The Molecular Evolution of Arthropod & Molluscan Hemocyanin

Evidence for Apomorphic origin and convergent evolution in O2 hinding sites

Anupam N&am, Jimmy Ng and Trustin Ennacheril

December 1, 1997

Abstract

The similarities between the oxygen binding sites in Arthropod and Molluscan hemocyanin have lead to debate in the biological community concerning how this molecule evolved. Evidence, such as sequence similarities as well as mRNA hybridization experiments suggest that of the two copper binding sites present in hemocyanin, the Cu-B site arose from apomorphic origins while the Cu-A site most probably arose independently and subsequently in the two different lineages. Furthermore, the study of hemocyanin has significant relevance for the field of evolutionary biology because it can serve as a model for studies in molecular evolutionary processes.

The similarities between oxygen carriers of arthropods and molluscs known as hemocyanins, have led to the suggestion that the molecules had a very early common ancestor and have diverged, while another site on the molecule experienced convergent evolution. Common ancestry may be suggested by high degrees of sequence similarity between the active sites or by cladistic analysis showing monophyly. Convergent evolution may be suggested by a different origin at the nucleotide or amino acid sequence in the molecule. Evidence from analysis of sequence similarities can illustrate such a different origin. This topic is important since it is a good model for investigating both common ancestry and convergence. The evolution of this molecule is a unique one since one cannot adopt the sole view of either common ancestry or convergence, one must encompass both views. Evidence suggests that of the two copper binding sites present in hemocyanins , designated Cu-A site and Cu-B site, one converged while the other diverged.

Background

Hemocyanins are proteins that use copper binding sites to bind and transport oxygen in a variety of arthropods and molluscs. Typically, hemocyanins have high molecular weights and multiple subunits. Also, pH, temperature, and ionic concentration modulate the oxygen affinity of the molecule. The subunits of hemocyanin tend to aggregate.

The major reason for this aggregation in arthropod hemocyanin is that the molecules are not encapsulated in blood cells. This aggregation allows a high concentration without a corresponding increase in osmolarity. The hemocyanin found in

crustaceans originates in hemocytes called cyanoblasts. These structures then mature into cyanocytes and release hemocyanin into the hemolymph(Gupta, 1979:7 17-724). In general, it is less efficient than hemoglobin in oxygen carrying capacity, but this suits it to the less active crustacea that live in aqueous environments. From an evolutionary standpoint, the emergence of a tracheal system made hemocyanin obsolete. This fact is well-illustrated in insects. For example, land animals with poor tracheal systems still use hemocyanin to a certain degree (Gupta,l979: 717-724).

Arthropod hemocyanin is comprised of 6 or multiples of 6 subunits (see Figure 3&3b). For example, in crayfish, lobsters and crabs there are hexameric and dodecameric structures. There is a highly heterogeneous subunit structure and each subunit is made up of three domains. Oxygen is bound to the active site in the second domain. The first and third domains are most likely responsible for hemocyanin binding only molecular oxygen (Ellerton and Husain,l991:37-48). There are two copper atoms for every molecule which is 75,000 daltons. So, any polypeptide chain of arthropod hemocyanin binds one molecule of 02 (Waxman, 1975).

The structure of molluscan hemocyanin is quite different from that of arthropods, in weight, subunit structure and oxygen binding capacity. The molluscan version is 290,000 daltons, with two copper atoms for every 50,000 daltons. So one polypeptide chain binds 6-8 02 molecules. There is some degree of sequence similarity within the structure, which would indicate that the sequence was duplicated 6 times within the 290,000 daltons (Waxmanl975). There are usually 8 functional units (8 oxygen binding sites). The structure of molluscan hemocyanin is that of a partly hollow cylinder with a collar, constructed from decamers of large polypeptide chains (see Figures 2&4). This

molecule can exist in a number of discrete states. For example in gastropods, there are three polypeptide chains. All mollusc hemocyanin domains are derived from a common ancestral gene(Ellerton and Husain,l991:37-48). The copper atoms are bound directly to the amino acid side chain as opposed to the metal being bound to a prosthetic group as in hemoglobin (Bonaventura and Bonaventura,l983:26-29).

Further differences between arthropod and molluscan hemocyanin subunits are revealed by hydrogen peroxide treatments. These experiments illustrate differences in the reactivity towards the active site ligands and therefore differences in the active site geometries of arthropod and molluscan hemocyanins. They differ in their molecular architecture (Topham et.al 1988). Removal of copper from hemocyanin by cyanide ions is different between arthropods and molluscs. The second copper from molluscan hemocyanin is both slower and more difficult to remove, whereas all the coppers come off at once in arthropods (Van Holde and Miller, 1995:47).

Evolution of Arthropod and Molluscan Hemocyanins

The evolution of arthropod hemocyanin encompasses the development of a hexameric structure as well as events required to make an effective oxygen transporter. Similarities in sequence and quarter-nary structure, X-ray diffraction and immunological cross reactivity illustrate that arthropods diverged about 600 million years ago and the hexameric structure of its hemocyanin occurred before this divergence. An interesting fact about the evolution of this molecule in arthropods is that there is a tendency to evolve a simpler structure, such as hexameric or dodecameric (6x2) structures as opposed to 6x4 structures common among primitive arthropods. There is sequence similarity between arthropod hemocyanin and larval storage proteins (LSP), also known as insect storage

proteins, which lack the copper binding site. This could mean that LSPs are relics of the precursor to arthropod hemocyanin or degenerate versions of it. When the tracheal system developed, it is possible that hemocyanin lost its transport function. Then LSPs subsequently acquired their new function as storage proteins. Two evolutionary events were required to get arthropod hemocyanin to be an effective oxygen transport protein. First, duplication and fusion of genetic elements of the ancestral binding protein generated an oxygen binding site that utilized two copper atoms. Evidence of this will be presented later on. The second was the development of subunit interaction that could lead to multisubunit structure and allosteric behavior. This subunit interaction was necessary for the molecule to allow for cooperative binding, which is important in uptake and delivery of oxygen. Evolution of hemocyanin in arthropods is different from that in molluscs evolution (Van Holde and Miller, 1995:66-68).

Molluscan evolution of hemocyanin shows dramatic changes in the internal organization since the multiunit polypeptide chain of molluscan hemocyanin first evolved. Three successive rounds of gene duplication can explain the 8 functional units mentioned previously. The time of origin of molluscan hemocyanin is about 700 million years ago. There is very little sequence similarity between molluscan and arthropod hemocyanin, which suggests that the divergence occurred early on between the ancestral lines of the two phyla.. This is evidence that the two proteins arose in independent events but at about the same time (Van Holde and Miller, 1995: 66-68).

Predictions

Arthropod and molluscan hemocyanins have two copper binding sites known as the Cu-A site and the Cu-B site; whether these sites evolved independently or diverged

from an ancestral oxygen transport protein is an important question. Sequence similarities lead to the prediction that there is a single common ancestor (see Figure la). However, due to significant differences in subunit structure, convergence might also be the answer (see Figure lb).

A third possible prediction is that the sites evolved differently such that the Cu-A site in the arthropod and molluscan hemocyanin converged while the Cu-B site is derived from a common ancestor.

Evidence

Sequence Similarity Evidence for the Common Ancestry in the Cu-B Site

A sequence study on the primary structure of the mollusc *Helix pomatia,* allowed for the comparison between molluscan hemocyanin, of which only pieces have been sequenced due to its size, and arthropod hemocyanin which has already been sequenced. The results showed that there were no similarities between arthropod and molluscans hemocyanins for most of the sequence. There was only one region where the helix was aligned easily, which represented the Cu-B site (Drexel et.al 1988: 255).

Sequence Similarity with the Molecule Tyrosinase

With regard to comparisons of the Cu-B site, it is important to consider another molecule, known as tyrosinase, which is similar to both arthropod and molluscan hemocyanins in different sequence similarity experiments. Tyrosinase is a copper-containing monooxygenase that catalyzes the hydroxylation of monophenols. It has a very similar copper active site, but the two proteins serve very different functions. One is for oxygen transport while the other is for monoxygenase activity. An important piece of evidence is that there is a high degree of sequence identity between the Cu-B sites in

arthropods and tyrosinases, and Cu-A sites from molluscs with tyrosinases (see figure 8) (Huber and Lerch,l986:265-276). The copper binding domains at the amino- and carboxy- terminal ends in tyrosinase make it believable that arthropod hemocyanin has evolved from a common ancestor of tyrosinase (Lerch, 1986:86-90).

Sequence comparison demonstrates that the hemocyanins of both molluscs and arthropods are at best distantly related to other classes of oxygen transporters. The molluscan hemocyanin and arthropod hemocyanin, differ in their quartemary structure and there is little sequence identity between molluscan functional units and arthropod subunits. It appears that only the Cu-B sites are derived from the same primitive sequences, whereas again the A-site looks like one of the binding sites in tyrosinases (see Figures 7 & 8). The conclusion from this comparison is that the two classes of proteins share one copy of the B-site sequence from some ancestral metal binding protein (Van Holde and Miller, 1995:64-66). Further evidence comes from an experiment where the amino acid composition of two tyrosinases, two arthropod hemocyanins and one mollusc hemocyanin were compared. The results were that all the proteins show a highly identical region of 56 amino acids in the C terminal, which contains three unchanging histidines. These histidines have been identified as the ligand of copper atoms for the Cu-B site (see Figure 7). This B site ligand of 3 histidines is arranged in the order of His – X-X-X- His – 37 amino acids – His, causing a loop to be formed when all three bind copper. It may be concluded that since all sequences contain such a region, it is highly conserved, and the B site of tyrosinases and hemocyanin must have originated during the earliest period of life (Lerch *et al.* Also, sequence identities indicate that molluscan

hemocyanin and tyrosinases diverged much later than arthropod and molluscan life. This is evident in the fact that they share a slightly similar A site.

There is evidence that arthropod hemocyanin preceded tyrosinase. This theory comes from the internal sequence identity of these molecules. Arthropods have an internal sequence identity of 26% and tyrosinases have one of 42%. Yet it is unspecified what parts of the genes are compared. If the rate of amino acid substitution is similar for both, then the arthropods have been subject to this substitution longer since it has a smaller sequence similarity. This disproves any theory that tyrosinase is the common ancestor for hemocyanins of the two different phyla (Volbeda and Ho1 1989). However it is difficult to prove that both molecule had the same rate of substitution.

This idea is in line with the molecular clock theory, which states that if the mutation rate is constant populations should evolve at a constant rate (Lewin ,1997 : 118). Even though there are many difficulties with the molecular clock theory, it seems that the evolutionary clock is neither as good nor as bad as people had suspected. Thus with the appropriate changes, it seems that a stochastic clock can become increasingly accurate as the time span increases (Lewin ,1997 : 118). Thus if the appropriate molecular clock can be identified, it can be demonstrated that the differences in sequence similarity between tyrosinase and arthropods sites results from the fact that arthropods diverged earlier in their lineages (Lewin, 1997: 118).

RNA Hybridization of Arthropod and Molluscan Hemocyanin

A similar experiment illustrated that the mRNA from both phyla would hybridize to a sequence of oligonucleotides that encodes for the copper binding site. The duplexes

that form between the oligonucleotide sequence with the arthropod mRNA and between the oligonucleotide sequence with the molluscan mRNA have similar melting temperatures which are 46.5 ‘C and 44.5’C respectively (Avissar, *et al.* 1986). In hybridization experiments bonds melt at higher temperatures in hybrids with few dissimilarities due to more hydrogen bonding, so the high melting points of the strands indicate a high degree of sequence identity(Lewin, 1997:22-23).

Evidence for Convergence of the Cu-A site

The convergence of the Cu-A sites in arthropod and molluscan hemocyanin is evident by very little sequence similarity, indicating large differences in amino acid composition, although they share a functional similarity. With regard to the experiment that compared two tyrosinases, two arthropod hemocyanins, and one molluscan hemocyanin, there was little sequence similarity for the A site. So three altered sequences of the site arose through evolution and while arthropod and molluscan B sites show no similarity, the molluscan and tyrosinase A sites do show a limited similarity. (Lerch *et al.* 1986:2 13-217). Unlike the high degree of sequence identity found in the Cu-B site, the Cu-A site has a different ligand environment in all the different molecules.

However, the A sites in each had different mechanisms that brought them into existence. In arthropods, there was probably a sequence duplication of the B site to make the A site. Evidence for this comes again from the fact that arthropods had an internal sequence similarity of 26%. If the Cu-A site and the Cu-B site were superimposed, the copper ions would be in the same places as well as 26% of the amino acids (8 out of 3 1) including the histidine ligands. Further evidence comes from a high degree of structural similarities between the two copper binding helical pairs in arthropod hemocyanin. A

factor known as the pseudo-2 fold axis of the helices relating the Cu-A site to the Cu-B site is observed (see Figure 5). This fold is unchanging between different kinds of arthropods. This pseudo-symmetry is a result of a gene duplication that had left a high degree of structural similarity but not sequential similarity (Volbeda and Hoi, 1989). It is likely that the A site came from the B site as explained, rather than the A site and the B site coming from a common predecessor (see Figure 6). The high amount of sequence identity, 26% over several hundred million years, strongly suggest that the B site has been duplicated to make the A site. This data decreases the possibility of a predecessor molecule having both sites because the two sites would show low sequence identity whereas here they show a moderate degree of similarity. If however the predecessor molecule did have a high degree of sequence similarity between the two sites, this would indicate that the B sites were already duplicated in this pre-ancestor molecule.

There was also an unidentified ancestral A protein, which fused with the B site to form the Cu-A site in molluscan hemocyanin. The best way to prove a gene fusion is to find the ancestral A protein and compare it with the modern Cu-A site. Tyrosinase inherited its A site from this ancestral A protein, which could explain the molluscan A sites ability to hybridize with tyrosinase segments. Thus the evidence suggests that the A site evolved independently and subsequently of the B site (VanHolde and Miller, 1995:64-66).

Conclusion

In summary, the arthropod hemocyanin inherited its Cu-B site from an ancestral protein and duplicated this site to make the Cu- A site. The molluscan hemocyanin inherited its B site from the same ancestral protein and fused it with an ancestral A

protein to make its Cu-A site (see Figure 9). The tyrosinase molecule inherited both its binding sites from ancestral proteins. The evidence presented thus far clearly illustrates that the Cu-B sites of arthropods and molluscs are similar in sequence. The vast differences between the two Cu-A sites in the different phyla make it unlikely that they shared an apomorphic origin, especially in light of the similarities between the molluscan Cu-A site and tyrosinase. Sequence similarities and the weak tyrosinase-like activity of the molluscan hemocyanins point to an apomorphic origin between tyrosinase and molluscan hemocyanins of the A sites.

Alternate theses are not supported by the evidence presented. For example, the thesis that states that both molecules evolved from a single common ancestor is incorrect since the Cu-A sites in arthropods and molluscs show little sequence identity but high functional similarity. This indicates convergence. Convergence is not the single explanation either since the Cu-B sites of both shows a very strong similarity in sequence. Another thesis concerning the sugars bound to hemocyanins, which are glycoproteins, being responsible for the difference in sizes is also implausible since the subunit structure of both molecules are vastly different(Van Holde and Miller 1995: 13

Proposed Future Experiments

Proposals for future investigations are to determine the identity of and function of the ancestral Cu-A site protein, and the factors that necessitated the development of two different copper binding sites. Also, it would be interesting to know which site has a stronger affinity for Cu and for oxygen. Further analysis between LSPs and arthropod hemocyanin or between tyrosinases and molluscan hemocyanins may be very helpful.

For the most part this has been a phenetic analysis using sequence similarity to illustrate ancestry. However sequence similarity does not conclusively prove ancestry conclusively. Even though “molecules can reconstruct the phylogeny with high degree of accuracy,” in order to prove homology, and not mere similarity, a cladistic analysis is necessary (Lewin, 1997: 36). Cladistics aims at reconstructing evolutionary histories, such as investigating shared apomorphies and morphologies to illustrate ancestry. Unfortunately, no such cladistic analysis has been performed.

REFERENCES

Avissar, I., Daniel, E., and Daniel, V. (1986). *Biochem. J., 233,253-257"*

Bonaventura, C and Bonaventura, J* (1983). 1n "The Mollusca Volume 2" (P_ W. Hochachka, ed.), pp.26-29. Academic Press, NewYork

Drexel, R., Schneider, H. -J., Sigmund, S., Linzen, B., Gielens, C., Lontie, R., Preaux, G., Lottspeich, F., and Henschen, A. (1986) 1n "Invertebrate Oxygen Carriers" (Linzen, B. Ed.), pp.255-258. Springer-Verlag, New York.

Ellerton, D. H., Husain, N.R. (1991). 1n "Structure and Function of Invertebrate Oxygen Carriers" (S. N. Vinogradov and 0. H. Kapp, eds.), pp. 37-48. Springer-Verlag, New York.

Gupta, A. P. (1979). 1n "Arthropod Phylogeny" (A. P. Gupta, ed.), pp.669-735. Van Nostrand Reinhold Company? New York.

Hamilton, M. G., Herskovits, T. T., and Wall, J+ S. (1991). 1n "Structure and Function of Invertebrate Oxygen Carriers" (S. N. Vinogradov and 0. H. Kapp, eds.), pp. 25-32. Springer-Verlag, New York.

Huber, M., Lerch, K. (1986). 1n "Invertebrate Oxygen Carriers" (Linzen, B. Ed.), pp.265-276. Springer-Verlag, New York.

Lambert, O., Taveau, J. -C., Boisset, N., and Lamy, J. N. (1995). *Arch. Biochem. Biophys. 319,231-243.*

Lerch, K. (1986). In "Advances in Cell Pigment Research" (J. T. Bagnara, ed.), pp.85-98. Alan R. Liss Inc., New York.

Lerch, K., Huber, M., Schneider, H.J., Drexel, R., and Linzen, B. (1996). *J. Inorg. Biochem. 26,213-217.*

Lewin, R. (1997). *In* "Patterns in Evolution" pp.22-23. Scientific American Library, New York.

Mangum, C. P., and Godette, G. (1986). 1n "Invertebrate Oxygen Carriers" (Linzen, B. Ed.), pp.277-280. Springer-Verlag, New York.

Topham, R. W., Tesh, S., Bonaventura, C., and Bonaventura, J. (1988). *Arch. Biochem. Biophys. 261, 299-3*11.

van Holde, K. E. and Miller, K.I. (1995). In "Advances in Protein Chemistry" (C. B. Anfinsen, J. T. Edsall, F. M. Richards and D. S. Eisenburg, eds.), pp. 1-81. Academic Press, New York.

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

Waxman, L. (1975). *J. Biol. Chem.*

Figure 1a

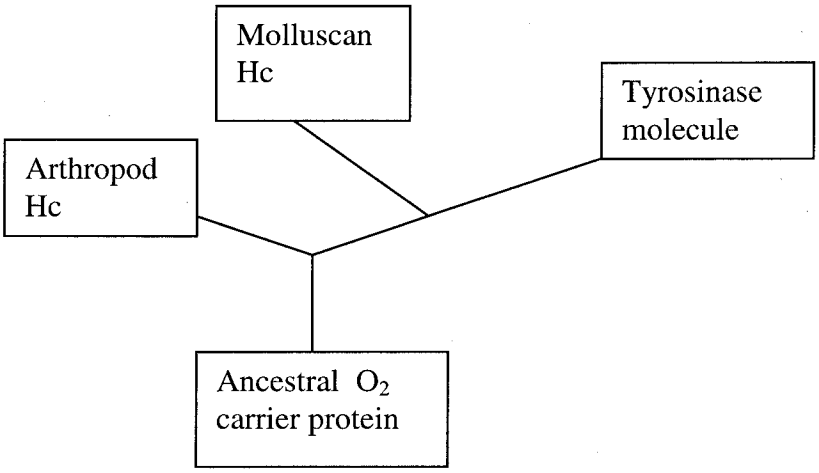
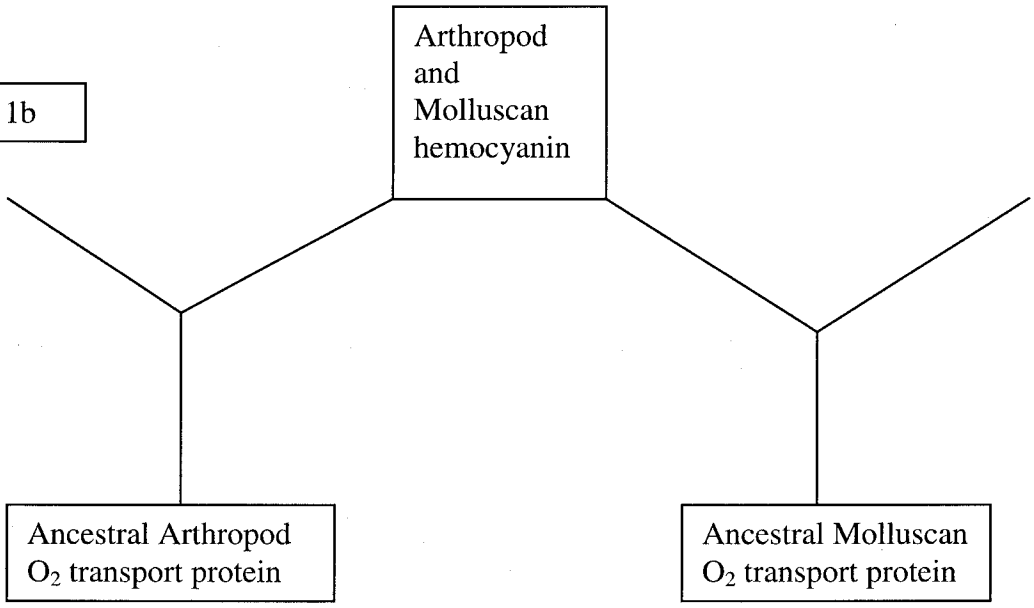


Figure 1b



Figures 1a is a tree for the possible thesis that arthropods and molluscan oxygen carriers come from a common ancestor.
Figures 1b is a tree for the possible thesis that arthropods and molluscan oxygen carriershave converged..

THREE-DIMENSIONAL RECONSTRUCTION OF *Nucula* HEMOCYANIN

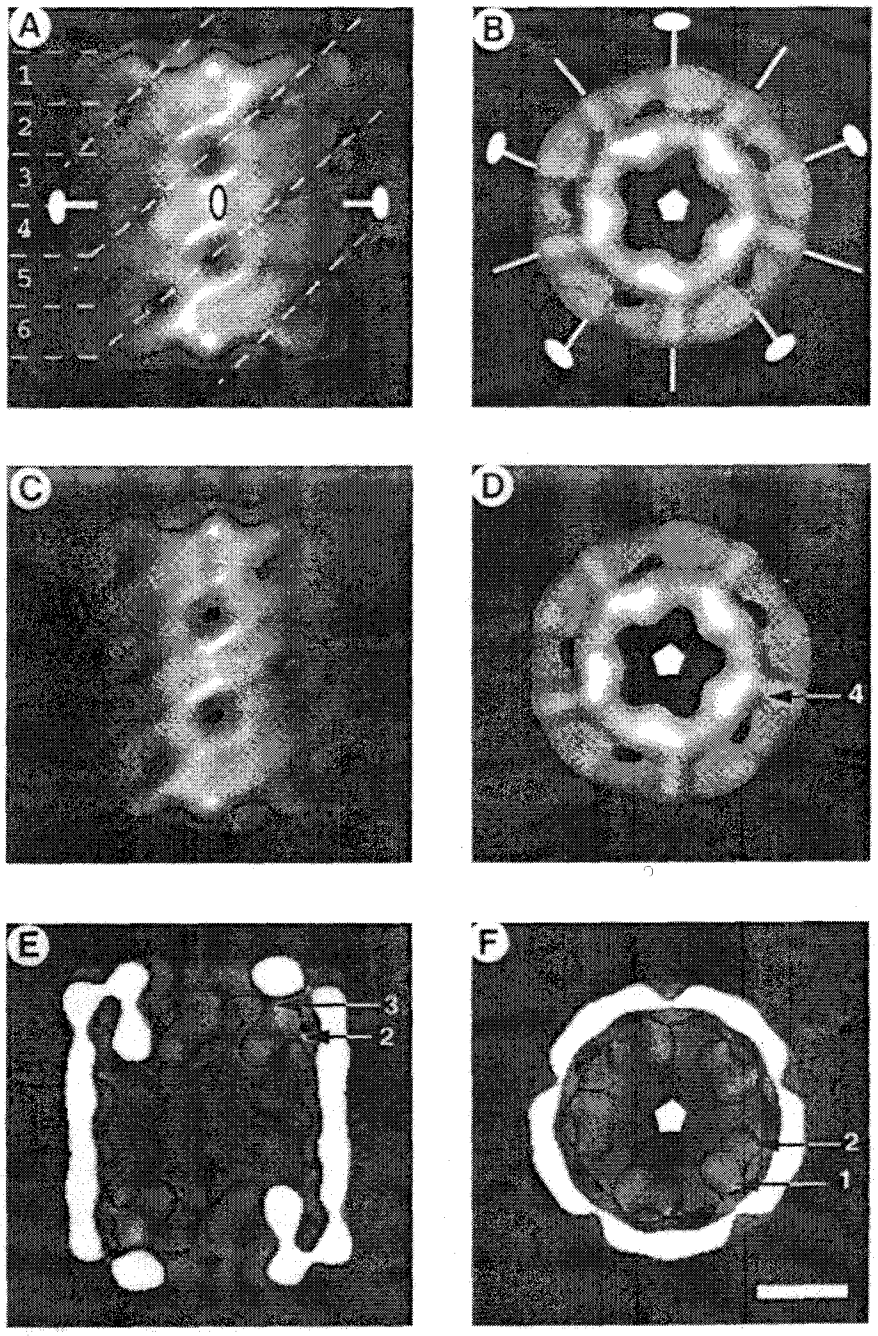


FIG. 2. Computer generated 3D reconstruction of *Nuclula hanleyi* hemocyanin seen from the side (A, C, E) and from the end (B, D, F). Arrows 1 and 2 indicate the arch-wall connections, arrow 3 indicates the arch-collar connection, and arrow 4 indicates the collar-wall connection. The scale bar in F is 10nm. (Lambert, et al., 1995)

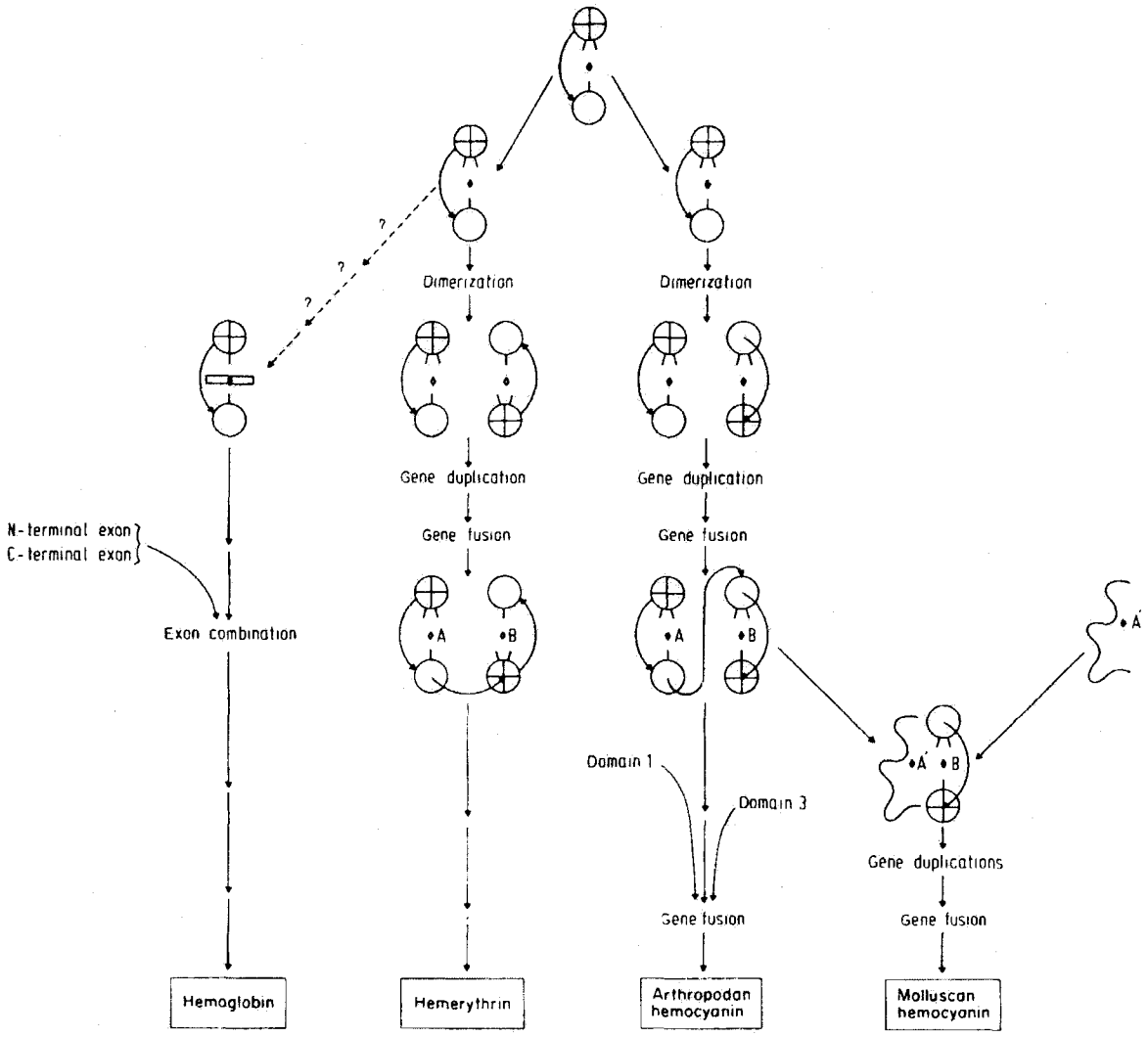


Figure 6. Possible evolutionary relationship of the 4 known classes of oxygen-transporting proteins. The postulated common ancestor is a metal-binding helical pair containing 2 histidine residues, separated by 3 residues, in the 1st helix and 1 histidine residue in the 2nd helix. Circles represent alpha helices. If a circle contains a cross, the helix is viewed from the N terminus, otherwise it is viewed from the C terminus. A diamond indicates the bound metal ion; open diamonds represent iron ions, filled diamonds represent copper ions. Further explanation is given in the text.

(Volbeda and Hol, 1989)