

AN ABSTRACT OF THE THESIS OF

Walter H. Lang for the degree of Doctor of Philosophy in

Biochemistry and Biophysics presented on July 13, 1990.

Title: cDNA cloning and Sequencing of *Octopus dofleini* hemocyanin

Signature redacted for privacy.

Abstract approved: _____

Hemocyanins are oxygen transport molecules found only in molluscs and arthropods. They are very large molecules with molecular weights in the range of several million Daltons. They are multisubunit aggregates and their oxygen binding site contains a binuclear copper center. Molluscan and arthropodan hemocyanins are fundamentally different in structure and it has been of interest for a long time whether both proteins are products of divergent or convergent evolution. It was clear that sequence data were needed from both proteins in order to resolve this question. Our laboratory is investigating the structure and function of *Octopus dofleini* hemocyanin. It consists of ten polypeptide chains of 350,000 Da each. Each subunit is composed of seven domains. In order to sequence hemocyanin recombinant DNA methods were chosen, because conventional protein sequencing methods seemed not feasible. Complementary DNA clones coding for three domains at the C-terminal end of hemocyanin were isolated and sequenced. Comparison of these sequences with those of arthropodan hemocyanins showed no similarity, except for a small region corresponding to the "Copper B" site. Molluscan hemocyanins are more closely related to tyrosinases than they are to arthropodan hemocyanins.

Sequence comparisons with domains of other recently published molluscan hemocyanins showed that molluscan hemocyanins already existed in the precambrian before the molluscan orders diverged from each other in the early cambrium. Sequence comparisons of molluscan hemocyanins with tyrosinases allowed us to identify potential ligands for the "Copper A" site, whose structure, unlike in arthropodan hemocyanins, is different from the "Copper B" site.

CDNA CLONING AND SEQUENCING OF
OCTOPUS DOFLEINI HEMOCYANIN

by

Walter H. Lang

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed July 13, 1990

Commencement June 1991

APPROVED:

Signature redacted for privacy.

Professor of Biochemistry and Biophysics in charge of major

Signature redacted for privacy.

Chairman of the department of Biochemistry and Biophysics

Signature redacted for privacy.

Dean of Graduate School

Date Thesis is presented July 13, 1990

To my daughter Tanya Katharina,
who I loved but never got to know

ACKNOWLEDGEMENTS

I would like to express my thanks to Professor Kensal van Holde for giving me the opportunity to carry out this research project and to Professor Gary Merrill for instructing me in some basic techniques used for recombinant DNA work.

TABLE OF CONTENTS

	page
A. INTRODUCTION	1
1. Overview	1
2. Structure of hemocyanins	4
2.1. Arthropodan hemocyanins	4
2.2. Molluscan hemocyanins	6
3. Biosynthesis of hemocyanins	9
4. Hemocyanin sequences	11
5. Active site structure	12
6. Structure of <i>Octopus dofleini</i> hemocyanin	17
 B. MATERIALS AND METHODS	 20
1. Materials	20
1.1. Chemicals	20
1.2. Reagents for molecular biology	20
1.3. Buffers and solutions	21
1.4. Plasmids and strains	22
1.5. Media for growth of bacteria	22
2. Methods	23
2.1. Routine Methods	23
2.1.1. Sterilization	23
2.1.2. Extraction with organic solvents	24
2.1.3. Ethanol precipitation	24
2.1.4. Enzymatic manipulations of DNA	25
2.1.5. Gel electrophoresis	26
2.1.6. Transformation of <i>E. coli</i>	28
2.1.7. Radiolabeling of DNA	29
2.1.8. Oligonucleotides	29
2.2. Purification of plasmid DNA	31
2.2.1. Small scale purification	31
2.2.2. Large scale purification	33
2.3. RNA purification	37
2.4. cDNA cloning	38
2.4.1. Homopolymeric tailing of plasmids and cDNA	38

2.4.2. cDNA cloning according to Heidecker and Messing (1983)	39
2.4.3. cDNA cloning according to Gubler and Hoffmann (1983)	41
2.5. Screening of cDNA libraries	42
2.6. Purification and subcloning of DNA fragments	43
2.7. Generation of unidirectional deletions	45
2.8. Sequencing	46
2.9. Sequence analysis	49
 C. RESULTS	 50
1. RNA purification and Northern blotting	50
2. cDNA cloning and sequencing	52
2.1. Isolation and characterization of pHCl	52
2.2. Isolation and characterization of pHCl	60
2.3. Isolation and characterization of XT3 and XT22	65
2.4. Presence of two cDNAs coding for <i>Octopus dofleini</i> hemocyanin	71
3. Properties of the polypeptide chain	73
 D. DISCUSSION	 91
1. Presence of two polypeptide chains	91
2. Comparison of sequences of molluscan hemocyanins	93
2.1. Comparison of <i>Octopus dofleini</i> domains with each other	93
2.2. Comparison to other molluscan hemocyanins	96
3. Structure of the active site	98
4. Evolution of molluscan hemocyanins	107
 E. REFERENCES	 112

LIST OF FIGURES

Figure 1	Structure of the active site in hemocyanins	15
Figure 2	Gel electrophoresis of <i>Octopus dofleini</i> RNA	51
Figure 3	Restriction map of the pHc1 cDNA insert	53
Figure 4	Sequence of the cDNA insert of pHc1 and derived protein sequence	55
Figure 5	Sequence comparison of the C-terminal domains of <i>Octopus dofleini</i> and <i>Paroctopus dofleini dofleini</i>	58
Figure 6	Restriction map of pHc2	61
Figure 7	Sequence of the cDNA insert of pHc2 and the translated protein sequence	62
Figure 8	Restriction maps of cDNA clones XT3, XT5, XT22 and XT53	67
Figure 9	Sequence of the cDNA insert of XT22	68
Figure 10	DNA sequence obtained with primer WL4	72
Figure 11	Restriction map of <i>Octopus dofleini</i> cDNA	74
Figure 12	Comparison of the G-type and A-type sequence at the DNA and protein level	76
Figure 13	Sequence of <i>Octopus dofleini</i> cDNA	82
Figure 14	Results of secondary structure prediction according to Chou and Fasman (1978)	88
Figure 15	Hydropathy plots for all <i>Octopus</i> domains according to the method of Kyte and Doolittle (1980)	89
Figure 16	Comparison of G-type and A-type sequences at the DNA and protein level in putative linker regions	92
Figure 17	Alignment of complete or almost complete sequences for domains of molluscan hemocyanins	94
Figure 18	Matrix of results of sequence comparisons of single domains with each other	95
Figure 19	Alignment of domains of three molluscan hemocyanins	97
Figure 20	Alignment of sequences around the putative "Copper B" site in hemocyanins and tyrosinases	100
Figure 21	Alignment of sequences around the putative "Copper A" site in molluscan hemocyanins and tyrosinases	103
Figure 22	Scheme for the evolution of hemocyanins and tyrosinases	111

LIST OF TABLES

Table 1	N-terminal sequences of individual domains of <i>Octopus dofleini</i> hemocyanin (from Lamy <i>et al.</i> , 1987)	19
Table 2	Sequences of oligonucleotides WL2, WL3 and WL4	30
Table 3	Amino acid composition of <i>Ode</i> , <i>Odf</i> , and <i>Odg</i>	87

CDNA CLONING AND SEQUENCING OF *OCTOPUS DOFLEINI* HEMOCYANIN

A. INTRODUCTION

1. Overview

Primitive organisms are able to supply the oxygen needs of their tissues by relying on uptake by simple diffusion processes. More complex organisms are dependent on oxygen transport from their body surface to the site of consumption. A number of ways have been developed to supply tissues with oxygen. For example, insects evolved a tracheal system which penetrates the whole body and enables oxygen to reach the remotest corners within the body. By far the most used means of supplying oxygen to tissues is by oxygen being bound to an extra-, or intracellular transport protein circulating throughout the body within an open or closed circulatory system. These proteins evolved to bind oxygen reversibly at high partial pressures and release it at low partial pressures. There are three different kinds of proteins found in organisms capable of performing this task: hemoglobins, hemerythrins and hemocyanins.

Hemoglobins are found in a wide range of organisms and are mostly enclosed within red blood cells. A great number of sequences are known (the latest release of the Protein Identification Resource sequence data bank lists 398 entries); all have heme as the prosthetic group. In vertebrates they are found in $\alpha_2\beta_2$ conformation and are always enclosed in red blood cells (Dickerson and Geis, 1983). Each polypeptide chain contains one heme.

Hemoglobins found in invertebrate organisms occur both extra- and intracellularly (Vinogradov, 1985). Extracellular hemoglobins are also called erythrocruorins or chlorocruorins. They form very large subunit aggregates with one heme per polypeptide chain. Sometimes they are associated with other non heme containing subunits which presumably are involved in quaternary structure formation. In a few invertebrate groups very large hemoglobins are found which contain two and more hemes per polypeptide chain. Chlorocruorins contain a heme variant that gives rise to a green color. They are found as large subunit aggregates in some annelids.

Members of the two minor phyla, *Brachiopoda* and *Sipunculida*, have a very peculiar protein, named hemerythrin as oxygen transport protein (Klotz and Kurtz, 1984). Some hemerythrins are sequenced and the crystal structure for one is solved at a resolution of 2 Å (Stenkamp *et al.*, 1984). The native form is composed of eight subunits; each contains a binuclear iron center at its active site. One iron atom has two histidine ligands, the other one three. One aspartate side chain, one glutamate side chain and one hydroxyl group serve as bridging ligands between the two iron atoms. One of those has only five ligands and oxygen is presumably bound to it as hydroperoxide and occupies the sixth coordination position.

The third group of oxygen transport proteins are the hemocyanins (van Holde and Miller, 1982; Ellerton *et al.*, 1983; Preaux and Gielens, 1984; Herskovits, 1988). These are always found as extracellular proteins in organisms belonging to two major phyla, the molluscs and arthropods, but not all representatives from these two phyla have hemocyanins as an oxygen transport protein. They were recognized early as very large proteins with molecular weights in the range of several million Daltons (Eriksson-Quensel and Svedberg, 1936). Their active site contains a binuclear copper center giving

rise to a blue color upon oxygen binding which gave these proteins their name. Although they superficially appear to be very similar, molluscan and arthropodan hemocyanins are fundamentally different structurally.

The main objective of this work is concerned with determination of the protein sequence of *Octopus dofleini* hemocyanin whose structure and function our laboratory is investigating. When this work was initiated very little sequence information was available about molluscan hemocyanins. In contrast several complete sequences were known from arthropodan hemocyanins (Linzen *et al.*, 1985). A number of questions about the structure and evolution of hemocyanins could be answered with the availability of more sequence information about molluscan hemocyanins:

(i) Is the sequence of molluscan hemocyanins repetitive? This could be expected from results of electron microscopy showing that subunits of molluscan hemocyanins are composed of globular domains, their number corresponding the the number of oxygen binding sites.

(ii) Is the structure of the active site identical in both molluscan and arthropodan hemocyanins? Results from spectroscopic investigations suggest that the active site in both classes might be very similar if not identical.

(iii) Did molluscan and arthropodan hemocyanins evolve from a common ancestor or are they products of convergent evolution? Results from spectroscopic data suggest the first, subunit structure and assembly of both classes suggest the latter. Sequence information from both molluscan and arthropodan hemocyanins is needed in order to gain insight on how both classes evolved and to determine whether they were of common ancestry or products of convergent evolution.

In the following paragraphs I will outline features of the subunit structure of both molluscan and arthropodan hemocyanins, results of

sequencing studies, the proposed structure of the active site, and hemocyanin biosynthesis. Then I will summarize results obtained in our laboratory on the structure of *Octopus dofleini* hemocyanin before presenting results of my own research.

2. Structure of hemocyanins

2.1. Arthropodan hemocyanins

In their native form arthropodan hemocyanins are found at various levels of aggregation depending on the source of the protein (van Holde and Miller, 1982; Ellerton *et al.*, 1983; Herskovits, 1988). The basic building block is a hexamer composed of nonidentical subunits with molecular weights in the range of 75 kDa. The native forms of arthropodan hemocyanins are composed of one, two, four, six, or eight of these hexameric units. Sometimes two forms can be found in the hemolymph. Between three and eight immunologically distinct polypeptide chains can be found in arthropodan hemocyanins. In general the higher the level of aggregation, the more different polypeptide chains are present. A minimum of eight different ones are found in the most complex arthropodan hemocyanins. Each one of those polypeptide chains contains a binuclear copper center which is part of the active site.

An example for a hemocyanin composed of hexamers only is found in *Panulirus interruptus*. It contains three different polypeptide chains named a, b, and c, with apparent molecular weights of 94, 90 and 80 kDa, respectively (Folkerts and van Eerd, 1981; Van Eerd and Folkerts, 1981). Sequence analysis however gave molecular weights of 77 kDa for subunit a and b and 80 kDa for subunit c (Bak and Beintema, 1987; Jekel *et al.*, 1988; Neutebom *et al.*, 1990). Not all of the polypeptide chains are required for hexamer formation. It can be reconstituted from chains a and b only. This form was used for X-ray

diffraction studies that resulted in solving the first three dimensional structure for a hemocyanin (Gaykema *et al.*, 1984). An example for a hemocyanin composed of eight hexamers is found in the horseshoe crab *Limulus polyphemus* (Bijlholt *et al.*, 1979). Eight immunologically distinct polypeptide chains with molecular weights of around 70 kDa are found (Brenowitz *et al.*, 1981) and up to 15 chains can be separated electrophoretically (Brenowitz *et al.*, 1984). It seems that some of these electrophoretically distinct forms are functionally identical and differences are due to microheterogeneity or post-translational modifications.

The individual polypeptide chains fulfill specific roles in the assembly of higher order aggregates of arthropodan hemocyanins. For example, certain polypeptide chains are essential for their formation. If these are omitted from a reconstitution mixture, higher order structures will not form. This implies that they are involved in forming specific contacts between hexamers. The roles of individual polypeptide chains in forming of higher order structures has been examined in detail in two systems, *Androctonis australis* (Lamy *et al.*, 1981) and *Eurypelma californicum* (Markl *et al.*, 1981), using immunological methods combined with electronmicroscopy.

The native hemocyanin of the scorpion *Androctonis australis* consists of four hexamers composed of eight different polypeptide chains named 2, 3A, B and C, 4, 5A and B, and 6. The hemocyanin was dissociated into its subunits which were then purified. The purified subunits were mixed in various combinations and the reassociation products were examined. It was observed that the tetrahexameric structure was only formed when subunits 3C and 5B were added to the reconstitution mixture. They form an inner ring holding the four hexamers together. Subunits 3A, 3B, and 5A are necessary for formation of structures of higher order than hexamers. The second system examined

was the hemocyanin of the tarantula *Eurypelma californica*. It is composed of seven different polypeptide chains named a through g. Results were similar to those obtained with *Androctonis* hemocyanin. Here subunit b and c form the inner ring and in addition the f subunit is thought to be involved in stabilizing the tetrahexameric structure.

As already mentioned above a crystal structure has been solved for the hemocyanin of *Panulirus interruptus* (Gaykema *et al.*, 1984). At a resolution of 3.2 Å the course of the polypeptide backbone within the subunit can be traced very well. Each subunit is composed of three domains of approximately 175, 230 and 250 residues each. Domains 1 and 2 are predominantly α -helical, domain 3 has extensive β -sheet structure, which is folded into a β -barrel. There are two arms extending from domain 3 which are in contact with domain 1 and 2. Domain 1 provides contacts between different subunits and domain 2 contains the binuclear copper center.

2.2. Molluscan hemocyanins

Molluscan hemocyanins are assembled from very large polypeptide chains with a molecular weight ranging from 350 to 450 kDa each (van Holde and Miller, 1982; Ellerton *et al.*, 1983; Preaux and Gielens, 1984; Herskovits, 1988). When viewed under the electron microscope these subunits are mostly composed of eight globular domains. The sole known exceptions are found in octopod hemocyanins whose subunits are composed of seven domains. It is possible to cleave whole subunits into single domains or multidomain fragments by partial proteolysis (Gielens *et al.*, 1975). Single domains are functional in oxygen binding and it appears that each domain contains one oxygen binding site. This is supported by studies of the copper content of

molluscan hemocyanins which consistently find about two copper atoms per 50 kDa .

The basic building block of a native molluscan hemocyanin has the shape of a hollow cylinder about 300 Å in diameter and 140 to 190 Å in height and is assembled from ten polypeptide chains. This structure is also termed the "60S" molecule, but the sedimentation coefficients actually range from 51S up to 65S. Native hemocyanin of this type is found in cephalopods and placophorans (Herskovits, 1988). The hemocyanin of most gastropods shows the next higher level of aggregation, in which two of these cylinders are stacked on top of each other. This form is also known as the 100S molecule. Higher levels of aggregation are reported from ophistobranch hemocyanins with reported S values of 130S and up. These hemocyanins can form long tubular structures under certain conditions (Ghiretti-Magaldi *et al.*, 1981). However it has been shown that tubular polymers can also be prepared by brief proteolytic treatment of *Helix pomatia* β_c -hemocyanin (Van Breemen *et al.*, 1975) and *Lymnea stagnalis* hemocyanin (Wood, 1977), which both have only 100S molecules in their native form. Therefore it remains to be shown whether molecules larger than 100S are actually native forms occurring in the hemolymph or possibly artefacts due to partial proteolysis during preparation.

Electronmicroscopy reveals one major structural difference between gastropod and cephalopod hemocyanins. The 60S gastropod hemocyanin "half molecule" shows a pronounced collar region at one end of the hollow cylinder when viewed from the side. This results in two different end views. Cephalopod hemocyanins don't have a collar and give identical end views. Model building supported by high resolution electron microscopy suggests that in gastropod hemocyanins the six N-terminal domains form the wall of the hollow cylinder and the two C-terminal domains the collar (Berger *et al.*, 1977).

The individual polypeptide chains are presumably arranged in parallel resulting in a ten-fold rotational symmetry. In eight domain cephalopod hemocyanins six domains form the wall of the cylinder and both the N- and C-terminal domain are folded toward the inside of the molecule (Wichertjes *et al.*, 1986) whereas in octopod hemocyanins only the C-terminal domain is folded towards the inside (Miller *et al.*, 1990). There is conflicting evidence whether the polypeptide chains are arranged parallel or antiparallel. In case of *Octopus dofleini* hemocyanin it seems certain that the polypeptide chains are arranged in an antiparallel fashion (Miller *et al.*, 1990).

Studies on the dissociation behavior of *Helix pomatia* hemocyanin led to the proposal that at least three different polypeptide chains are present in this hemocyanin (Lontie, 1983). This finding has been further substantiated by studies using partial proteolytic digestion of whole subunits. The existence of three different components named α , α' and β was confirmed. These differences in sensitivity towards proteases make it conceivable that this subunit heterogeneity is due to sequence differences rather than post-translational modifications. However there are no sequence data yet available to substantiate this hypothesis because all the sequence information to date has been obtained from the β component.

Other molluscan hemocyanins have been separated into different components (Bonaventura *et al.*, 1981; Lips *et al.*, 1981), but it is unclear whether the observed heterogeneity is due to differences in primary structure or to post-translational modifications. However other molluscan hemocyanins are apparently composed of only one polypeptide chain like *Sepia officinalis* hemocyanin (Preaux *et al.*, 1979). Immunological studies and sedimentation behavior seemed to indicate the presence of only one polypeptide chain in *Octopus dofleini* hemocyanin (Lamy *et al.*, 1986; 1987). However recent

sequencing results showed that there are apparently two different chains present.

3. Biosynthesis of hemocyanins

Although knowledge about biosynthesis is limited, some general features emerge (reviewed in detail by Preaux and Gielens, 1983). At first glance it seems that there is as much variability in the location of hemocyanin biosynthesis, as there is in the hemocyanin structure itself. An early hypothesis was that hemocyanins are synthesized in the hepatopancreas (or midgut gland), an assumption based solely on the high copper content of this tissue. Subsequent studies, however, showed that this assumption is only correct for some species.

Arthropodan hemocyanins are synthesized in a specialized cell type called hemocytes, cyanocytes or cyanoblasts, which are found accumulated in various tissues and sometimes also freeflowing in the bloodstream in the same organism. These cells are very rich in polysomes which is typical for cells involved in synthesizing large amounts of proteins. The hemocyanin is deposited intracellularly in form of a quasicrystalline array. The deposited hemocyanin is not surrounded by a membrane and is presumably released into the blood stream by holocrine secretion. In *Limulus polyphemus* cyanocytes are found accumulated behind the compound eye although the cells probably do not originate from there (Fahrenbach, 1970). *Eurypelma californicum* hemocyanin is synthesized in hemocytes lining the heart. After bleeding the animal, these cells proliferate in order to replenish the depleted hemocyanin (Kempter, 1983). In the crab *Carcinus maenas* cyanocytes were identified in the reticular connective tissue surrounding the ophtalmic artery, the gizzard and the hepatopancreas (Ghiretti-Magaldi *et al.*, 1977). Results of

these histological studies were confirmed by experiments employing *in vitro* translation of mRNA fractions and *in vitro* incorporation of radiolabeled amino acids into hemocyanin. Radiolabeled amino acids were incorporated into immuno-precipitable hemocyanin in the hepatopancreas of *Astacus leptodactylus* (Hennecke *et al.*, 1990) and *Homarus americanus* (Senkbeil and Wriston, 1981). Messenger RNA isolated from the hepatopancreas of *Astacus leptodactylus* and *Carcinus maenas* (Preaux *et al.*, 1986) as well as from hemocytes of *Eurypelma californicum* heart (Voit and Schneider, 1986) directed the synthesis of hemocyanin in rabbit reticulocyte lysate.

In contrast, cells involved in synthesis of molluscan hemocyanins are very rich in rough endoplasmatic reticulum. Hemocyanin is deposited in vacuoles and released into the blood stream by exocytosis. In gastropods hemocyanin is synthesized in specialized cells, called pore cells. They can be found at various locations within the body and often accumulate in mantle tissue (Reviewed in detail by Sminia, 1977). Pore cells are much more abundant in mantle tissue than in the foot of the freshwater snail *Lymnea stagnalis*. Consequently radiolabeled leucine was incorporated at a sixfold greater rate into hemocyanin in mantle tissue, compared with foot tissue (Wood *et al.*, 1981). Translation of the total mRNA fraction isolated from the mantle tissue yielded immuno-precipitable high MW bands on SDS-PAGE gels migrating closely to the native hemocyanin (Wood and Siggins, 1981).

Only cephalopod hemocyanins are actually synthesized in a special organ. The paired branchial glands have long been implicated to be involved in hemocyanin synthesis. Ultrastructural investigations showed that this highly vascularized organ is very rich in rough endoplasmatic reticulum. Vacuoles were found to contain material that resembled in shape molluscan hemocyanins (Dilly and Messenger, 1972; Schipp *et al.*, 1973). This material is

sometimes present in a crystallike form (Muzii, 1981). Other studies showed that radiolabelled leucine only became incorporated into hemocyanin when the branchial glands were intact (Messenger *et al.*, 1974) and mRNA isolated from these tissues directed hemocyanin synthesis in *Xenopus* oocytes (Preaux *et al.*, 1986). However, recent studies showed that not all cephalopod hemocyanins are synthesized in branchial glands. Nautiloids do not possess those and their hemocyanin is apparently synthesized in the hepatopancreas (Ruth *et al.*, 1988).

4. Hemocyanin sequences

A wealth of sequence information is now available for subunits of arthropodan hemocyanins. The first complete sequences reported were from *Eurypelma californicum* chains d and e (Schneider *et al.*, 1983, Schartau *et al.*, 1983) and *Tachypleus tridendatus* α -chain (Nemoto and Takagi in: Linzen *et al.*, 1985). The sequence of *Limulus polyphemus* component II (Nakashima *et al.*, 1986) and *Panulirus interruptus* chains a and b are also completed (Bak and Beintema, 1987; Jekel *et al.*, 1988). From the latter the sequence of chain c is 97% complete (Neutebom *et al.*, 1990). A partial sequence is also known from *Astacus leptodactylus* (Schneider *et al.*, 1986). Already in 1985 (Linzen *et al.*, 1985) enough sequences were known in order to make extensive comparisons and present hypotheses about hemocyanin evolution. All this sequence work was done by protein sequencing. Only recently have recombinant DNA methods been employed for sequence determination of arthropodan hemocyanin subunits (Voit and Schneider, 1986).

In contrast there was not very much sequence information known for molluscan hemocyanins until a few years ago. The first partial sequences were reported in 1985 (Drexel *et al.*, 1986; Takagi, 1986) and the first complete

sequence of a domain did not appear until 1987 (Drexel *et al.*, 1987). The determination of the sequence of molluscan hemocyanins provides a formidable task due to the size of the protein. The organization of the proteins in domains and the fact that subunits could be cleaved proteolytically into single domains which can be unambiguously identified by immunological methods, provides a way to attack the problem with conventional protein sequencing methods. There is one problem with this approach, however. The proteolytic cleavage between domains occurs in the linker region connecting two domains. It is conceivable that a small peptide, a part of the linker, could be missed during purification and its sequence would remain unknown. It would be difficult to prove that there is no missing sequence after determining the sequence of two adjacent domains. This problem can be overcome by employing recombinant DNA methods and sequencing the cDNA rather than the protein itself.

More and more sequences for domains of molluscan hemocyanins are now becoming available. The sequences of domains *Ode*, *Odf*, and *Odg* from *Octopus dofleini* (this Thesis) and *Helix pomatia* β_c -hemocyanin functional unit *d* (*Hpd*) (Drexel *et al.*, 1987) are now known completely. Furthermore the sequence of *Sepia officinalis* functional unit *h* (*Soh*) (Declerc *et al.*, 1990) and *Helix pomatia* functional unit *g* (*Hpg*) (Xin *et al.*, 1990) are almost complete. The sequence of *Sepia officinalis* functional unit *f* (*Sof*) (Ton *et al.*, 1990) and *g* (*Sog*) (Declerc *et al.*, 1990) are partially known.

5. Active site structure

The structure of the active site of hemocyanins has been investigated intensely by spectroscopic methods and chemical modifications (reviewed in detail by Solomon, 1981, and Ellerton *et al.*, 1983). Judging from these

investigations, the active site appears to be similar in hemocyanins from both phyla despite their big structural differences. However there are a few differences between both types: (i) molluscan, but not arthropodan, hemocyanins show a catalase like activity towards hydrogenperoxide (Felsenfeld and Printz, 1959); (ii) molluscan hemocyanins also exhibit a low intrinsic tyrosinase activity (Winkler *et al.*, 1981); and (iii) the kinetics of copper removal from the protein is entirely different between both types of hemocyanins (Himmelwright *et al.*, 1978). Extensive similarities in spectroscopic properties were also noted between hemocyanins and tyrosinases.

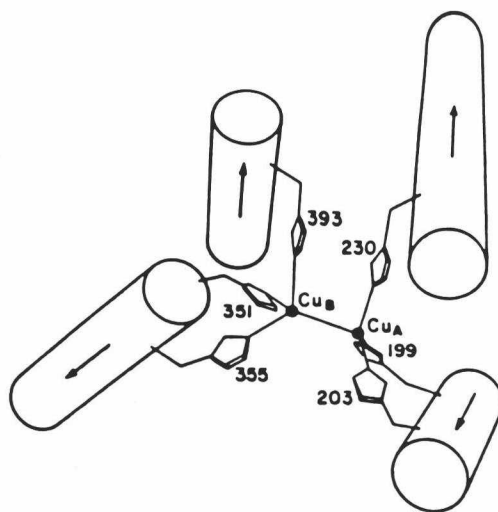
The active site of hemocyanins contains a binuclear copper center, which is also found in tyrosinases and the multi copper oxidases laccase, ascorbate oxidase and caeruloplasmin. Chemical evidence (reviewed in Ellerton *et al.*, 1983) suggests that both copper atoms are coordinated by histidine sidechains and suggests at least two nitrogen ligands per copper atom. The absence of an absorption spectrum in the near UV and in the visible suggests that oxidation state of the copper in deoxyhemocyanin is +I. This is consistent with the lack of EPR signals. Extended X-ray Absorption Fine Structure (EXAFS) studies confirmed the oxidation state of copper in deoxyhemocyanin as +I (Brown *et al.*, 1980; Co and Hodgson, 1981; Woolery *et al.*, 1984). Those studies also confirmed that each copper atom is coordinated by two imidazoles with a Cu-N distance of 1.95 Å. No other low-Z atoms were observed at a short distance which does not exclude the presence of weakly bound ligands at a longer distance. The Cu-Cu distance was determined to be approximately 3.4 Å.

Oxyhemocyanin, however, shows absorption maxima at 345 and 570 nm. Oxygen is bound to the active site as peroxide (Freedman *et al.*, 1977) and the

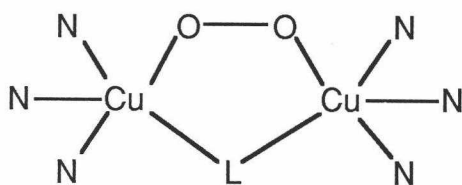
oxidation state of copper therefore changes to +II upon oxygen binding, which was confirmed by EXAFS studies (Brown *et al.*, 1980; Co and Hodgson, 1981; Woolery *et al.*, 1984). These experiments also showed that the Cu-Cu distance increases to approximately 3.55 Å upon oxygen binding. The Cu-N distance remains unchanged, but the coordination number of each copper atom is increased by two. Results from resonance Raman spectroscopy showed that oxygen is bound symmetrically and presumably forms a μ -peroxo bridge between the two copper atoms (Figure 1b) (Thamann *et al.*, 1977). Oxyhemocyanin does not give an EPR signal despite the presence of two Cu(II), which suggests a strong antiferromagnetic coupling through an endogenous ligand bridge in addition to the μ -peroxo bridge (Himmelwright *et al.*, 1980).

A three dimensional structure has been determined for *Panulirus interruptus* hemocyanin to a resolution of 3.2 Å (Gaykema *et al.*, 1984). The structure was solved for deoxyhemocyanin. Both copper atoms were found to be complexed by three histidine sidechains. Two of these were located on one α -helix separated from each other by three other residues, the third was furnished by a second helix running antiparallel to the first one (Figure 1a). Both helices are connected by a loop. The Cu-Cu distance was estimated to be about 3.7 Å, which is in fairly good agreement with results from EXAFS studies. There is a pseudo two-fold symmetry between the two copper binding sites - termed "Copper A" and "Copper B" - which are very similar in structure (Volbeda and Hol, 1989). There is no evidence for the existence of a bridging ligand at the current level of resolution.

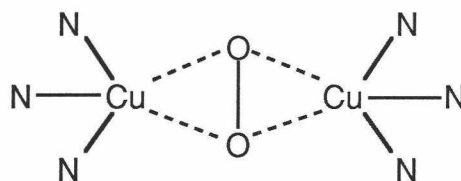
The sequences around both copper binding sites are highly conserved in all arthropodan hemocyanins sequenced so far (Linzen *et al.*, 1985). However only the "Copper B" site is conserved between molluscan and arthropodan



(a)



(b)



(c)

Figure 1: Structure of the active site in hemocyanins: a) Structure of the active site in *Panulirus interruptus* hemocyanin as determined by X-ray diffraction (from Gaykema *et al.*, 1984), b) Proposed μ -peroxo bridge in oxyhemocyanin after Freedman *et al.* (1977), c) Structure of the active site as proposed by Kitajima *et al.* (1989).

hemocyanins (Drexel *et al.*, 1987). It is also conserved in tyrosinases. This indicates that despite the strong spectroscopic similarities between hemocyanins and tyrosinases, the structure of the "Copper A" site appears to be completely different from the "Copper B" site in molluscan hemocyanins and in tyrosinases.

A great deal of effort has been invested in synthesizing copper complexes that mimic the active site of hemocyanins both in structure and spectroscopic properties (reviewed by Sorrel, 1989). Recently the synthesis of a μ -1,2 peroxo copper(II) complex was reported, but X-ray diffraction studies of this complex report a Cu-Cu distance of 4.359 Å (Jacobson *et al.*, 1988). This is not in agreement with EXAFS studies of oxyhemocyanin which report a distance of 3.55 Å (see above). Also the absorption spectra of oxyhemocyanin and this μ -1,2 peroxo copper(II) complex are significantly different. Shortly thereafter, the synthesis of a μ -peroxo complex whose physicochemical properties were remarkably similar to oxyhemocyanin was reported. Crystallographic studies revealed that oxygen is bound in form of μ - η^2 : η^2 -peroxo binuclear complex (Figure 1c) (Kitajima *et al.*, 1989). Moreover the crystal structure of this model complex gives a Cu-Cu distance of 3.560 Å, which is in good agreement with the results of EXAFS studies of oxyhemocyanin. It appears therefore that oxygen is bound in oxyhemocyanin in form of a μ - η^2 : η^2 peroxo complex rather than a μ -1,2 peroxo bridge. This model is also in agreement with the following observations: (i) the oxygen is bound symmetrically between both copper atoms, (ii) the coordination number of each copper atom increases by two upon oxygen binding, and (iii) the strong antiferromagnetic coupling between both Cu(II) atoms in oxyhemocyanin. It is no longer necessary to postulate the presence of an additional endogenous ligand bridge to account for the lack of EPR signals in oxyhemocyanin.

6. Structure of *Octopus dofleini* hemocyanin

The hemocyanin of the giant Pacific octopus (*Octopus dofleini*) is a 51S molecule composed of ten 11.1S subunits (Miller and van Holde, 1982). The whole molecule can be dissociated into its subunits and completely reassociated in the presence of sufficient divalent cations (Ca^{++} , Mg^{++}). The subunits of the dissociated hemocyanin appeared to be homogeneous in molecular weight as shown by SDS gel electrophoresis and sedimentation equilibrium. This and the complete homogeneity of sedimentation behaviour as calculated by the integral distribution of sedimentation coefficients show that there is apparently only one type of polypeptide chain with a molecular weight of 350 kDa. Crossed immunoelectrophoresis of the dissociated molecule shows only one homogeneous precipitation peak (Lamy *et al.*, 1986) and protein sequencing of the whole subunit shows only one N-terminal sequence (Lamy *et al.*, 1987). Electron microscopy reveals seven globular domains which are immunologically distinct (Lamy *et al.*, 1986; Lamy *et al.*, 1987). The whole subunit can be cleaved into 50 kDa domains by partial proteolytic digestion. These domains were arbitrarily numbered in their order of identification. The order of the domains within the whole subunit was determined and found to be Od7 - Od4 - Od3 - Od6 - (Od5 - Od2) - Od1 with some uncertainty regarding the order of domains Od2 and Od5 (Lamy *et al.*, 1987). One domain (Od1) can be cleaved off very easily upon tryptic digestion. This was shown to be the C-terminal domain. The N-terminal sequences of domain Od1, Od2 and the domain Od(2-5) fragment could be determined and are shown in Table 1 (Lamy *et al.*, 1987).

The nomenclature originally used for the domains of *Octopus dofleini* hemocyanin is different from the standard nomenclature for molluscan hemocyanins, which uses italic letters starting with *a* for the N-terminal

domain, *b* for the second domain, *h* for the C-terminal domain and so on. We therefore have renamed the domains *a* through *g* in order to comply with the standard nomenclature, using the prefix *Od* for *Octopus dofleini*. The N-terminal domain Od7 (according to the old nomenclature) is now domain *Oda*, the C-terminal domain Od1 is now domain *Odg* and so on. It should be noted however that the C-terminal domain in most other molluscan hemocyanins, which have eight, carries the letter *h*.

In the following chapters the three letter abbreviations for domains or functional units of molluscan hemocyanins will be used throughout for simplicity of discussion.

Table 1: N-terminal sequences of individual domains of *Octopus dofleini* hemocyanin (from Lamy *et al.*, 1987).

Domain	Sequence
Od1 (<i>Odg</i>)	TVGDAIIRKVNSLTPSDIKELRDAMA
Od2 (<i>Odf</i>)	XPPSNEDADIDTPLNHIRRN
Od(2-5) (<i>Od(e-f)</i>)	a: (SEEGNQ/EYLVMA) b: (A/GXSAPSLGRK)
Od7 (<i>Oda</i>)	NLIKNVDA L(D) (E)

Amino acid residues in brackets are of uncertain identity, a slash indicates two possibilities for one position.

B. MATERIALS AND METHODS

1. Materials

1.1 Chemicals

Unless noted otherwise, laboratory chemicals were purchased from Mallinckrodt, Sigma or Research Organics.

Acrylamide	Biorad
Bisacrylamide	Biorad
Ultrogel A2	LKB
SDS	BDH

1.2. Reagents for molecular biology

Restriction enzymes	New England Biolabs, BRL, Pharmacia, IBI, Promega
Bacterial alkaline phosphatase	IBI
T4 polynucleotide kinase	BRL, New England Biolabs
DNA polymerase I, large fragment (Klenow)	New England Biolabs, Stratagene
Terminal deoxynucleotidyl transferase	IBI, Pharmacia
AMV reverse transcriptase	Boehringer, Life Sciences
DNA polymerase I	Boehringer
Ribonuclease H (RNaseH)	Boehringer
<i>E. coli</i> DNA ligase	Boehringer
Mungbean nuclease	Stratagene, Boehringer, Pharmacia
Exonuclease III	Pharmacia, BRL, New England Biolabs

T4 DNA polymerase	BRL, New England Biolabs
T4 DNA ligase	BRL, Promega
RNasin	Promega
λ DNA markers	New England Biolabs
M13/pUC sequencing primers	New England Biolabs, Stratagene
Brewers yeast tRNA	Boehringer
5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside (X-Gal)	Research Organics
Isopropyl- β -D-galactopyranoside (IPTG)	Research Organics
Ampicillin	Sigma
Tetracyclin	Sigma
Chloramphenicol	Sigma
Yeast extract	Difco
Tryptone	Difco
α - 32 P dNTPs	NEN, ICN, Amersham
γ - 32 P ATP	NEN, ICN

1.3. Buffers and solutions

TE7.4	10 mM TrisHCl pH 7.4, 1 mM EDTA
TE8.0	10 mM TrisHCl pH 8.0, 1 mM EDTA
TES	10 mM TrisHCl pH 7.4, 1 mM EDTA, 1 % SDS
GTE	25 mM TrisHCl pH 8.0, 50 mM glucose, 10 mM EDTA
STE	10 mM TrisHCl pH 8.0, 1 mM EDTA, 100 mM NaCl
1x TAE	40 mM Tris base, 40 mM acetic acid, 1 mM EDTA
1x TBE	89 mM Tris base, 89 mM boric acid, 1 mM EDTA
Medium Salt Buffer	20 mM TrisHCl pH 8.0, 10 mM MgCl ₂ , 1 mM EDTA, 1mM DTT, 50 mM NaCl

Mungbean Nuclease

Buffer	30 mM sodium acetate pH 5.0, 50 mM NaCl, 1 mM ZnCl ₂ , 5% glycerol
Denhart's	0,04% BSA, 0.04% polyvinylpyrrolidone, 0.04% Ficoll-400™

1.4. Plasmids and strains

pBR 322	Our collection
pUC18/19	Our collection
JM 103	Our collection
DH5α	BRL
XL1 blue	Stratagene

1.5. Media for growth of bacteria

Liquid media:

LB	1% Tryptone, 0.5% yeast extract, 0.5% NaCl
SOB	2% Tryptone, 0.5% Yeast extract, 10 mM MgCl ₂ , 10 mM MgSO ₄ , 2.5 mM KCl, 10 mM NaCl

Antibiotic stock solutions were prepared as described in Maniatis *et al.* (1983) and used at the following final concentrations:

ampicillin (amp)	25 µg/ml
tetracyclin (tet)	15 µg/ml
chloramphenicol (cla)	170 µg/ml

Agar plates:

For preparation of agar plates, Bacto agar was added to the medium at 1.5% (w/v) before autoclaving. If needed antibiotics were added after the agar

solution cooled to below 50°C. Plates were poured on an even surface and used immediately after solidifying or kept at room temperature overnight to dry out the condensation before storage at 4°C.

LBamp	LB + 50 µg/ml ampicillin
LBtet	LB + 15 µg/ml tetracyclin
LBcla	LB + 170 µg/ml chloramphenicol
SOBtet	SOB + 15 µg/ml tetracyclin
X-Gal plates	LBamp + 1mM IPTG and 0.004% X-Gal

X-Gal stock: 2 % (w/v) in Dimethylformamide, filter sterilized and stored in freezer

IPTG stock: 1 M in water, filter sterilized and stored in freezer

2. Methods

2.1. Routine methods

2.1.1. Sterilization

All glassware and plasticware used for manipulating DNA were sterilized by autoclaving for 15 min at 121°C. Solutions were autoclaved for 30 min. Non-autoclavable solutions were sterilized by filtration through a 0.22 µ membrane filter. Glassware used for RNA work was baked for at least 4 hours at 250°C. Water and solutions used in RNA work were treated with Diethylpyrocarbonate as described in Maniatis *et al.* (1983) or Berger and Kimmel (1987).

2.1.2. Extraction with organic solvents

Crystalline redistilled phenol was equilibrated with buffer as described in Maniatis *et al.* (1983) and stored at 4°C. A 24:1 mixture (v/v) of Chloroform and Isoamylalcohol was prepared and stored in a brown bottle at room temperature. Phenol and chloroform prepared in this way were used for extracting DNA and RNA solutions.

For extraction with phenol the sample was mixed with an equal volume of phenol and vortexed for 15 to 30 sec. The phases were separated by centrifugation for 5 min in an appropriate centrifuge. The aqueous phase (usually the top phase) was then transferred to a fresh tube. Care was taken not to remove any of the interface. For extraction with phenol/chloroform the sample was first mixed with 0.5 volumes phenol and vortexed for 15 to 30 sec, 0.5 volumes chloroform was then added and the sample vortexed again for 15 to 30 sec. The phases were separated by centrifugation for 5 min in an appropriate centrifuge and the aqueous phase was transferred to a fresh tube without disturbing the interphase. For extraction with chloroform the sample was mixed with an equal volume of chloroform, vortexed for 15 to 30 sec, the phases separated by centrifugation for 1 min and the supernatant transferred to a fresh tube without disturbing the interface.

2.1.3. Ethanol precipitation

Nucleic acids were concentrated by precipitation with ethanol. The sample was adjusted to either 300 mM sodium acetate pH 5.2, 250 mM NaCl, or 2.5 M ammonium acetate. DNA was precipitated by addition of 2 volumes of ice cold 100% ethanol. For low amounts of DNA 2.5 volumes ethanol were added. DNA fragments of less than 200 bp were precipitated by adjusting the sample to 10 mM MgCl₂. RNA was precipitated with 2.5 volumes 100% ethanol. After

addition of ethanol the samples were stored for about one hour at -20°C and the precipitates were collected by centrifugation for 15 min in a microfuge or at 12,000 g in an appropriate centrifuge.

2.1.4. Enzymatic manipulations of DNA

Digestion with restriction enzymes:

In a standard digest, 1 - 2 μg DNA were digested with 2 U of enzyme in the buffer recommended by the manufacturer in a total volume of 20 μl for 2 hours at 37°C . For digests of large amounts of DNA the final volume was adjusted so that the final concentration of DNA did not exceed 0.2 $\mu\text{g}/\mu\text{l}$ and the final glycerol concentration did not exceed 5% (v/v). Digests were then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol or loaded onto a gel directly for analysis.

5'-Dephosphorylation:

For labeling purposes, 5' overhangs were dephosphorylated with 0.01 U/pmole 5'-ends Bacterial Alkaline Phosphatase (BAP) for 1 hour at 65°C . For cloning purposes, 0.1 U/pmole ends were used. For dephosphorylation of blunt ends twice as much BAP was used and incubation time was increased to 90 min. All reactions were carried out in 50 mM Tris8.0/50mM NaCl as recommended by the manufacturer (IBI). BAP was inactivated by addition of Proteinase K to 100 $\mu\text{g}/\text{ml}$ and incubation for 30 min at 37°C . The reaction mixture was then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol.

Fill In of 3'-recessed ends:

This reaction was normally done during a standard restriction enzyme digest by adding 2 U Klenow and 1 μl 10 mM dNTPs. Otherwise 1 - 2 μg digested DNA were filled in with 2 U Klenow in Medium Salt Buffer (see

Section B.1.3. for composition) containing 0.5 mM dNTPs in a final reaction volume of 20 μ l. After incubation for 30 min, the reaction mixture was extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol.

Removal of single-stranded extensions:

DNA (1 - 2 μ g) with single stranded extensions (3' or 5' overhangs) were treated with 2 - 3 U freshly diluted Mungbean Nuclease in 20 μ l Mungbean Nuclease Buffer (see Section B.1.3. for composition) for 30 min at 30°C. The reaction was then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol.

2.1.5. Gel electrophoresis

Agarose gel electrophoresis

DNA fragments were separated mostly in flatbed agarose gels with dimensions of 20 x 25 x 0.4 cm (BRL model H4, requires 200 ml agarose) and 6.5 x 10 x 0.4 cm (LKB minigel apparatus, requires 35 ml agarose) prepared in 1x TAE. Percentages between 1% and 2% were used depending on the size of DNA fragments to be separated (see Maniatis *et al.*, 1983). The agarose was boiled in 1x TAE until it was completely dissolved and cooled to below 50°C before pouring in order to avoid distortion of gel moulds. For analysis of radiolabeled DNA mini slab gels (IDEA Scientific, size 10 x 15 x 0.1 cm) were used. To pour mini slab gels the glass plates of the gel mould were prewarmed to 50°C and the agarose solution was poured while still hot. Gels were run at Voltage gradients between 1 and 15 V/cm, depending on the desired resolution or quickness of separation. After running, the gels were stained with ethidium bromide and the bands visualized on a UV Transilluminator operating at a wavelength of 300 nm.

RNA gel electrophoresis and Northern Blotting:

RNA was electrophoresed on 1.1% agarose gels containing 2.2 M formaldehyde as described in Maniatis *et al.* (1983). After electrophoresis gels were soaked in six changes of water for one hour to remove formaldehyde, either stained with ethidium bromide and destained for three hours to overnight, or transferred to BA 85 nitrocellulose (Schleicher and Schuell) without staining (Thomas, 1980).

Blots were prehybridized overnight in 50% formamide/5x SSC/1x Denhardtts/250 ug/ml heat denatured sonicated salmon sperm DNA/50mM sodium phosphate pH 6.5, and hybridized overnight in the same mixture including 10% dextran sulfate and 2×10^6 cpm probe. After hybridization the blots were washed twice for 5 min in 2x SSC/0.1% SDS at room temperature and twice for 15 min at 55°C. Washing at higher stringency was done if necessary. Blots were exposed overnight at -80°C using intensifying screens.

Alkaline agarose gel electrophoresis

Alkaline agarose gels were used for analysis of first strand and second strand products during cDNA synthesis. Gel analyses were done according to McDonnell *et al.* (1977), as described in Maniatis *et al.* (1983), except they were prepared as slab gels (BRL model V16, dimensions: 16 x 18 x 0.15 cm, volume: 50 ml). The gel mould was assembled without the bottom spacer, sealed on three sides by taping and prewarmed to about 40°C. The agarose was melted in 40 ml 50 mM Na Cl/1 mM EDTA. After cooling to about 40°C 10 ml 150 mM NaOH/50 mM NaCl/1 mM EDTA was added, the final solution mixed rapidly and poured into the gel mould as quickly as possible before solidifying. The samples were prepared and loaded as described in Maniatis *et al.* (1983) and the gel was run at 80V until the marker dye (bromocresol green) had migrated

halfway down. The gel was then fixed in icecold 7% trichloroacetic acid, dried onto nitrocellulose and exposed to film.

2.1.6. Transformation of *E.coli*

Competent cells were prepared according to protocol 2 for frozen storage described by Hanahan (1983, 1985). In brief: cells were streaked on SOB plates (SOBtet for strain XL1blue) and grown for 16 - 20 hours at 37°C (colonies reached diameters of 2 - 3 mm). Twenty colonies were picked with a sterile loop and dispersed in 1 ml SOB. The cell suspension was used to inoculate 200 ml SOB in a 2.8 l Fernbach flask. Cells were grown to an OD₆₀₀ of 0.42 in a 37°C shaker at 300 rpm. The culture was collected into four 50 ml Falcon 2070 tubes and chilled on ice for 15 min. Cells were pelleted at 2500 rpm for 10 min in a Beckmann J6B centrifuge. The supernatant was removed completely and cells were gently resuspended in 12.5 ml FSB (100 mM KCl/45 mM MnCl₂/10 mM CaCl₂/3 mM hexamine cobalt trichloride/10 mM potassium acetate/10% (v/v) glycerol, pH 6.4) and incubated on ice for 15 min. Cells were then pelleted again at 2500 rpm for 10 min. The supernatant was removed completely and cells were gently resuspended in 4 ml FSB and 150 µl DMSO were added, mixed by swirling and the suspension incubated for 5 min on ice. Then another 150 µl of DMSO were added and incubation was continued for 15 min. The cells were then pipetted in 200 µl aliquots into chilled microfuge tubes and then stored at -80°C until use.

For transformation tubes were removed from the freezer as needed and thawed on ice. The DNA solution was added in a volume of up to 20 µl if possible and mixed well with the cell suspension by tapping the tube at the bottom. The tube was incubated on ice for 45 - 60 min. The cells were then heatshocked for 90 sec at 42°C and immediately chilled on ice for 90 sec. Then

800 μ l of prewarmed SOB was added and 100 - 200 μ l of cells plated on appropriate plates. The plates were incubated overnight until colonies appeared.

2.1.7. Radiolabeling of DNA

DNA dephosphorylated on the 5' end (1 - 20 pmoles of ends) was labeled in a final volume of 10 μ l with 10 U T4 polynucleotide kinase and 150 μ Ci γ - 32 P ATP (≥ 7000 Ci/mmmole) in 50 mM TrisHCl pH7.6/10 mM MgCl₂/5 mM DTT/0.1 mM spermidine/0.1 mM EDTA for 30 min at 37°C. In order to inactivate the kinase the reaction was either incubated for 10 min at 65°C or extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol, depending on the planned use of the labelled DNA. DNA with 3'-recessed ends was labeled in a fill in reaction as described in section B.2.1.3., except one or several of the unlabeled dNTPs were substituted with 50 μ Ci of the appropriate α - 32 P dNTP.

Random primed labelled probes (Feinberg and Vogelstein, 1983) were prepared using commercially available kits from Boehringer or US Biochemical. Labelling reactions were performed according to the protocol supplied by the manufacturer.

2.1.8. Oligonucleotides

All oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer at the Center for Gene Research and Biotechnology at OSU. All oligonucleotides were gel purified by electrophoresis on denaturing 20% acrylamide gels. The bands were located by UV shadowing, cut out, and the DNA recovered from the gel by HAP recovery (Tabak and Flavell, 1978, see section B.2.6.). The sequences of all oligonucleotides are shown in Table 2. The

Table 2: Sequences of oligonucleotides WL2, WL3 and WL4

Name	Sequence
WL2	5' GT CCA RTC CCA RTA RTC 3'
WL3	5' CAA CTG AGG GGA ATT CCA TG 3'
WL4	5' AAT CCA TGG TCG CTT TC 3'

sequence of WL2 was derived from a portion of the known amino acid sequence of the C-terminal domain of *Paroctopus dofleini dofleini* (Takagi, 1986). WL3 was synthesized complementary to a sequence portion near the 5'-terminal end of pHC1 (underlined in Figure 4), WL4 to a sequence near the 5'-terminal end of pHC2 (underlined in Figure 7).

2.2. Purification of plasmid DNA

2.2.1. Small scale purification

Initially, small scale purification of plasmid DNA was performed according to the alkaline lysis method of Birnboim and Doly (1979) as described in Maniatis *et al.* (1983). A 5 ml culture containing the appropriate antibiotic was inoculated with a single colony from a freshly streaked plate and incubated overnight in a 37°C shaker. A 1.5 ml aliquot of the culture was transferred to a microfuge tube and the cells pelleted by centrifugation, resuspended in 100 µl GTE and incubated for 5 min at room temperature. Then, 200 µl of freshly prepared 0.2 M NaOH/1% SDS was added, the contents of the tube mixed by inversion and incubated on ice for 5 min. Then 150 µl 3M potassium acetate pH 4.8 (prepared as described in Maniatis *et al.*, 1983) was added, the tube vortexed upside down for 15 sec, incubated for 5 min on ice and centrifuged for 5 min in the microfuge. The supernatant was transferred to a fresh tube, extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol at -20°C for one hour. The tube was spun for 15 min in the microfuge, the supernatant discarded and the pellet resuspended in 50 µl of TE8.0. Five to ten µl were used for digestion with restriction enzymes in presence of 100µg/ml RNase A.

Plasmid DNA prepared in this manner was satisfactory for most purposes but could not be digested with some restriction enzymes that were

sensitive to contaminants. It was also not clean enough to be used for sequencing. In order to obtain plasmid DNA suitable for sequencing I included the following modifications in the initial protocol: After the potassium acetate precipitation, the supernatant was first extracted once with phenol before extraction with phenol/chloroform. After ethanol precipitation, the pellet was resuspended in 100 μ l TE8.0, RNaseA was added to 100 μ g/ml, and the sample was incubated at 37°C. After two hours the sample was adjusted to 2 M ammonium acetate by addition of 25 μ l 10 M ammonium acetate, Proteinase K was added to 80 μ g/ml, and incubation was continued for two hours at 37°C. The sample was then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol.

Later I simplified this somewhat tedious procedure greatly by adaptation of the large scale protocol as follows: After centrifugation the cell pellet of the entire 5 ml overnight culture was resuspended in 200 μ l GET containing 4mg/ml lysozyme. After transfer to a microfuge tube and incubation at room temperature for 5 min, 400 μ l 0.2 M NaOH/1% SDS were added and the contents of the tube mixed by flicking. After 5 min on ice 200 μ l 7 M ammonium acetate/ 3 M potassium acetate were added and the tube was vortexed upside down for 5 sec. Incubation was continued for 10 min on ice. The sample was then centrifuged for 5 min in a microfuge, the supernatant transferred to a fresh tube and nucleic acids were precipitated by addition of 0.6 volumes isopropanol followed by incubation for 10 min at room temperature. After centrifugation for 5 min, the pellet was resuspended in 100 μ l TE8.0. RNase A was added to 100 μ g/ml and 5 μ l were used for restriction analysis. The remainder was incubated for 30 min at 37°C, then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol.

Plasmid DNA prepared in this manner was equal in quality to CsCl-purified DNA for sequencing.

2.2.2. Large scale purification

The volumes given in the different procedures for large scale purification of plasmid DNA were used for processing of 500-ml chloramphenicol-amplified cultures or 250-ml cultures grown to saturation. If different volumes were processed the procedure was scaled up or down accordingly.

Cell growth:

Five milliliters medium with appropriate antibiotic in a Falcon 2006 tube was inoculated with a single colony from a plate which was not older than a week, and grown overnight. On the next day the main culture which contained no antibiotic was inoculated with the overnight culture and grown to an OD₆₀₀ of approximately 0.6. At this point, chloramphenicol was added to a final concentration of 15 µg/ml (Frenkel and Bremer, 1986) and incubation was continued overnight. The next morning the cells were harvested by centrifugation at 4200 rpm for 20 min in a Beckmann J-6B centrifuge.

Alternatively, one or two colonies from a plate were dispersed in 1 ml medium and added to the main culture which contained the appropriate antibiotic. Cells were grown to saturation for 18 to 20 hours, harvested and processed.

Standard protocol:

Following the standard procedure a crude lysate was prepared using the alkaline lysis method (Birnboim and Doly, 1982). Linear and nicked DNA as well as the bulk of the contaminating RNA was removed by CsCl banding (Clewell and Helinski, 1969) and in a final step the supercoiled plasmid DNA

was purified from the remaining contaminating low molecular weight RNA by column chromatography on Ultrogel A2 (LKB) or an equivalent gel filtration matrix (Maniatis et al., 1983) as follows.

The cell pellet was resuspended in 7 ml GTE (see Section B.1.3.). After transfer to a 50-ml polysulfone tube, 1 ml freshly prepared GTE containing 16 mg/ml Lysozyme was added. The suspension was mixed and incubated at room temperature for 10 to 15 min in order to break the cell walls. Then 16 ml of 0.2M NaOH/1% SDS was added, the solution vortexed in order to shear bacterial chromosomal DNA, and kept on ice for 10 min. Then 12 ml of 3 M potassium acetate pH 4.8 were added in order to precipitate denatured protein and bacterial chromosomal DNA. After mixing, the solution was kept on ice for 10 min and then centrifuged for 15 min in a Sorvall SS 34 rotor at 10,000 rpm. The supernatant was then transferred to two fresh polysulfone centrifuge tubes and nucleic acids were precipitated with ethanol.

After centrifugation the pellet was resuspended in TE7.4 to a final volume of 8.2 ml. Then 8.8 g solid CsCl were added and dissolved completely. After addition of 0.82 ml 10 mg/ml ethidium bromide, the density of the solution was checked. It should be approximately 1.54 g/ml. Two 5.1 ml Quickseal tubes (Beckman) were filled with the solution, loaded into a VTi 65.2 rotor and spun at 45,000 rpm for 17 hrs. The plasmid bands were located by shining UV light from the side and removed with a syringe by puncturing a needle through the wall of the tube underneath the band. The ethidium bromide was removed by multiple extractions with CsCl saturated isopropanol until the organic phase did not show any more traces of pink color. Three volumes TE7.4 were added to the aqueous phase to dilute the CsCl and the plasmid DNA was precipitated with ethanol.

After centrifugation the pellet was resuspended in 0.5 ml TES, loaded on a Ultrogel A2 column (1 x 20 cm) and size fractionated to remove the remaining small molecular weight RNA. One milliliter fractions were collected and 10 µl of each fraction used for A₂₆₀ determinations. The leading peak contained the plasmid DNA. Those fractions were pooled and the DNA precipitated with ethanol.

Alternative method:

This procedure described by Micard *et al.* (1985) avoids the use of the time consuming CsCl banding. After alkaline lysis the crude lysate was extracted with acid phenol (Zasloff *et al.*, 1978) to remove residual proteins and chromosomal and open circular DNA. High molecular weight RNA is precipitated out of 2.5 M ammonium acetate and small molecular weight RNA is removed by gel filtration as described in the standard method.

Alkaline lysis was carried out as described in the standard method. After ethanol precipitation the pellet was resuspended in 5 ml 50 mM sodium acetate pH 4.0/75 mM NaCl and transferred to a Falcon 2059 tube. The solution was extracted with 0.5 volumes of acid phenol (freshly prepared before use by equilibrating with 50 mM sodium acetate pH 4.0 for at least three times until the pH of aqueous phase was < 4.1) by inversion of the tube several times. Then 0.5 volumes chloroform was added and the solution again extracted by inverting the tube several times. The phases were separated by centrifugation in a tabletop centrifuge for 5 min at 2600 rpm. The aqueous phase was transferred to a fresh tube and extracted once more with acid phenol and chloroform. The top phase was transferred to a 30-ml Corex tube and nucleic acids were precipitated with ethanol.

After centrifugation the pellet was resuspended in 5 ml TE8.0. Solid ammonium acetate (0.96 g) was added (2.5 M final concentration), dissolved,

and the solution was incubated on ice for 20 min to precipitate high molecular weight RNA. After centrifugation in a Sorvall SS34 or HB4 rotor for 15 min at 8500 rpm, the supernatant was transferred to a fresh tube and nucleic acids precipitated with ethanol. The low molecular weight RNA was then removed from the plasmid DNA by gel filtration as described in the standard method.

Modified method:

In order to streamline the purification of plasmid DNA, I introduced a number of modifications into the alkaline lysis procedure. The cell pellet was resuspended in 6 ml GTE. After addition of 1 ml GTE containing 14 mg/ml lysozyme and mixing, the suspension was incubated for 15 min at room temperature. Then 14 ml 0.2 M NaOH/1%SDS was added and the solution vortexed and kept on ice for 10 min. Then 7 ml 7 M ammonium acetate/3M potassium acetate was added and mixed. By adjusting the lysate to a final concentration of 1.75 M ammonium acetate/0.75 M potassium acetate, all high molecular weight RNA was precipitated in addition to neutralizing the pH and precipitating denatured protein and bacterial chromosomal DNA. After 20 min incubation on ice the tube was centrifuged in a Sorvall HB4 rotor at 10,000 rpm for 15 min and the supernatant transferred to a fresh 50-ml polysulfone tube. Nucleic acids were precipitated at room temperature for 10 min by addition of 0.6 volumes isopropanol. This has the advantage of precipitating all DNA but only part of the low molecular weight RNA. After centrifugation for 15 min at 10,000 rpm the precipitate was resuspended in TE7.4 for further purification by CsCl banding. Due to the greatly reduced amounts of RNA present only half the volumes described in the standard method are needed for CsCl banding and after centrifugation usually all RNA was removed and no further purification by gel filtration was needed. Alternatively the pellet was resuspended in 5 ml 50 mM sodium acetate pH 4.0/75 mM NaCl for acid phenol

extraction (see alternative method) for removal of residual protein and subsequent purification by gel filtration. The purification by gel filtration was easier due to the greatly reduced amount of low molecular weight RNA, and better separation of plasmid DNA from RNA could be achieved.

2.3. RNA purification

Purification of total RNA:

Total RNA was purified according to the method of Chirgwin *et al.* (1979) from *Octopus dofleini* branchial gland, which is the site of hemocyanin biosynthesis in most cephalopods (Preaux and Gielens, 1984). A 50-pound animal was obtained through Professor Art Martin from the University of Washington. The animal was anesthetized with ethanol and bled to obtain hemolymph for hemocyanin preparation. Then the branchial glands were removed (30 g total weight of tissue), cut in small pieces and homogenized immediately in a total of 150 ml 4 M guanidinium thiocyanate/20 mM sodium citrate pH 7.0/0.5% sarkosyl. One gram of CsCl was added per ml of homogenate, dissolved and layered on top of a 8 ml 5.7 M CsCl cushion in a 38.5-ml polyallomer ultracentrifuge tube. The tubes were spun in a Beckmann SW28 rotor for 24 hrs at 23,000 rpm. After centrifugation the layer above the CsCl cushion was removed carefully, the top of the tube was cut off and the CsCl cushion was removed. The gel-like RNA pellet was resuspended in TES (see Section B.1.3.). The RNA was then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol. After centrifugation the pellets were resuspended in water and the yield and purity determined by A_{260}/A_{280} readings. The RNA was stored in 10-mg aliquots as an ethanol precipitate at -80°C .

PolyA RNA purification:

PolyA⁺ RNA was purified by affinity chromatography on oligo(dT)-cellulose (Aviv and Leder, 1977) as described in Maniatis *et al.* (1983). Total RNA (10 mg) was dissolved in 5ml water and heated to 65°C for 5 min. An equal volume of 2x loading buffer (1x loading buffer is 20 mM Tris7.6/0.5 M NaCl/1mM EDTA/0.1% SDS), preheated to 65°C, was added and cooled to room temperature. The RNA solution was then passed two times over an oligo(dT)-cellulose column of 1-ml bed volume, which was prepared by successive washes with 3 ml water, 3 ml 0.1M NaOH/5mMEDTA, 3 ml water and 10 ml 1x loading buffer. After application of the RNA, the column is washed with 10 ml 1x loading buffer and 3 ml loading buffer containing 0.1M NaCl. The polyA⁺ RNA was eluted with 3 ml 10 mM TrisHCl7.4/1 mM EDTA/0.1% SDS and precipitated with ethanol.

2.4. cDNA cloning

2.4.1. Homopolymeric tailing of plasmids and cDNA

Homopolymeric tailing of pUC19 was performed according to Deng and Wu (1981) with the enzyme terminal transferase (TdT), which adds deoxynucleotides to the 3'-hydroxyl group of DNA. Note that 3'-protruding ends are preferred over blunt and 3'-recessed ends, and that the substrate has to be at least a three residue oligonucleotide.

For tailing, 9 µg of pUC19 cut with *KpnI* or *PstI* (Both enzymes generate 3'-protruding ends) were incubated with 25 U TdT in CoCl₂ buffer (140mM potassium cacodylate pH6.9/1 mM CoCl₂/1mM DTT) in a final volume of 50 µl containing 1 µl 0.2 mM dCTP or 1 µl 0.2 mM dTTP for 1 hour at 37°C. The reaction was terminated by extracting once with phenol/chloroform, once with chloroform and nucleic acids were precipitated with ethanol. The appropriate

α - ^{32}P dNTP was added for monitoring the reaction. Under these conditions an average of 20 (in case of dCTP) or 40 (in case of dTTP) nucleotides were added to the 3'-end. Each preparation was analyzed by digesting the tailed plasmid with *Hae*II, which generates four fragments that were separated on a 2% agarose gel. Only the two smallest fragments should have radiolabel incorporated. The length of the tail was analyzed by digesting the tailed plasmid with *Eco*RI or *Hind*III, and separating the fragments on a sequencing gel with a sequencing ladder as a size standard.

Complementary DNA was tailed in a 20 μl reaction in CoCl_2 buffer with 25 U TdT and 1 μl 1mM dGTP for 30 min at 37°C. Addition of dGTP terminates after approximately 20 nucleotides are added to the 3' ends (Berger and Kimmel, 1987). The reaction was then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol.

2.4.2. cDNA Cloning according to Heidecker and Messing (1983)

In this procedure first strand synthesis was carried out with *Pst*I cut pUC19, tailed with dTTP, as primer. The cDNA was tailed with dGTP and size fractionated on an alkaline sucrose gradient. The separated strands were annealed with the complementary strands of pUC19 tailed with dCTP and the second strand synthesized in a fill in reaction. The cDNA library was transformed into *E.coli* JM 103. Although this method is laborious and time consuming it selects in principle for full length cDNA clones at two points: tailing of the cDNA with dGTP and size fractionation on an alkaline sucrose gradient.

First strand synthesis was carried out under the following conditions: 0.8mM dNTPs, 70 mM KCl, 50 mM Tris8.2, 10 mM MgCl_2 , 2 mM DTT, 40 U RNasin, 1 μg pUC19 tailed with dTTP, 4 μg polyA⁺ RNA and 27 U reverse

transcriptase in a final volume of 30 μ l. The reaction was incubated for 90 min at 42°C, extracted once with phenol/chloroform, once with chloroform and precipitated twice with ethanol. The cDNA was then tailed with dGTP as described above, except 1 mM CoCl₂ was substituted with 2 mM MnCl₂. Under these conditions blunt ends are preferentially tailed over 3' recessed ends (Deng and Wu, 1981). After 30 min at 37°C, the reaction was extracted once with phenol/chloroform, once with chloroform and precipitated twice with ethanol.

An alkaline sucrose gradient was prepared (5% -20% sucrose in 0.2 M NaOH/0.8 M NaCl/1 mM EDTA with a 0.5 ml 60% sucrose cushion) and the sample loaded in 50 μ l of the 5% sucrose solution. Centrifugation was carried out in SW 50.1 rotor for 17 hours at 4°C. The gradient was collected in 300 μ l fractions and the profile established by Cherenkov counting. The fractions containing the front edge of the leading peak were pooled and a 10 fold molar excess of pUC19 tailed with dCTP was added. The mixture was dialysed against TNE (10 mM Tris7.6/10 mM NaCl/1mM EDTA) in the cold overnight and concentrated by ethanol precipitation. The DNA strands were annealed in 32% (v/v) formamide/50 mM NaCl/10 mM Tris7.6 at a plasmid DNA concentration of 5 μ g/ml at 37°C for 24 hrs and dialysed against STE (see Section B.1.3.) overnight.

After concentration by ethanol precipitation the DNA was taken up in Medium Salt Buffer (see Section B.1.3.) containing 1 mM DTT and 0.1 mM dNTPs. Five units Klenow were added and the mixture incubated for 1 hour at 15°C and 1 hour at room temperature. The DNA was then extracted once with phenol/chloroform, once with chloroform and precipitated with ethanol. After resuspension in 50 μ l TNE, the DNA was transformed into *E. coli* JM103 and plated on X-Gal plates.

2.4.3. cDNA cloning according to Gubler and Hoffmann (1983)

This method is used the most for cDNA cloning due to its simplicity. First strand synthesis can be primed with oligo-dT₍₁₂₋₁₈₎, random hexanucleotide primers or specific primers (single or mixed sequence oligonucleotides). Second strand synthesis is done with DNA polymerase I (Pol I) synthesizing the second strand, RNaseH generating RNA primers for Pol I and *E.coli* DNA ligase for ligating the synthesized DNA strands together. The double stranded cDNA can be inserted into cloning vectors by homopolymeric tailing or linker/adaptor addition, which requires more enzymatic manipulations.

Synthesis of the first strand was done in 50 mM Tris8.3, 10 mM MgCl₂, 10 mM DTT, 4 mM sodium pyrophosphate, 1 mM dNTPs, 50 U RNasin, 25 U reverse transcriptase, 5 µg polyA⁺RNA and 0.25 µg oligo dT₍₁₂₋₁₈₎ (or 0.25 µg random primer, or 0.2 µg specific primer). The RNA and primer was first heated at 95°C for 5 min in a volume of 25 µl and cooled to 42°C. All other components were added with sodium pyrophosphate last to a final volume of 50 µl. Five µl were removed and added to 1 µl α-³²P labelled dCTP or dGTP used for analysis of first strand products by alkaline gel electrophoresis and quantization by incorporation assays (DE 85 filterbinding, see Maniatis *et al.*, 1983). After 45 min at 42°C EDTA was added to 20 mM and the reaction was extracted once with phenol/chloroform, once with chloroform and precipitated twice with ethanol out of 2 M ammonium acetate to remove unincorporated nucleotides.

Second strand synthesis was carried out in a volume of 100 µl of 20 mM Tris7.5, 5 mM MgCl₂, 10 mM (NH₄)₂SO₄, 100mM KCl, 0.15 mM NAD, 50 µg/ml BSA, 40 µM dNTPs with 10 µCi α-³²P labeled dGTP or dCTP, 23 U DNA polymerase I, 0.8 U RNaseH, 1 U *E. coli* DNA ligase and up to 500 ng first

strand product. If more first strand product had to be processed the reaction was scaled up accordingly. The reaction was done for 1 hour at 14°C and 1 hour at 22°C. After extracting once with phenol/chloroform and once with chloroform, the sample was size fractionated on an Ultrogel A2 column prepared in a siliconized pasteur pipette plugged with siliconized glasswool. Fractions of 200 µl were collected and assayed by Cherenkov counting. The leading peak was pooled and concentrated by ethanol precipitation or lyophilization in a Speedvac.

The double stranded cDNA was then tailed with dGTP as described above (Section B.2.4.1.) and annealed to a 3 fold molar excess of pUC19 tailed with dGTP in 10 mM Tris7.6/1 mM EDTA/100 mM NaCl at 55°C for 1.5 hours in a volume of 100 µl. The DNA was then transformed into *E. coli* DH5α or XL1 blue.

2.5. Screening of cDNA libraries

After transformation, the libraries were plated out on Millipore HATF filters laid on X-Gal plates (200 µl of cell suspension per filter) and incubated at 37°C overnight. On the next day two replicas were made from each filter according to Hanahan and Meselson (1983). The replica filters were grown for about 4 - 5 hours so the colonies could be seen on them. The replicas were then transferred to LBcl plates and incubated overnight for plasmid amplification. The next morning the replica filters were processed in three steps (Wood, 1985). The filters were first placed in a tray with Whatman 3MM paper soaked with 0.5 M NaOH for 5 min to lyse the cells. Then they were placed in a second tray with Whatman3MM soaked with 1 M Tris8.0 to neutralize the NaOH. After 5 min they were placed on a third tray with Whatman 3MM paper soaked with 1 M Tris8.0/1.5 M NaCl for 15 min to bind DNA to the filter. The processed

filters were dried under a heat lamp and baked in a 80°C vacuum oven for 2 hours. The cell debris was removed by washing in 3x SSC/0.1% SDS at 65°C. The liquid was changed several times with the last wash going overnight. The filters were sealed in plastic bags for hybridization.

Prehybridization was done in 6x SSC/1x Denhardt's/0.5% SDS/100µg/ml heat denaturated sonicated salmon sperm DNA/0.05% sodium pyrophosphate for 4 hours to overnight at the required hybridization temperature. Hybridization was done in 6xSSC/1x Denhardt's/0.5% SDS/0.05% sodium pyrophosphate/20 µg/ml tRNA/1 ng/ml labelled oligonucleotide probe. The hybridization temperatures used was for probe WL2: 37°C, WL3: 55°C, WL4: 47°C. Hybridization was done overnight in a shaker bath under shaking so bubbles wouldn't interfere with hybridization. Filters were dipped briefly in 6xSSC/0.05% sodium pyrophosphate, then washed in the same solution once for 10 min at room temperature and once for 1 hour at the respective hybridization temperature. The filters were blotted dry and mounted for autoradiography. Exposure was done at room temperature for 4 hours to overnight or at -80°C overnight with intensifying screens.

Positive clones were located on the master filters and all colonies in the area containing them were picked and stabbed on a fresh plate in a grid pattern for rescreening. Colony purification was done, if needed. Rescreening was done as described above. Positive clones were grown and the plasmids purified in small scale, analyzed for insert size and sequenced.

2.6. Purification and subcloning of DNA fragments

DNA was digested with restriction enzymes and fragments were separated on agarose gels. If regular grade agarose was used, fragments were purified from agarose by hydroxylapatite (HAP) recovery (Tabak and Flavell,

1978). After electrophoresis the desired bands were located by staining with ethidium bromide. A small well was cut with a scalpel in front of the band and filled with HAP which was equilibrated with 1x TAE (section B.1.3.). Electrophoresis was continued until the entire band had migrated into the DNA. The HAP was then loaded on top of a Sephadex G-50 column (swollen in water) poured in a siliconized pasteur pipette plugged on the bottom with siliconized glasswool. The column was never allowed to run dry. The bound DNA was eluted from HAP with 400 μ l 1 M NaPO_4 pH 7.0, then water was added on top to separate DNA from salts. After discarding the void volume (800 μ l), the next 400 μ l were collected, which contained the DNA. The column run can easily be monitored with UV light if necessary. The DNA fragment was concentrated by ethanol precipitation or directly used for ligation into the desired cloning vector.

Ligations were performed at a final DNA concentration of approximately 5 to 10 ng/ μ l in a final reaction volume of 20 μ l. The molar ratio of vector to insert DNA was 1 : 2. Buffer conditions for ligation as recommended by the manufacturer, were used. The vector was cut with (an) appropriate restriction enzyme(s), to provide compatible ends and was dephosphorylated. For blunt end ligations 0.5 to 1 U T4 DNA ligase was used per reaction and 0.1 U for sticky end ligations. Ligations were performed at room temperature for three hours. The reaction was then stopped by incubation at 65°C for ten minutes. Half of the sample was used for transformation of *E. coli*.

Subcloning was considerably sped up by doing "in gel" ligations (Struhl, 1985). Restriction digests were separated on a low melting agarose gel (Seaplaque agarose, FMC). After briefly staining with ethidium bromide, the region containing the desired restriction fragment was precisely cut out of the

gel. The gel slice was melted at 68°C and 10 µl of the molten gel were added to 10 µl 2x ligation buffer and kept at 37°C to keep the agarose melted. After addition of ligase the reaction was incubated for four hours at room temperature. The reaction was heated at 68°C for ten minutes to remelt the agarose. Eighty µl STE was added to prevent the agarose from resolidifying, and 50 µl was used for transformation of *E. coli*.

2.7. Generation of unidirectional deletions

This procedure (Henikoff, 1984) was developed to sequence large DNA fragments by generating a set of deletions from one end so that a set of overlapping sequences spanning the entire fragment can be obtained. It takes advantage of the fact that Exonuclease III (ExoIII) cannot digest DNA from 3'-ends that are protruding at least four bases or are protected with α -thio nucleotides. Although originally developed for isolated DNA fragments this procedure can also be used on fragments cloned into plasmids. A combination of two unique restriction sites must be found which are situated between a primer site and the insert DNA. One of these sites is needed to protect the primer site and must yield either a 4 base 3'-overhang or a 5'-overhang that can be blocked with α -thio nucleotides. The second site must lie between the protecting site and the insert DNA and be either a 5'-overhang or a blunt end. It is usually no problem to find such a combination in modern polylinker plasmids like pUC19. A similar method was also described for single stranded DNA (Dale *et al.*, 1985).

Fifty µg of the clone to be sequenced are first cut to completion at the protecting site and then at the second site in a second reaction. The ExoIII reaction was done at 30°C under the following conditions: 50 mM Tris8.0, 5 mM MgCl₂, 10 mM β -mercapto ethanol, 0.4 µg/µl double digested DNA, 20U

ExoIII/ μ g DNA. The final reaction volume was 5 μ l multiplied by the number of time points desired plus an extra 10 μ l reaction volume. For example: for ten timepoints the final reaction volume is $10 \times 5 \mu\text{l} + 10 \mu\text{l} = 60 \mu\text{l}$. All components were prewarmed at 30°C before ExoIII was added to initiate the reaction. The deletion rate of ExoIII at 30°C under the given conditions is on the average 230 bp/min. Time points were usually taken at 50 sec intervalls and resulted in 150 to 200 bp deletions.

Aliquots of 5 μ l were removed at each timepoint and added to a previously prepared tube containing 5 μ l 10x Mungbean Nuclease Buffer (Section B.1.3.) and 31 μ l water. Each tube was immediately placed at 65°C to inactivate ExoIII. After 15 min all tubes were put on ice and 3 U freshly diluted Mungbean nuclease (Mungbean nuclease dilution buffer: 10 mM sodium acetate pH 5.0, 0.1 mM ZnCl_2 , 1 mM DTT, 0.1% TX100 and 50% glycerol) was added to each tube. After mixing the reactions were placed at 30°C for 30 min. Sample loading buffer (Maniatis *et al.*, 1983) was added to each reaction prior to loading onto a low melting agarose gel. Electrophoresis was done at low voltage for high resolution. After electrophoresis the gel was briefly stained with ethidium bromide and the progressively shorter bands for each timepoint were cut out precisely. The gel slices were melted at 68°C and 10 μ l was added to 10 μ l of a 2x ligation mix. Ligation and transformation was carried out as described in section B.2.6.

2.8. Sequencing

Sequencing was done according to the dideoxy method (Sanger *et al.*, 1977) using denatured supercoiled plasmid DNA as template (Chen and Seeburg, 1985). Klenow fragment (Chen and Seeburg, 1985), Reverse Transcriptase (Zagursky *et al.*, 1985) and SequenaseTM (Tabor and Richardson,

1986) were used to generate sequencing ladders. The purity of the template DNA was found to be crucial. Most problems encountered were due to impure plasmid DNA. The modified procedure for small scale preparation of plasmid DNA described in section B.2.2.1. gave satisfactory results. The clarity of the sequencing ladders generated was equal to those obtained from CsCl purified DNA. This was important since many different plasmids had to be prepared for this sequencing project and having to purify each template by CsCl banding would have been very time consuming. Mostly reverse transcriptase and Sequenase™ were used for sequencing. In my hands Sequenase™ generated the longest ladders, reverse transcriptase could better read through homopolymeric GC regions.

The template DNA was prepared by alkaline denaturation as follows: plasmid DNA (1 to 2 µg) was added to 2 µl 2M NaOH, then water was added to 10 µl. After 5 min at room temperature 1 µl primer (15 ng/µl), 3 µl 3 M potassium acetate pH 4.8 and 6 µl water was added. The primer anneals to the template under these conditions. The DNA was precipitated with 50 µl ethanol.

For sequencing with Klenow the DNA pellet was resuspended in 1.5 µl 16.7x Annealing Buffer (833 mM Tris7.5, 83 mM MgCl₂, 83mM DTT, 16.7 mM EDTA), 10.5 µl water, 2 µl α-³²P dATP (800 Ci/mmol) and 1 µl Klenow (5 U). 3µl were added to 2 µl of each of the four dideoxy mixes (purchased from Stratagene or New England Biolabs) and the reactions were incubated at 42°C for 20 min. Then 1 µl chase solution (2 mM of each dNTP) was added to each tube. After further incubation for another 20 min, 6 µl Stop Solution (98% deionized formamide, 10 mM EDTA, 0.3% (w/v) xylene cyanole, 0.3% (w/v) bromophenol blue) was added to each reaction and all tubes were heated at 95°C for 5 min to denature the DNA. From each sample 2 µl was loaded on a sequencing gel.

For sequencing with reverse transcriptase (protocol obtained from Pharmacia) the DNA pellet was resuspended in 2 μ l 10x sequencing buffer (600 mM Tris8.3, 750 mM NaCl, 75 mM MgCl₂, 5 mM DTT), 15 μ l water, 2 μ l α -³²P dATP (3000 Ci/mmol) and 20 U reverse transcriptase. Four μ l of the reaction mixture was added to 1 μ l of each of the four dideoxy mixes and incubated at 42°C for 10 min. Then 1 μ l chase solution (2 mM of each dNTP) was added to each tube and incubation continued for another 5 min. After addition of 4 μ l Stop Solution (see above) to each reaction, all tubes were heated at 95°C for 5 min and 2 μ l of each sample was loaded on a sequencing gel.

For sequencing with Sequenase™ (protocol obtained through US Biochemical) the DNA pellet was resuspended in 2 μ l 5x Sequenase™ buffer (200 mM Tris7.5, 250 mM NaCl, 100 mM MgCl₂), 8 μ l water, 0.5 μ l α -³²P dATP (3000 Ci/mmol), 2 μ l labeling mix (7.5 μ M dGTP, 7.5 μ M dCTP, 7.5 μ M dTTP), diluted 1 : 15 with water, and 2 μ l Sequenase™ (13 U/ μ l) diluted 1 : 8 with dilution buffer. The labeling reaction was incubated at room temperature for 5 min. Then 3.5 μ l of the labeling reaction was added to 2.5 μ l of each of the four prewarmed dideoxy mixes (purchased from US Biochemical) and incubated at 37°C for 5 min. After addition of 4 μ l Stop Solution (see above) to each reaction, all tubes were heated at 95°C for 5 min and 2 μ l of each sample was loaded on a sequencing gel.

If six or more templates were sequenced, I performed the reactions in 72 well Terasaki plates. Incubations were done as described above, except the denaturation step was done at 80°C for 10 min. Sequencing gels were made according to Maniatis *et al.* (1983). Mostly 7% gels of 0.4 mm thickness were used, which were prerun at constant current until they reached a temperature of about 50°C. Samples were then loaded and run into the gel at 1000 V. After the two dyes separated and formed sharp bands the current was adjusted to

running conditions. As electrophoresis buffer 1x TBE was used. Gels of 25 cm width were prerun at 40 mA and run at 32 mA, gels of 34 cm width were prerun at 65 mA and run at 55 mA, all at constant current. Under these conditions the temperature of the gelplates usually stayed at about 50°C. Otherwise the current was adjusted accordingly. After running the bromophenol blue band to the bottom, gels were fixed in 10% methanol/10% acetic acid, dried onto Whatmann 3 MM paper and autoradiographed overnight at room temperature without intensifying screens. Up to 200 bases could be read on a 43-cm long gel from a single loading and up to 350 bases could be read from a single loading on a 52-cm long gel, when a field gradient was used (Sheen and Seed, 1988). In this case the gel was prerun as usual, but after loading the samples 0.5 buffer volumes 3 M potassium acetate were added to the lower buffer chamber. The gel was set to the standard parameters and a field gradient established during electrophoresis. This resulted in compression of the bands in the lower portion of the gel and expansion in the upper portion and therefore in increased resolution.

2.9. Sequence analysis

The secondary structure of *Ode*, *Odf*, and *Odg* was analyzed according to the method of Chou and Fasman (1978), and the hydrophobicity profile of the same sequences was analyzed according to Kyte and Doolittle (1982). Both analyses were done using programs of the Intelligenetics Suite.

C. RESULTS

1. RNA purification and Northern blotting.

Total RNA and polyA⁺ RNA were purified as described in section B.2.3. from *Octopus dofleini* branchial gland. The yield was 170 mg total RNA with an A₂₆₀/A₂₈₀ ratio of 2.0. It was analyzed for intactness on a nondenaturing agarose gel. Two intense bands, corresponding to large and small ribosomal RNAs could be seen after staining with ethidium bromide. No signs of degradation could be seen. From 10 mg total RNA between 87 and 170 µg polyA⁺ RNA were obtained by affinity chromatography on oligo-(dT) cellulose.

Total RNA and polyA⁺ RNA were electrophoresed on a formaldehyde agarose gel as shown in Figure 2A. RNA markers (available from BRL) were included for size determination. After ethidium bromide staining I observed five bands in the polyA⁺ RNA lane. Their sizes were determined to be approximately 9.5, 5.0, 3.0, 2.1, and 1.1 kb. The two larger and the smallest bands were enriched by oligo-(dT) selection, and therefore have polyA tails or polyA-rich regions. The other two bands are apparently the two ribosomal RNAs since these two bands correspond in size to the two most abundant RNA species in total RNA. They were still present in considerable amounts after two cycles of oligo-(dT) selection. The large ribosomal RNA, with a size of 3.0 kb, would be the smallest of all large eukaryotic rRNAs, whose size has been determined (Huysmans and deWachter, 1988). The small rRNA with a size of 2.1 kb on the other hand would be one of the largest.

To be able to code for *Octopus dofleini* hemocyanin the mRNA must have a size of approximately 10 kb. The 9.5 kb RNA species would be of sufficient length. In order to find out whether it actually is the hemocyanin mRNA, northern blotting experiments were done. The probe was obtained from the

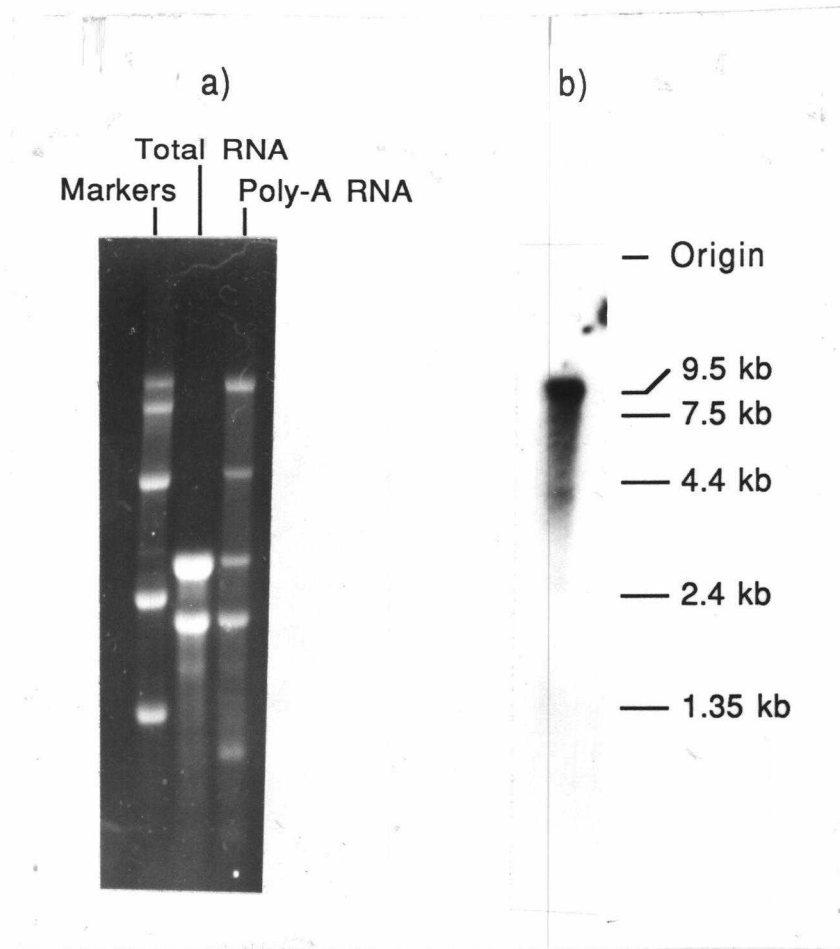


Figure 2: Gelelectrophoresis of *Octopus dofleini* RNA. (A) Ethidium bromide stained 1.1% formaldehyde agarose gel of RNA samples from *Octopus dofleini* branchial gland. Loaded from left to right: RNA markers (3 μ g), total RNA (4 μ g), polyA⁺ RNA (4 μ g). Sizes of RNA markers from top: 9.5, 7.5, 4.4, 2.4, 1.35 kb. (B) Northern blot of polyA⁺ RNA probed with labeled *Eco*RI/*Eco*RV fragment of pHC1 cDNA insert. Indicated on the right are positions of RNA markers electrophoresed on the same gel.

EcoRI/EcoRV fragment of clone pHC1 (see below) by random priming; it hybridized only to the 9.5 kb polyA⁺ RNA species (Figure 2 B). Experiments using the oligonucleotide probe WL2 (see Table 2) were not successful.

2. cDNA Cloning and Sequencing

2.1. Isolation and characterization of pHC1

A cDNA library was constructed according to the method of Heidecker and Messing (1983) as described in section B.2.4.2. This method was chosen for two reasons: first, it includes a size fractionation step (alkaline sucrose gradient) and second, during the tailing step with dGTP in presence of manganese, blunt ends are preferentially tailed and therefore full length transcripts are preferentially cloned. After transformation of the library, I obtained about 5000 clones. They were screened with probe WL2 (Table 2) as described in section B.2.5. The probe hybridized to 30 clones which were rescreened. Plasmid DNA was prepared in small scale from all positives and the size of the inserts was determined by restriction analysis. The clone with the longest insert size was called pHC1. It had an insert size of approximately 1200 bp as judged by agarose gel electrophoresis (not shown). A restriction map was determined and the results of the restriction analysis of pHC1 are shown in Figure 3. Sequencing using the reverse primer for M13/pUC gave only one open reading frame longer than 100 bp. Translation of it confirmed that this clone indeed codes for hemocyanin based on comparison with the partially known sequence of the C-terminal domain of *Paroctopus doleini dofleini* hemocyanin (Takagi, 1986) (see Figure 5). Restriction fragments were subcloned and sequenced as indicated in Figure 3 and the sequence of the entire cDNA clone was determined.

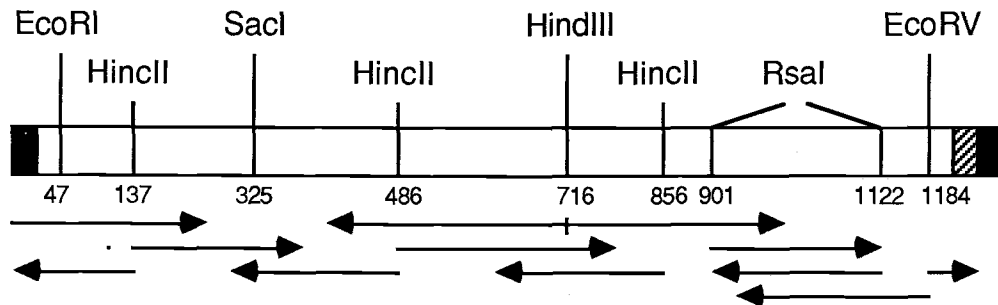


Figure 3: Restriction map of the pHCl cDNA insert. Nucleotide 1 is the first nucleotide of the cDNA insert. The black areas represent the pUC19 polylinker, the hatched area the poly-dA region. Arrows indicate direction of sequencing.

The cDNA insert of clone pHCl contains an open reading frame of 1071 bp, coding for 357 amino acids. The nucleotide sequence and its translated open reading frame are shown in Figure 4. There is a 3' untranslated region of 80 bp and a polyadenylation signal 16 bp upstream of the polydA region (double underlined in Fig. 4). The polydA region is approximately 50 bp long. It was not possible to sequence through it, so the exact size could not be determined. The presence of the polydA region as well as a stop signal in the open reading frame confirm that pHCl codes for the C-terminal domain. In fact clone pHCl coded for almost the entire C-terminal domain. However none of the N-terminal 27 amino acids determined by protein sequencing (Lamy *et al.*, 1987) are found. Judging from sequence comparison (Figure 5) with the N-terminal sequence of the C-terminal domain from *Paroctopus dofleini dofleini* determined by Takagi (1986), about 40 residues at the N-terminal end are not coded for by pHCl. I also examined the protein sequence for the presence of the peptide, the 17mer oligonucleotide probe was made for. Surprisingly the corresponding peptide was slightly different: instead of the sequence DYWDWT, PYWDWT was found (residues 95 to 100 in Figure 5). A 17 bp match was therefore not possible, but a look at the DNA sequence showed that the probe could form a 15 basepair contiguous match, which accounts for the fact that I could detect positive clones under the hybridization conditions used for screening.

Figure 4 (pages 56 and 57): Sequence of the cDNA insert of pHCl and derived protein sequence. Clone pHCl codes for most of domain *Odg*. The underlined stretches show the complementary sequence of WL2 and WL3; the double underlined stretch shows the polyadenylation signal.

1
Poly-dC...TTGCTTCCTATCATGGAATTC~~CCCCTCAGTT~~GTCATTATGAGAATGGCACAGCTTATGCATGTTGC
A S Y H G I P L S C H Y E N G T A Y A C C

66
CAACATGGAATGGTAACTTTCCCTAACTGGCATAGATTGTTGACTAAACAAATGGAAGATGCACTAGTTGCTAAG
Q H G M V T F P N W H R L L T K Q M E D A L V A K

141
GGATCTCACGTTGGTATTCC~~TACTGGGACTGGAC~~CACTACCTTCGCTAATTTACCTGTACTGGTCACTGAGGAG
G S H V G I P Y W D W T T T F A N L P V L V T E E

216
AAAGATAACTCCTTCCACCATGCCCATATTGATGTTGCCAATACGGACACAACCAGATCACCAAGAGCTCAACTT
K D N S F H H A H I D V A N T D T T R S P R A Q L

291
TTTGATGATCCAGAAAAAGGAGATAAATCATTCTTCTATCGCCAGATAGCACTTGCTCTGGAACAAACAGATTTC
F D D P E K G D K S F F Y R Q I A L A L E Q T D F

366
TGTGATTTTGAAATTCAGTTTGAAATCGGTCACAATGCTATTCATTCATGGGTTGGCGGTAGTAGCCCATATGGT
C D F E I Q F E I G H N A I H S W V G G S S P Y G

441
ATGTCAACTCTTCACTATACTTCCTATGATCCTCTCTTCTACCTTCATCACTCTAATACTGATCGTATTGTTCT
M S T L H Y T S Y D P L F Y L H H S N T D R I W S

516
GTATGGCAAGCATTACAGAAGTATCGAGGATTACCTTACAACACTGCGAACTGTGAAATCAATAAACTTGTTAAA
V W Q A L Q K Y R G L P Y N T A N C E I N K L V K

591
CCACTAAAACCATTCAACTTAGATACCAATCCTAACGCCGTTACAAAAGCCCATTCTACTGGTGCTACATCTTTC
P L K P F N L D T N P N A V T K A H S T G A T S F

666
GATTACCACAAGCTTGGTTATGATTATGATAATCTTAATTTCCATGGAATGACCATTCTGAGTTGGAAGAACAC
D Y H K L G Y D Y D N L N F H G M T I P E L E E H

741
CTTAAAGAAATACAACACGAAGACAGGGTATTTGCTGGTTTCTTACTTCGTACCATTGGTCAATCTGCTGATGTT
L K E I Q H E D R V F A G F L L R T I G Q S A D V

816
AACTTCGACGTCTGCACCAAAGATGGTGAATGTACATTTGGAGGTACGTTCTGTATTCTCGGTGGAGAACATGAA
N F D V C T K D G E C T F G G T F C I L G G E H E

891
ATGTTTTGGGCATTTGACCGTCCTTTTAAATATGATATTACCACAAGTCTGAAACATCTCCGCTTAGATGCGCAT
M F W A F D R L F K Y D I T T S L K H L R L D A H

966
GATGATTTTCGATATAAAGGTTACTATTAAAGGTATCGATGGTCATGTCTTGTTCCAATAAATACCTCAGTCCTCCG
D D F D I K V T I K G I D G H V L S N K Y L S P P

Figure 4.

1041

ACGGTTTTCTCGCTCCAGCCAAAACAACGCATTAATTATGTGTACTAATCATTTATATGATTTAATATTTGTGTT
T V F L A P A K T T H stop

1116

TTTTTGTA(TTTTCATTCAATAAATGATATCTCAAACAATAAAAAA.....

Figure 4 continued.

Figure 5 (page 59): Sequence comparison of the C-terminal domains of *Octopus dofleini* and *Paroctopus dofleini dofleini*. The sequence for *Odg* (O) derived from pHCl is compared with known sequence fragments of the C-terminal domain of *Paroctopus dofleini dofleini* (P) (Takagi, 1986). Boxed residues show differences between both sequences.

1 10 20 30
 P: E A V R G T I I R K N V N S L T P S I K E L R D A M A K V Q A D T S

40 50 60 70
 O: A S Y H G I P L **S C** H Y E N G T A Y A C **C** Q H G M V T F
 P: D D G Y Q K I A S Y H G I P L **C S** H Y E N G T A Y A C **G** Q H G M V T F

80 90 100
 O: P N W H R L L T K Q M E D A L V A K G S H V G I P Y W D W T T **T** F A N
 P: P N - H R L L T K Q M E D A L V A K G S H V G I P Y W D W T T **W** F A N

110 120 130 140
 O: L P V L V T E E K D N S F H H A **H** I D V A N T D T T R S P R A Q L F D
 P: L P V L V T E E K D N S F H H A **V** I D ----- T T R S P R A Q L F D

150 160 170 *
 O: D P **E** K G D K S F F Y R Q I A L A L E Q T D F C D F E I Q F E I G H N
 P: D P **D** K G - K S F F Y R

* 180 190 200 * 210
 O: A I H S W V G G S S P Y G M S T L H Y T S Y D P L F Y L H H S N T D R
 P:

220 230 240
 O: I W S V W Q A L Q K Y R G L P Y N T A N C E I N K L V K P L K P F **N** L
 P: I W S V W Q A L Q K Y R G L P Y N T A N C E I - K L V K P L K P F **D** L

250 260 270 280
 O: D T N P N A V T K A H S T G A T S F D Y H K L G Y D Y D N L N F H G M
 P: D A H S T G A T S F D Y H K L G Y D

290 300 310
 O: T I P E L E E H L K E I Q H E D R V F A G F L L R T I G **Q** S A D V N F
 P: E I Q H E D R V F A G F L L R T I G **V** S A D -----

320 330 340
 O: A D V C T **K D G E** C T F G G T F C I L G G E H E M F W **A F** D - R L F K
 P: - D V C T **S N F K** M F W **A W** D F R L F K

350 360 370
 O: Y D I T T S L K H L R L D A H D ----- D **F** D **I** K V T I K G I D G H V
 P: Y D I T T S L K H L R L D A H D P I A V D **T** D **F** K V T I K G I D G H V

380 390
 O: L S **N** K Y L S **P** P T **V** F L A P A K T T H
 P: L S **D** K Y L S **V** P T **P** F L A P A K T T H

Figure 5.

2.2 Isolation and characterization of pHC2.

In order to obtain the remainder of the sequence of the C-terminal domain, and possibly the whole subunit, a new cDNA library was constructed according to the method of Gubler and Hoffmann (1983) (see section B.2.4.3.) using oligonucleotide WL3 (see Table 2) as a specific primer. It was made complementary to the underlined region near the 5' end of pHC1 cDNA insert (see Figure 4). Specific primers have been used on several occasions to clone 5' regions of cDNAs which could otherwise not be obtained, for example for Troponin cDNA (Breitbart *et al.*, 1985). WL3 contains an internal *Eco*RI site, so it is possible to obtain information about the orientation of the cDNA insert by cutting recombinants hybridizing to the probe with *Eco*RI. After transformation 3000 clones were obtained. About 10% of the library was initially screened with probe WL3. From this 98 positives were isolated, plasmid DNA prepared in small scale and inserts analyzed for orientation by digesting with *Eco*RI and size by cutting with *Eco*RI and *Bam*HI to release the insert from the vector. This also gave a characteristic restriction pattern for one group of clones which was composed of a 160 bp, a 400 bp and a third fragment of variable size. Some of the positives were sequenced using the reverse and forward M13/pUC primers. The obtained sequence showed, adjacent to the poly-dC region, the primer sequence followed by the remaining 11 nucleotides of the 5' portion of the pHC1 coding sequence. This shows that these clones are really extensions of the original cDNA clone. An other group of clones was also identified that showed a 550 bp fragment and an additional piece. No readable sequence could be obtained initially and a more in depth analysis was postponed.

The clone with the longest insert belonged to the first group and was termed pHC2. A restriction map was assembled and its sequence determined

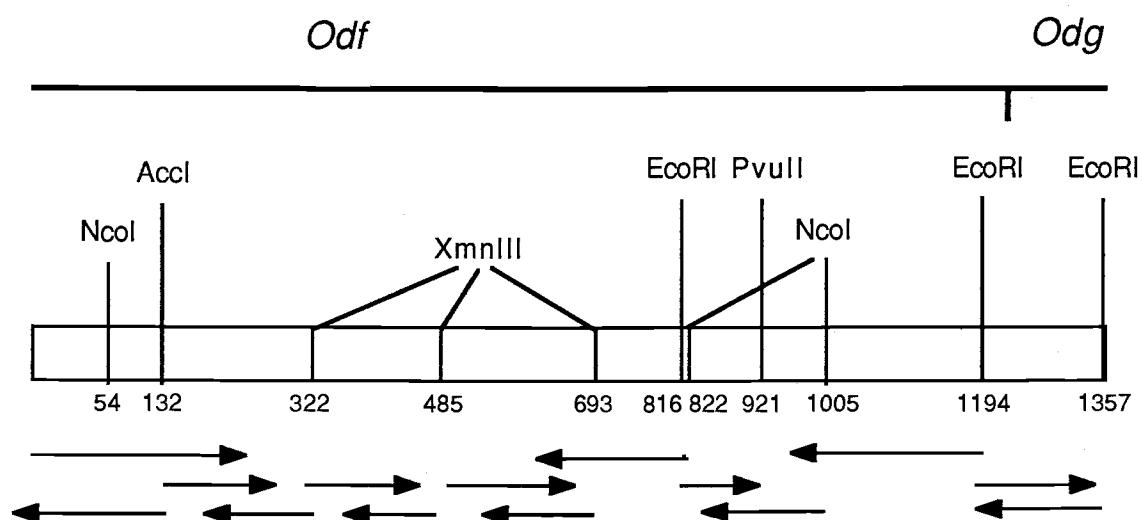


Figure 6: Restriction map of pHC2. Nucleotide number 1 is the first nucleotide of the cDNA insert. Arrows show directions of sequencing. The regions coding for *Odf* and *Odg* are indicated.

Figure 7 (pages 63 and 64): Sequence of the cDNA insert of pHC2 and the translated protein sequence. The underlined stretch shows the complementary sequence of WL4 and the three internal *Eco*RI sites.

1
AGATATTGAGAATCTTATGGCAGCTCTTACTTGGGTTAAGAAAGATGAAAGCGACCATGGATTCCAGACTATT
D I Q N L M A A L T W V K K D E S D H G F Q T I

74
GCTAGTTATCATGGTTCAACGCTGTGTCCGAGTCCAGAGGAGCCCAAATATGCCTGTTGTCTACATGGAATGCCC
A S Y H G S T L C P S P E E P K Y A C C L H G M P

149
GTCTTCCCACATTGGCACCGTGTCTTACTTATTACATTTTGAAGATTCTATGCGCCGGCATGGCTCCAGTGTGCG
V F P H W H R V Y L L H F E D S M R R H G S S V A

224
ACTCCTTATTGGGATTGGACACAACCTGGTACGAAACTGCCTAGACTTTTAGCAGATTCTGACTACTATGATGCT
T P Y W D W T Q P G T K L P R L L A D S D Y Y D A

299
TGGACTGATAATGTGACTGAGAATCCATTCCCTGAGGGGCTACATTACATCTGAAGACACCTACACAGTCAGGGAC
W T D N V I E N P F L R G Y I T S E D T Y T V R D

374
GTAAAGCCAGAGCTATTTGAAATCGGTGGAGGAGAGGGATCTACTCTTTACCAACAAGTACTACTGATGCTTGAA
V K P E L F E I G G G E G S T L Y Q Q V L L M L E

449
CAAGAAGACTACTGTGACTTCGAAGTTCAGTTTGAAGTAGTTTCATAACTCTATTCACTACCTTGTGGTGGTCAC
Q E D Y C D F E V Q F E V V H N S I H Y L V G G H

524
CAGAAATATGCCATGTCCAGTTTGGTCTATAGTTCCTTTGATCCTATCTTCTACGTTTCATCATTCAARGGTTGAT
Q K Y A M S S L V Y S S F D P I F Y V H H S M V D

599
CGTCTCTGGGCTATTTGGCAAGCTCTCCAGGAACACAGACATTTGCCGTTTGATAAGGCTTACTGTGCCCTGGAA
R L W A I W Q A L Q E H R H L P F D K A Y C A L E

674
CAACTGTCATTCCCAATGAAGCCTTTTCGTTTGGGAGTCCAACCCAAATCTACATACACGAGCTGCATCAACACCA
Q L S F P M K P F V W E A N P N L H T R A A S T P

749
CAACATCTATTTGACTACAATAAACTCGGTTACAAATATGATGACCTCGAATTCATGGAATGAATATAGATCAA
Q H L F D Y N K L G Y K Y D D L E F H G M N I D Q

824
TTGGAAAATGCTATTTCATAAAACGCAGAACAAAGACAGAGTTTTTGTCTTCCTTCTTACTCTTTGGTATTAAAACC
L E N A I H K T Q N K D R V F A S F L L F G I K T

899
TCAGCTGATGTCCATTTGAAACTTTGCAAAGATGAAACTTGTGAAGATGCTGGTGTAGTCTTTGTACTTGGTGGT
S A D V H L K L C K D E T C E D A G V V F V L G G

Figure 7.

974

GACAATGAAATGCCATGGCCCTTCGATAGAACGTACAAGATGGATATTACCAATGTTCTACATAAAATGCACATT
D N E M P W P F D R T Y K M D I T N V L H K M H I

1049

CCTTTGGAAGATCTGTATGTTTCATGGAAGCACTATTCACCTTGAAGTAAAGATTGAATCTGTAGATGGAAAAGTA
P L E D L Y V H G S T I H L E V K I E S V D G K V

1124

TTAGATTCTAGCTCATTGCCAGTTCCTTCAATGATATATGTCCCAGCTAAAGAATTCACAAAAGAGATAGAAAAG
L D S S S L P V P S M I Y V P A K E F T K E I E K

1199

GAAGCCGTGAGAGGTACTATAATCAGGAAGAATGTCAATAGTTTGACTCCATCTGATATTAAAGAACTAAGAGAC
E A V R G T I I R K N V N S L T P S D I K E L R D

1274

GCCATGGCTAAAGTACAAGCTGACACATCAGATAATGGTTACCAGAAAATTGCTTCCTATCATGGAATTCCTCCTC
A M A K V Q A D T S D N G Y Q K I A S Y H G I P L

1349

AGTTG
S

Figure 7 continued.

completely. The restriction map and the sequencing strategy are shown in Figure 6. Clone pHC2 has an insert of 1350 bp and contains a single open reading frame spanning the entire length. It codes for the N-terminal portion of the C-terminal domain *Odg* not coded for by pHC1, and for most of domain *Odf*, as judged by sequence comparison with *Odg*. The complete cDNA sequence and the derived protein sequence are shown in Figure 7. However there is a disagreement between the N-terminal sequence for *Odg* reported earlier (Lamy *et al.*, 1987; see Table 1) and the one shown in Figure 7. The first five residues obtained by protein sequencing were TVGDA, the first six residues obtained by cDNA sequencing were EAVRGT. However, the following residues in both sequences are identical. At first it seemed likely that the two sequences have been obtained from different *Octopus* species. However, results presented below show that this was apparently not the case (see section D. 1.).

2.3. Isolation and characterization of clones XT3 and XT22

From analysis of first and second strand synthesis products (not shown) it was known that clones with inserts of up to 2.0 kb could be expected. It is also possible that second strand synthesis was incomplete and hemocyanin cDNA clones could have been missed during the screen with WL3. Therefore the entire library was rescreened with WL4 (see table 2), which was synthesized complementary to a portion of the pHC2 sequence close to the 5' end (underlined in Figure 7). Sixty-five clones hybridized to the probe, which were picked and rescreened. The sizes of the inserts and their *EcoRI/BamH1* restriction patterns were analyzed as described for pHC2. Several of these clones gave a restriction pattern similar to that of pHC2: it showed, in addition

to the 160bp and 400 bp fragments, a 900 bp fragment plus an additional one of variable size. Of those clones, XT3 had the longest insert with about 1.8 kb. Another clone (XT22) had an insert of about 1900 bp but showed a different restriction pattern. Sequence analysis using WL4 as sequencing primer showed that XT22 overlapped with XT3 but missed the 160 bp, 400 bp and part of the 900 bp fragment. Both clones were sequenced entirely as follows: restriction fragments of the regions of XT3 not overlapping with pH2 were subcloned and sequenced, the same was done with XT22. XT22 was also sequenced entirely in both orientations by creating a series of nested deletions (Henikoff, 1984). A restriction map from both clones and the sequencing strategy is shown in Figure 8. The sequence of clone XT22 is shown in Figure 9.

Like for *Odg*, there was also a difference in sequencing results for the N-terminus of *Odf*. By protein sequencing (Lamy *et al.*, 1987) the N-terminal portion was found to be IPPSKQDADIDTP, but by sequencing the cDNA it was found to be IPPSNEDADIDIP. Two different sequences were determined for the N-terminus of *Ode*, but both were somewhat uncertain (Lamy *et al.*, 1987). Sequence a (Table 1) is in good agreement with the DNA sequence and it seems that XT22 codes for the entire domain *Ode*, except for two residues at the N-terminus, and for the first 221 amino acids of domain *Odf*. Clone XT3 codes for the C-terminal 196 amino acids of *Ode*, the entire domain *Odf* and the N-terminal 42 amino acids of *Odg*.

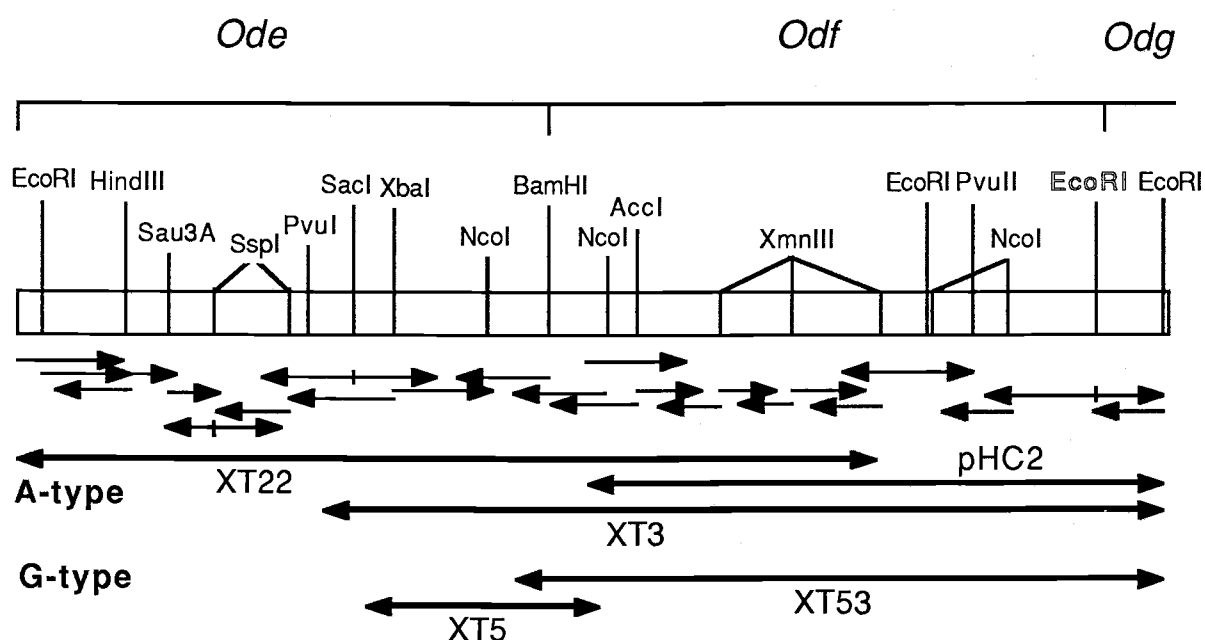


Figure 8: Restriction map of cDNA clones XT3, XT5, XT22 and XT53. The arrows show the direction of sequencing of subcloned DNA fragments. The *EcoRI* site shown in outlined letters is present only in pHC2 and XT3, both clones code for the A-type sequence. XT5 and XT53 code for the G-type sequence as is indicated in this figure.

Figure 9 (pages 69 and 70): Sequence of the cDNA insert of XT22.

1
GAGAAGGCAATGAATATCTGGTACGAAAGAATGTCGAAAGACTTTCTCTGTCTGAAATGAATTCTTTGATACAC
E G N E Y L V R K N V E R L S L S E M N S L I H

75
GCTTTTCAGAAGAATGCAGAAGGACAAGTCATCTGACGGTTTTGAGGCAATCGCTTCATTCCATGCTCTTCCTCCT
A F R R M Q K D K S S D G F E A I A S F H A L P P

150
CTCTGTCCAAGCCCAACTGCCAAACATAGGCACGCTTGTTGTCTTCATGGTATGGCTACGTTCCCTCACTGGCAC
L C P R P T A K H R H A C C L H G M A T F P H W H

225
AGGCTCTACGTTGTTCAAGTTGAACAAGCTTTACATAGACATGGAGCTACGGTTGGCGTTCCTTACTGGGATTGG
R L Y V V Q F E Q A L H R H G A T V G V P Y W D W

300
ACCGTTCCTATTTCAAAGATCCCTGATTTTCATAGCGTCGGAAAAGTATTCTGATCCTTTCACTAAAAATAGAGGTT
T R P I S K I P D F I A S E K Y S D P F T K I E V

375
TATAACCCATTTAATCATGGTCATATTTCTTTCATTAGTGAGGACACTACGACTAAACGAGAGGTCAGCGAATAT
Y N P F N H G H I S L I S E D T T T K R E V S E Y

425
TTGTTTGAACATCCTGTACTGGGAAAACAGACGTGGCTCTTTGATAACATCGCTTTAGCTTTGGAACAGACCGAT
L F E H P A L G K Q T W L F D N I A L A L E Q T D

500
TATTGCGATTTTCGAAATACAATTAGAGATTGTTTACAATGCCATTCACTCGTGGATTGGCGGGAAAGAAGAGCAT
Y C D F E I Q L E I A H N A I H S W I G G K E E H

575
TCCTTGAACCATTACATTATGCAGCCTACGACCCAATATTCTATCTACATCATTCTAATGTCGATCGTTTGTGG
S L N H L H Y A A Y D P I F Y L H H S N V D R L W

650
GTTATTTGGCAAGAATTGCAGAAATTGAGAGGTCTCAATGCTTATGAATCCCATTGTGCTCTTGAACCTTATGAAA
V I W Q E L Q K L R G L N A Y E S H C A L E L M K

725
GTTCCGTTGAAACCGTTCTCTTTTCGGAGCTCCTTACAATTTGAATGATCTAACTACCAAATGTCTAAACCTGAA
V P L K P F S F G A P Y N L N D L T T K L S K P E

800
GATATGTTTAGATACAAAGACAACTTCCATTACGAATACGACATTCTAGATATTAACAGTATGTCTATTAATCAG
D M F R Y K D N F H Y E Y D I L D I N S M S I N Q

875
ATAGAGTCGTCGTACATCAGACACCAAAAAGACCACGATCGTGTTTTTGCCGGCTTTTTGTTGAGTGGATTGGT
I E S S Y I R H Q K D H D R V F A G F L L S G F G

950
TCATCAGCTTATGCAACCTTTGAAATCTGTATTGAAGGAGGAGAATGTCATGAAGGAAGTCACTTTGCTGTGTTG
S S A Y A T F E I C I E G G E C H E G S H F A V L

Figure 9.

1025

GGAGGCAGCACAGAAATGCCATGGGCCTTCGACCGTCTCTATAAGATAGAAATTACTGATGTACTTTCTGATATG
G G S T E M P W A F D R L Y K I E I T D V L S D M

1100

CACTTAGCGTTCGATTGAGCTTTCACTATTAAAAACGAAAATAGTTGCTCAAAAATGGAACTGAACTGCCAGCTAGC
H L A F D S A F T I K T K I V A Q N G T E L P A S

1175

ATTCTACCAGAAGCAACTGTAATAAGGATCCACCTTCCAAGCAAGACGCCGATATTGACATCCCACTAAATCAT
I L P E A T V I R I P P S K Q D A D I D I P L N H

1250

ATCCGACGAAATGTAGAGTCTTTGGACGAAAGAGATATTCAGAATCTTATGGCAGCTCTTACTCGGGTTAAGAAA
I R R N V E S L D E R D I Q N L M A A L T R V K K

1325

GATGAAAGCGACCATGGATTCCAGACTATTGCTAGTTATCATGGTTCAACGCTGTGTCCGAGTCCAGAGGAGCCC
D E S D H G F Q T I A S Y H G S T L C P S P E E P

1400

AAATATGCCTGTTGTCTACATGGAATGCCCGTCTTCCACATTGGCACCGTGTTTACTTATTACATTTTGAAGAT
K Y A C C Y H G M P V F P H W H R V Y L L H F E D

1475

TCTATGCGCCGGCATGGCTCCAGTGTTGCCACTCCTTATTGGGATTGGACACAACCTGGTACGAAACTGCCTAGA
S M R R H G S S V A T P Y W D W T Q P G T K L P R

1550

CTTTTAGCAGATTCTGACTACTATGATGCTTGGACTGATAATGTGATCGAGAATCCATTCCCTGAGGGGCTACATT
L L A D S D Y Y D A W T D N V I E N P F L R G Y I

1625

ACATCTGAAGACACCTACACAGTCAGGGACGTAAAGCCAGAGCTATTTGAAATCGGTGGAGGAGAGGGATCTACT
T S E D T Y T V R D V K P E L F E I G G G E G S T

1700

CTTTACCAACAAGTACTACTGATGCTTGAACAAGAAGACTACTGTGACTTCGAAGTTCAGTTTGAAGTAGTTCAT
L Y Q Q V L L M L E Q E D Y C D F E V Q F E V V H

1775

AACTCTATTCACTACCTTGTTGGTGGTCACCAGAAATATGCCATGTCCAGTTTGGTCTATAGTTCCTTTGATCCT
N S I H Y L V G G H Q K Y A M S S L V Y S S F D P

1850

ATCTTCTACGTTTCATCATTCAATGGTTGATCGTCTCTGGGCTATTTGGCA
I F Y V H H S M V D R L W A I W

Figure 9, continued.

2.4. Presence of two cDNAs coding for *Octopus dofleini* hemocyanin.

When the cDNA clone XT3 was sequenced using oligonucleotide WL 4 as a primer, one nucleotide difference was obtained: position 32 in pHc 2 was changed from a T to a C in XT3. In order to see which one of the two represented a cloning artifact, several other cDNA clones were sequenced and it turned out that the C is the correct nucleotide, which means that the correct amino acid in this position is arg instead of trp. The T at position 32 in pHc2 possibly arose from a misincorporation during first strand synthesis, since reverse transcriptase is much more error-prone than DNA polymerase I (Loeb, 1986). However, a closer look at all sequences obtained in this manner revealed the presence of several other nucleotide changes. It appeared that there were two different cDNA sequences present (Figure 10). These were termed the G-type and A-type sequence after the first nucleotide difference observed in the sequence obtained with primer WL4. An analysis of a total of 27 sequences obtained in this manner revealed that 13 sequences were of the G-type and 14 were of the A-type. Clones XT3 and XT22 coded for the A-type sequence and XT5 and XT53, besides several others, for the G-type sequence.

Clone XT53 was the G-type clone with the longest cDNA insert (about 1.6 kb). After digesting this clone with *Eco*RI and *Bam*HI, a 900 bp and a 550 bp fragment could be seen on a gel. A re-examination of a few of those clones obtained after the WL3 screen, which also gave a 550 bp fragment during analysis, confirmed that they were also G-type clones, which have one less internal *Eco*RI site (see Figure 8). Clone XT5 overlaps with XT53 (see Figure 8) and was sequenced from subcloned restriction fragments. XT53 was sequenced by creating a set of nested deletions in both orientations. Together they code for the C-terminal portion of *Ode*, the entire domain *Odf* and the N-terminal 42 residues of *Odg*.

```

      *
G: GAATATCTCTTTCGTCCAAAGACTCTACATTTCTGTCGGATATGATTTAGTGGGGTGTCAATATCAGCGTCTT
A:                                     A           G

      Odf | Ode
G: CATTGGAAGGTGGGATCCTTATTACAGTTGTTTCTGGAAGAATGCTAGCTGGCAGTTCAGT
A: GC                                     T

```

Figure 10: DNA sequence obtained with primer WL4. It is complementary to the coding sequence and begins in the N-terminal region of *Odf* and stretches into the C-terminal parts of *Ode*. The start of clone pH2 is marked with an asterisk. G stands for G-type sequence, A for A-type sequence. "|" marks the first nucleotide of the *Odf* coding sequence. The nucleotide differences are indicated.

It was now unclear whether pHCl codes for the A-type or G-type sequence. Longer cDNA clones, which code for the C-terminal domain, were needed in order to obtain overlaps. A new oligo(dT)-primed cDNA library was constructed and screened with the oligonucleotide probe WL3. Among the positive clones identified and analyzed, clone pHCl1 codes for the A-type sequence and overlaps with pHCl2, and clone p17/90 codes for the G-type sequence and overlaps with XT53 (see Figure 11). Clone p17/90 has an internal *HindIII* site like pHCl, pHCl1 does not. Therefore pHCl is a G-type clone. Sequence analysis of pHCl1 and p17/90 is still in progress.

Figure 12 shows a comparison between the G-type and A-type sequence both at the DNA and protein level. From the sequence towards the C-terminal end, only those parts which have been unambiguously determined are included. There are 60 nucleotide changes leading to 34 amino acid substitutions. These sequencing results suggest that there are two polypeptide chains present in *Octopus dofleini* hemocyanin.

3. Properties of the polypeptide chain

The cDNA sequence coding for 1235 amino acid residues has been obtained so far which accounts for three domains out of seven of the whole subunit. The cDNA sequence and the translated protein sequence are shown in Figure 13. The polypeptide chain of *Ode* is 410 residues, of *Odf* 428 and *Odg* 399 residues long. Their amino acid composition and molecular weights of the pure polypeptide chains are shown in Table 3. There are a total of 167 acidic (asp and glu) and 110 basic side chains (arg and lys) in the portion of the sequence determined so far. This results in an overall acidic character of the protein.

Figure 11 (page 75): Restriction map of *Octopus dofleini* cDNA. Only a few restriction sites are shown. The sequences covered by individual clones are shown by double arrows. The *Eco*RI site shown in outlined letters is present only in the A-type sequence, the *Hind*III site marked in outlined letters only in the G-type sequence. The hatched area represents the polyA region.

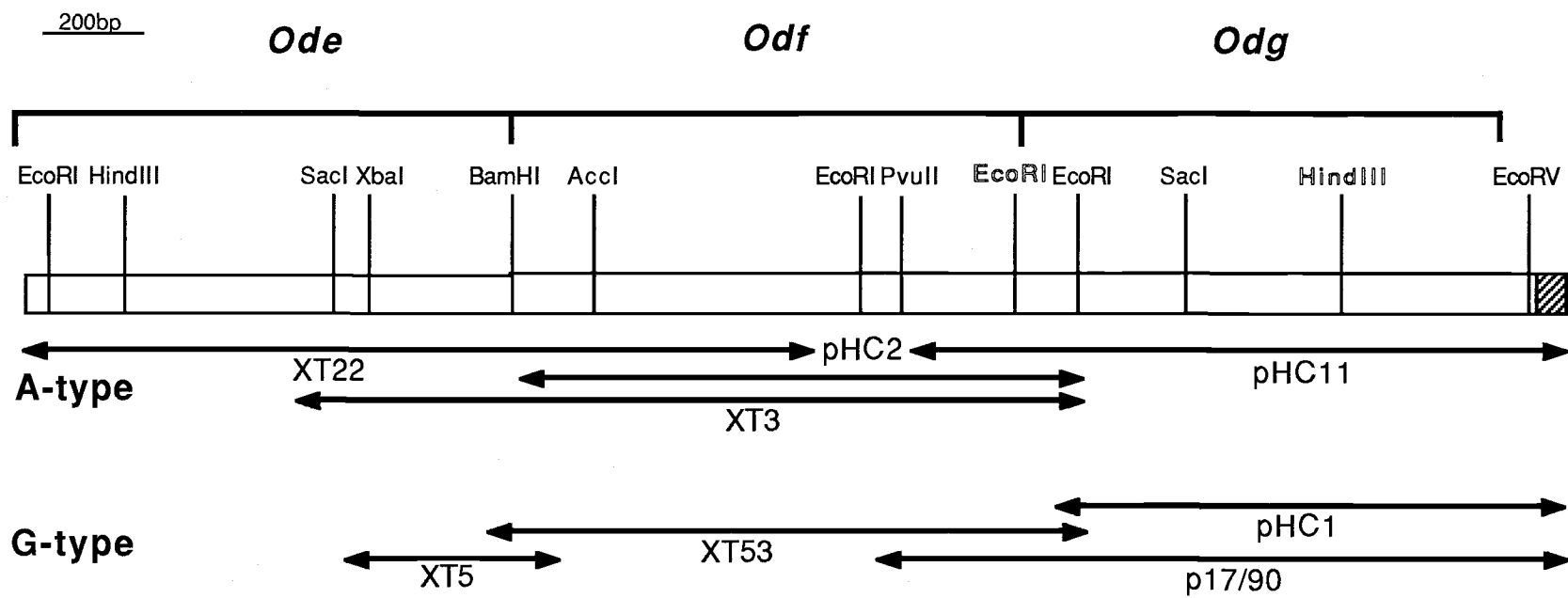


Figure 11.

Figure 12 (pages 77 to 80): Comparison of the G-type and A-type sequence at the DNA and protein level. G stands for G-type sequence, A for A-type sequence, "|" marks the start of a domain.

1
G: ACCAAAGAGATCACGATTGTGTTTTGCCGGCTTTTGTGAGGGGATTGGTTCATCAGCTTATGCAACC
A: A C C T
G: Q R D H D C V F A G F L L R G F G S S A Y A T
A: K R S

72
TTTGAAATCTGTATTGAAGGAGGAGAATGTCATGAAGGAAGTCACTTTTCTGTGTTGGGAGGCAGCACAGAAATG
 G
F E I C I E G G E C H E G S H F S V L G G S T E M
 A

147
CCATGGGCCTTCGACCGTCTCTATAAGATAGAAATTACTGATATACTTTCTGATATGAACTTAGCGTTCGATTCA
 G C
P W A F D R L Y K I E I T D I L S D M N L A F D S
 V H

222
GCTTTCACTATTAAAACGAAACTGGTTGCTCAAAATGGAAGTGAAGTCCAGCTAGCATTCTTCCAGAAGCAACT
 A A A
A F T I K T K L V A Q N G T E L P A S I L P E A T

297 *Ode | Odf*
GTAATAAGGATCCCTCCTTCCAATGAAGACGCTGATATTGACACCCCACTAAATCATATCCGACGAAATGTAGAG
 A GC C T
V I R I P P S N E D A D I D T P L N H I R R N V E
 K Q I

372
TCTTTGGACGAAAGAGATATTCAGAATCTTATGGCAGCTCTTACTCGGGTTAAGGAAGATGAAAGCGACCATGGA
 A
S L D E R D I Q N L M A A L T R V K E D E S D H G
 K

447
TTCCAGACTATTGCTAGTTATCATGGTTCAACGCTTTGTCCGAGTCCAGAGGAGCCCAAATATGCCTGTTGTCTA
 G
F Q T I A S Y H G S T L C P S P E E P K Y A C C Y

522
CATGGAATGCCCGTCTTCCACATTGGCACCGTGTTTACTTATTACAATTTGAAGATTCTATGCGCCGGCATGGC
 T
H G M P V F P H W H R V Y L L Q F E D S M R R H G
 H

597
TCCAGTGTTGCCACTCCTTATTGGGATTGGACACAACCTGGTACGAAACTGCCTAGACTTTTAGCAGATTCTGAC
S S V A T P Y W D W T Q P G T K L P R L L A D S D

Figure 12.

672

TACTATGATGCTTGGACTGATAATGTGATCGAGAATCCATTCCCTTAGGGGCTACATTAAAACTGAAGACACCTAC
 Y Y D A W T D N V I E N P F L R G Y I K T E D T Y
 G C T
 T S

747

ACAGTCAGGGACGTAAAGCCAGAGCTATTTGAAATCGGTGGAGGAGAGGGATCTACTCTTTACCAACAAGTACTA
 T V R D V K P E L F E I G G G E G S T L Y Q Q V L

822

CTGATGCTTGAACAAGAAGACTACTGTGACTTCGAAGTTCAGTTTGAAGTAGTTCATAACTCTATTCACTACCTT
 L M L E Q E D Y C D F E V Q F E V V H N S I H Y L

897

GTTGGTGGTCACCAGAAATATGCCATGTCCAGTTTGGTCTATAGTTCCTTTGATCCTATCTTCTACGTTTCATCAT
 V G G H Q K Y A M S S L V Y S S F D P I F Y V H H

972

TCAATGGTTGATCGTCTCTGGGCTATTTGGCAAGCTCTCCAGGAACACAGACATTTACCGTTTGATAAGGCTTAC
 S M V D R L W A I W Q A L Q E H R H L P F D K A Y
 G

1047

TGTGCCCTGGAACAACTGTCATTCCCAATGAAGCCTTTCGTTTGGGAGTCCAACCCAAATCTACATACACGAGCT
 C A L E Q L S F P M K P F V W E S N P N L H T R A

1122

GCATCAACACCACAACATCTATTTGACGACAATAAACTGGGTTACAAATATGACAACCTCGAATTCCATGGAATG
 A S T P Q H L F D D N K L G Y K Y D N L E F H G M
 T C TG
 Y D

1197

AATATAGATCAATTGGAATAATGCTATTCATAAACAGCAGAACAAAGACAGAGTTTTTGCTTCCTTCTTACTCTTT
 N I D Q L E N A I H K Q Q N K D R V F A S F L L F
 AC
 T

1272

GGTATCAAAACCTCAGCTGATGTCCATTTGAACTTTGCAAAGATGAACTTGTGAAGATGCTGGTGTAGTCTTT
 G I K T S A D V H L K L C K D E T C E D A G V V F
 T

Figure 12 continued.

1347

ATACTTGGTGGTGACAATGAAATGCCATGGCATTTCGATAGAACGTACAAGAAAGATATTACTCATGTTCTACAT
 G CC TG CA
 I L G G D H E M P W H F D R T Y K K D I T H V L H
 V P M N

1422

CAAATGCATATTCCTTTGGAAGATCTGTATGTTTCATGGAAGCACTATTCTACTTGAAGTAGAAAATTGAAACTGTA
 A C AC A G T
 Q M H I P L E D L Y V H G S T I L L E V E I E T V
 K H K S

1497

GATGGAAAAGTATTAGATTCTAGCTCATTGCCAGCTCCTTCAATGATATATGTCCCAGCTAAAGATTTCAAAAGA
 T A C A
 D G K V L D S S S L P A P T M I Y V P A K D F K R
 V E T K

1572 *Odf* | *Odg*

GAGGTGCACAAGAAAAGTGTGGGTGATGCTATAATCAGGAAGAATGTCAATAGTTTGACTCCATCTGATATTAAA
 A A G A G G C A A G A
 E V H K K T V G D A I I R K N V N S L T P S D I K
 I E E A R G T

1647

GAACTAAGAGACGCCATGGCTAAAGTACAAGCTGACACATCAGATAATGGTTACCAGAAAATTGCTTCCTATCAT
 E L R D A M A K V Q A D T S D N G Y Q K I A S Y H

1722

GGAATTCCCCTCAGTTGTCTATTATGAGAAATGGCACAGCTTATGCATGTTGCCAACATGGAATGGTAACTTTCCCT
 G I P L S C H Y E N G T A Y A C C Q H G M V T F P

1797

AACTGGCATAGATTGTTGACTAAACAAATGGAAGATGCACTAGTTGCTAAGGGATCTCACGTTGGTATTCCTTAC
 N W H R L L T K Q M E D A L V A K G S H V G I P Y

1872

TGGGACTGGACCACTACCTTCGCTAATTTACCTGTACTGGTCACTGAGGAGAAAAGATAAATCCTTCCACCATGCC
 W D W T T T F A N L P V L V T E E K D N S F H H A

1947

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

Figure 12 continued.

2022

AAATCATTCTTCTATCGCCAGATAGCACTTGCTCTGGAACAAACAGATTTCTGTGATTTTGAAATTCAGTTTGAA

K S F F Y R Q I A L A L E Q T D F C D F E I Q F E

2097

ATCGGTCACAATGCTATTCATTCATGGGTTGGCGGTAGTAGCCCATATGGTATGTCAACTCTTCACTATACTTCC

I G H N A I H S W V G G S S P Y G M S T L H Y T S

2172

TATGATCCTCTCTTCTACCTTCATCACTCTAATACTGATCGTATTTGGTCTGTATGGCAAGCATTACAGAAGTAT

Y D P L F Y L H H S N T D R I W S V W Q A L Q K Y

2247

CGAGGATTACCTTACAACACTGCGAACTGTGAAATCAATAAACTTGTTAAACCACTAAAACCATTCAACTTAGAT

R G L P Y N T A N C E I N K L V K P L K P F N L D

2322

ACCAATCCTAACGCCGTTACAAAAGCCCATTCTACTGGTGCTACATCTTTCGATTACCACAAGCTTGGTTATGAT

T N P N A V T K A H S T G A T S F D Y H K L G Y D

2397

TATGATAATCTTAATTTCCATGGAATGACCATTCTGAGTTGGAAGAACACCTTAAAGAAATACAACACGAAGAC

Y D N L N F H G M T I P E L E E H L K E I Q H E D

2472

AGGGTATTTGCTGGTTTCTTACTTCGTACCATTGGTCAATCTGCTGATGTTAACCTTCGACGTCTGCACCAAAGAT

R V F A G F L L R T I G Q S A D V N F D V C T K D

2547

GGTGAATGTACATTTGGAGGTACGTTCTGTATTCTCGGTGGAGAACATGAA

G E C T F G G T F C I L G G E H E

Figure 12 continued.

The secondary structure of the individual domains has been predicted according to the method of Chou and Fasman (1978). The probabilities for a particular sequence attaining a given secondary structure were first calculated by a computer program. Assignments for ambiguous regions, which had probabilities for more than one type of structure, were done according to the rules published by Chou and Fasman (1978). The results are shown in Figure 14. From the X-ray structure of *Panulirus interruptus* hemocyanin is known that the copper ligands are furnished by α -helices. The positions of the putative copper ligands (see section D.3.) are marked with "▲". Only a few of them are actually situated within predicted α -helices. However it should be noted that secondary structure predictions are very unprecise, so this result should be treated with caution.

The hydropathy of each domain was predicted according to the method of Kyte and Doolittle (1980) (Figure 15). Hydrophobic regions have assigned positive values and hydrophilic regions negative values. The copper ligands of *Panulirus interruptus* hemocyanin are buried inside of the protein and are within hydrophobic regions. It seems that in *Octopus* hemocyanin the ligands of "Copper B" are in a hydrophobic region, however the ligands of "Copper A" seem to be situated in a hydrophilic part and may be very close to the surface of

Molluscan hemocyanins are glycoproteins and it is known that the C-terminal domain of both *Octopus doleini* and *Paroctopus dofleini dofleini* contains carbohydrate (Takagi, 1986; Miller *et al.*, 1988). I examined the sequence for presence of the consensus sequence asn-x-ser/thr for carbohydrate attachment sites of N-linked carbohydrate chains. There are two such sites present starting with Asn 310 and 889. Both sites are also located in

Figure 13 (pages 83 to 86): Sequence of *Octopus dofleini* cDNA. Shown as cloned so far together with translated protein sequence. The underlined asparagine residues are potential attachment sites for carbohydrate sidechains. The starts of *Odf* and *Odg* are marked with "|". The sequence shown is the combined sequence of clones XT22, pHCl and pHCl2. The underlined DNA sequences are the stop codon and the polyadenylation signal.

1
GAGAAGGCAATGAATATCTGGTACGAAAGAATGTCGAAAGACTTTCTCTGTCTGAAATGAATTCTTTGATACAC
E G N E Y L V R K N V E R L S L S E M N S L I H

75
GCTTTTCAAGAATGCAGAAGGACAAGTCATCTGACGGTTTTGAGGCAATCGCTTCATTCCATGCTCTTCCTCCT
A F R R M Q K D K S S D G F E A I A S F H A L P P

150
CTCTGTCCAAGCCCAACTGCCAAACATAGGCACGCTTGTGTCTTCATGGTATGGCTACGTTCCCTCACTGGCAC
L C P R P T A K H R H A C C L H G M A T F P H W H

225
AGGCTCTACGTTGTTCAAGTTTGAACAAGCTTTACATAGACATGGAGCTACGGTTGGCGTTCCTTACTGGGATTGG
R L Y V V Q F E Q A L H R H G A T V G V P Y W D W

300
ACCCGTCCTATTTCAAAGATCCCTGATTTTCATAGCGTCGGAAAAAGTATTCTGATCCTTTCACTAAAAATAGAGGTT
T R P I S K I P D F I A S E K Y S D P F T K I E V

375
TATAACCCATTTAATCATGGTCATATTTCTTTCATTAGTGAGGACACTACGACTAAACGAGAGGTCAGCGAATAT
Y N P F N H G H I S L I S E D T T T K R E V S E Y

450
TTGTTTGAACATCCTGTACTGGGAAAAACAGACGTGGCTCTTTGATAACATCGCTTTAGCTTTGGAACAGACCGAT
L F E H P A L G K Q T W L F D N I A L A L E Q T D

525
TATTGCGATTTTCGAAATACAATTAGAGATTGTTTACAATGCCATTCACTCGTGGATTGGCGGGAAAGAAGAGCAT
Y C D F E I Q L E I A H N A I H S W I G G K E E H

600
TCCTTGAACCATTTACATTATGCAGCCTACGACCCAATATTCTATCTACATCATTCTAATGTCGATCGTTTGTGG
S L N H L H Y A A Y D P I F Y L H H S N V D R L W

675
GTTATTTGGCAAGAATTGCAGAAATTGAGAGGTCTCAATGCTTATGAATCCCATTGTGCTCTTGAACTTATGAAA
V I W Q E L Q K L R G L N A Y E S H C A L E L M K

750
GTTCCGTTGAAACCGTTCTCTTTTCGGAGCTCCTTACAATTTGAATGATCTAACTACCAAATTGTCTAAACCTGAA
V P L K P F S F G A P Y N L N D L T T K L S K P E

825
GATATGTTTAGATACAAAGACAACCTTCCATTACGAATACGACATTCTAGATATTAACAGTATGTCTATTAATCAG
D M F R Y K D N F H Y E Y D I L D I N S M S I N Q

900
ATAGAGTCGTCGTACATCAGACACCAAAAAAGACCACGATCGTGTTTTTGCCGGCTTTTTTGTGAGTGGATTGGT
I E S S Y I R H Q K D H D R V F A G F L L S G F G

975
TCATCAGCTTATGCAACCTTTGAAATCTGTATTGAAGGAGGAGAATGTCATGAAGGAAGTCACTTTGCTGTGTTG
S S A Y A T F E I C I E G G E C H E G S H F A V L

Figure 13.

1050
GGAGGCAGCACAGAAATGCCATGGGCCTTCGACCGTCTCTATAAGATAGAAATTACTGATGTACTTTCTGATATG
G G S T E M P W A F D R L Y K I E I T D V L S D M

1125
CACTTAGCGTTCGATTTCAGCTTTCACTATTAAAAACGAAAAATAGTTGCTCAAAATGGAACTGAACTGCCAGCTAGC
H L A F D S A F T I K T K I V A Q N G T E L P A S

1200 *Ode* | *Odf*
ATTCTACCAGAAGCAACTGTAATAAGGATCCACCTTCCAAGCAAGACGCCGATATTGACATCCCACTAAATCAT
I L P E A T V I R I P P S K Q D A D I D I P L N H

1275
ATCCGACGAAATGTAGAGTCTTTGGACGAAAGAGATATTCAGAATCTTATGGCAGCTCTTACTCGGGTTAAGAAA
I R R N V E S L D E R D I Q N L M A A L T R V K K

1350
GATGAAAAGCGACCATGGATTCCAGACTATTGCTAGTTATCATGGTTCAACGCTGTGTCCGAGTCCAGAGGAGCCC
D E S D H G F Q T I A S Y H G S T L C P S P E E P

1425
AAATATGCCTGTTGTCTACATGGAATGCCCGTCTTCCCACATTGGCACCGTGTTTACTTATTACATTTTGAAGAT
K Y A C C Y H G M P V F P H W H R V Y L L H F E D

1500
TCTATGCGCCGGCATGGCTCCAGTGTTGCCACTCCTTATTGGGATTGGACACAACCTGGTACGAACTGCCTAGA
S M R R H G S S V A T P Y W D W T Q P G T K L P R

1575
CTTTTAGCAGATTCTGACTACTATGATGCTTGGACTGATAATGTGATCGAGAATCCATTCTGAGGGGCTACATT
L L A D S D Y Y D A W T D N V I E N P F L R G Y I

1650
ACATCTGAAGACACCTACACAGTCAGGGACGTAAAGCCAGAGCTATTTGAAATCGGTGGAGGAGAGGGATCTACT
T S E D T Y T V R D V K P E L F E I G G G E G S T

1725
CTTTACCAACAAGTACTACTGATGCTTGAACAAGAAGACTACTGTGACTTCGAAGTTCAGTTTGAAGTAGTTCAT
L Y Q Q V L L M L E Q E D Y C D F E V Q F E V V H

1800
AACTCTATTCACTACCTTGTGGTGGTCACCAGAAATATGCCATGTCCAGTTTGGTCTATAGTTCCTTTGATCCT
N S I H Y L V G G H Q K Y A M S S L V Y S S F D P

1875
ATCTTCTACGTTTCATCATTCAATGGTTGATCGTCTCTGGGCTATTTGGCAAGCTCTCCAGGAACACAGACATTTG
I F Y V H H S M V D R L W A I W Q A L Q E H R H L

1950
CCGTTTGATAAGGCTTACTGTGCCCTGGAACAACGTGTCATTCCCAATGAAGCCTTTTCGTTTGGGAGTCCAACCCA
P F D K A Y C A L E Q L S F P M K P F V W E S N P

2025
AATCTACATACACGAGCTGCATCAACACCACAACATCTATTTGACTACAATAAACTCGGTTACAAATATGATGAC
N L H T R A A S T P Q H L F D Y N K L G Y K Y D D

Figure 13, continued.

2100
 CTCGAATTCATGGAATGAATATAGATCAATTGGAAAATGCTATTCAAAAACGCAGAACAAAGACAGAGTTTTT
 L E F H G M N I D Q L E N A I H K T Q N K D R V F

2175
 GCTTCCTTCTTACTCTTTGGTATTAAAACCTCAGCTGATGTCCATTTGAAACTTTGCAAAGATGAACTTGTGAA
 A S F L L F G I K T S A D V H L K L C K D E T C E

2250
 GATGCTGGTGTAGTCTTTGTACTTGGTGGTGACAATGAAATGCCATGGCCCTTCGATAGAACGTACAAGATGGAT
 D A G V V F V L G G D N E M P W P F D R T Y K M D

2325
 ATTACCAATGTTCTACATAAAATGCACATTCCTTTGGAAGATCTGTATGTTTCATGGAAGCACTATTCACCTTGAA
 I T N V L H K M H I P L E D L Y V H G S T I H L E

2400
 GTAAAGATTGAATTCGTAGATGGAAAAGTATTAGATTCTAGCTCATTGCCAGTTCCTTCAATGATATATGTCCCA
 V K I E S V D G K V L D S S S L P V P S M I Y V P

2475
Odf | *Odg*
 GCTAAAGAATTCACAAAAGAGATAGAAAAGGAAGCCGTGAGAGGTACTATAATCAGGAAGAATGTCAATAGTTTG
 A K E F T K E I E K E A V R G T I I R K N V N S L

2550
 ACTCCATCTGATATTAAAGAACTAAGAGACGCCATGGCTAAAGTACAAGCTGACACATCAGATAATGGTTACCAA
 T P S D I K E L R D A M A K V Q A D T S D N G Y Q

2625
 AAAATTGCTTCCTATCATGGAATTCCTCAGTTGTCATTATGAGAATGGCACAGCTTATGCATGTTGCCAACAT
 K I A S Y H G I P L S C H Y E N G T A Y A C C Q H

2700
 GGAATGGTAACTTTCCCTAACTGGCATAGATTGTTGACTAAACAAATGGAAGATGCACTAGTTGCTAAGGGATCT
 G M V T F P N W H R L L T K Q M E D A L V A K G S

2775
 CACGTTGGTATTTCCTTACTGGGACTGGACCACTACCTTCGCTAATTTACCTGTACTGGTCACTGAGGAGAAAGAT
 H V G I P Y W D W T T T F A N L P V L V T E E K D

2850
 AACTCCTTCCACCATGCCCATATTGATGTTGCCAATACGGACACAACCAGATCACCAAGAGCTCAACTTTTTTGAT
 N S F H H A H I D V A N T D T T R S P R A Q L F D

2925
 GATCCAGAAAAAGGAGATAAATCATTCTTCTATCGCCAGATAGCACTTGCTCTGGAACAAACAGATTTCTGTGAT
 D P E K G D K S F F Y R Q I A L A L E Q T D F C D

3000
 TTTGAAATTCAGTTTGAAATCGGTCACAATGCTATTCAATTCATGGGTTGGCGGTAGTAGCCCATATGGTATGTCA
 F E I Q F E I G H N A I H S W V G G S S P Y G M S

3075
 ACTCTTCACTATACTTCCTATGATCCTCTCTTCTACCTTCATCACTCTAATACTGATCGTATTTGGTCTGTATGG
 T L H Y T S Y D P L F Y L H H S N T D R I W S V W

Figure 13, continued.

3150

CAAGCATTACAGAAGTATCGAGGATTACCTTACAACACTGCGAACTGTGAAATCAATAAACTTGTTAAACCACTA
Q A L Q K Y R G L P Y N T A N C E I N K L V K P L

3225

AAACCATTCAACTTAGATACCAATCCTAACGCCGTTACAAAAGCCCATTCTACTGGTGCTACATCTTTCGATTAC
K P F N L D T N P N A V T K A H S T G A T S F D Y

3300

CACAAGCTTGGTTATGATTATGATAATCTTAATTTCCATGGAATGACCATTCCCTGAACACCTTAAAGAAATACAA
H K L G Y D Y D N L N F H G M T I P E H L K E I Q

3375

CACGAAGACAGGGTATTTGCTGGTTTCTTACTTCGTACCATTGGTGAGTTGGAACAATCTGCTGATGTAACTTC
H E D R V F A G F L L R T I G E L E Q S A D V N F

3450

GACGTCTGCACCAAAGATGGTGAATGTACATTTGGAGGTACGTTCTGTATTCTCGGTGGAGAACATGAAATGTTT
D V C T K D G E C T F G G T F C I L G G E H E M F

3525

TGGGCATTTGACCGTCCTTTTAAATATGATATTACCACAAGTCTGAAACATCTCCGCTTAGATGCGCATGATGAT
W A F D R L F K Y D I T T S L K H L R L D A H D D

3600

TTCGATATAAAGGTTACTATTAAAGGTATCGATGGTCATGTCTTGTCCAATAAAATACCTCAGTCCTCCGACGGTT
F D I K V T I K G I D G H V L S N K Y L S P P T V

3675

TTCCTCGCTCCAGCCAAAACAACGCATTAATATGTGTACTAATCATTTATATGATTTAATATTTGTGTTTTTTTG
F L A P A K T T H stop

3750

TACTTTTTTCATTCAATAAATGATATCTCAAACAAT....PolyA

Figure 13, continued.

Table 3: Amino acid composition of *Ode*, *Odf*, and *Odg*.

	<i>Odg</i>	<i>Odf</i>	<i>Ode</i>
gly	25	20	20
ala	27	20	29
cys	8	7	7
pro	7	26	19
asp	31	34	22
glu	19	30	31
his	24	24	26
lys	23	22	20
arg	13	15	17
ile	21	20	28
leu	33	41	38
val	19	29	15
met	6	11	9
asn	19	15	16
gln	12	14	11
ser	22	29	30
thr	32	22	19
phe	23	20	24
trp	7	8	7
tyr	17	21	18
total	399	426	408
MW	45316	48159	47027

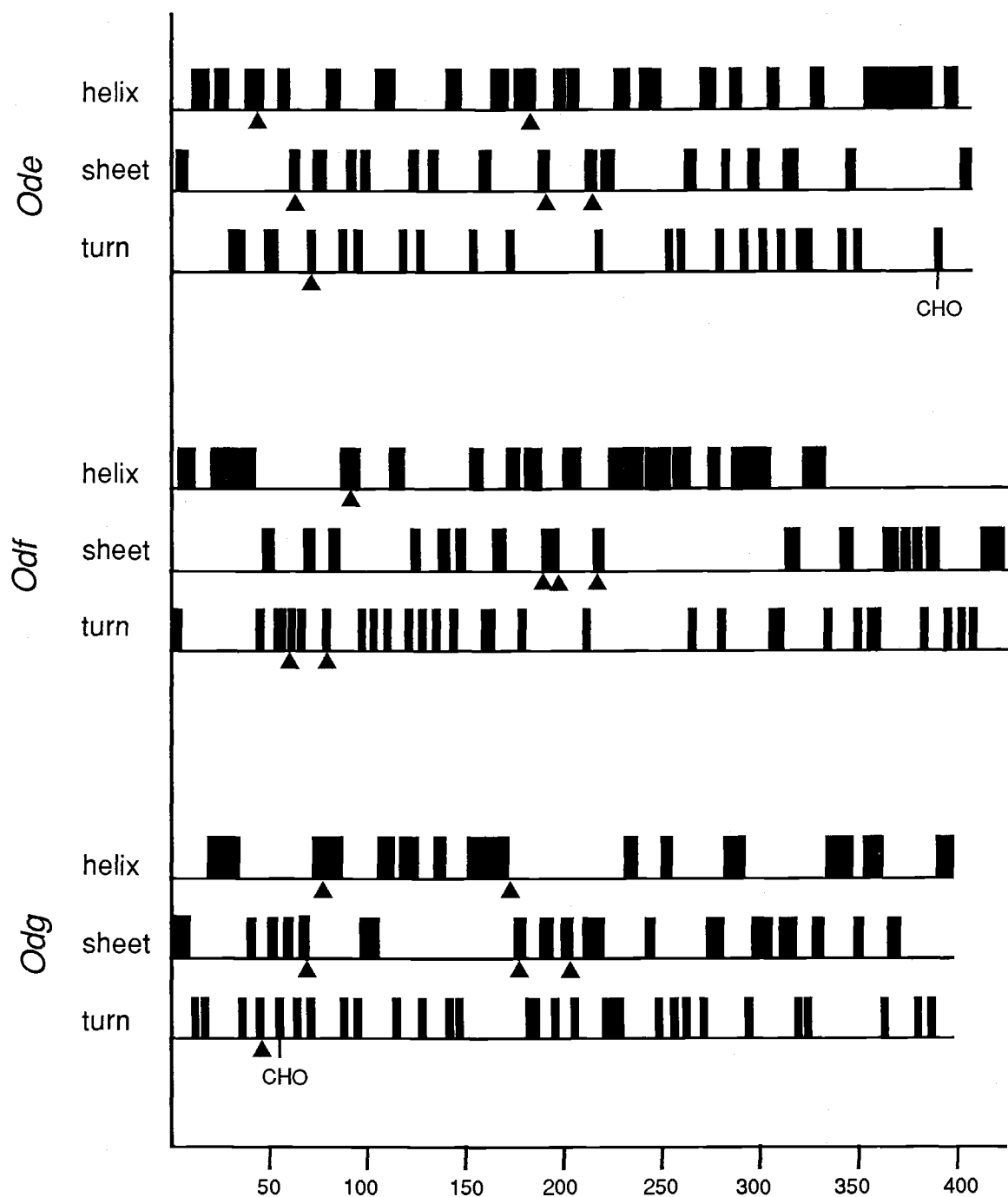


Figure 14: Results of secondary structure prediction according to Chou and Fasman (1978).

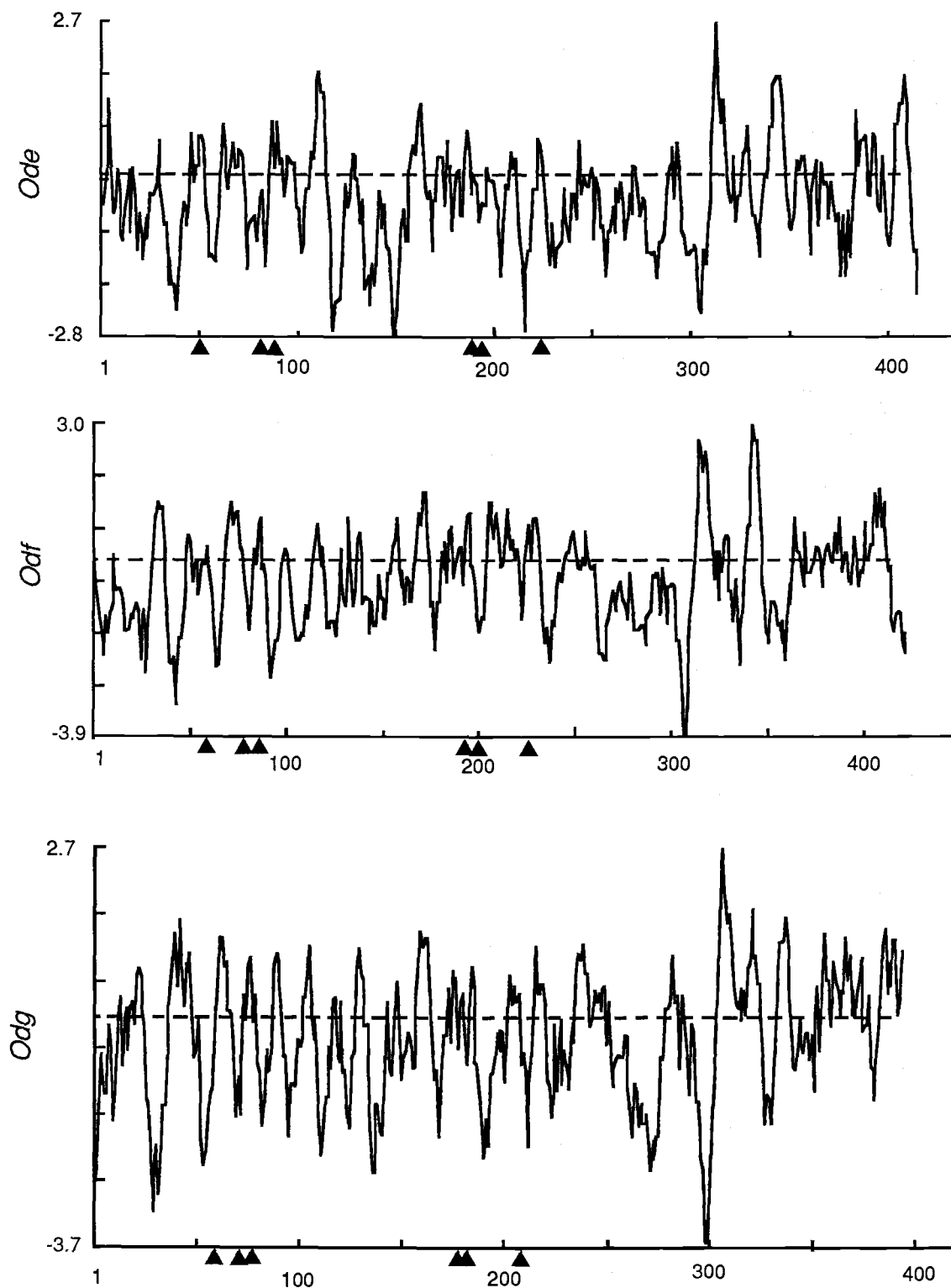


Figure 15: Hydropathy plots for all *Octopus* domains according to the method of Kyte and Doolittle (1980). A window of seven residues was used.

a predicted β -turn (see Figure 14) and are therefore likely candidates for carbohydrate attachment. The homologue of Asn 310 of *Sepia officinalis* functional unit *f* carries a carbohydrate sidechain. The corresponding site for asn 889 in the C-terminal domain of *P. dofleini dofleini* has been shown to have carbohydrate attached to it (Takagi, 1986). Since the C-terminal domain has been shown to contain carbohydrate (Miller *et al.*, 1988) this site is likely to be glycosylated as well in *Octopus dofleini* hemocyanin. Takagi also reported that possibly Ser 890 carries a carbohydrate chain too, so we have to consider the possibility of additional O-linked carbohydrate chains being present in the C-terminal domain and possibly others. However this would be in contrast to results of Kamerling *et al.* (1990). These researchers analyzed a series of hemocyanins, both of arthropodan and molluscan origin, for carbohydrate content and did not find any evidence for the presence O-linked carbohydrate chains in hemocyanins.

D. DISCUSSION

1. Presence of two polypeptide chains

As described in section A.6., the available experimental evidence suggested the presence of only one polypeptide chain in *Octopus dofleini* hemocyanin. However two slightly different cDNAs coding for *Octopus dofleini* hemocyanin were found, which were called G-type and A-type sequences. In a stretch of about 2500 nucleotides which is available from both sequences, only 60 nucleotide substitutions were found which cause 34 amino acid substitutions. Large parts are identical in both sequences, especially those coding for domain *Odg*; others have few changes. Nucleotide substitutions are clustered in the putative linker regions between domains, especially in the *Odf/Odg* linker (Figure 16). These sequencing results suggest that there are two polypeptide chains present in *Octopus dofleini* hemocyanin. As can be seen in Figure 16, the N-terminal sequences for *Odf* and *Odg* determined by protein sequencing are exactly those for the G-type sequence. The A-type sequence was never found. However, two different protein sequences have been found for the N-terminus of *Ode*, one of which is in fairly good agreement with the A-type sequence. The other one may represent the the G-type sequence, but no DNA sequence data are available yet to support this possibility. Since the differences between both polypeptide chains in the region sequenced so far is very small, it might be almost impossible to separate both chains from each other due to the size of the polypeptide chain. Moreover, despite the fact that several charged sidechains are substituted, their number and the overall net charge remains the same. The presence of two polypeptide chains also would explain the difficulties encountered in obtaining unambiguous N-terminal peptide sequences of the various domains. However the N-terminal peptide sequences

Linker region Ode/Odf

G: GCAACTGTAATAAGGATCCCTCCTTCCAATGAAGACGCTGATATTGACACCCCACTAAATCATATCCGACGA
 A: A GC C T
 G: A T V I R I P P S N E D A D I D T P L N H I R R
 A: K Q I

Linker region Odf/Odg

G: ATATATGTCCCAGCTAAAGATTTCAAAAGAGAGGTGCACAAGAAAAGTGTGGGTGATGCTATAATCAGGAAG
 A: A C A A A G A G G C A A G A
 G: I Y V P A K D F K R E V H K K T V G D A I I R K
 A: E T K I E E A R G T

Figure 16: Comparison of G-type and A-type sequence at the DNA and protein level in putative linker regions. The putative linker is underlined and connects conserved sequence elements from the C-terminal end of one domain with ones at the N-terminus of the following domain.

both from *Odf* and *Odg* represented the G-type. It is possible that the two polypeptide chains have a different susceptibility towards proteases, in this case trypsin. It seems that one of the two, possibly the G-type chain, is more susceptible to trypsin than the A-type chain and therefore may be cleaved at a faster rate. After very brief treatment, only *Odg* from the G-type chain would be cleaved off which can be unambiguously sequenced. But after longer treatment both chains are cleaved and the liberated *Odg* is a mixture of both types, which would give an uninterpretable sequencing result.

2. Comparison of sequences of molluscan hemocyanins

2.1 Comparison of *Octopus dofleini* domains with each other

The electron microscopic view of molluscan hemocyanin subunits shows that they appear to be composed of eight globular domains (Van Holde and Miller, 1982; Ellerton *et al.*, 1983; Preaux and Gielens, 1984) with the exception of *Octopus* hemocyanins which are composed of seven (Lamy *et al.*, 1986; 1987). The number of domains also coincides with the number of copper pairs per subunit (Van Holde and Miller, 1982; Ellerton *et al.*, 1983; Preaux and Gielens, 1984). This structural feature suggests that the polypeptide sequence of molluscan hemocyanins might be repetitive, the repeating unit being one oxygen binding domain. The availability of the complete sequence of three domains from one molluscan hemocyanin subunit now provides an opportunity to test this hypothesis. When the sequences of the individual *Octopus* domains are compared with each other (see Figure 18), they show a degree of similarity ranging between 51% and 54%. All three sequences show conserved sequence elements at corresponding positions and contain putative copper binding sites (Figure 17). The sequence obtained so far shows a repetitive pattern with a single oxygen binding domain as the repeating unit,

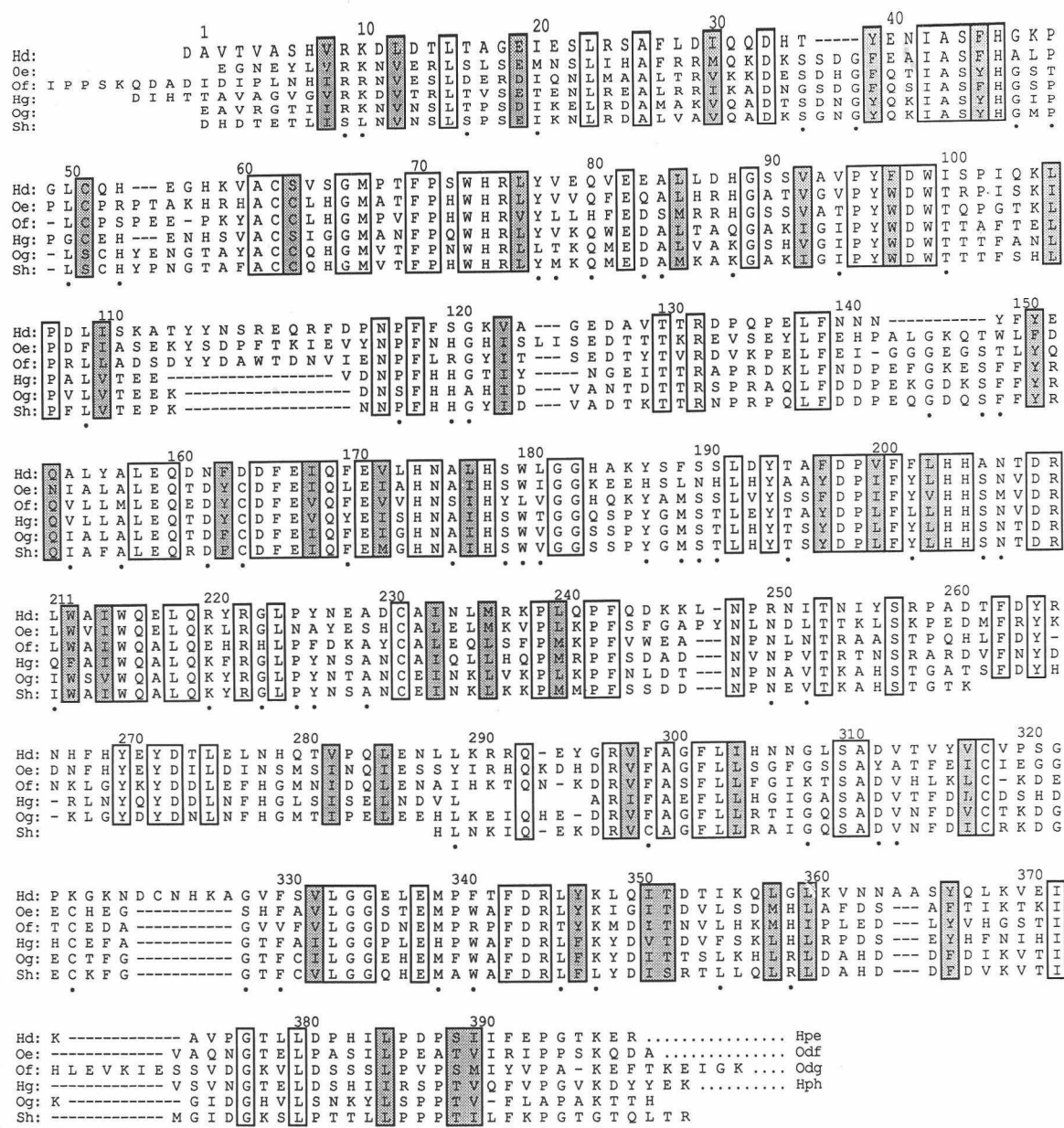
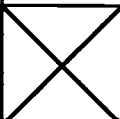

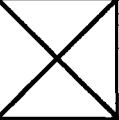
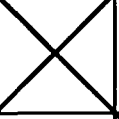
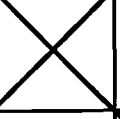

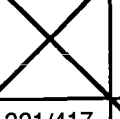



Figure 17: Alignment of complete or almost complete sequences for domains of molluscan hemocyanins. The sequences are from domains *Ode* (Oe), *Odf* (Of), *Odg* (Og), *Soh* (Sh), *Hpd* (Hd) and *Hpg* (Hg). Boxed residues show identical residues, shaded boxes similar residues shared by all six sequences. Identical or similar residues shared by five out of the six sequences are marked with ".". The following groups of amino acid residues are considered as similar: E and D; N and Q; S and T; S and C; M, L, I and V; F, Y and W; H, K and R.

Percent Identity

	<i>Ode</i>	<i>Odf</i>	<i>Odg</i>	<i>Sof</i>	<i>Sog</i>	<i>Soh</i>	<i>Hpd</i>	<i>Hpg</i>	
<i>Ode</i>		174/417 42%	175/414 42%	87/116 75%	62/164 38%	170/389 44%	178/423 42%	182/409 44%	<i>Ode</i>
<i>Odf</i>	218/417 52%		178/421 42%	51/115 44%	104/167 62%	176/393 45%	170/407 42%	171/421 41%	<i>Odf</i>
<i>Odg</i>	216/414 51%	224/421 53%		49/115 43%	66/162 41%	283/376 75%	174/424 41%	214/405 53%	<i>Odg</i>
<i>Sof</i>	96/116 83%	66/115 57%	65/115 57%		n.a.	46/115 40%	47/114 41%	57/114 50%	<i>Sof</i>
<i>Sog</i>	77/164 47%	116/167 70%	86/162 53%	n.a.		47/144 33%	57/169 34%	58/161 36%	<i>Sog</i>
<i>Soh</i>	211/389 54%	221/393 56%	303/367 82%	63/115 55%	65/144 45%		165/395 42%	196/372 53%	<i>Soh</i>
<i>Hpd</i>	226/423 53%	217/426 51%	229/424 54%	59/114 52%	76/169 45%	216/395 55%		176/417 42%	<i>Hpd</i>
<i>Hpg</i>	220/409 54%	219/421 52%	250/405 62%	65/115 57%	74/161 46%	225/372 61%	221/417 53%		<i>Hpg</i>
	<i>Ode</i>	<i>Odf</i>	<i>Odg</i>	<i>Sof</i>	<i>Sog</i>	<i>Soh</i>	<i>Hpd</i>	<i>Hpg</i>	

Percent Similarity

Figure 18: Matrix of results of sequence comparisons of single domains with each other. The three letter abbreviations used are the same as in text. Both percent similarity and percent identity are given. The numbers in each square are: number of similar (identical) residues / number of positions compared; percent similarity (identity).

as could be expected from electron micrographs of subunits of molluscan hemocyanins. At both the N-terminal and the C-terminal ends are conserved sequence elements (see Figure 17) which can be tentatively designated as the putative N-terminus and C-terminus of a domain. Since to date there is no N-terminal protein sequence information available from *Odb*, *Odc* and *Odd*, this designation would be useful in defining where a domain ends and a new domain starts as more sequence information becomes available.

2.2. Comparison to other molluscan hemocyanins

Four complete and two almost complete sequences are now known for domains of molluscan hemocyanins. The sequences overall are remarkably conserved. When they are compared with each other they show a high degree of similarity (Figure 17). Ninety residues are identical and forty-four are isofunctional. Sixty-seven additional positions are occupied by identical or isofunctional residues in five out of the six sequences. Single sequences were compared with each other and the results are summarized in Figure 18. Also included in this comparison were the sequences of *Sof* and *Sog*, which are only partially known (110 and 140 amino acid residues respectively).

When we look at the results of the comparison of the *Octopus* domains with each other and with *Hpd* we find a degree of similarity of 51 % to 54 %. Not only do these sequences occupy different positions in the whole polypeptide chain but also belong to representatives of two molluscan classes whose ancestors diverged about 550 to 600 million years ago. It could be expected that domains occupying corresponding positions in their respective polypeptide chains show a higher degree of similarity. If we look at Figure 18, we find that this is indeed the case. For example *Sof* shows a degree of similarity with *Ode* of 82%, but only 51% and 52% with *Odf* and *Odg* respectively. This indicates

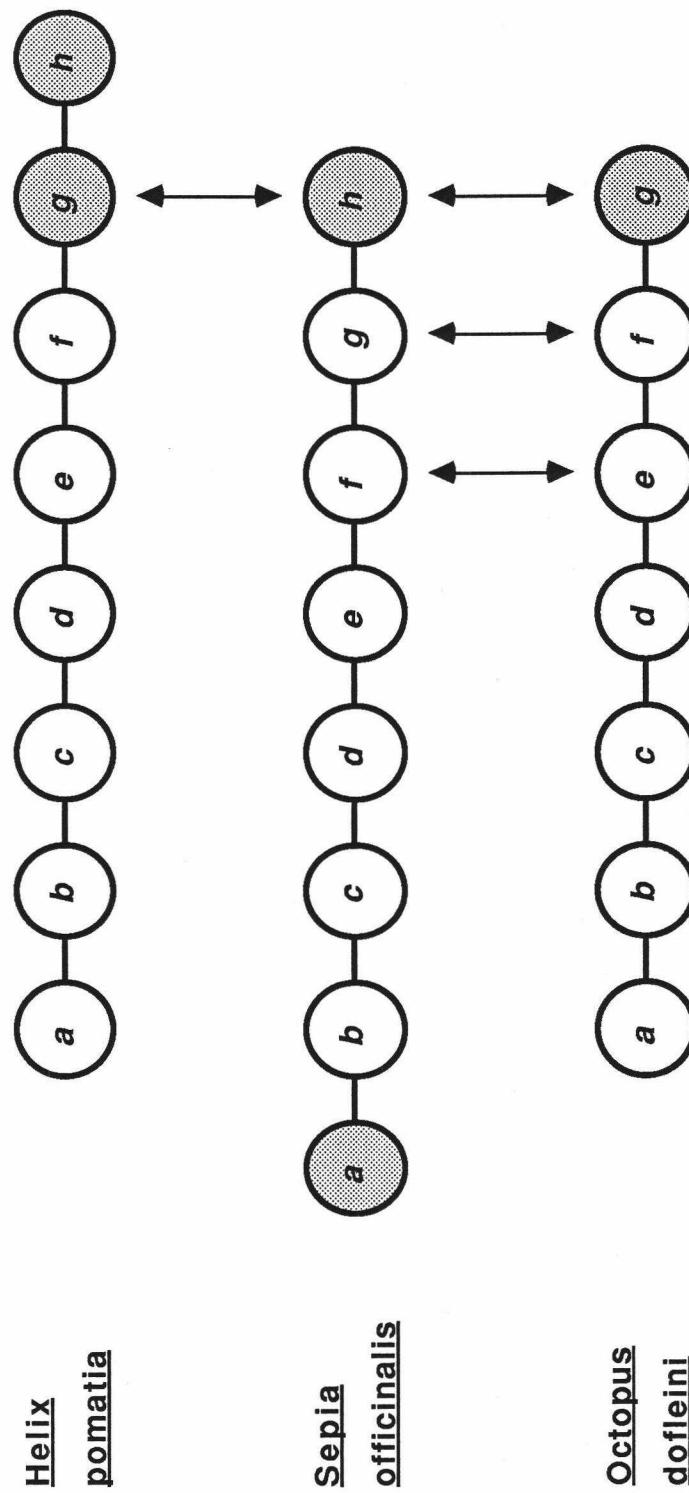


Figure 19: Alignment of domains of three different molluscan hemocyanins. Homologies are indicated by double arrows. Domains which form the wall are white, domains which form the collar are shaded.

that *Ode* and *Sof* are homologous. The same is the case for *Sog* and *Odf* and also for *Soh* and *Odg*. The very high degree of similarity of homologous domains of *Sepia* and *Octopus* reflects the close evolutionary relationship of the two groups. *Hpg* and *Odg* show a degree of similarity of 65%. Both domains are clearly homologues and their lower percentage reflects the fact that the ancestors of both groups diverged a much longer time ago. Figure 19 shows the hemocyanins of *Octopus dofleini*, *Sepia officinalis* and *Helix pomatia* in the "pearls on a string" representation with the homologies between single domains indicated by arrows.

3. Structure of the active site

After the determination of the crystal structure of *Panulirus interruptus* hemocyanin (Gaykema *et al.*, 1984), the structure of the copper binding sites in arthropodan hemocyanins is very well known. Both copper binding sites have a very similar architecture: each copper atom has three histidine ligands, two separated by three amino acid residues with the side chains projecting from the same side of an α -helix, the third furnished by a second α -helix running antiparallel to the first one. The sequences around both sites are highly conserved (Linzen *et al.*, 1985), so it can be safely assumed that the copper binding sites have the same architecture in all arthropodan hemocyanins sequenced so far.

Drexel *et al.* (1987) did not find any similarities between molluscan and arthropodan hemocyanins when they compared the sequence of *Hpd* with the sequences of arthropodan hemocyanins, except for a region which corresponds to the "Copper B" binding site in *Panulirus interruptus* hemocyanin. The sequence in this region is also conserved in tyrosinases. They also reported fairly extensive similarities between the *Helix* sequence and sequences of

tyrosinases around the "Copper B" site and in other parts of the sequence. When I compared the *Octopus* sequences with those proteins I arrived at the same result. It is therefore likely that histidines 174, 178 and 205 (numbered as in Figure 17) serve as ligands for copper in the "Copper B" site. They are marked with asterisks in Figure 20, which shows an alignment of sequences around the "Copper B" site in tyrosinases and molluscan and arthropodan hemocyanins. Site-directed mutagenesis of His 174 and 205 (Figure 20) in *Streptomyces glaucescens* tyrosinase led to inactivation of the enzyme (Huber and Lerch, 1988; Jackman *et al.*, 1989). These mutations affect binding of copper as it has been demonstrated for the His 174 mutant. It contains only one mole of copper per mole of protein (Huber and Lerch, 1988). Mechanism based inactivation of *Neurospora crassa* tyrosinase leads to destruction of the residue corresponding to His 205 and the loss of one copper atom (Dietler and Lerch, 1982). Both results provide fairly solid evidence for the involvement of His 174 and 205 in copper binding in tyrosinases. Since both residues, as well as the regions around them, are well conserved in both tyrosinases and molluscan hemocyanins, they most likely also serve as copper ligands in the latter.

This leaves the question as to what the ligands of "Copper A" are? Avissar *et al.* (1986) probed Northern blots of polyA⁺ RNA isolated from *Levantina hierosolima* and *Sepia officinalis* with an oligonucleotide probe complementary to the sequence coding for the peptide his-his-trp-his-trp-his. This sequence is commonly referred to as the "Copper A" binding site peptide in arthropods, but is actually restricted to chelicerate hemocyanins (Linzen *et al.*, 1985). Avissar *et al.* (1986) observed hybridization of their probe to northern blots of *Sepia* and *Levantina* polyA⁺ RNA and concluded that this sequence is part of the active site in molluscan hemocyanins as well. However this peptide

Figure 20 (page 101): Alignment of sequences around the putative "Copper B" site in hemocyanins and tyrosinases. Histidine residues serving as copper ligands are marked with "*".

Code:

Ysg: *Streptomyces glaucescens* tyrosinase (Huber and Lerch, 1985)

YNc: *Neurospora crassa* tyrosinase (Lerch, 1982)

YM1: mouse tyrosinase 1 (Shibahara *et al.*, 1986)

YM2: mouse tyrosinase 2 (Kwon *et al.*, 1988)

YHs: human tyrosinase (Kwon *et al.*, 1987)

Ece: *Eurypelma californicum* hemocyanin chain e (Schartau *et al.*, 1983)

Ecd: *Eurypelma californicum* hemocyanin chain d (Schneider *et al.*, 1983)

Pia: *Panulirus interruptus* hemocyanin chain a (Bak and Beintema, 1987)

Lp2: *Limulus polyphemus* hemocyanin component II (Nakashima *et al.*, 1986)

Abbreviations used for molluscan hemocyanin functional units same as used in text.

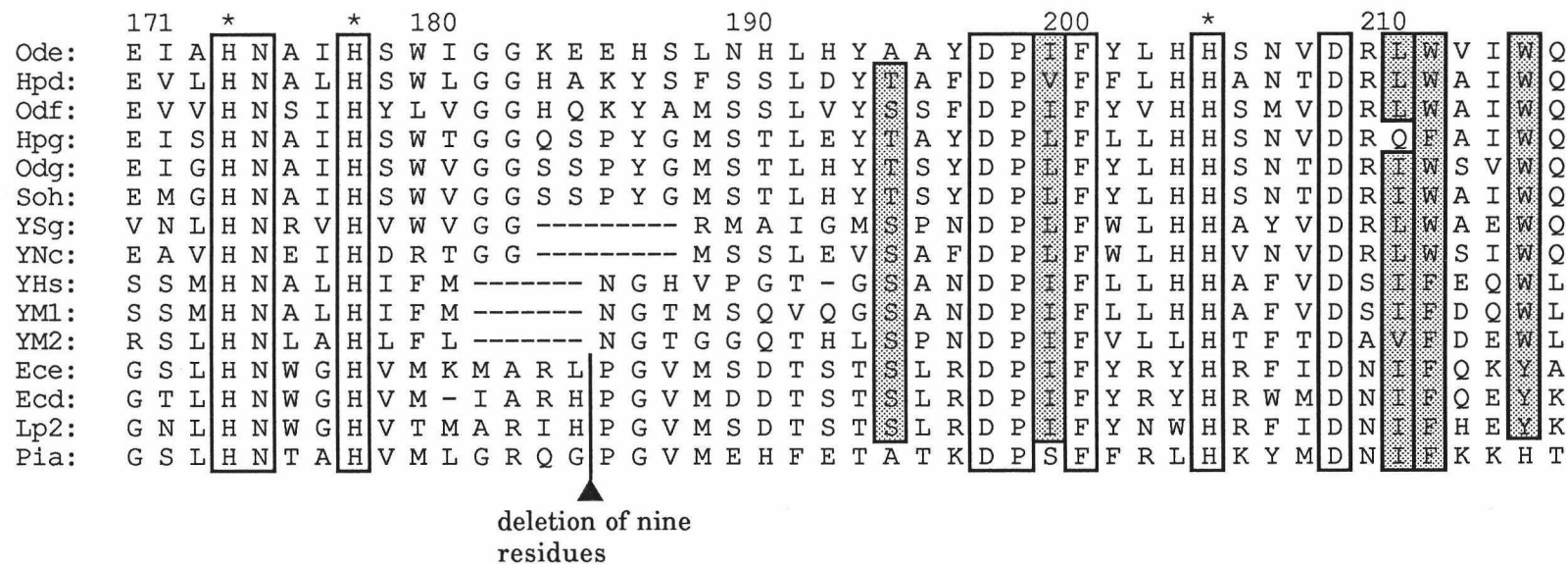


Figure 20.

cannot be found in any molluscan hemocyanin sequence available to date. Nor is a histidine pattern identical to that of the "Copper B" site present elsewhere in those molecules. Drexel *et al.* (1987) found that in the known sequences of *Hpd* and tyrosinases the aminoacid triplet WHR (positions 73 - 75 in Figure 17) is conserved. It is also conserved in all other molluscan hemocyanins and tyrosinases sequenced so far (Figure 21). Site-directed mutagenesis of this histidine in *Streptomyces glaucescens* tyrosinase leads to inactivation of the enzyme (Huber and Lerch, 1986). Determination of the copper content of the mutant protein shows the loss of one copper per mole of protein (Huber and Lerch, 1988). The current spectroscopic evidence (Van Holde and Miller, 1982; Ellerton *et al.*, 1983) points out that the copper atoms in all hemocyanins are liganded by the imidazole nitrogen of two, or possibly three, histidines. It is reasonable to assume that the sidechains involved in copper binding would be conserved in all sequences of molluscan hemocyanins and, therefore, the obvious thing to do is to look for histidines that are conserved in all sequences. Besides His 74 there is only one more conserved histidine residue, His 46, upstream of the "Copper B" site. This residue is also conserved in tyrosinases, except *Neurospora crassa* tyrosinase (Figure 21). His 53 was suggested earlier as another possible ligand (Lang, 1988), but this residue is not conserved. Only one other histidine residue is conserved in all molluscan hemocyanins and tyrosinases: His 204 (see Figure 20). It is positioned next to the putative "Copper B" ligand His 205 and therefore would be very close to the copper binding site, but appears not to be involved in copper binding. This residue is apparently not destroyed by mechanism based inactivation in *Neurospora crassa* tyrosinase (Dietler and Lerch, 1982), which argues against its role as ligand for "Copper A". His 65, which is conserved in all tyrosinases but not all molluscan hemocyanins (Figure 21), was also shown to be involved in copper

Figure 21 (page 104): Alignment of sequences around the putative "Copper A" site in molluscan hemocyanins and tyrosinases. The code is the same as in Figure 20.

	40		+		50		60		+		70		*																															
Hpd:	Y	E	N	I	A	S	F	H	G	K	P	G	L	C	Q	H	-----	E	G	H	K	V	A	C	S	V	S	G	M	P	T	F	P	S	W	H	R	L	Y	V				
Ode:	F	E	A	I	A	S	F	H	A	L	P	P	L	C	P	R	P	-----	T	A	K	H	R	H	A	C	C	L	H	G	M	A	T	F	P	H	W	H	R	L	Y	V		
Sof:	F	E	S	I	A	S	F	H	A	L	P	P	L	C	P	N	P	-----	T	A	K	H	R	Y	A	C	S	L	-	G	M	A	T	F	P	Q	W	H	R	L	Y	V		
Odf:	F	Q	T	I	A	S	Y	H	G	S	T	-	L	C	P	S	P	-----	E	E	P	K	-	Y	A	C	C	L	H	G	M	P	V	F	P	H	W	H	R	V	Y	L		
Hpg:	F	Q	S	I	A	S	F	H	G	S	P	P	G	C	E	H	-----	E	N	H	S	V	-	A	C	S	I	G	G	M	A	N	F	P	Q	W	H	R	L	Y	V			
Odg:	Y	Q	K	I	A	S	Y	H	G	I	P	L	S	C	H	Y	-----	E	N	G	T	A	Y	A	C	C	Q	H	G	M	V	T	F	P	N	W	H	R	L	L	T			
Soh:	Y	Q	K	I	A	S	Y	H	G	M	P	L	S	C	H	Y	-----	P	N	G	T	A	F	A	C	C	Q	H	G	M	V	T	F	P	H	W	H	R	L	Y	M			
YSg:	Y	D	E	F	V	T	T	H	N	A	F	I	I	G	D	-----	T	D	A	G	E	R	T	G	H	R	S	P	S	F	L	P	W	H	R	R	Y	L						
YM1:	F	E	N	I	S	V	T	H	Y	Y	S	V	K	K	T	F	L	G	T	G	Q	E	S	F	G	D	V	D	F	S	H	E	G	P	A	F	L	T	W	H	R	Y	H	L
YM2:	Y	D	L	F	V	W	M	H	Y	Y	V	S	R	D	T	L	L	G	-	G	S	E	I	W	R	D	I	D	F	A	H	E	A	G	P	F	L	P	W	H	R	L	F	L
YHs:	Y	D	L	F	V	W	M	H	Y	Y	V	S	M	D	A	L	L	G	-	G	Y	E	I	W	R	D	I	D	F	A	H	E	A	P	A	F	L	P	W	H	R	L	F	L
YNc:	H	G	M	P	F	K	P	W	A	G	V	P	S	D	T	D	W	S	Q	P	G	S	S	G	F	G	G	Y	C	T	H	S	S	I	L	F	I	T	W	H	R	P	Y	L

	80	▼			90		100																				
Hpd:	E	Q	V	E	E	A	L	L	D	H	G	---	S	S	V	A	V	P	Y	F	D	W	I	S	P	I	
Ode:	V	Q	F	E	Q	A	L	H	R	H	G	---	A	T	V	G	V	P	Y	W	D	W	T	R	P	I	
Sof:	V	Q	F	E	Q	S	L	N	R	H	G	---	A	T	V	G	V	P	Y	T	D	W	T	Y	P	M	
Odf:	L	H	F	E	D	S	M	R	R	H	G	---	S	S	V	A	T	P	Y	W	D	W	T	Q	P	G	
Hpg:	K	Q	W	E	D	A	L	T	A	Q	G	---	A	K	I	G	I	P	Y	W	D	W	T	T	A	F	
Odg:	K	Q	M	E	D	A	L	V	A	K	G	---	S	H	V	G	I	P	Y	W	D	W	T	T	T	F	
Soh:	K	Q	M	E	D	A	M	K	A	K	G	---	A	K	I	G	I	P	Y	W	D	W	T	T	T	F	
YSg:	L	E	F	E	R	A	L	Q	---	S	V	D	A	S	V	A	L	P	Y	W	D	W	S	A	D	R	
YM1:	L	Q	L	E	R	D	M	Q	E	M	L	Q	E	P	S	F	S	L	P	Y	W	N	F	A	T	G	K
YM2:	L	L	W	E	Q	E	I	R	E	L	T	G	D	E	N	F	T	V	P	Y	W	D	W	R	D	A	E
YHs:	L	R	W	E	Q	E	I	R	E	L	T	G	D	E	N	F	T	I	P	Y	W	D	W	R	D	A	E
YNc:	A	L	Y	E	Q	A	L	Y	A	-	S	V	Q	A	D	F	R	A	P	Y	F	D	W	A	S	Q	P

Figure 21.

binding in *Streptomyces glaucescens* tyrosinase by site-directed mutagenesis (Jackman et al., 1989). However, this histidine is linked to a cysteine residue via a thioether bridge in *Neurospora crassa* tyrosinase and is not involved in copper binding in this enzyme. In this enzyme two residues down stream from His 74 have been shown to be involved in copper binding by photoinactivation (Pfiffner and Lerch, 1981). These residues are not conserved in other tyrosinases and molluscan hemocyanins and therefore this tyrosinase could very well be a special case.

Because there are no other conserved histidine residues in molluscan hemocyanins we have to consider now three possibilities:

(i) First, the third ligand for "Copper A" may be a histidine, but different histidines are used in different domains in gastropod and cephalopod hemocyanins. It is difficult to guess which histidine then would be a good candidate since there are several in the vicinity. However, it is hard to accept that in an otherwise highly conserved protein the structure of the active site that is critical for its function would not be conserved.

(ii) The second possibility one has to consider is that the third sidechain involved in "Copper A" binding is another amino acid residue. Cysteine and methionine sidechains are known to serve as copper ligands in other copper proteins like plastocyanin (Colman *et al.*, 1978). Cysteine has been investigated earlier as possible copper ligand, but has been ruled out subsequently (Lontie, 1958). EXAFS studies have excluded cysteine as copper ligand in hemocyanin (Woolery *et al.*, 1984; Brown *et al.*, 1980; Co and Hodgson, 1981).

This leaves methionine as the best possible candidate and indeed there is one methionine (Met 67) which is conserved in all six complete or almost complete sequences known for molluscan hemocyanin domains to date (see Figure 17). It is also conserved in the partially known sequence of *Sof*. As a

close neighbor of His 65 which has been implicated in copper binding in tyrosinases, it could be substituting as a "Copper A" ligand in molluscan hemocyanins. The copper protein plastocyanin was shown to have a cysteine and a methionine sidechain as copper ligands in addition to two histidine sidechains (Colman *et al.*, 1978). The original data have now been refined to a resolution of 1.6 Å. However EXAFS studies are only able to detect cysteine sulfur as a copper ligand (Scott *et al.*, 1982). The lack of evidence for a methionine ligand could be explained by a large Debye-Waller factor due to the much weaker copper thioether bond. Therefore methionine cannot yet be excluded as a possible copper ligand in molluscan hemocyanins. Aspartate and glutamate side chains (there is one conserved glutamic acid residue (Glu 82) present in all sequences in Figure 21) are also known as metal ligands in metalloproteins, for example in hemerythrin, but have not yet been observed as ligands of copper.

(iii) The third possibility would be that there are only two "Copper A" ligands, both of them histidines. There is some experimental evidence that this could be the case. It is found that one of the two coppers in molluscan hemocyanins is removed much more easily by treatment with cyanide than the second one (Himmelwright *et al.*, 1978). This is not the case in arthropodan hemocyanins. It could be thought that because one copper is bound only by two ligands, it is much more easily removed than the other copper which is held by three. However, it cannot be ruled out that folding of the peptide backbone leads to distortion of the coordination geometry which also would weaken the bonding of copper. Also the presence of one weak and two strong ligands instead of three strong ligands could conceivably give the same result. Results from ligand substitution reactions (summarized by Preaux *et al.*, 1988) can

also be interpreted to mean the presence of only two ligands for the "Copper A" in molluscan hemocyanins.

In summary, it can be said that the structure of the active site in molluscan hemocyanin remains an unsolved question. X-ray studies which are currently underway will hopefully provide an answer to this problem in the near future (Cuff *et al.*, 1990).

4. Evolution of Molluscan Hemocyanins

When the sequence of *Hpd* was published, Drexel *et al.* (1987) did extensive comparisons of it with sequences of tyrosinases and arthropodan hemocyanins. The authors of this paper suggested that molluscan and arthropodan hemocyanins evolved independently from a common ancestral mononuclear copper protein. They hypothesized that a gene duplication of this ancestral mononuclear copper protein corresponding to the "Copper B" site, led to evolution of arthropodan hemocyanins and a fusion with a different type of copper binding structure led to the evolution of tyrosinases. Molluscan hemocyanins probably arose from several gene duplications of the new binuclear copper structure.

The results of sequence comparisons of the *Octopus* domains with arthropodan hemocyanins and tyrosinases are consistent with the hypothesis of Drexel *et al.* (1987). The degree of similarity between single domains of molluscan hemocyanins is 51% to 54% when domains are compared with each other that are not homologues. This is consistent with the hypothesis that molluscan hemocyanins arose from a monomeric binuclear copper protein by a series of gene duplications. This event must have happened before gastropods and cephalopods diverged in the early Cambrium about 500 to 600 million years ago (Yochelsen, 1979; Runnegar and Pojeta, 1985). Using the

fraction identity values of Figure 18, the time of the origin of the molluscan eight domain chain can be estimated as 800 to 1000 million years ago. Since it was present before the general diversification of molluscan organisms, it means that all molluscs should potentially be able to produce this protein. It seems that some molluscan organisms like scaphopods lost this ability entirely and others replaced hemocyanin by an other oxygen carrier. For example, most bivalves have red blood cells (Terwilliger *et al.*, 1988), but primitive bivalves, which are also ancient orders, still have hemocyanin (Morse *et al.*, 1986; Terwilliger *et al.*, 1988).

Since the "Copper A" region is somewhat conserved between molluscan hemocyanins and tyrosinases and several strongly conserved residues are found close by in both tyrosinases and hemocyanins, it is likely that a tyrosinase like protein is the ancestor of molluscan hemocyanins. An additional argument in favor of this hypothesis is the fact that molluscan hemocyanins still have a weak tyrosinase activity (Winkler *et al.*, 1981).

However, there are also quite significant differences in the structure of the putative "Copper A" site in tyrosinases and molluscan hemocyanins which overall is less strongly conserved than the "Copper B" site. Therefore it is also possible that tyrosinases and molluscan hemocyanins arose on two different occasions by fusion of a "Copper A" structure to the ancient "Copper B" site. One of these events then led to the evolution of tyrosinases and the other to the evolution of molluscan hemocyanins. The first one of these events must have happened very early since tyrosinases are ubiquitously distributed. But because tyrosinases are such ancient proteins and because it is also easier to postulate only one "Copper A"-"Copper B" fusion event, it is more likely that molluscan hemocyanins evolved from an ancestral tyrosinase-like protein.

Octopod hemocyanins are distinct from other cephalopod hemocyanins in that they have seven oxygen binding domains per subunit instead of eight. It appears that one domain was lost in the course of evolution. Partial proteolytic digestion of *Sepia officinalis* hemocyanin with trypsin yields a six-domain and a two-domain fragment. The latter is composed of domains *Sog* and *Soh* (Gielens *et al.*, 1983). The same treatment done with *Octopus* hemocyanin gives a six-domain fragment and a single domain fragment which was identified as the C-terminal domain (Lamy *et al.*, 1987). It might be thought that therefore the C-terminal domain of the eight domain structure of other cephalopods became deleted to form the seven domain structure. My sequence comparisons, however, show that *Soh* is homologous to *Odg* and also *Sof* to *Ode* (see D.2.1.). Therefore, it cannot be the C-terminal domain of the eight domain structure which became deleted, but instead one that is situated towards the N-terminus, if not the N-terminal domain. Deletion of the N-terminal domain could potentially be without consequence for the assembly of the 60S structure. If the C-terminal domain is deleted in *Octopus dofleini* hemocyanin, the remaining six-domain fragment still can assemble into a ringstructure which is similar to the native structure (Miller *et al.*, 1990). However, recent comparative immunological and spectroscopic studies on *Sepia officinalis* and *Octopus vulgaris* hemocyanin suggest that the analog of *Soe* is missing in *Octopus vulgaris* hemocyanin (Loncke *et al.*, 1990). There is one problem with this result. As depicted in Figure 19, *Soe* is one of the domains that form the wall of the hollow cylinder of the 60S native structure. Precise contacts between domains may be necessary in order to stabilize this structure. One can easily imagine that if *Soe* would be deleted, a large number of these contacts between domains in the wall would have to be modified at the same time in order to allow assembly of the modified chains into the 60S structure. More sequence

information is needed in order to confirm which one of the domains is "missing" in the Octopod hemocyanins. A scheme for the evolution of hemocyanins and tyrosinases is presented in Figure 22.

It should also be noted that in gastropod hemocyanins domains *g* and *h* form the collar and domains *a* through *f* the wall of the 60S structure, whereas in cephalopod hemocyanins domains *b* through *g* form the wall (Figure 19). It could be thought that the ancestral molluscan hemocyanin was composed of nine domains (or maybe more) and that deletion of the N-terminal domain of the nine domain structure led to evolution of gastropod hemocyanins and deletion of the C-terminal one to cephalopod hemocyanins. Alternatively it could be thought the difference in the structure of cephalopodan and gastropodan hemocyanins arose by transposition of one domain from the N-terminus to the C-terminus or vice versa.

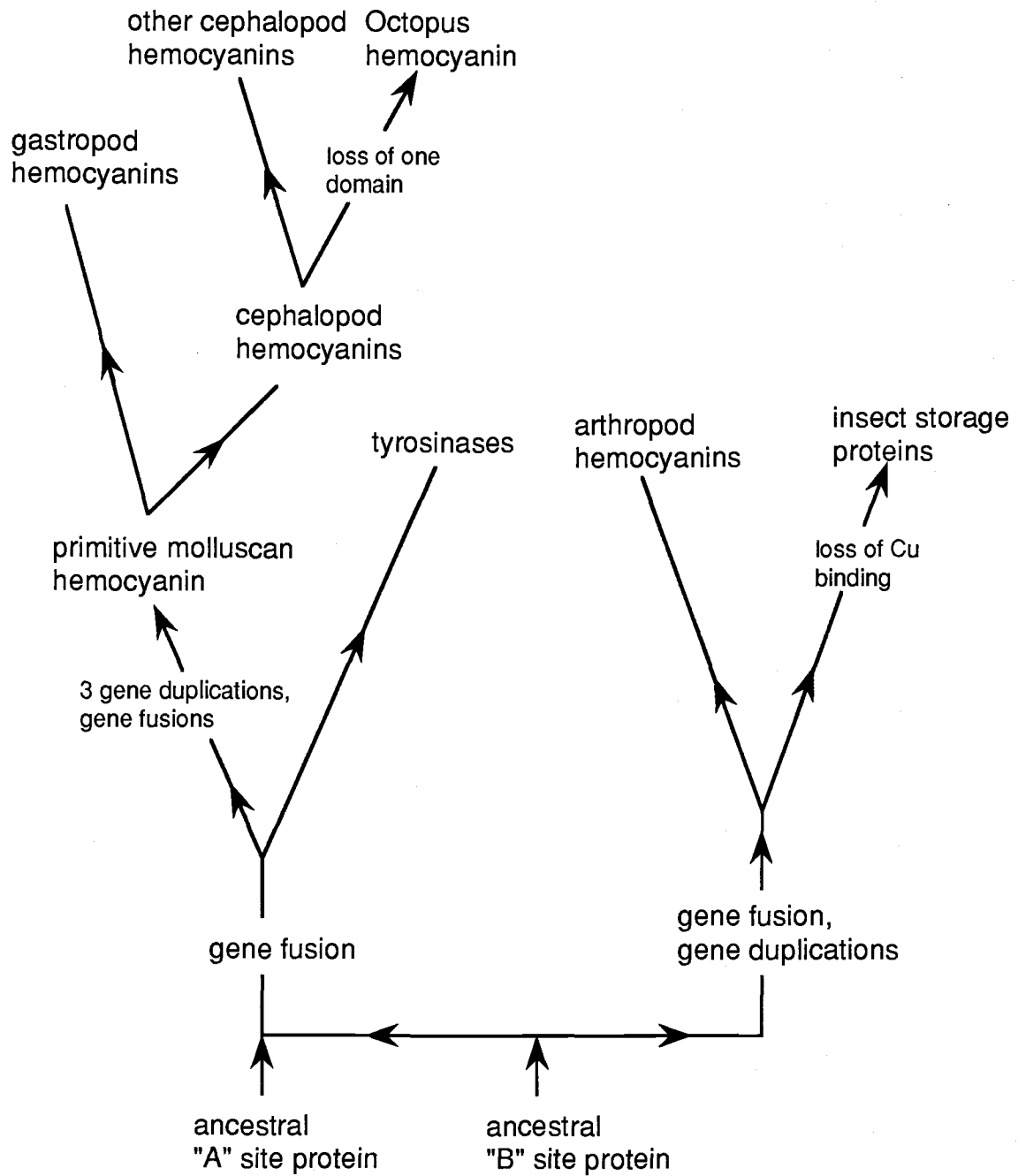


Figure 22: Scheme for the evolution hemocyanins and tyrosinases. Insect storage proteins, which strongly resemble arthropodan hemocyanins (Willott *et al.*, 1989), but are deficient in copper binding, are also included.

E. REFERENCES

- Avissar, I., Daniel, E. and Daniel, V. (1986) *Biochem. J.* **233**, 253-257.
- Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci.* **69**, 1408-1412.
- Bak, H. J. and Beintema, J. J. (1987) *Eur. J. Biochem.* **169**, 333-348.
- Berger, J., Pilz, I., Witters, R. and Lontie, R. (1977) *Eur. J. Biochem.* **80**, 79-82.
- Berger, S. L. and Kimmel, A. R. (1987) *Meth. in Enzym.* **152**.
- Bijlholt, M. M. C., van Bruggen, E. F. J. and Bonaventura, J. (1979) *Eur. J. Biochem.* **95**, 399-405.
- Birnboim, H. C. and Doly, J. (1979) *Nucleic Acids Res.* **7**, 2251-2266.
- Bonaventura, C., Bonaventura, J., Miller, K. I. and van Holde, K. E. (1981) *Arch. Biochem. Biophys.* **211**, 589-598.
- Breitbart, R. E., Nguyen, H. T., Medford, R. M., Destree, A. T., Mahdavi, V. and Nadal-Ginard, B. (1985) *Cell* **41**, 67-82.
- Brenowitz, M., Bonaventura, C., Bonaventura, J. and Gianazza, E. (1981) *Arch. Biochem. Biophys.* **210**, 748-761.
- Brenowitz, M., Bonaventura, C. and Bonaventura, J. (1984) *Arch. Biochem. Biophys.* **230**, 238-249.
- Brown, J. M., Powers, L., Kincaid, B., Larrabee, J. A. and Spiro, T. G. (1980) *J. Am. Chem. Soc.* **102**, 4210-4216.
- Chen, E. Y. and Seeburg, P. H. (1985) *DNA* **4**, 165-170.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) *Biochemistry* **18**, 5249-5299.
- Chou, P. Y. and Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45-148.
- Clewell, D. B. and Helinski, D. R. (1969) *Proc. Natl. Acad. Sci.* **62**, 1159-1166.
- Co, M. S. and Hodgson, K. O. (1981) *J. Am. Chem. Soc.* **103**, 3200-3201.
- Colman, P. M., Freeman, H. C., Guss, J. M., Murata, M., Norris, V. A., Ramshaw, J. A. M., and Venkatappa, M. P. (1978) *Nature* **272**, 319 - 323.
- Cuff, M. E., Hendickson, W. A., Lamy, J., Lamy, J., Miller, K. I., van Holde, K. E. (1990) in *"Invertebrate Dioxygen Carriers"* (Preaux, G. and Lontie, R. Eds) in press.

- Dale, R. M., McClure, B. A. and Houchins, J. P. (1985) *Plasmid* **13**, 31-41.
- Declercq, L., Witters, R. and Preaux, G. (1990) in "*Invertebrate Dioxygen Carriers*" (Preaux, G. and Lontie, R., Eds.), in press.
- Deng, G. and Wu, R. (1981) *Nucleic Acids Res.* **9**, 4173-4188.
- Dickerson, R. E., and Geis, I. (1983) "*Hemoglobins: Structure, Function, Evolution and Pathology*", Benjamin/Cummings Publishing Company, Menlo Park, CA.
- Dietler, C. and Lerch, K. (1982) in "Oxidases and Related Redox Systems" (King, T. E., Morrison, M., Mason, H. S., Eds.) pp. 305 - 317, Pergamon Press, Oxford and New York.
- Dilly, P. N. and Messenger, J. B. (1972) *Z. Zellforsch.* **132**, 193-201.
- Drexel, R., Schneider, H. J., Sigmund, S., Linzen, B., Gielens, C., Lontie, R., Preaux, G., Lottspeich, F. and Henschen, A. in (1986) "*Invertebrate Oxygen Carriers*" (Linzen, B., Ed.) pp 255-258, Springer Verlag, Heidelberg.
- Drexel, R., Siegmund, S., Schneider, H. J., Linzen, B., Gielens, C., Preaux, G., Kellerman, J. and Lottspeich, F. (1987) *Biochem. HS* **368**, 617-635.
- Ellerton, H. D., Ellerton, N. F. and Robinson, H. A. (1983) *Prog. Biophys. Mol. Biol.* **41**, 143-248.
- Eriksson-Quensel, I.-B. and Svedberg, T. (1936) *Biol. Bull.* **71**, 996-998.
- Fahrenbach, W. A. (1970) *J. Cell Biol.* **44**, 445-453.
- Feinberg, A. P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6-13.
- Felsenfeld, G. and Printz, M. P. (1959) *J. Am. Chem. Soc.* **81**, 6259-6264.
- Folkerts, A. and Van Eerd, J.-P. (1981) in "*Invertebrate Oxygen binding proteins: Structure, Active Site and Function*" (Lamy, J. and Lamy, J., Eds.) pp. 153-163, Marcel-Dekker, New York.
- Freedman, T. B., Loehr, J. S. and Loehr, T. S. (1976) *J. Am. Chem. Soc.* **98**, 2809-2815.
- Frenkel, L. and Bremer, H. (1986) *DNA* **5**, 539-544.
- Gaykema, W. P. J., Hol, W. G. J., Vereijken, J. M., Soeter, N. M., Bak, H. J. and Beintema, J. J. (1984) *Nature(London)* **309**, 23-29.
- Ghiretti-Magaldi, A., Milanese, C. and Tognon G. (1977) *Cell Differ.* **6**, 167-186.

- Ghiretti-Magaldi, A., Salvato, B., Tognon, G., Mammi, M. and Zanotti, G. (1981) In *"Invertebrate Oxygen binding Proteins: Structure, Active Site and Function"* (Lamy, J. and Lamy, J., Eds.) pp. 393-404, Marcel-Dekker, New York.
- Gielens, C., Preaux, G., and Lontie, R. (1975) *Eur. J. Biochem.* **60**, 271-280.
- Gielens, C., Bosman, F., Preaux, G. and Lontie, R. (1983) *Life Chem. Reports*, Suppl. **1**, 121-124.
- Gubler, U. and Hoffman, B. J. *Gene* **25**, 263-269.
- Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580.
- Hanahan, D. (1985) in *"DNA Cloning Vol. 1"* (Clover, D., Ed.) pp. 109-135, IRL Press, Oxford, Washington DC.
- Hanahan, D. and Meselson, M. (1983) *Meth. in Enzym.* **100**, 333-342.
- Heidecker, G. and Messing, J. (1983) *Nucleic Acids Res.* **11**, 4891-4907).
- Henikof, S. (1984) *Gene* **28**, 351- 359.
- Hennecke, R., Gellissen, G., Spindler-Barth, M. and Spindler, K.-D. (1990) in *"Invertebrate Dioxygen Carriers"* (Preaux, G. and Lontie, R. Eds.) in press.
- Herskovits, T. T. (1988) *Comp.Biochem.Physiol.* **91B**, 597-611.
- Himmelwright, R. S., Eickman, N. C. and Solomon, E. I. (1978) *Biochem. Biophys. Res. Comm.* **81**, 243- 247.
- Himmelwright, R. S., Eickman, N. C., LuBien, C. D. and Solomon, E. I. (1980) *J. Am. Chem. Soc.* **102**, 5377-5388.
- Huber, M. and Lerch, K. (1985) *Biochemistry* **24**, 6038-6044.
- Huber, M. and Lerch, K. (1986) in *"Invertebrate Oxygen Carriers"* (Linzen, B., Ed.) pp 265-276, Springer Verlag, Heidelberg.
- Huber, M. and Lerch, K. (1988) *Biochemistry* **27**, 5610-5615.
- Huysmans, E. and De Wachter, R. (1988) *Nucleic Acids Res.* **16**, r88-r179.
- Jackman, M., Huber, M. and Lerch, K. (1989) *Highlights Modern Biochem.* **1**, 281-290.
- Jacobson, R. R., Tyeklar, Z., Farooq, A., Karlin, K. D., Liu, S. and Zubieta, J. J. (1988) *J. Am. Chem. Soc.* **110**, 3690- 3692.

- Jekel, P. A., Bak, H. J., Soeter, N. M., Vereijken, J. M. and Beintema, J. J. (1988) *Eur. J. Biochem.* **178**, 403-412.
- Kempter, B. (1983) *Naturwissenschaften* **70**, 255-256.
- Kitajima, N., Fujisawa, K. and Moro-oka, Y. (1989) *J. Am. Chem. Soc.* **111**, 8975-8976.
- Klotz, I. M. and Kurtz, D. M. Jr. (1984) *Acc. Chem. Res.* **17**, 16-22.
- Kwon, B. S., Haq, A. K., Pomerantz, S. H. and Halaban, R. (1987) *Proc. Natl. Acad. Sci.* **84**, 7473-7477.
- Kwon, B. S., Wakulchik, M., Haq, A. K., Halaban, R. and Kestler, D. (1988) *Biochem. Biophys. Res. Comm.* **153**, 1301-1309.
- Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105-132.
- Linzen, B., Soeter, N. M., Riggs, A. F., Behrens, P. Q., Nakashima, H., Takagi, T., Nemoto, T., Vereijken, J. M., Bak, H. J., Beintema, J. J., Volbeda, A., Gaykema, W. P. J. and Hol, W. G. J. (1985) *Science* **229**, 519-524.
- Lamy, J., Bijlholt, M. M. C., Sizaret, P.-Y., Lamy, J. and van Bruggen, E. F. J. (1981), *Biochemistry* **20**, 1849-1856.
- Lamy, J., Lamy, J. N., Leclerc, M., Compin, S., Miller, K. I. and van Holde, K. E. (1986) in *"Invertebrate Oxygen Carriers"* (Linzen, B., Ed.) pp 231-234, Springer Verlag, Heidelberg.
- Lamy, J., Leclerc, M., Sizaret, P.-Y., Lamy, J. N., Miller, K. I., McParland, R. and van Holde, K. E. (1987) *Biochemistry* **26**, 3509-3518.
- Lang, W. H. (1988) *Biochemistry* **27**, 7276-7282.
- Lerch, K. (1982) *J. Biol. Chem.* **257**, 6414-6419.
- Lips, D., Gielens, C., Preaux, G. and Lontie, R. (1981) *Arch. Int. Physiol. Biochem.* **90**, B128.
- Loeb, L. R. (1986) *Ann. Rev. Biochem.* **52**, 482-512.
- Loncke, P., Vanderzande, M., Gielens, C. and Preaux, G. (1990) *"Invertebrate Dioxygen Carriers"* (Preaux, G. and Lontie, R. Eds.) in press.
- Lontie, R. (1958) *Clin. Chim. Acta* **3**, 68-71.
- Lontie, R. (1983) *Life Chem. Reports*, Suppl. **1**, 109-121.

- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1983) *"Molecular Cloning. A laboratory manual"*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Markl, J., Kempter, B., Linzen, B., Bijlholt, M. M. C. and van Bruggen, E. F. J. (1981) *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 1631-1641.
- Messenger, J. B., Muzii, E. O., Nardi, G. and Steinberg, H. (1974) *Nature(London)* **250**, 154-155.
- Micard, D., Sorbier, M. L., Couderc, J. L. and Dastugue, B. (1985) *Anal.Biochem.* **148**,121-126.
- Miller, K. I. and van Holde, K. E. (1982) *Comp.Biochem.Physiol.* **73B**, 1013-1018.
- Miller, K. I., van Holde, K. E., Toumadje, A., Johnson, W. C. Jr., Lamy, J. (1988) *Biochemistry* **27**, 7282-7288.
- Morse, M. P., Meyerhoff E., Otto, J. J. and Kuzirian, A. M. (1986) *Science* **231**, 1302-1304.
- Muzii, E. (1981) *Cell. Tissue Res.* **220**, 435-438.
- Nakahara, A., Shinnichiro, S. and Kino J. (1983) in *Life Chem. Rep.*, Suppl. 1 (Wood, E.J., Ed.), pp 319-322.
- Nakashima, H., Behrens,P. Q., Moore,M. D., Yokota, E., and Riggs, A. F. (1986) *J.Biol.Chem.* **261**, 10526-10533.
- Neutebom, B., Jekel, P. A., Hofstra, R. M. W., Sierdsema, S. J. and Beintema, J. J. (1990) in *"Invertebrate Dioxygen Carriers"* (Preaux, G. and Lontie, R. Eds.) in press.
- Pfiffner, E. and Lerch, K. (1981) *Biochemistry* **20**, 6030-6035.
- Preaux, G., Gielens, C. and Lontie, R. (1979) in *"Metalloproteins: Molecular Function and Clinical Aspects"* (Weser, U. Ed.). pp 73-77, Thieme Verlag, Stuttgart.
- Preaux, G. and Gielens, C. (1984) in *"Copper Proteins and Copper Enzymes"* (Lontie, R., Ed.) Vol. 2, pp 159-205, CRC Press.
- Preaux, G. Vandamme, A., de Bethune, B., Jacobs, M.-P. and Lontie, R. (1986) in *"Invertebrate Oxygen Carriers"* (Linzen, B., Ed.) pp 485-488, Springer Verlag, Heidelberg.
- Preaux, G., Gielens, C., Witters, R. and Lontie, R. (1988) *Bull. Soc. Chim. Belga.* **97**, 1073-1044.

- Runnegar, B. and Pojeta, J. Jr. (1985) in "The Mollusca Vol. 10: Evolution" (Trueman, E. R., Clarke M. R., Eds.) pp 1 - 57, Academic Press, Orlando, Fla.
- Ruth, P., Schipp, R. and Klussendorf, B. (1988) *Zoomorphol.* **108**, 1-11.
- Salvato, B., Jori, G., Piazzese, A., Ghiretti, F., Beltramini, M. and Lerch, K. (1983) in *Life Chem. Rep.*, Suppl. **1** (Wood, E.J., Ed.), pp 313-318.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc.Natl.Acad.Sci.USA* **74**, 5463-5467.
- Schartau, W., Eyerle, F., Reisinger, P., Geisert, H., Storz, H. and Linzen, B. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 1383-1397.
- Schipp, R., Hoehn, P. and Ginkel, G. (1973) *Z. Zellforsch.* **139**, 253-269.
- Schneider H.-J., Drexel, R., Feldmaier, G., Linzen, B., Lottspeich, F. and Henschen, A. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 1357-1381.
- Schneider, H.-J., Voll, W., Lehmann, R., Grisshammer, R., Goettges, A. and Linzen, B. (1986) in "*Invertebrate Oxygen Carriers*" (Linzen, B., Ed.) pp 173-176, Springer Verlag, Heidelberg.
- Scott, R. A., Hahn, J. E., Doniach, S., Freeman, H. C. and Hodgson, K. O. (1982) *J. Am. Chem. Soc.* **104**, 5364-5369.
- Senkbeil, E. G. and Wriston, J. C., Jr. (1981) *Comp. Biochem. Physiol.* **68B**, 163-171.
- Sheen, J.-Y. and Seed, B. (1988) *BioTechniques* **6**, 942-944.
- Shibahara, S., Tomita, Y., Sakahura, T., Nager, C., Chaudhuri, B. and Mueller, R. (1986) *Nucleic Acids Res.* **14**, 2413-2427.
- Sminia, T. (1977) in "*Structure and Function of Haemocyanin*" (Bannister, J. V. Ed.) pp. 279-288, Springer Verlag, Heidelberg.
- Solomon, E. I. (1981) in "*Copper Proteins*" (Spiro, T. G. Ed.) pp. 41-108, Wiley Interscience, New York.
- Sorrel, T. N. (1989) *Tetrahedron* **45**, 3- 68.
- Stenkamp, R. E., Sieker, L. C. and Jensen, L. H. (1984) *J. Am. Chem. Soc.* **106**, 618-622.
- Struhl, K. (1985) *Biotechniques* **3**, 2-3 .
- Tabak, H. F. and Flavell, R. A. (1978) *Nucl. Acids Res.* **5**, 2321-2332.

- Tabor, S. and Richardson, C. C. (1986) *Proc. Natl. Acad. Sci.* **84**, 4767-4771.
- Takagi, T. (1986) in *"Invertebrate Oxygen Carriers"* (Linzen, B., Ed.) pp 259-262, Springer Verlag, Heidelberg.
- Terwilliger, N. B., Terwilliger, R. C., Meyerhoff, E. and Morse, M. P. (1988) *Comp. Biochem. Physiol.* **89B**, 189-195.
- Thaman, T. J., Loehr, J. S. and Loehr, T. M. (1977) *J. Am. Chem. Soc.* **99**, 4187-4189.
- Thomas, P. S. (1980) *Proc. Natl. Acad. Sci.* **77**, 5201-5205.
- Top, A., Gielens, C., Witters, R., van Beumen, J. and Preaux, G. (1990) in *"Invertebrate Dioxygen Carriers"* (Preaux, G. and Lontie, R. Eds) in press.
- Van Breemen, J. F. L., Wichertjes, T., Muller, M. F. J., van Driel, R. and van Bruggen, E. F. J. (1975) *Eur. J. Biochem* **60**, 129-378.
- Van Eerd, J.-P., and Folkerts, A. (1981) in *"Invertebrate Oxygen binding proteins: Structure, Active Site and Function"* (Lamy, J. and Lamy, J., Eds.) pp. 139-149, Marcel Dekker, New York.
- Van Holde, K. E. and Miller, K. I. (1982) *Q. Rev. Biophys.* **15**, 1-129.
- Vinogradov, S. N. (1985) *Comp. Biochem. Physiol.* **82B**, 1-15.
- Voit, R. and Schneider, H.-J. (1986) *Eur. J. Biochem.* **159**, 23-29.
- Wichertjes, T., Gielens, C., Schutter, W.G., Preaux, G., Lontie, R. and van Bruggen, E. F. J. (1986) *Biochim. Biophys. Acta* **872**, 183-194.
- Williot, E., Wang, X.-Y. and Wells, M. A. (1989) *J. Biol. Chem.* **264**, 19052-19059.
- Winkler, M. E., Lerch, K. and Solomon, E. I. (1981) *J. Am. Chem. Soc.* **103**, 7001-7003.
- Wood, D. J. (1984) *Focus* **4**, 1-3.
- Wood, E. J. (1977) *"Structure and Function of Haemocyanin"* (Bannister, J. V, Ed.) pp. 77-84, Springer Verlag, Heidelberg.
- Wood, E. J., Corfield, G. C. and Siggins, K. W. (1981) *Comp. Biochem. Physiol.* **69B**, 877-882.
- Wood, E. J. and Siggins, K. W. (1981) in *"Invertebrate Oxygen binding proteins: Structure, Active Site and Function"* (Lamy, J. and Lamy, J., Eds.) pp. 825-829, Marcel Dekker, New York.

- Woolery, G. L., Powers, L., Winkler, M., Solomon, E. I. and Spiro, T. G. (1984) *J. Am. Chem. Soc.* **106**, 86-92.
- Xin, X.-Q., Gielens, C., Witters, R. and Preaux, G. (1990) in *"Invertebrate Dioxygen Carriers"* (Preaux, G. and Lontie, R., Eds.), in press.
- Yochelsen, E. L. (1979) in *"The Origin of Major Invertebrate Groups"* (House, M. R., Ed.) pp 323-358, Academic Press, New York.
- Zagursky, R. J., Baumeister, K., Lomax, N. and Berman, M. L. (1985) *Gene Anal. Techn.* **2**, 89-94.
- Zasloff, M., Ginder, G. D. and Felsenfeld, G. (1978) *Nucleic Acids Res.* **5**, 1139-1152.