
Pintc
Penv1
Pinta
BombA
MsexA
Tni_M
BombM
PapPO
DrpPO
Euryd
LimII
Anda6
Eurye
200
400
600

= 7) in Fig.5 is calculated according to Kyte and Doolittle (1982). As could be expected for a globular protein occurring freely dissolved in the hemolymph, it shows no extensive hydrophilic or hydrophobic domains. Surface probability based on amino acid side-chain solvent accessibilities (Emini et al., 1985) and regional backbone flexibility (Jameson and Wolf, 1988) are plotted in figures 6 and 7. All 3 indices are strikingly similar among all crustacean Hcs as well as among other subgroups (the chelicerate Hcs Euryd, Eurye, Anda6 and LimII, the methionine-rich storage proteins BombM and Tni_M, the arylphorins BombA and MsexA and the prophenoloxidases PapPO and DrpPO). Some motifs appear to be conserved in all arthropod Hc-type proteins. None of the three indicators suggests any structural homology between the arthropod proteins mentioned above on the one hand and the molluscan Hcs and tyrosinases on the other. The high degree of sequence similarity among arthropod Hcs (30%-70% sequence identity, see Figs.4 and 9) suggests a common tertiary structure. X-ray crystallography of Hc from *Panulirus* and *Limulus* (Volbeda and Hol, 1989b; Hazes et al., 1993) has shown that arthropod Hcs consist of 3 domains (Fig.8). Domain 1 (residues 1-174 in *C. magister* subunit 6) is quite variable and mainly α-helical in structure. Domain 2 (residues 175-399 in *C. magister*) contains the oxygen binding CuA and CuB sites and is the most conserved part of

Fig.8: Structure of *Limulus* Hc according to X-ray crystallography at 2.18 Å resolution (Hazes, 1993, with permission). A, whole subunit. B, domain 1. C, domain 2. D, domain 3. Solid black spheres in center: Cu^{+} ions. Black sphere to the left: Cl^{-} ion. Shaded sphere: Ca^{2+} ion.

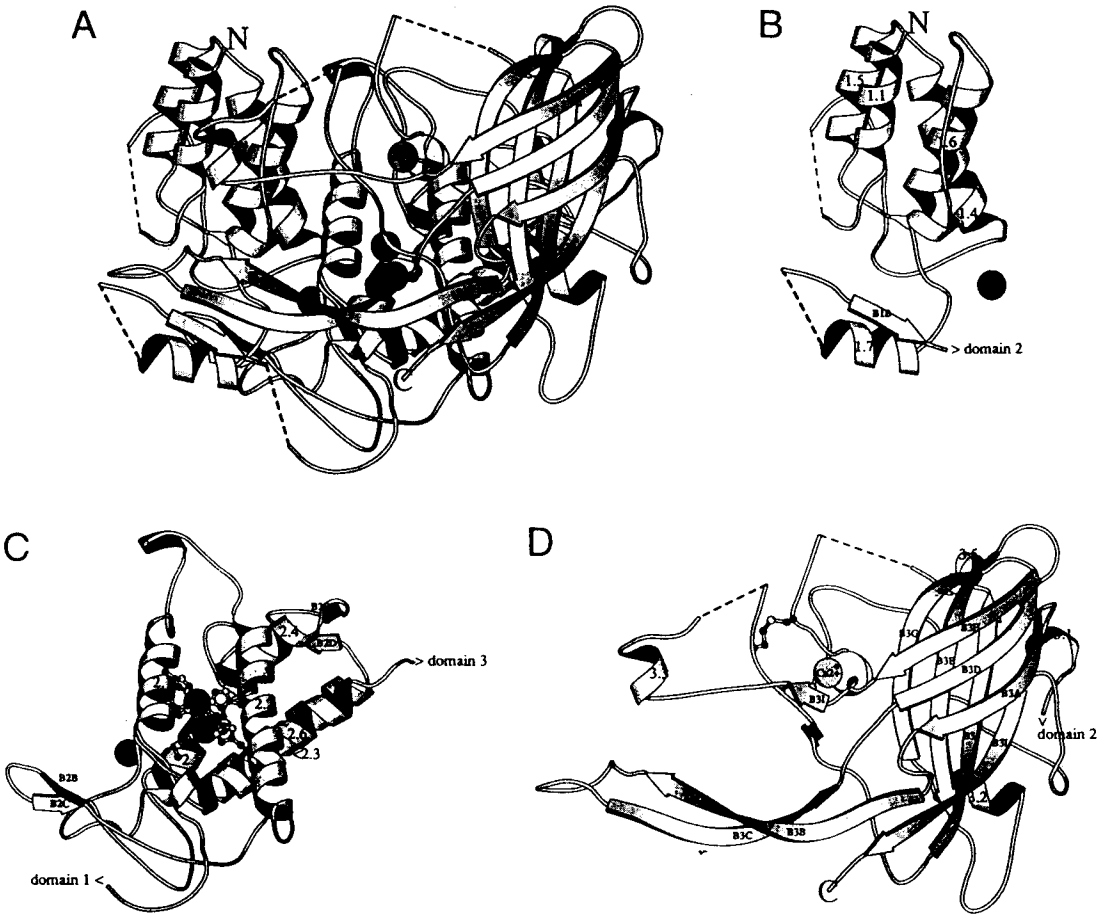

A
N
B
N
> domain 2
C
> domain 3
domain 1 <
D
domain 2

Fig.9: Pairwise distance matrix of taxa aligned in Fig.4. Below diagonal: Absolute distances. Above diagonal: Mean distance index (adjusted for missing data). Gaps treated as missing data.

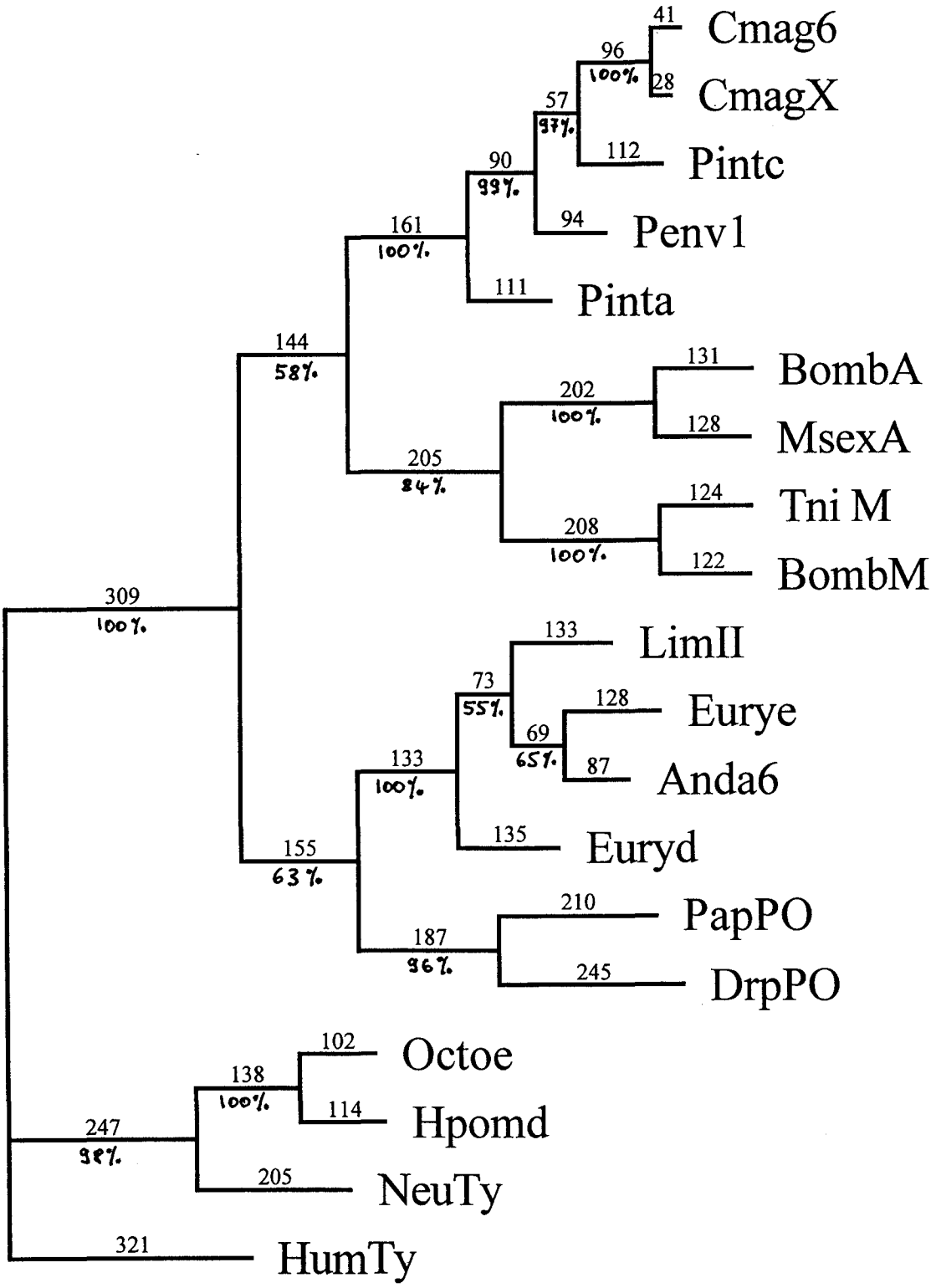
	Cmag6	CmagX	Pintc	Pinta	Penv1	Limll	Euryd	Eurye	Anda6	BombA	MsexA	Tni M	BombM	PapPO	DrpPO	Octoe	Hpomd	NeuTy	HumTy
Cmag6		0.151	0.36	0.434	0.343	0.686	0.699	0.678	0.676	0.743	0.736	0.765	0.753	0.757	0.759	0.931	0.925	0.934	0.913
CmagX	69		0.317	0.367	0.316	0.669	0.676	0.667	0.659	0.726	0.732	0.787	0.753	0.758	0.76	0.885	0.877	0.93	0.906
Pintc	233	146		0.401	0.328	0.672	0.688	0.688	0.666	0.731	0.728	0.768	0.74	0.75	0.77	0.917	0.905	0.929	0.911
Pinta	281	168	262		0.379	0.67	0.682	0.672	0.645	0.733	0.741	0.748	0.722	0.741	0.768	0.918	0.911	0.927	0.915
Penv1	219	149	212	245		0.664	0.682	0.676	0.649	0.719	0.706	0.743	0.731	0.733	0.76	0.909	0.901	0.911	0.915
Limll	423	289	413	415	408		0.465	0.428	0.4	0.773	0.748	0.77	0.751	0.681	0.708	0.913	0.911	0.92	0.91
Euryd	429	292	423	422	419	288		0.437	0.409	0.753	0.725	0.769	0.765	0.687	0.716	0.918	0.916	0.932	0.923
Eurye	415	288	421	414	413	264	272		0.346	0.753	0.742	0.753	0.744	0.7	0.717	0.917	0.91	0.934	0.913
Anda6	416	284	408	398	397	248	255	215		0.752	0.741	0.756	0.748	0.678	0.698	0.924	0.925	0.926	0.923
BombA	462	336	461	461	455	464	451	449	449		0.374	0.707	0.705	0.827	0.824	0.918	0.93	0.957	0.943
MsexA	460	341	462	468	449	451	436	444	444	259		0.688	0.675	0.804	0.828	0.934	0.924	0.952	0.943
Tni M	489	374	499	483	483	475	474	461	465	490	477		0.332	0.804	0.835	0.925	0.932	0.953	0.932
BombM	482	359	481	467	476	465	472	457	460	490	469	246		0.806	0.83	0.933	0.937	0.949	0.921
PapPO	467	341	467	464	462	408	410	415	404	531	516	534	533		0.68	0.906	0.9	0.933	0.928
DrpPO	468	346	479	480	478	425	428	426	417	525	530	550	548	455		0.913	0.928	0.925	0.93
Octoe	338	201	332	335	328	337	335	333	339	335	342	358	360	337	345		0.546	0.844	0.894
Hpomd	332	193	324	329	318	328	327	323	332	332	330	355	356	332	346	218		0.852	0.883
NeuTy	455	305	455	456	440	449	455	453	451	488	481	530	523	488	470	347	345		0.88
HumTy	419	299	420	422	421	405	408	401	409	445	445	449	443	449	437	295	280	365	

the protein. CuA and CuB each consist of an antiparallel helix pair containing 3 Cu-binding histidine residues. In *C. magister* subunit 6, the CuA helix pair extends from residue 186-200 (helix 2.1) and 215-239 (helix 2.2). The Cu binding histidines are located at positions 192 and 196 (helix 2.1) as well as 224 (helix 2.2). The CuB helix pair extends from residue 341-353 (helix 2.5) and 378-396 (helix 2.6). Its Cu binding histidines are located at positions 343 and 347 (helix 2.5) as well as 383 (helix 2.6). Domain 3 is rich in ß-sheets and forms a ß-barrel structure (Hazes and Hol, 1992).

Phylogenetic analysis using parsimony was performed with the PAUP program (Swofford, 1991). A pairwise distance matrix of the taxa aligned in Fig.4 is presented in Fig.9. Sequence comparison (Figs.4 and 9) showed a high degree of homology among several taxa. Amino acid sequence identity between Cmag6 and CmagX was 85%. Among any two crustacean Hcs it was approximately 60%, chemical similarity over 80%. Sequence identity among chelicerate Hcs ranged from 53% to 65%, among molluscan Hcs it was 42%.

The single most parsimonious phylogenetic tree consistent with the dataset (total size = 5045 substitutions) is shown on Fig.10. It was obtained through a heuristic search algorithm treating gaps as missing data. Various search options (simple and random addition, branch and tree swapping) gave the same result. The resulting tree

Fig.10: Single most parsimonious unrooted tree of the proteins aligned in Fig.4. Gaps are treated as missing data. Total tree size: 5045 substitutions. Branch lengths proportional to number of substitutions (indicated above branches). Bootstrap values (500 replicates) indicated below branches.


41
Cmag6
96
100%
28
CmagX
57
97%
112
Pintc
90
99%
94
Penv1
161
100%
111
Pinta
144
58%
131
BombA
202
100%
128
MsexA
205
84%
124
Tni M
208
100%
122
BombM
309
100%
133
LimII
73
55%
128
Eurye
69
65%
87
Anda6
133
100%
135
Euryd
155
63%
210
PapPO
187
96%
245
DrpPO
102
Octoe
138
100%
114
Hpomd
247
98%
205
NeuTy
321
HumTy

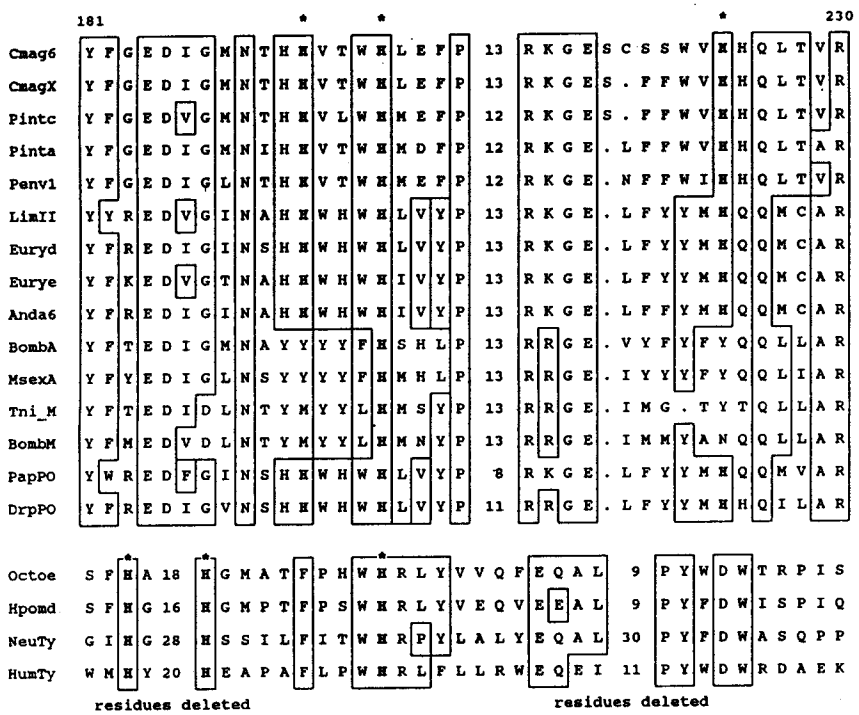
represents a molecular phylogeny of Hc-class proteins, not a phylogeny of the involved species.

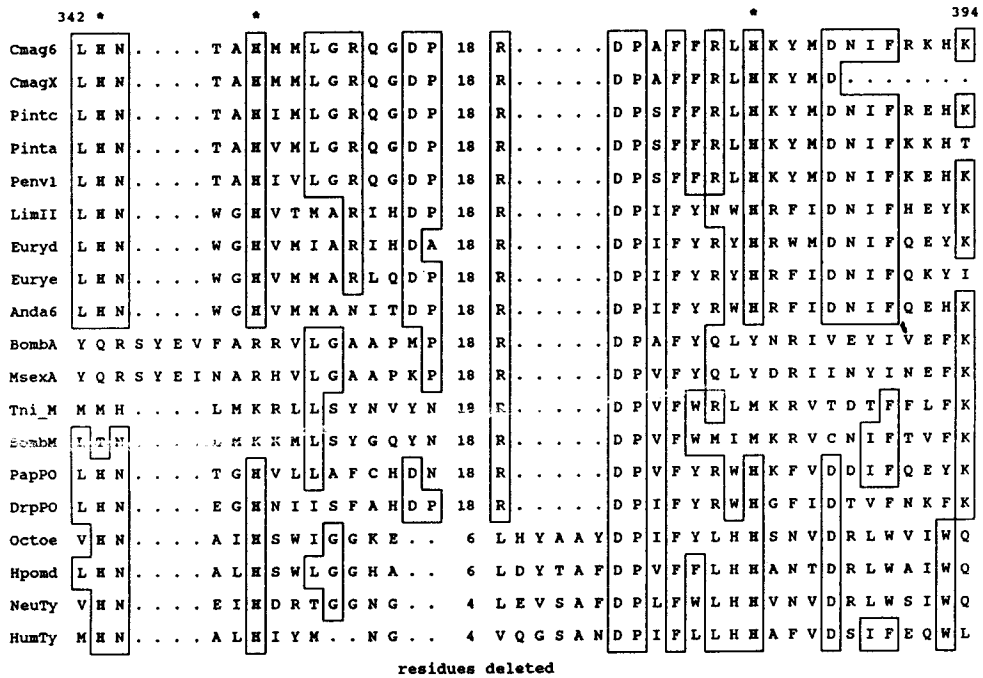
Sequence alignment of the functionally important CuA and CuB sites (Fig.11) illustrates several points: (1) The histidine ligands are conserved in those proteins that bind Cu, i.e., the arthropodan and molluscan Hcs, the tyrosinases and the prophenoloxidases. In the non-Cu-binding insect hexamerins these residues are not conserved, although the overall sequence similarity of the hexamerins to crustacean Hcs is high. (2) The CuB site is the only region that exhibits significant homologies in all taxa surveyed, including the molluscan Hcs and tyrosinases. This suggests a common origin for at least part of the molecule. (3) The CuA site is either of the arthropodan or the molluscan type. Sequence homology between these types is marginal at best. All arthropod proteins in this study form a monophyletic group relative to molluscan Hcs and tyrosinases.

DISCUSSION

The phylum Arthropoda is composed of 3 major taxa: The Chelicerata, the Insecta and the Crustacea. Traditionally, the latter two have been considered to be more closely related and were grouped together as Mandibulata (Remane et al., 1980). This relationship is supported by 18S rRNA

Fig.11: Sequence conservation in the Cu binding sites of Hc-type proteins. Top, CuA site; Bottom, CuB site. Numbering of residues according to Cmag6. Residues conserved in more than half of the taxa or in one complete group of taxa are boxed. *, conserved histidine, presumably acting as Cu ligand.


181
230
Cmag6
CmagX
Pintc
Pinta
Penv1
LimII
Euryd
Eurye
Anda6
BombA
MsexA
Tni_M
BombM
PapPO
DrpPO
Octoe
Hpomd
NeuTy
HumTy
residues deleted
residues deleted


342
394
Cmag6
CmagX
Pintc
Pinta
Penv1
LimII
Euryd
Eurye
Anda6
BombA
MsexA
Tni_M
BombM
PapPO
DrpPO
Octoe
Hpomd
NeuTy
HumTy
residues deleted

(Turbeville et al., 1991; Garey et al., 1996) and mitochondrial 12S rRNA sequence comparisons (Ballard et al., 1992). Two minor taxa, the Myriapoda and Onychophora, are placed at the base of the arthropod lineage by both studies, a placement that also is supported by their greater morphological similarity to the annelids. However, this phylogeny is by no means certain, and the problem is compounded by the question of whether the arthropods form a monophyletic group at all or arose from annelid-like ancestors in several independent lineages [for a discussion of this problem see Ruppert and Barnes (1994), pp. 611 and 801-802, and Brusca and Brusca (1990), p.681-691].

A common misconception about phylogenetic reconstruction is the idea that taxa should be grouped according to overall similarity (Stewart, 1993). It is often difficult, however, to decide whether a character state shared between two taxa, a "similarity", represents a synapomorphy (a shared trait, ancestral or derived, directly attributable to common ancestry) or a homoplasy (a shared non-derived trait, i.e., a similarity not attributable to common ancestry). The latter case means that evolutionarily very distant taxa may look quite similar due to convergent evolution, which tends to obscure the true phylogenetic relationship. By the same token, closely related taxa may develop greatly dissimilar phenotypes when exposed to different selective pressures.

Parsimony analysis evaluates character states, identifies informative sites (i.e., sites that favor one potential phylogenetic tree over another) and evaluates them based on the notion that a simple explanation is superior to a more complex one. The resulting most parsimonious tree is the one requiring the least number of character state changes consistent with the dataset. Compared to other potential trees, it implies the least amount of homoplasy.

The most parsimonious phylogenetic tree (Fig.10) of the 19 taxa aligned in Fig.4 indicates four monophyletic groups within the arthropods: the crustacean Hcs, the insect hexamerins, the chelicerate Hcs and the prophenoloxidases. These arthropod proteins are clearly monophyletic with respect to the mollusc Hcs and tyrosinases. These conclusions are supported by very robust nodes in the phylogenetic tree as indicated by bootstrap values well over 80%. They are also in agreement with the comparison of predicted structural parameters (Figs. 5, 6 and 7) that suggests a significant degree of structural conservation among the arthropod proteins, but not between the arthropodan and the molluscan groups. However, parsimony analysis fails to resolve the relative arrangement of crustacean and chelicerate Hcs, hexamerins and prophenoloxidases within the arthropod lineage, as indicated by low bootstrap values (58% and 63%) for the 2 major arthropod branches in Fig.10. These data suggest that (1)

the common ancestor of all arthropod Hcs, hexamerins and prophenoloxidases was a Cu binding arthropod Hc-type protein and (2) that the insect hexamerins lost their copper binding capabilities after the insects diverged from the crustaceans, presumably due to the development of the tracheal system that made respiratory proteins unnecessary.

In this context the recent discovery of a molt cycle-regulated Hc-type protein in the Dungeness crab is of particular interest (Otoshi, 1994). This protein, named cryptocyanin, shares many features with Hc but does not contain Cu. Its concentration in the hemolymph changes in synchrony with the crab's molt cycle, suggesting a role as storage protein, analogous in function to the hexamerins of holometabolous insects.

Aspan et al. (1995) recently discovered certain sequence similarities between arthropod prophenoloxidases and arthropod Hcs. The prophenoloxidases represent closely related copper binding non-hemocyanin proteins that occur in both insects (DrpPO) and crustaceans (PapPO). Although identical in function to tyrosinases (NeuTy and HumTy), their sequences show only a slight resemblance to them. Instead, prophenoloxidases appear most closely related to the hexamer-type family of arthropod proteins, sporting a typical arthropodan CuA site. Tyrosinases from most non-arthropod phyla of the animal kingdom as well as from plants, fungi and procaryotes contain a CuA site of the

molluscc-type (van Holde and Miller, 1995). It is therefore reasonable to assume that arthropod prophenoloxidases arose through gene duplication from an ancestral arthropod-type binuclear Cu protein after the arthropods diverged from the molluscs. This ancestral protein then evolved into four protein types: The crustacean Hcs, chelicerate Hcs, arthropod prophenoloxidases and - through loss of Cu - insect hexamerins. Prophenoloxidases are structurally more similar to arthropod Hcs than are the hexamerins; they even have two functional Cu sites. Their detection in the tracheal cuticle of insects (who are thought not to have respiratory proteins) has suggested the fascinating possibility of a respiratory function for prophenoloxidases in that taxon (Kawabata et al., 1995). The length of the branches leading to both the insect and crustacean prophenoloxidases illustrates the long independent evolutionary history of these proteins and suggests they diverged from the other lineages early in arthropod evolution.

Comparison of the distance matrix (Fig.9) with the cladogram of crustacean Hcs based on maximum parsimony (Fig.10) illustrates that distance matrix and parsimony analysis do not always agree: Although *Pennaeus vannamei* Hc subunit 1 (Penv1) displays a higher overall sequence similarity with *C. magister* subunit 6 (Cmag6) than does subunit c from *Panulirus interruptus* (Pintc), 66% vs. 64%,

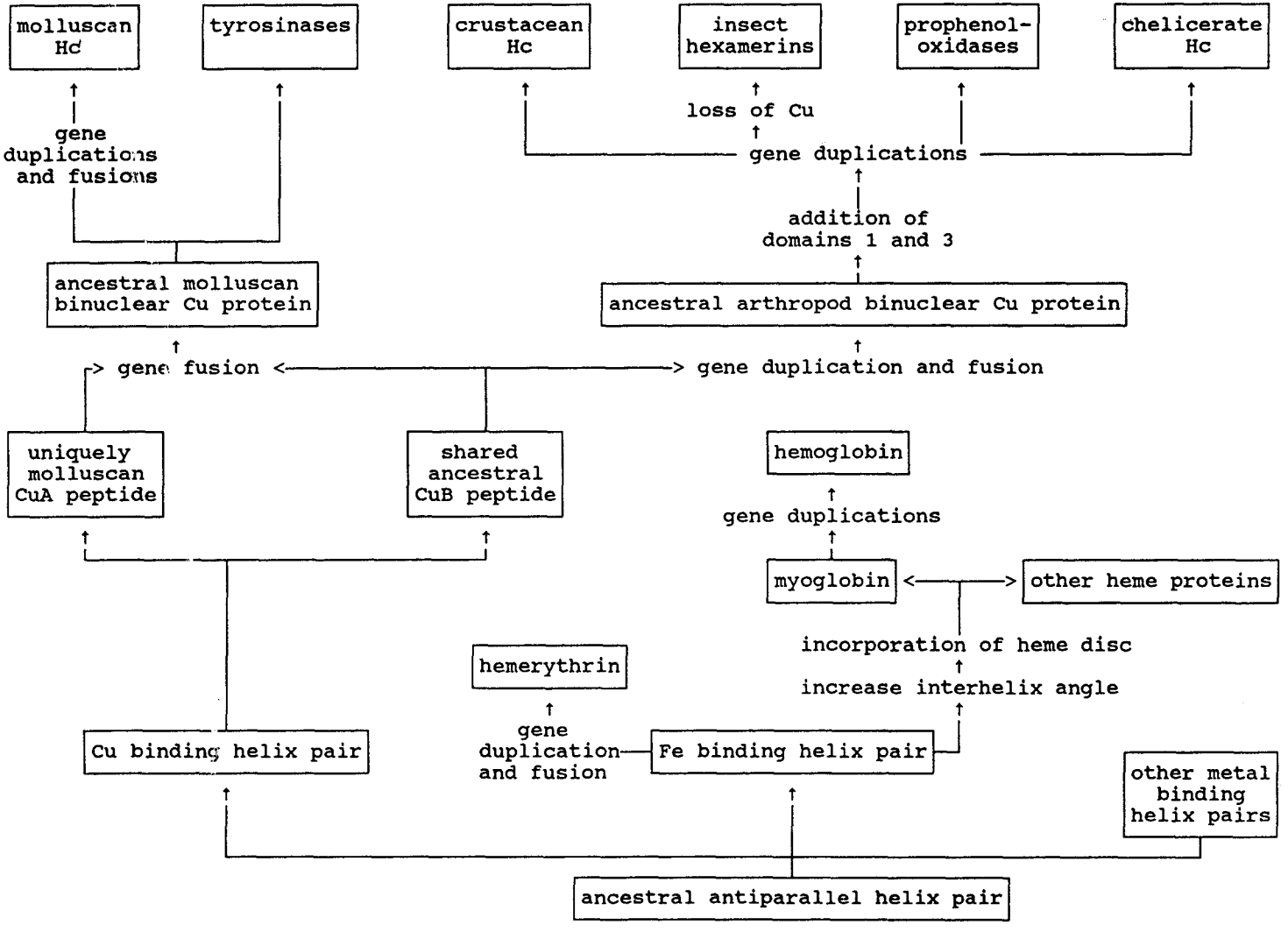
the most parsimonious tree groups Cmag6 with Pintc and not with Penv1.

The multitude of different subunit types found in crustacean and chelicerate Hcs is probably the result of gene duplications that occurred independently after these taxa diverged (Neuteboom et al., 1990). This split occurred about 600 mya, during the early Cambrian, after the arthropods diverged from the molluscs.

The absence of a true outgroup (Hc occurs only in arthropods and molluscs) makes speculation about the relationship of arthropodan and molluscan Hcs difficult. The CuB sites are homologous and not the result of convergence (van Holde and Miller, 1995). This means all arthropodan Hcs, hexamerins and prophenoloxidases share a common ancestor with the mollusc Hcs and tyrosinases. The tyrosinases in particular appear to be phylogenetically very old, because they are found in animals, plants, fungi and even procaryotes, and the degree of sequence similarity between human and procaryotic *(Streptomyces)* tyrosinase, for example, is remarkable. Since there apparently are no procaryotic Hcs, it seems reasonable to assume that molluscan Hcs arose from tyrosinase-like ancestors.

A speculative model of Hc evolution is given in Fig.12. Our analysis supports the view that both arthropodan and molluscan Hcs arose from a common ancestral Cu protein. Whether this ancestor was mono- or binuclear cannot be

Fig.12: Possible evolutionary relationships between major classes of respiratory proteins (based on Volbeda and Hol, 1989a; van Holde and Miller, 1995, and Durstewitz).


molluscan Hc
tyrosinases
crustacean Hc
insect hexamerins
prophenol-oxidases
chelicerate Hc
gene duplications and fusions
loss of Cu
gene duplications
addition of domains 1 and 3
ancestral molluscan binuclear Cu protein
ancestral arthropod binuclear Cu protein
gene fusion
gene duplication and fusion
uniquely molluscan CuA peptide
shared ancestral CuB peptide
hemoglobin
gene duplications
myoglobin
other heme proteins
incorporation of heme disc
increase interhelix angle
hemerythrin
gene duplication and fusion
Cu binding helix pair
Fe binding helix pair
other metal binding helix pairs
ancestral antiparallel helix pair

decided from our data. van Holde and Miller (1995) assume a common origin for the arthropodan and molluscan CuB site and consider the arthropodan CuA site a result of gene duplication and fusion in that lineage. This notion is supported by the fact that in arthropods, the CuA site is very similar in sequence and structure (HxxxH for the first two Cu ligands) to the CuB site (Volbeda and Hol, 1989a), while in molluscs it is not. The ancestral arthropodan Hc would, then, be a binuclear Cu-binding protein, corresponding roughly to domain 2 of today's arthropodan Hc. Domains 1 and 3 would have been added later, following an evolutionary trend to provide sites for allosteric regulation and multisubunit cooperativity. In this scenario, the CuA site of molluscan Hcs and tyrosinases is of separate origin from the CuA site of arthropods, and the molluscan Hcs are the fusion product of this uniquely molluscan CuA peptide and a CuB site shared with the arthropods. The weak tyrosinase activity of molluscan, but not arthropodan, Hcs (Salvato et al., 1983; Nakahara et al., 1983; Markl and Decker, 1992) is further evidence for a common origin of tyrosinases and molluscan Hcs. The huge multidomain Hcs of modern day molluscs would have arisen from this monomeric binuclear Cu-protein through a series of gene duplication and fusion events.

There is some evidence for homology between all respiratory proteins at an even more basic level. Clearly,

the CuB peptide is an extremely ancient molecule. Its presence in procaryotic tyrosinases would put its age in the neighborhood of at least 2 billion years, the early stages of life on this planet. However, the antiparallel helix pair of the CuB site is not unique: Volbeda and Hol (1989a) cite structural similarities between the helix pairs responsible for metal binding in the three classes of respiratory proteins, the hemocyanins, hemerythrins and hemoglobins. They consider it possible that an ancestral antiparallel helix pair was able to bind various metal ions at its active site. Over time this helix pair evolved into a Cu-binding and a Fe-binding variety, the prior giving rise to the Hc family, the latter to the hemerythrins and hemoglobins. In both Hcs and hemerythrins, 4 α-helices surround the active site. Each of their two metal ions is complexed directly by histidine ligands provided by two of the helices, and the interhelix angle is almost the same. Myoglobins and their multisubunit cousins, the hemoglobins, have widened the inter-helix angle to accommodate a heme disc with the Fe^{2+} ion at its core. Each helix still provides one histidine ligand to complex the iron ion.

Although this hypothesis is highly speculative, the presence of antiparallel helix pairs in all classes of respiratory proteins indicates severe structural constraints on metal binding sites in these proteins. The *de novo* development of a metal binding site in a globular protein is

an extremely complex and hence unlikely and rare event. The beauty of the concept of a common ancestral metal binding helix pair is that the metal binding site would only have to be invented once.

CHAPTER V

CONCLUDING SUMMARY

This study investigates developmentally regulated changes in the expression of the copper based respiratory protein hemocyanin. Hemocyanins are large multisubunit oxygen transport proteins that occur extracellularly in the hemolymph of many arthropods and molluscs. In the Dungeness crab *Cancer magister*, Hc subunit composition and functional characteristics change during development, much like hemoglobin does during the fetal-adult shift in mammalian development.

In chapter II we showed how conserved functional domains in the respiratory protein Hc can be used to develop subunit-specific primers and cDNA probes as tools for the study of gene expression in the Dungeness crab. All six protein subunits of *Cancer magister* Hc were purified and their amino-terminal sequences determined. Subunit-specific oligonucleotide primers were designed based on these amino-terminal sequences and on a conserved region near the active site. Using these primers, stretches of cDNA coding for the developmentally regulated Hc subunit 6 were amplified with the polymerase chain reaction.

In chapter III those cDNA fragments were used as subunit-specific probes to investigate developmental changes in hemocyanin expression during the life cycle of the Dungeness crab at the molecular level. Animals were raised under controlled conditions, and total RNA was isolated from 13 developmental stages and 6 tissue types, run on denaturing formaldehyde agarose gels, blotted onto nylon membranes and probed with radioactive ^{32}P-labeled adult Hc-specific cDNA probes. We showed that synthesis of adult-type Hc occurs in the hepatopancreas only and is initiated during the 6th juvenile instar stage as indicated by the appearance of subunit 6 mRNA. This is the first described case of an ontogenetic change in a copper-based respiratory protein, and a model was presented to explain the observed subunit stoichiometries in juvenile and adult hemocyanin.

In chapter IV we presented the complete cDNA- and protein sequences of the developmentally regulated Hc subunit 6 investigated in chapter III, as well as the sequence of another putative Hc subunit obtained from a cDNA library screen. Functional domains within each protein were identified, and both *Cancer magister* Hc sequences were aligned with proteins displaying apparent sequence similarities (other crustacean Hcs, chelicerate Hcs, molluscan Hcs, tyrosinases, insect hexamerins and prophenoloxidases). The comparison of computer-assisted predictions of hydrophilicity, surface probability and

regional backbone flexibility among the taxa showed a remarkable degree of structural conservation among the proteins of the arthropod branch. Parsimony analysis of the aligned sequences allows a phylogenetic reconstruction of their evolutionary history. Confidence limits were established with the bootstrap approach. The most parsimonious phylogenetic tree consistent with the dataset identified four monophyletic groups on the arthropod branch of the genetree: Crustacean Hc, insect hexamerins, chelicerate Hc and arthropod prophenoloxidases. Consistent with the comparison of sequence-based structure predictions, they form a monophyletic group relative to molluscan Hc and non-arthropod tyrosinases. Results for individual clades are evaluated and discussed in the light of the evolutionary history of the Hc gene family.

APPENDIX

APPENDIX A

REVERSE TRANSCRIPTION (RT) AND PCR AMPLIFICATION OF HEMOCYANIN mRNA

1. Reverse transcription of total RNA: 1st strand synthesis.
   a. Dilute 1µl total RNA (1 µg/µl) with 10.65 µl autoclaved H_2O and add 0.75 µl of oligo-dT primer (0.27 µg/µl).
   b. Incubate at 65°C for 3′.
   c. Cool slowly to room temperature and spin briefly in a microcentrifuge.
   d. Add 4 µl 5x RT-buffer [250mM Tris-HCl (pH 8.5), 200mM KCl, 30mM MgCl], 1 µl 20mM DTT, 1 µl 25mM dNTPs, 1 µl RNAsin (10 units/µl) and 0.6 µl AMV reverse transcriptase (17 units/µl).
   e. Vortex, spin briefly and incubate at 42°C for 1.5 h.
   f. Dilute to a total volume of 500 µl with autoclaved water and store at -20°C.

2. PCR-amplification of 1st strand cDNA.
   a. Add 10 µl of cDNA from the RT reaction described above to 18.5 µl H_2O, 5 µl 10x PCR-buffer (670mM Tris-HCl), 4 µl 2.5mM dNTPs, 5 µl 10x BSA (1 µg/µl), 1 µl of each primer (0.2 µg/µl), 0.5 µl Taq polymerase (5 units/µl) and 5 µl 40mM $MgCl_2$.

b. Mix well, spin and overlay with a drop of PCR-oil.

c. Carry out PCR reaction using the following protocol:

denature: 94°C for 40"

anneal: 55°C for 40"

polymerize: 72°C for 1′

Repeat 35 cycles, then 5′ at 72°C and hold at 4°C.

d. Analyze 10 μl aliquots of each reaction on 1.2% agarose TAE-minigels.

APPENDIX B

cDNA CLONING OF PCR PRODUCTS

a. Separate PCR products (40 µl) according to size by electrophoresis on 1.2% agarose TAE-maxigels.

b. On a UV-lightbox, excise bands of interest (in our case, 1.2 - 2kb) from the gel and purify in a glassmilk-procedure (Geneclean II kit, Bio 101). Note: yield is low for PCR products < 500 bp.

c. Repair ends with Klenow polymerase.

d. Phosphorylate with T4-polynucleotide kinase.

e. Cut Bluescript SK vector with restriction endonuclease SmaI and dephosphorylate with CIP.

f. Blunt-end-ligate phosphorylated PCR products into Bluescript vector in a molar ratio insert/vector of 3:1. Use 1 Weiss unit T4 DNA ligase in a total volume of 20 µl. Incubate at 16°C over night.

g. Transform competent E. coli XL-1 Blue cells with 50 ng DNA from the ligation mix.

h. Excise inserts from positive clones with restriction enzymes EcoRV/XbaI and analyze on 1.2% agarose minigels.

APPENDIX C

NORTHERN BLOTS USING TOTAL RNA FROM DIFFERENT TISSUES AND DEVELOPMENTAL STAGES OF *CANCER MAGISTER*

a. To 0.3 μg RNA from each tissue or developmental stage (in a volume of up to 15 μl H_2O) add RNA sample buffer (75% formamide, 7.8% formaldehyde, 15mM MOPS, 6mM NaOAc, 0.75mM EDTA) to a total volume of 20 μl.
b. Denature 5′ at 65°C, then chill on ice.
c. Spin briefly, then add 2μl 10x RNA loading dye (25mg Xylene cyanole FF in 6ml glycerol / 2mM EDTA).
d. Electrophorese samples for 2.5 h at 160 V on a 1.2% agarose formaldehyde gel (6% HCHO, Sambrook et al., 1989)
e. Soak gel 40′ in 20x SSC (Sambrook et al, 1989).
f. Pressure blot gel onto a nitrocellulose or nylon membrane (Hybond, Amersham) for 1 h at 75 mm Hg.
g. Crosslink 2′ with UV (Stratalinker, Stratagene).
h. Bake 1 h at 80°C.
i. Prehybridize blots with agitation for 2 h at 42°C in 50% formamide, 5x SSPE, 2x Denhardt's, 0.1% SDS.
j. Hybridize with agitation overnight at 42°C with a ^{32}P random prime labeled 750 bp 5′ probe that had been previously amplified by PCR (described above).

k. Wash 4x for 20′ at 45°C in 2x SSC, 0.1% SDS. Monitor activity of blot with hand-held Geiger counter.

l. Evaluate by autoradiography with intensifying screen overnight.

APPENDIX D

CREATING NESTED DNaseI DELETIONS IN A HEMOCYANIN cDNA

based on Lin et al. (1985) and Christine and Bernard Thisse (pers. comm.):

a. Precipitate 50 μg maxiprep DNA (1800 bp hemocyanin cDNA cloned into a Bluescript SK vector) with 2 volumes ethanol and 0.1 volume 3M NaOAc.

b. Resuspend in 300 μl DNase I/Mn^{2+} buffer (0.2 M Tris-HCl, 10 mM $MnCl_2$, 1 mg\ml BSA).

c. Divide into 5 aliquots and digest each for 5′ with 4 μl of the following serial dilutions of DNase I: 0.2 ng/μl, 0.1 ng/μl, 0.05 ng/μl, 0.02 ng/μl and 0.01 ng/μl.

d. After 5′ terminate digests by phenol/chloroform extraction (Sambrook et al, 1989).

e. Run 6 μl aliquots of each reaction on a 0.6% agarose gel. Choose the reaction showing the best linearization (no supercoiled or nicked DNA, no smear), discard the others.

f. Ethanol precipitate DNA (see above) and spin, wash and dry pellet.

g. Resuspend in 150 μl TE (pH 8.0) and digest with 20-50 units of enzyme 1 (SmaI) in a total volume of 200 μl at

37°C for 2 h.

h. Add 200 μl PEG (13% PEG_{8000} in 1.6 M NaCl), mix and store on ice for 90′.

i. Spin 15′ at 16000 g at 4°C in microcentrifuge. Resuspend pellet in 400 μl TE (pH 8.0).

j. Extract with phenol, phenol-chloroform and chloroform.

k. Ethanol precipitate (see above).

l. Spin, wash and dry pellet. To repair ends of fragments, add 10 μl 0.25 mM dNTPs, 5 μl [0.1 M Tris-HCl (pH 7.8) 0.1 M $MgCl_2$], 34 μl H_2O and 1 μl (5 units) Klenow fragment. Incubate 15′ at room temperature.

m. Stop reaction by incubating 15′ at 68°C.

n. To religate fragments, add 40 μl 5x blunt end ligation buffer [250 mM Tris-HCl (pH 7.5), 50 mM $MgCl_2$, 25% PEG_{8000}, 5 mM ATP, 5 mM DTT], 1 ul T4 DNA ligase (1:10 dil., 40 biolabs units) and 109 μl H_2O. Incubate overnight at 16°C.

o. Stop ligation by incubation at 68°C for 15′.

p. Digest DNA with 40 units enzyme 2 (EcoRI) in a total volume of 400 μl. Incubate at 37°C for 2 h. Ethanol precipitate and resuspend in 50 μl TE.

q. Transform 200 μl competent E. coli XL-1 Blue cells with 1 and 10 μl of the mixture (Sambrook et al., 1989) and plate out on LB-Amp plates. Grow overnight at 37°C.

r. Select 60 positive clones and isolate plasmid DNA by alkaline lysis miniprep (Sambrook et al., 1989).

s. Determine insert size by restriction analysis with enzymes 3 and 4 (KpnI/XbaI; Sambrook et al., 1989).

t. Select 15 clones with useful insert sizes (150 to 1800 bp) and sequence by dideoxy-method (SEQUENASE kit, US Biochemical) according to manufacturer's instructions.

BIBLIOGRAPHY

Chapter I

Brown, A.C. (1991) Dissertation, University of Oregon.

Brown, A.C. and Terwilliger, N.B. (1992) Biol. Bull. 182, 270-277.

Ellerton, H.D., Carpenter, D.E. & Van Holde, K.E. (1970) Biochemistry 9, 2225-2232.

Fredericq, L. (1878) Arch. Zool. Exptl. J. 7, 535-583.

Hazes, B., Magnus, K.A., Bonaventura, C., Bonaventura, J., Dauter, Z., Kalk, K. and Hol, W.G.J. (1993) Protein Science 2, 597-619.

Heip, J., Moens, L., Hertsens, R., Wood, E.J., Heyligen, H., Van Broeckhoven, A., Vrints, R., de Chaffoy, D. and Kondo, M. (1980) in *The Brine Shrimp Artemia*, vol. 2, eds. Personne, G.T., Sorgeloos, P., Rods, O. and Jaspers, E. (Universa press, Belgium), pp. 427-448.

Hobbs, R.C., Botsford, L.W. and Thomas, A. (1992) Can. J. Fish. Aquat. Sci. 49, 1379-1388.

Ingermann, R. (1992) in *Adv. Comp. Environ. Physiol.*, ed. Mangum, C.P. (Springer Verlag, Berlin, Heidelberg, New York), 13, pp. 411-431.

Kitajima, N., Fujisawa, K., Fujimoto, C., Moro-oka, Y., Hashimoto, S., Kitagawa, T., Toriumi, K., Tatsumi, K. and Nakamura, A. (1992) J. Am. Chem. Soc. 114, 1277-1291.

Krogh, A. (1929) Am. J. Physiol. 90, 243-251.

Kurtz, D.M. (1986) in *Invertebrate Oxygen Carriers*, ed. Linzen, B. (Springer Verlag, Berlin, Heidelberg, New York), pp. 9-21.

Larson, B.A., Terwilliger, N.B. & Terwilliger, R.C. (1981) Biochim. Biophys. Acta 667, 294-302.

Lontie, R., DeLey, M., Robberecht, H. and Witters, R. (1973) Nature New Biol. 242, 180-182.

MacKay, D.C.G. (1942) Bull. Fish. Res. Bd. Canada 62, 1-32.

Magnus, K.A., Hazes, B., Ton-That, H., Bonaventura, C., Bonaventura, J. and Hol, W.G.J. (1994) Prot. Struct. Funct. Genet. 19, 302-309.

Otoshi, C. (1994): Masters Thesis, University of Oregon.

Schin, K., Laufer, H. & Clark, R.M. (1979) J. Exp. Zool. 210, 265-275.

Shanks, A.L. (1983) Mar. Ecol. Prog. Ser. 13, 311-315.

Shanks, A.L. (1986) Mar. Biol. 92, 189-199.

Terwilliger, N.B. and Terwilliger, R.C. (1982) J. Exp. Zool. 221, 181-191.

Terwilliger, N.B., Terwilliger, R.C. & Graham, R. (1986) in *Invertebrate Oxygen Carriers*, ed. Linzen, B. (Springer Verlag, Berlin), pp. 333-335.

Terwilliger, N.B. and Brown, A.C. (1993) J. Exp. Biol. 183, 1-13.

Terwilliger, N.B. and Otoshi, C. (1994) The Physiol. 37, A-67.

Van Holde, K.E. & Miller, K.I. (1995) Adv. Prot. Chem. 47, 1-81.

Volbeda, A. and Hol, W.G.J. (1989a) J. Mol. Biol. 209, 249-279.

Volbeda, A. and Hol, W.G.J. (1989b) J. Mol. Biol. 206, 531-546.

Wittenberg, J.B. (1992) in: Advances in comparative and environmental physiology 13, ed. Mangum, C.P., 59-85.

Chapter II

Alliel, P.M., Dautigny, A., Lamy, J., Lamy, J.-N., Jolles, P. (1983) Eur. J. Biochem. 134, 407-414.

Bak, H.J., Beintema, J.J. (1987) Eur. J. Biochem. 169, 333-348.

Beintema, J. J., Stam, W.T., Hazes, B., Smidt, M.P. (1994) Mol. Biol. Evol. 11, 493-503.

Brown, A.C. (1991) Dissertation, University of Oregon.

Brown, A.C. and Terwilliger, N.B. (1992) Biol. Bull. 182, 270-277.

Bunn, H.F., Forget, B.G., Ranney, H.M. (1977) "Human Hemoglobins." W.B. Saunders Company, Philadelphia, London, Toronto.

Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucl. Ac. Res. 12 (1), 387-395.

Drexel, R., Siegmund, S., Schneider, H.-J., Linzen, B., Gielens, C., Preaux, G., Lontie, R., Kellermann, J. and Lottspeich, F. (1987) Biol. Chem. Hoppe Seyler 368: 617-635.

Fahrenbach, W.H. (1970) J. Cell. Biol. 44, 445-453.

Gaykema, W.P.J., Hol, W.G.J., Vereijken, J.M., Soeter, N.M., Bak, H.J. and Beintema, J.J. (1984): Nature 309, 23-29.

Ghiretti-Magaldi, A., Milanesi, C. and Salvato, B. (1973) Experientia 29, 1265-1267.

Ghiretti-Magaldi, A., Milanesi, C. and Tognon, G. (1977) Cell. Differ. 6, 167-186.

Graham, R.A., Mangum, C.P., Terwilliger, R.C. and Terwilliger, N.B. (1983) Comp. Biochem. Physiol. 74A, 45-50.

Hazes, B., Magnus, K.A., Bonaventura, C., Bonaventura, J., Dauter, Z., Kalk, K. and Hol, W.G.J. (1993) Protein Science 2, 597-619.

Hennecke, R., Gellissen, G., Spindler-Barth, M. and Spindler, K.-D. (1990) In: Preaux, G. and Lontie, R. (eds): "Invertebrate Dioxygen Carriers." Leuven: Leuven Univ. Press, pp 503-506.

Kempter, B. (1986) In: Linzen, B. (ed): "Invertebrate Oxygen Carriers." Berlin, Heidelberg, New York: Springer, pp 489-494.

Krogh, A. (1929) Am. J. Physiol. 90, 243-251.

Lang, W. and van Holde, K.E. (1991) P. Natl. Acad. Sci. USA, 88, 244-248.

Lin, H.C., Lei, S. and Wilcox, G. (1985) Anal. Biochem. 147, 114-119.

Linzen, B., Soeter, N.M., Riggs, A.F., Schneider, H.-J., Schartau, W., Moore, M.D., Yokota, E., Behrens, P.Q., Nakashima, H., Takagi, T., Nemoto, T., Vereijken, J.M., Bak, H.J., Beintema, J.J., Volbeda, A., Gaykema, W.P.J., Hol, W.G.J. (1985): Science 229, 519-524.

Markl, J. and Decker, H. (1992) In Mangum CP (ed): "Advances In Comparative and Environmental Physiology." Berlin: Springer-Verlag, Vol 13, pp 325-376.

Markl, J., Stumpp, S., Bosch, F.X. and Voit, R. (1990) In Preaux G, Lontie R (eds): "Invertebrate Dioxygen Carriers." Leuven: Leuven University Press, pp 497-502.

McMahon, B.R., McDonald, D.G. and Wood, C.M. (1979) J. Exp. Biol. 80, 271-285.

Morris, S. and McMahon, B.R. (1989) Physiol. Zool. 62, 654-667.

Otoshi, C. (1994): Masters Thesis, University of Oregon.

Préaux, G., Vandamme, A., De Bethune, B., Jacobs M.-P. and Lontie, R. (1986) In Linzen, B. (ed): "Invertebrate Oxygen Carriers" Berlin, Heidelberg, New York: Springer Verlag, pp 485-488.

Sambrook, J., Maniatis, T. and Fritsch, E.F. (1989): "Molecular Cloning." (2nd ed.) New York: Cold Spring Harbor Laboratory Press.

Senkbeil, E.G. and Wriston, J.C. (1981) Comp. Biochem. Physiol. 68B, 163-171.

Sullivan, B., Bonaventura, J. and Bonaventura, C. (1974) Proc. Natl. Acad. Sci. USA 71, 2558-2562.

Terwilliger, N.B. and Terwilliger, R.C. (1982) J. Exp. Zool. 221, 181-191.

Terwilliger, N.B., Terwilliger, R.C. and Graham, R. (1985) In Linzen, B., (ed.): "Invertebrate Oxygen Carriers" Berlin, Heidelberg, New York: Springer-Verlag pp 333-335.

Terwilliger, N.B. and Brown, A.C. (1993) J. Exp. Biol. 183, 1-13.

Terwilliger, N.B. and Otoshi, C. (1994) The Physiol. 37, A-67.

Truchot, J.P. (1992) In Magnum, C.P. (ed.): "Advances In Comparative and Environmental Physiology." Berlin: Springer-Verlag, Vol. 13, pp 377-410.

van Holde, K.E. and Miller, K.J. (1982) Q. Rev. Biophys. 15, 1-129.

van Holde, K.E., Miller, K.I. and Lang, W.H. (1992) In Mangum, C.P. (ed.): "Advances In Comparative and Environmental Physiology." Berlin: Springer-Verlag, Vol 13, pp 258-300.

Volbeda, A. and Hol, W.G.J. (1989a) J. Mol. Biol. 209, 249-279.

Volbeda, A. and Hol, W.G.J. (1989b) J. Mol. Biol. 206, 531-546.

Wood, E.J. and Bonaventura, J. (1981) Biochem. J. 196, 653-656.

Chapter III

Alliel, P.M., Dautigny, A., Lamy, J., Lamy, J.-N. & Jolles, P. (1983) Eur. J. Biochem. 134, 407-414.

Bak, H.J. & Beintema, J.J. (1987) Eur. J. Biochem. 169, 333-348.

Beintema, J.J., Stam, W.T., Hazes, B. & Smidt, M.T. (1994) Mol. Biol. Evol. 11, 493-503.

Brown, A.C. & Terwilliger, N.B. (1992) Biol. Bull. 182, 270-277.

Cleveland, D.W., Fischer, S.G., Kirschner, M.W. & Laemmeli, U.K. (1977) J. Biol. Chem. 252, 1102-1106.

de Haas, F., Bijlhout, M.C. & van Bruggen, E.F.J. (1991) J. Struct. Biol. 107, 86-94.

Ellerton, H.D., Carpenter, D.E. & Van Holde, K.E. (1970) Biochemistry 9, 2225-2232.

Fahrenbach, W.H. (1970) J. Cell. Biol. 44, 445-453.

Freedman, T.B., Loehr, J.S. & Loehr, T.M. (1976) J. Am. Chem. Soc. 98, 2809-2815.

Ghiretti-Magaldi, A., Milanesi, C. & Salvato, B. (1973) Experientia 29, 1265-1267.

Ghiretti-Magaldi, A., Milanesi, C. & Tognon, G. (1977) Cell Differ. 6, 167-186.

Heip, J., Moens, L., Hertsens, R., Wood, E.J., Heyligen, H., Van Broeckhoven, A., Vrints, R., de Chaffoy, D. and Kondo, M. (1980) in *The Brine Shrimp Artemia*, vol. 2, eds. Personne, G.T., Sorgeloos, P., Rods, O. and Jaspers, E. (Universa press, Belgium), pp. 427-448.

Hennecke, R., Gellissen, G., Spindler-Barth, M. & Spindler, K.-D. (1990) in *Invertebrate Dioxygen Carriers*, eds. Preaux, G. & Lontie, R. (Leuven University Press, Leuven), pp. 503-506.

Ingermann, R. (1992) in *Adv. Comp. Environ. Physiol.*, ed. Mangum, C.P. (Springer Verlag, Berlin, Heidelberg, New York), 13, pp. 411-431.

Kempter, B. (1983) Naturwissenschaften 5, 255.

Larson, B.A., Terwilliger, N.B. & Terwilliger, R.C. (1981) Biochim. Biophys. Acta 667, 294-302.

Linzen, B., Soeter, N.M., Riggs, A.F., Schneider, H.-J., Schartau, W., Moore, M.D., Yokota, E., Behrens, P.Q., Nakashima, H., Takagi, T., Nemoto, T., Vereijken, J.M., Bak, H.J., Beintema, J.J., Volbeda, A., Gaykema, W.P.J. & Hol, W.G.J. (1985) Science 229, 519-524.

Markl, J., Schmid, R., Czichos-Tiedt, S. and Linzen, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1713-1725.

Markl, J., Stumpp, S., Bosch, F.X. & Voit, R.(1990) in *Invertebrate Dioxygen carriers*, eds. Preaux, G. & Lontie, R. (Leuven University Press, Leuven), pp. 497-502.

Markl, J. & Decker, H. (1992) in *Advances in Comparative and Environmental Physiology*, ed. Mangum, C.P., (Springer Verlag, Berlin, Heidelberg, New York), 13, pp. 325-376.

Pickett, S.M., Riggs, A.F. and Larimer, J.L. (1966) Science 151, 1005-1007.

Preaux, G., Vandamme, A., De Bethune, B., Jacobs, M.-P. & Lontie, R. (1986) in *Invertebrate Oxygen Carriers*, ed. Linzen, B. (Springer Verlag, Berlin, Heidelberg, New York), pp. 485-488.

Rainer, J. & Brouwer, M. (1993) Comp. Biochem. Physiol. 104B, 69-73.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning* (Cold Spring Harbor Laboratory Press, New York), pp. 7.43-7.50.

Schin, K., Laufer, H. & Clark, R.M. (1979) J. Exp. Zool. 210, 265-275.

Schönenberger, N., Cox, J.A. & Gabbiani, G. (1980) Cell Tissue Res. 205, 397-409.

Senkbeil, E.G. & Wriston, J.C. (1981) Comp. Biochem. Physiol. 68B, 163-171.17.

Terwilliger, N.B. & Terwilliger, R.C. (1982) J. Exp. Zool. 221, 181-191.

Terwilliger, N.B., Terwilliger, R.C. & Graham, R. (1986) in *Invertebrate Oxygen Carriers*, ed. Linzen, B. (Springer Verlag, Berlin), pp. 333-335.

Terwilliger, N.B. & Brown, A.C. (1993) J. exp. Biol. 183, 1-13.

Terwilliger, N.B. & Bremiller, R. (1995) Am. Zool. 35, 65A.

Terwilliger, N.B. & Durstewitz, G. (1996) in: *Molecular Zoology: Advances, Strategies and Protocols*, ed. Ferraris, J. & Palumbi, S., (Wiley-Liss, New York), pp. 353-368.

Van Holde, K.E. & Miller, K.I. (1982) Quart. Rev. Biophys. 15, 1-129.

Van Holde, K.E. & Miller, K.I. (1995) Adv. Prot. Chem. 47, 1-81.

Voit, R. & Schneider, H.-J. (1986) Eur. J. Biochem. 159, 23-29

Volbeda, A. & Hol, W.G.J. (1989) J. Mol. Biol. 209, 249-279.

Wilkinson, M., Doskow, J. & Lindsey, S. (1990) Nucl. Acid Res. 19 (3), 679.

Wood, E.J. & Bonaventura, J. (1981) Biochem. J. 196, 653-656.

Chapter IV

Aspan, A., Huang, T.-S., Cerenius, L. and Söderhäll, K. (1995) Proc. Natl. Acad. Sci. USA 92, 939-943.

Ballard, J.W.O., Olsen, G.J., Faith, D.P., Odgers, W.A., Rowell, D.M. and Atkinson, P.W. (1992) Science 258, 1345-1348.

Beintema, J.J., Stam, W.T., Hazes, B. and Smidt, M.P. (1994) Mol. Biol. Evol. 11 (3), 493-503.

Brusca, R.C. and Brusca, G.J. (1990) *Invertebrates* (Sinauer, Sunderland), 681-691.

Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucl. Acids Res. 12 (1), 387-395.

Drexel, R., Siegmund, S., Schneider, H.J., Linzen, B., Gielens, C., Preaix, G., Kellermann, J. and Lottspeich, F. (1987) Biochem. HS 368, 617-635.

Durstewitz, G. and Terwilliger, N.B. (1996)

Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger, J. (1985) J. Virol. 55, #3, 836-839.

Fredericq, L. (1878) Arch. Zool. Exptl. J. 7, 535-583.

Fujii, T., Sakurai, H., Izumi, S. and Tomino, S. (1989) J. Biol. Chem. 264, # 19, 11020-11025.

Garey, J.R., Krotec, M., Nelson, D.R. and Brooks, J. (1996) Invert. Biol. 115 (1), 79-88.

GCG Sequence Analysis Software Package (1994) University of Wisconsin, Madison.

Hazes, B. and Hol, W.G.J. (1992) Proteins 12, 278-298.

Hazes, B., Magnus, K.A., Bonaventura, C., Bonaventura, J., Dauter, Z., Kalk, K.H. and Hol, W.G.J. (1993) Protein Science 2, 597-619.

Jameson, B.A. and Wolf, H. (1988) CABIOS 4, #1, 181-186.

Kawabata, T., Yusahara, Y., Ochiai, M., Matsuura, S. and Masaaki, A. (1995) Proc. Natl. Acad. Sci. USA 92, 7774-7778.

Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.

Larson, B.A., Terwilliger, N.B. and Terwilliger, R.C. (1982) Biochim. biophys. acta 667, 294-302.

Lerch, K., Huber, M., Schneider, H.-J., Drexel, R. and Linzen, B. (1986) J. Inorg. Chem. 26, 213-217.

Lin, H.C., Lei, S. and Wilcox, G. (1985) Anal. Biochem. 147, 114-119.

Linzen, B., Soeter, N.M., Riggs, A.F., Schneider, H.-J., Schartau, W., Moore, M.D., Yokota, E., Behrens, P.Q., Nakashima, H., Takagi, T., Nemoto, T., Vereijken, J.M., Bak, H.J., Beintema, J.J., Volbeda, A., Gaykema, W.P.J. & Hol, W.G.J. (1985) Science 229, 519-524.

Markl, J. and Decker, H. (1992) in *Advances in Comparative and Environmental Physiology*, ed. Mangum, C.P., (Springer Verlag, Berlin, Heidelberg, New York), 13, pp. 325-376.

Munn, E.A. and Greville, G.D. (1969) J. Insect Physiol. 15, 1935-1950.

Nakahara, A., Suzuki, S. and Kino, J. (1983) in: *Structure and Function of Invertebrate Respiratory Proteins*, ed. Wood, E.J., Life Chem. Rep. Suppl. 1, (Harwood, London), 319-322.

Neuteboom, B., Jekel, P.A., Hofstra, R.M.W., Sierdsema, S.J. and Beintema, J.J. (1990) in *Invertebrate Dioxygen Carriers*, eds. Preaux, G. and Lontie, R., (Leuven University Press, Leuven), pp. 85-88.

Ochiai, E.I. (1983) Biosystems 16, 81-86.

Otoshi, C. (1994) Masters Thesis, University of Oregon.

Remane, A., Storch, V. and Welsch, U. (1980) Systematische Zoologie, Gustav Fischer Verlag, Stuttgart and New York, 227.

Ruppert, E.E. and Barnes, R.D. (1994) *Invertebrate Zoology* (Saunders College Publ., Fort Worth), 611, 801-802.

Salvato, B., Jori, G., Piazzese, A., Ghiretti, F., Beltramini, M. and Lerch, K. (1983) in: *Structure and Function of Invertebrate Respiratory Proteins*, ed. Wood, E.J., Life Chem. Rep. Suppl. 1, (Harwood, London), 313-317.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning*, 2nd ed., (Cold Spring Harbor Laboratory Press, New York)

Stewart, C.B. (1993) Nature 361, 603-607.

Swofford, D. L. (1991) PAUP: Phylogenetic Analysis Using Parsimony, version 3.1, Illinois Natural History Survey, Champaigne.

Telfer, W.H. and Massey, H.C. (1987) *UCLA Symp. Mol. Cell. Biol. New Ser. 49*, ed. Law, J.H., Alan R. Liss, New York, 305-314

Telfer, W.H. and Kunkel, J.G. (1991) Annu. Rev. Entomol. 36, 205-228.

Terwilliger, N.B. and Brown, A.C. (1993) J. exp. Biol. 183, 1-13.

Terwilliger, N.B. and Durstewitz, G. (1996) in: *Molecular Zoology: Advances, Strategies and Protocols*, eds: Ferraris, J.D. and Palumbi, S.R., (Wiley-Liss), 353-368.

Turbeville, J.McC., Pfeifer, D.M., Field, K.G. and Raff, R.A. (1991) Mol. Biol. Evol. 8 (5), 669-686.

Van Holde, K.E. and Miller, K.I. (1995) Adv. Prot. Chem. 47, 1-81.

Volbeda, A. and Hol, W.G.J. (1989a) J. Mol. Biol. 206, 531-546.

Volbeda, A. and Hol, W.G.J. (1989b) J. Mol. Biol. 209, 249-279.