H., & Puigdomenech, P., Eds.) Vol. 101, pp 153-170, Plenum, New York.
Payvar, F., & Schimke, R. T. (1979) *J. Biol. Chem. 254*, 7636-7642.
Petropoulos, C. J., Yaswen, P., Panzica, M., & Fausto, N. (1985) *Cancer Res. 45*, 5762-5768.
Rosenfeld, M. G., Mermod, J. J., Amara, S. G., Swanson, L. W., Sawchenko, P. E., Rivier, J., Vale, W. W., & Evans, R. M. (1983) *Nature (London) 304*, 129-133.
Ruoslahti, E., Pihko, H., & Seppälä, M. (1974) *Transplant. Rev. 20*, 38-60.
Sakai, M., Morinaga, T., Urano, Y., Watanabe, K., Wegmann, T. G., & Tamaoki, T. (1985) *J. Biol. Chem. 260*, 5055-5060.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Schibler, U., Hagenbuchle, O., Wellauer, P. K., & Pittet, A. C. (1983) *Cell (Cambridge, Mass.) 33*, 501-508.
Sellem, C. H., Frain, M., Erdos, T., & Sala-Trepat, J. M. (1984) *Dev. Biol. 102*, 51-60.
Turcotte, B., Guertin, M., Chevrette, M., & Belanger, L. (1985) *Nucleic Acids Res. 13*, 2387-2398.
Wilson, W., Malim, M. H., Mellor, J., Kingsman, A. J., & Kingsman, S. M. (1986) *Nucleic Acids Res. 14*, 7001-7016.

cDNA Cloning of the *Octopus dofleini* Hemocyanin: Sequence of the Carboxyl-Terminal Domain[†,‡]

Walter H. Lang
Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331
Received February 22, 1988; Revised Manuscript Received May 17, 1988

ABSTRACT: A cDNA library was constructed in pUC 19, using poly(A+) RNA purified from *Octopus dofleini* branchial gland, which is the site of hemocyanin biosynthesis in cephalopods. The library was screened with an oligonucleotide probe derived from a portion of the partially known sequence of the C-terminal domain of *Paroctopus dofleini dofleini*. The clone with the longest insert—called pHC1—was sequenced and used as a probe for Northern blotting. It hybridized to a 9.5-kb RNA species, which was also visible as a band after ethidium bromide staining. The cDNA insert (approximately 1200 bp) of pHC1 contained an open reading frame of 1071 bp coding for 357 amino acids. In this insert, a region coding for 42 amino acids from the N-terminal end of the C-terminal domain is missing. These were obtained by sequencing a cloned primer extension product. By comparing our sequence with *Helix pomatia* β_c-hemocyanin unit D, we found 42.9% identical and 11.5% similar residues. One putative copper binding site (site B) was identified by homology to *Helix* hemocyanin and arthropodan hemocyanin. The location of a second possible site was identified.

Hemocyanins are copper-containing oxygen transport found in several orders of molluscs and arthropods (van Holde & Miller, 1982; Ellerton et al., 1983; Preaux & Gielens, 1984). Although molluscan and arthropodan hemocyanins are similar in amino acid composition, they differ greatly in molecular architecture: Arthropodan hemocyanins are composed of hexamers, which are found as (1 × 6)-, (2 × 6)-, (4 × 6)-, and (8 × 6)-meric aggregates, depending on the taxonomic group studied. There is one oxygen-binding site per subunit, which has a M_r of about 75000. On the other hand, molluscan hemocyanins are composed of 10 or 20 very large subunits (M_r about 350000–450000) that are arranged to form hollow cylinders. There are seven or eight oxygen-binding domains per subunit (M_r per domain of 45000–55000).

It has been of great interest whether these two hemocyanins evolved from a common ancestor or independently after the two phyla diverged. Their similarity in amino acid composition has been cited in favor of the first hypothesis, but this question can only be resolved by comparing amino acid sequences of both types. Currently, a number of amino acid sequences are known for subunits of arthropodan hemocyanins (Linzen et al., 1985; Nakashima et al., 1986; Bak & Beintema, 1987). Recently, the complete amino acid sequence for functional unit D of *Helix pomatia* β_c-hemocyanin was reported from amino acid sequencing of the isolated fragment (Drexel et al., 1987). This is the first sequence published for a functional domain of a molluscan hemocyanin.

When this sequence was compared to sequences of arthropodan hemocyanins, Drexel and co-workers (1987) found no similarities between the two types of hemocyanins except a region that corresponds to the "copper B" site in the hemocyanin of the arthropod *Panulirus interruptus*. This region is also found in known tyrosinase sequences. The N-terminal part of the *Helix* sequences contains sections that are clearly homologous to the tyrosinase sequences, but overall similarity is limited. The second copper-binding site could not be identified and must be completely different from the "copper A" site in arthropodan hemocyanins. The authors of this paper suggested on the basis of these findings that molluscan and arthropodan hemocyanins evolved independently from a common ancestral mononuclear copper protein. It is hypothesized that a gene duplication led to evolution of arthropodan hemocyanins, and a fusion with a different type of copper-binding structure led to the evolution of molluscan hemocyanins and tyrosinases.

[†] This work was supported by Grant DMB 85 10310 (to Kensal E. van Holde) from the National Science Foundation.
[‡] The nucleic acid sequence in this paper has been submited to GenBank under Accession Number J02835.

0006-2960/88/0427-7276$01.50/0

17mer Oligonucleotide:

Aminoacid sequence:		Asp	Tyr	Trp	Asp	Trp	Thr	
Coding sequence:	5'	GAY	TAY	TGG	GAY	TGG	ACX	3'
Oligonucleotide:	3'	CTR	ATR	ACC	CTR	ACC	TG	5'

20mer Primer/Probe:

5' CAA CTG AGG GGA ATT CCA TG 3'

FIGURE 1: Sequences of the synthesized oligonucleotides used in the course of the experiments. Y stands for C and T; R stands for A and G.

The hemocyanin of the giant Pacific octopus (*Octopus dofleini*) is a 51S molecule composed of 10 identical 350-kDa polypeptide chains as shown by sedimentation methods, N-terminal sequence homology, and immunological criteria (Miller & van Holde, 1982; Lamy et al., 1986, 1987). Electron microscopy reveals seven globular domains (Lamy et al., 1986). One domain can be cleaved off very easily upon tryptic digestion. This was shown to be the C-terminal domain (Lamy et al., 1987).

It is known that the branchial gland is the site of hemocyanin biosynthesis in cephalopods (Preaux et al., 1984). This finding and the availability of a partial sequence of the C-terminal domain of *Paroctopus dofleini dofleini* (Takagi, 1986) enabled us to approach the problem of sequencing *O. dofleini* hemocyanin with the methods of recombinant DNA technology. Considering the size of this molecule, this seems to be the method of choice compared to protein sequencing. It is greatly facilitated in this case by the homogeneity of the *Octopus* protein, since only one mRNA of the order of 10 kb is to be expected.

In this paper I present the complete sequence for the C-terminal domain (Od-1) of this hemocyanin. The sequence is compared with that of unit D of *H. pomatia* β_c-hemocyanin (Drexel et al., 1987) and to the published sequences of tryptic fragments of the C-terminal domain of *P. dofleini dofleini* hemocyanin (Takagi, 1986). The following paper (Miller et al., 1988) describes physical and functional properties of the isolated domain Od-1.

Materials and Methods

RNA Preparation. Total RNA was isolated from *O. dofleini* branchial gland by the guanidine thiocyanate/CsCl procedure (Chirgwin et al., 1979). After centrifugation, the RNA pellets were dissolved in 10 mM Tris[1]-HCl, pH 7.4, 1 mM EDTA, and 1% SDS, extracted once with phenol/chloroform and once with chloroform, and then precipitated with ethanol after addition of 0.1 volume of sodium acetate, pH 5.2. The RNA was resuspended in DEP-treated water, quantitated by measuring the A_{260} of a small aliquot, and stored in 10-mg portions at −80 °C.

Poly(A+) RNA was purified by two cycles of affinity chromatography on oligo(dT)–cellulose (New England Biolabs) as described in Maniatis et al. (1983).

Oligonucleotides. An 8-fold degenerate 17-mer oligonucleotide was prepared for initial screening of a cDNA library. Its sequence was derived from a portion of the partially known amino acid sequence of the C-terminal domain of *P. dofleini dofleini* (Takagi, 1986) (see Figure 1). A 20-mer primer was synthesized complementary to base pairs 43–62 of the pHC1 cDNA inset (underlined in Figure 4). All oligonucleotides were synthesized on an Applied Biosystems synthesizer.

Preparation and Screening of cDNA Library. Tailed vectors were prepared according to the method of Deng and Wu (1981). A cDNA library was constructed in pUC19 according to the method of Heidecker and Messing (1983), using 3 μg of poly(A+) RNA and 30 units of AMV reverse transcriptase (Life Sciences) in a reaction volume of 30 μL. After transformation into *Escherichia coli* JM 103, the library was plated out on Millipore HATF filters laid on YT plates containing 100 μg/mL ampicillin, 1 mM IPTG, and 0.004% X-Gal. Bacterial colonies were grown overnight, and then filter replicas were made and processed according to the method of Hanahan and Meselson (1983). The baked and washed filters were screened with the mixed 17-mer oligonucleotide probe (see Figure 1) as described by Wood (1985). Positive clones were picked, rescreened if necessary, and checked for inserts by cutting with restriction enzymes.

A primer extension library was prepared essentially according to the method of Gubler and Hoffman (1983), except that the first-strand synthesis contained 50 mM KCl and 200 ng of 20-mer primer (see Figure 1) instead of oligo(dT). After second-strand synthesis, the duplex DNA was filled in with reverse transcriptase and size fractionated on a small Ultrogel A2 column (LKB). The cDNA was then tailed with dGTP and annealed to *Kpn*I cut pUC19 DNA, tailed with dCTP. The annealing mixture was transformed into *E. coli* DH5α. The library was screened by using the 20-mer primer as a probe.

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 1 h to remove formaldehyde, stained with ethidium bromide, and destained for 3 h to overnight or transferred to BA 85 nitrocellulose (Schleicher & Schuell) without staining.

Random primed probes (Feinberg & Vogelstein, 1983) were prepared from the *Eco*RI/*Eco*RV fragment of pHC1 by using a commercially available kit (Boehringer). Blots were prehybridized overnight in 50% formamide, 5× SSC, 1× Denhardt's, 250 μg/mL sssDNA, and 50 mM 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 two times for 5 min in 2× SSC and 0.1% SDS at room temperature and two times for 15 min at 55 °C. Washing at higher stringency was done if necessary. Blots were exposed overnight at −80 °C by using intensifying screens.

Subcloning, Plasmid Purification, and Sequencing. Restriction fragments from pHC1 were subcloned into pUC19 for direct sequencing by using a simple protocol (Struhl, 1985). Large-scale plasmid purifications were done by a method employing gel filtration (Micard et al., 1985). Small-scale plasmid preparations were made by using the alkaline lysis procedure (Birnboim & Doly, 1979) with the following modifications: The supernatant after the high-salt precipitation was extracted once with phenol, once with phenol/chloroform, and once with chloroform, followed by ethanol precipitation. After centrifugation, the pellet was dissolved in 100 μL of 10 mM Tris-HCl, pH 7.4, and 1 mM EDTA, and RNase A was added to 200 μg/mL. After 2 h at 37 °C, proteinase k was

[1] Abbreviations: Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; DEP, diethyl pyrocarbonate; sssDNA, sonicated salmon sperm DNA; 1× SSC, 0.15 M NaCl, 15 mM sodium citrate; IPTG, isopropyl β-D-thiogalactopyranoside; X-Gal, 5-bromo-4-chloro-3-indoyl β-D-galactoside; Denhart's, 0.1% Ficoll 400, 0.1% bovine serum albumin, 0.1% poly(vinylpyrrolidone).

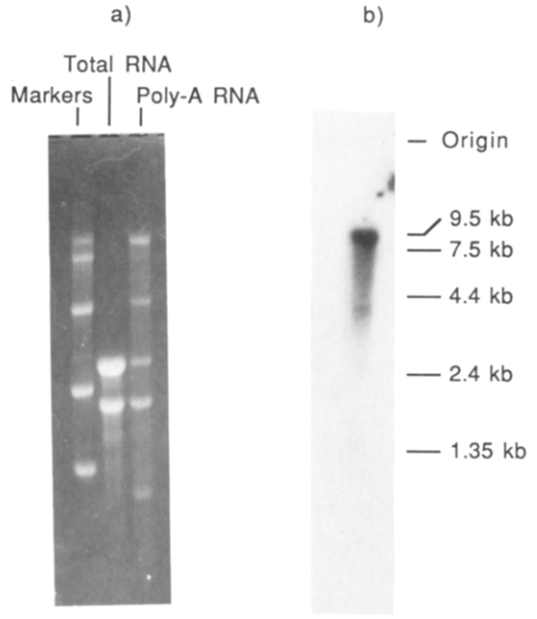

FIGURE 2: (a) Ethidium bromide stained 1.1% agarose gel of RNA samples from *O. dofleini* branchial gland. Loaded from left to right: RNA markers, total RNA, poly(A+) RNA. Sizes of RNA markers from top: 9.5, 7.5, 4.4, 2.4, and 1.35 kb. (b) Northern blot of poly(A+) RNA probed with labeled *Eco*RI/*Eco*RV fragment of pHC1 cDNA insert. Indicated on the right are the positions of RNA markers, electrophoresed on the same gel.

added to 100 μg/mL. After 1–2 more h at 37 °C, the DNA was extracted once with phenol/chloroform and once with chloroform and precipitated with ethanol out of 2.5 M ammonium acetate. The plasmid DNA was then sufficiently pure for DNA sequencing.

Plasmid DNA was sequenced directly (Chen & Seeburg, 1985) by the dideoxy method (Sanger et al., 1977) and the standard M13/pUC primers in a protocol (available from Pharmacia) adapted for the use of reverse transcriptase. The samples were labeled by using [α-^{32}P]dATP, and electrophoresis was carried out on 7% polyacrylamide/8 M urea gels.

Protein Sequence Analysis. Prediction of secondary structure was done according to the method of Chou and Fasman (1978) for the sequences of *O. dofleini* C-terminal domain, *H. pomatia* β_c-hemocyanin unit D, and *P. interruptus* hemocyanin α-chain.

Results

RNA Electrophoresis and Northern Blotting. Total RNA and poly(A+) RNA were electrophoresed on a formaldehyde gel as shown in Figure 2. RNA markers (available from BRL) were included for size determination. After ethidium bromide staining, I observed five bands in the poly(A+) 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 poly(A) tails or poly(A)-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.

In the Northern blotting experiments the probe, obtained from the *Eco*RI/*Eco*RV fragment of pHC1 by random priming, hydridized *only* to the 9.5-kb poly(A+) RNA species. An mRNA this size could code for hemocyanin.

cDNA Cloning and Sequencing. After transformation with our oligo(dT) primed cDNA library, I obtained about 5000

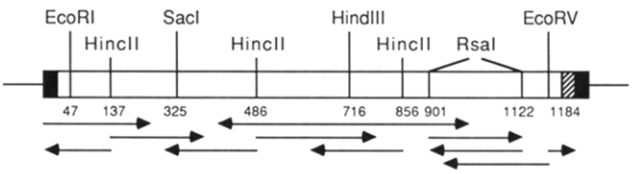

FIGURE 3: Restriction map of pHC1 cDNA insert. Nucleotide 1 is the first A of the *Hin*dIII site of the pUC19 polylinker. The black areas represent the pUC19 polylinker, the hatched area the poly(A) region. Arrows indicate direction of sequencing.

Table I: Amino Acid Composition of the C-Terminal Domain of *O. dofleini* Hemocyanin

name	no.	name	no.
Ala	27	Asn	19
Cys	8	Pro	7
Asp	31	Gln	12
Glu	19	Arg	13
Phe	23	Ser	22
Gly	25	Thr	32
His	24	Val	19
Ile	21	Trp	7
Lys	23	Tyr	17
Leu	33		
Met	6	total	399

clones. Thirty of these hybridized to the probe and were checked for insert size. The clone with the longest insert size—called pHC1—was used for initial sequencing. Restriction fragments were subcloned and sequenced as indicated in Figure 3. Clone pHC1 has an insert size of approximately 1200 bp as judged by agarose gel electrophoresis (not shown). The poly(A) region is approximately 50 bp long. It was not possible to sequence through it, so the exact size could not be determined. The cDNA insert contains an open reading frame of 1071 bp coding for 357 amino acids. It has a 3′ untranslated region of 80 bp and a polyadenylation signal 16 bp upstream of the poly(A) region (underlined in Figure 4).

After transformation with the primer extension library, I obtained approximately 3000 clones. About 1000 of these hybridized to the probe. While screening with the oligonucleotide was in progress, 24 clones were screened for inserts, and 5 clones with insert sizes of about 600 to 1200 base pairs were chosen for sequencing. In the sequence of one of these clones, there appeared, 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 the isolated clone is really an extension of the original cDNA clone. The DNA sequence obtained was used to deduce the amino acid sequence for the N-terminal portion of the C-terminal domain not coded for by pHC1. The complete cDNA sequence and the derived protein sequence are shown in Figure 4.

I examined the protein sequence for the presence of the peptide the 17-mer oligonucleotide probe was coding for. Surprisingly, the corresponding peptide was slightly different: Instead of the sequence DYWDWT, I found PYWDWT. A 17-bp match was therefore not possible, but the DNA sequence showed that a 15-bp contiguous match was possible, which accounts for the fact that it was possible to pick up positive clones under the hybridization conditions I used for screening [described by Wood (1985)].

Properties of the Polypeptide Chain. The polypeptide chain of the C-terminal domain is 399 residues long. Its amino acid composition is shown in Table I. The molecular weight of the pure polypeptide chain calculated from its amino acid composition is 45 316. There are 50 acidic (Asp and Glu) and 36 basic side chains (Arg and Lys). This results in an overall

```
1
GAAGCCGTGAGAGGTACTATAATCAGGAAGAATGTCAATAGTTTGACTCCATCTGATATTAAAGAA
E  A  V  R  G  T  I  I  R  K  N  V  N  S  L  T  P  S  D  I  K  E

67
CTAAGAGACGCCATGGCTAAAGTACAAGCTGACACATCAGATAATGGTTACCAAAAAATTGCTTCC
L  R  D  A  M  A  K  V  Q  A  D  T  S  D  N  G  Y  Q  K  I*  A  S

133
TATCATGGAATTCCCCTCAGTTGTCATTATGAGAATGGCACAGCTTATGCATGTTGCCAACATGGA
Y  H  G  I  P  L  S  C  H  Y  E  N  G  T  A  Y  A  C  C  Q  H  G

199
ATGGTAACTTTCCCTAACTGGCATAGATTGTTGACTAAACAAATGGAAGATGCACTAGTTGCTAAG
M  V  T  F  P  N  W  H  R  L  L  T  K  Q  M  E  D  A  L  V  A  K

265
GGATCTCACGTTGGTATTCCTTACTGGGACTGGACCACTACCTTCGCTAATTTACCTGTACTGGTC
G  S  H  V  G  I  P  Y  W  D  W  T  T  T  F  A  N  L  P  V  L  V

331
ACTGAGGAGAAAGATAACTCCTTCCACCATGCCCATATTGATGTTGCCAATACGGACACAACCAGA
T  E  E  K  D  N  S  F  H  H  A  H  I  D  V  A  N  T  D  T  T  R

397
TCACCAAGAGCTCAACTTTTTGATGATCCAGAAAAAGGAGATAAATCATTCTTCTATCGCCAGATA
S  P  R  A  Q  L  F  D  D  P  E  K  G  D  K  S  F  F  Y  R  Q  I

463
GCACTTGCTCTGGAACAAACAGATTTCTGTGATTTTGAAATTCAGTTTGAAATCGGTCACAATGCT
A  L  A  L  E  Q  T  D  F  C  D  F  E  I  Q  F  E  I  G  H  N  A

529
ATTCATTCATGGGTTGGCGGTAGTAGCCCATATGGTATGTCAACTCTTCACTATACTTCCTATGAT
I  H  S  W  V  G  G  S  S  P  Y  G  M  S  T  L  H  Y  T  S  Y  D

595
CCTCTCTTCTACCTTCATCACTCTAATACTGATCGTATTTGGTCTGTATGGCAAGCATTACAGAAG
P  L  F  Y  L  H  H  S  N  T  D  R  I  W  S  V  W  Q  A  L  Q  K

661
TATCGAGGATTACCTTACAACACTGCGAACTGTGAAATCAATAAACTTGTTAAACCACTAAAACCA
Y  R  G  L  P  Y  N  T  A  N  C  E  I  N  K  L  V  K  P  L  K  P

727
TTCAACTTAGATACCAATCCTAACGCCGTTACAAAAGCCCATTCTACTGGTGCTACATCTTTCGAT
F  N  L  D  T  N  P  N  A  V  T  K  A  H  S  T  G  A  T  S  F  D


793
TACCACAAGCTTGGTTATGATTATGATAATCTTAATTTCCATGGAATGACCATTCCTGAGTTGGAA
Y  H  K  L  G  Y  D  Y  D  N  L  N  F  H  G  M  T  I  P  E  L  E

859
GAACACCTTAAAGAAATACAACACGAAGACAGGGTATTTGCTGGTTTCTTACTTCGTACCATTGGT
E  H  L  K  E  I  Q  H  E  D  R  V  F  A  G  F  L  L  R  T  I  G

925
CAATCTGCTGATGTTAACTTCGACGTCTGCACCAAAGATGGTGAATGTACATTTGGAGGTACGTTC
Q  S  A  D  V  N  F  D  V  C  T  K  D  G  E  C  T  F  G  G  T  F

991
TGTATTCTCGGTGGAGAACATGAAATGTTTTGGGCATTTGACCGTCCTTTTAAATATGATATTACC
C  I  L  G  G  E  H  E  M  F  W  A  F  D  R  L  F  K  Y  D  I  T

1055
ACAAGTCTGAAACATCTCCGCTTAGATGCGCATGATGATTTCGATATAAAGGTTACTATTAAAGGT
T  S  L  K  H  L  R  L  D  A  H  D  D  F  D  I  K  V  T  I  K  G

1121
ATCGATGGTCATGTCTTGTCCAATAAATACCTCAGTCCTCCGACGGTTTCCTCGCTCCAGCCAAA
I  D  G  H  V  L  S  N  K  Y  L  S  P  P  T  V  F  L  A  P  A  K

1187
ACAACGCATTAATATGTGTACTAATCATTTATATGATTTAATATTTGTGTTTTTTTGTACTTTTTC
T  T  H  stop

1253
ATTCAATAAATGATATCTCAAACAATAAAAAAA.....
```

FIGURE 4: cDNA sequence and derived protein sequence of *O. dofleini* C-terminal domain. The underlined stretches show the complementary sequence of the 20-mer primer, the stop codon, and the polyadenylation signal. An asterisk marks the position of the first nucleotide of the coding sequence of the pHC1 cDNA insert. The preceding nucleotide sequence was obtained from a sequenced primer extension cDNA clone.

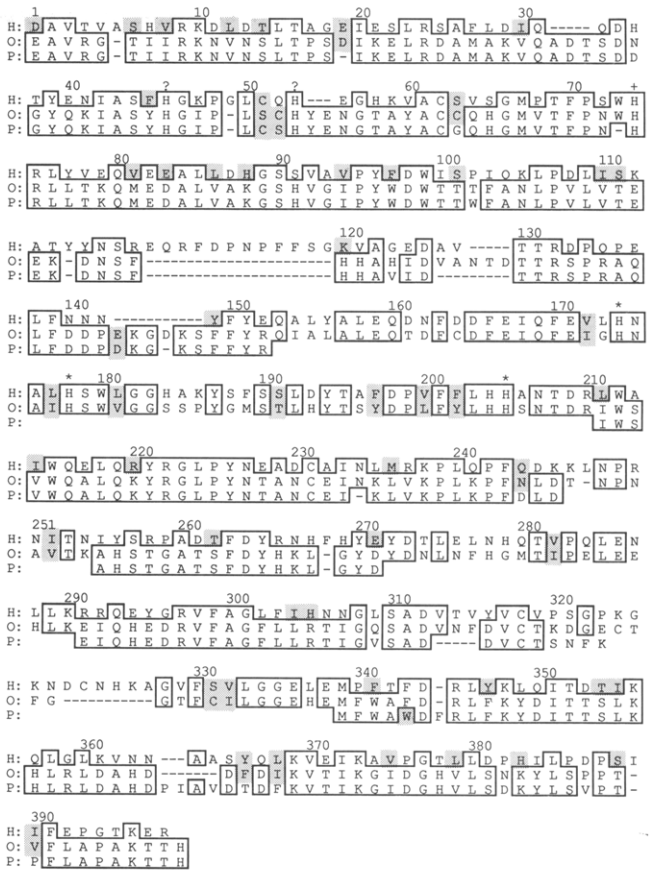

FIGURE 5: Sequence comparison between protein sequences of *H. pomatia* β_c-hemocyanin unit D (H), *O. dofleini* C-terminal domain (O), and *P. dofleini dofleini* (P). The numbers refer to the amino acid sequence of *O. dofleini* C-terminal domain. The putative copper B ligands are marked with an asterisk, the histidine in the WHR group is marked with a plus, and the other conserved histidines are marked with a question mark (see text for further explanation). Boxed areas show identical residues and shaded areas similar residues. The following groups of amino acids are considered as similar: Ser, Thr; Leu, Ile, Val, Met; Phe, Tyr, Trp; Glu, Asp; Gln, Asn; Arg, Lys, His; and Ser, Cys.

acidic character of the protein.

In earlier work from this laboratory (Lamy et al., 1987) we reported a sequence for the 27 N-terminal residues of the C-terminal domain of *O. dofleini*, which was obtained by protein sequencing. However, there is a disagreement between that sequence and the one reported here. The first five residues obtained by protein sequencing were TVGDA; the first six residues obtained by cDNA sequencing are EAVRGT, but the following residues of the two sequences are identical. It now seems likely that the two sequences have been obtained from different *Octopus* species. The blood sample from which the C-terminal domain was prepared that was ultimately used for sequencing was obtained from a small animal, which was thought to be a juvenile *O. dofleini*. However, there occurs another small *Octopus* species on the Oregon coast called *Octopus rubrescens*. Apparently, an adult *O. rubrescens* and a juvenile *O. dofleini* can be easily mistaken; thus, it is possible that the initial protein sequencing studies were carried out with a sample from the former species. All work done subsequently in this laboratory has used protein purified from blood of very large animals, which cannot be mistaken for another species. It appears that Takagi had a similar problem. His first published partial sequence of *P. dofleini dofleini dofleini* (Takagi, 1986) is identical with our cDNA-derived sequence. This was later revised [see Lamy et al. (1987) and Drexel et al. (1987)] and is now identical with the sequence we determined by protein sequencing. Our laboratory is currently in the process of repeating the protein sequencing work to resolve this dilemma. There is no question, however, concerning the identification of the *O. dofleini* used for mRNA preparation in this work.

Using the amino acid sequence, I attempted to predict the amount of secondary structure in the domain protein. According to the Chou–Fasman procedure the protein contains 31% α-helix, 27% β-sheet, and 30% β-turns. The folding of the polypeptide chain is sketched in Figure 6a.

Molluscan hemocyanins are glycoproteins, and it is known that the C-terminal domain of both *Octopus* species contains carbohydrate (Takagi, 1986; Miller et al., 1988). I examined

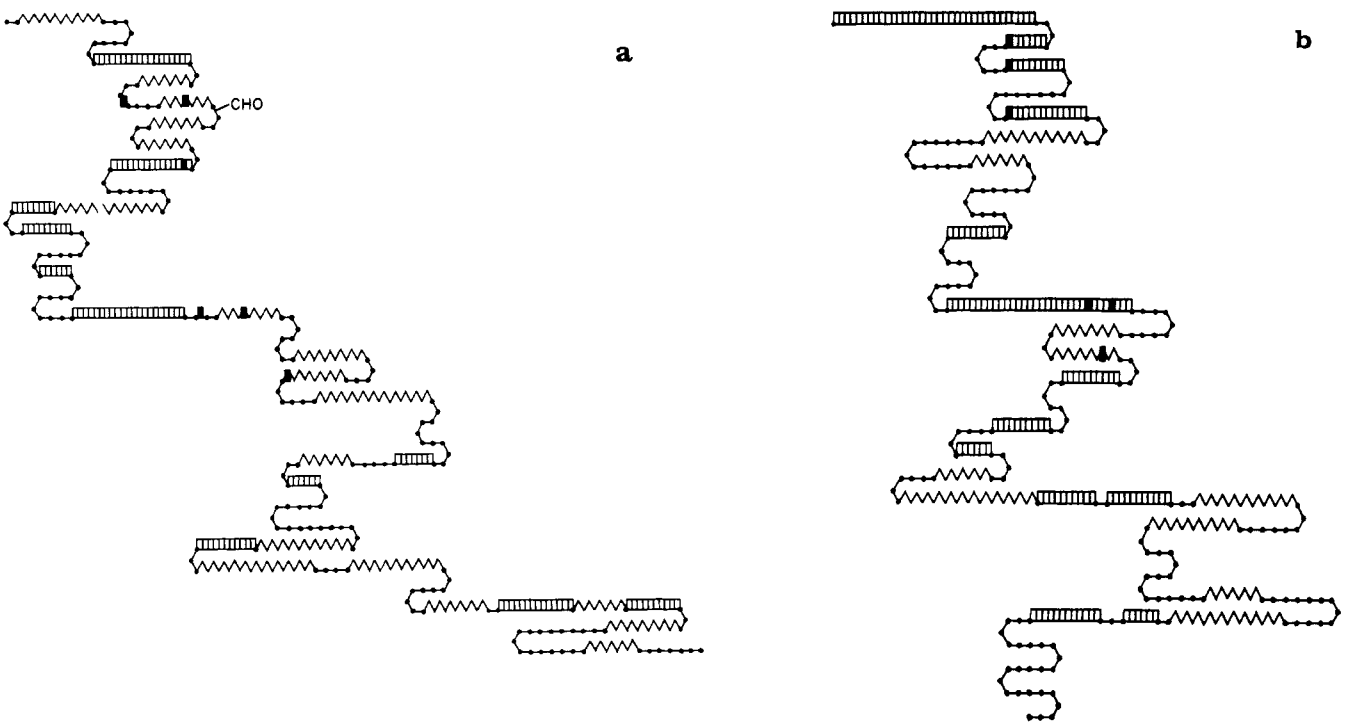

FIGURE 6: Sketches of secondary predictions according to Chou and Fasman (1978) for (a) *O. dofleini* C-terminal domain and (b) *H. pomatia* β_c-hemocyanin unit D: (■) histidine residues that potentially serve as copper ligands; (CHO) potential carbohydrate attachment site in *O. dofleini* C-terminal domain; (□) one α-helix residues; (△) one β-sheet residue; (●–●) a random structure residue. β-Turns are indicated by chain reversals.

the sequence for the presence of the consensus sequence Asn-X-Thr for carbohydrate attachment sites of N-linked carbohydrate chains. There is one such site present starting with Asn 56. This site is also located in a predicted β-turn (see Figure 6a) and is therefore a likely candidate for carbohydrate attachment. The corresponding site in the *P. dofleini dofleini* domain has been shown to have carbohydrate attached to it (Takagi, 1986). The same author also reported that possibly Ser 52 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.

Discussion

Comparison to Sequences of Molluscan Hemocyanins. The availability of the sequence of the *H. pomatia* β_c-hemocyanin unit D allows the comparison of two complete sequences for functional domains of molluscan hemocyanins. In Figure 5 the sequence of *O. dofleini* is aligned with the recently published sequence of *H. pomatia* β_c-hemocyanin unit D (Drexel et al., 1987) and to the published tryptic fragments of the C-terminal domain of *P. dofleini dofleini* hemocyanin (Takagi, 1986). Between the two *Octopus* proteins, 300 positions are compared; of these 91.0% are identical and 1.3% more are similar. From 424 positions compared between *H. pomatia* and *O. dofleini* hemocyanin, 42.2% are identical and 11.5% are similar. The high degree of similarity between the protein of the snail and the giant octopus is intriguing. Not only do the two functional units occupy different positions on the whole polypeptide chain, but the animals also belong to two molluscan classes, whose ancestors diverged about 550–600 million years ago (Yochelsen, 1979). It appears that molluscan hemocyanins are considerably more conserved than arthropodan hemocyanins. The degree of similarity between crustacean and cheliceratan chains is 29% (Linzen et al., 1985), compared to 53.7% between the two molluscan hemocyanins. Crustaceans and cheliceratans also diverged about 550–600 million years ago. It can be concluded that molluscan hemocyanins are composed of very similar functional units fused into a single polypeptide chain. It also can be expected that functional units occupying the same position on the polypeptide chain exhibit an even higher degree of similarity.

Structure of Copper-Binding Site. After the determination of the crystal structure of *P. interruptus* hemocyanin (Gaykema et al., 1984), the structure of the copper-binding sites in anthropodan 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.

Like Drexel and co-workers (Drexel et al., 1987) I did not find any similarities between the two types of hemocyanins when the *Octopus* sequence was compared to arthropodan sequences, except for a region that corresponds to the copper B binding site in *P. interruptus* hemocyanin. Histidines 174, 178, and 205 are likely to serve as ligands for copper and are marked with asterisks in Figure 5. It is of interest to compare the secondary structure predicted for the peptides containing the copper B ligands. In the *Octopus* protein His 174 is predicted to be part of a random structure. His 178 and 205 are predicted to occur in β-sheets (see Figure 6a). In the *Helix* protein the first two of the corresponding histidines are predicted to be part of an α-helix, the third one is also part of a β-sheet (see Figure 6b). However, such comparisons must be viewed with skepticism, since secondary structure predictions are not very reliable. In fact, the peptides containing the copper ligands in *P. interruptus* hemocyanin are predicted as β-sheets (not shown), but the crystal structure shows them to be part of α-helices (Gaykema et al., 1984).

This leaves the question as to what are the ligands of copper A? Avissar et al. (1986) probed Northern blots of poly(A+)

RNA isolated from *Levantina* and *Sepia* 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* poly(A+) RNA and concluded that this sequence is part of the active site of molluscan hemocyanins as well. However, this peptide cannot be found in the so far available molluscan hemocyanin sequences. Nor is a histidine pattern identical to that of the copper B site present elsewhere in the molecule. Drexel and co-workers found that in the known sequences of molluscan hemocyanins and tyrosinases the amino acid triplet WHR (position 73–75 in Figure 5) is highly conserved. There is good evidence that this histidine is part of the active site in *Streptomyces glaucescens* tyrosinase (Huber & Lerch, 1986). The current spectroscopic evidence [reviewed by van Holde and Miller (1982)] points out that the copper atoms are liganded by the imidazole nitrogen of three histidines and possibly one other bridging ligand. Therefore, the obvious thing to do is to look for histidines that are conserved in all three available sequences of molluscan hemocyanins. Besides His 74 there are only two more conserved histidine residues: His 46 and His 53 (see Figure 5). These histidines are spaced by six other residues in *Octopus* hemocyanin and seven in the functional unit of *Helix*. If they are part of an α-helix, they would be separated by two consecutive turns and possibly too far apart to both serve as ligands for a single copper atom. A proline residue is located between the first two histidines in all three sequences. This would place it within the postulated inner helix in these peptides, which cannot be the case. Secondary structure prediction for the *H. pomatia* unit D suggests a β-turn between the two histidines, which are themselves predicted part of two adjacent α-helices (see Figure 6b). This arrangement would bring the two histidines closer together. The structure prediction for the *Octopus* C-terminal domain does not, however, fit this pattern. His 46 is predicted to be part of a β-turn, and His 53 is located in a β-sheet (see Figure 6a). Random structure is predicted between the turn and the sheet, so there is a good possibility for the polypeptide chain to fold in a way that the two possible ligands come close enough to be able to serve as copper ligands.

Another interesting feature is the close proximity of the potential carbohydrate attachment site Asn 56 to one of the proposed copper ligands. Since carbohydrate chains always attached to the surface of a glycoprotein, the copper A site would be very close to the surface and more easily accessible for bulky ligands. This could be an explanation for the finding that molluscan hemocyanins exhibit tyrosinase activity, although the reaction rate is 50-fold lower for hemocyanins compared to that for the tyrosinases (Salvato et al., 1983; Nakahara et al., 1983). In contrast, arthropodan hemocyanins do not exhibit tyrosinase activity, and their copper-binding site is buried inside the molecule (Gaykema et al., 1984).

Altogether, the above arguments provide an attractive model for a second copper-binding site in molluscan hemocyanins. Sequencing of further domains in the *O. dofleini* hemocyanin may provide confirmatory or contradictory evidence. However, a final decision must surely await X-ray diffraction studies.

ACKNOWLEDGMENTS

I would like to thank Dr. K. E. van Holde for his support of this research and for critically reading the manuscript, Dr. A. W. Martin of the University of Washington and Ken Yates of the Hatfield Marine Science Center for their assistance in obtaining *Octopus*, Reginald McParland for synthesizing the oligonucleotides, and Jack Kramer for performing the computer analysis.

Registry No. Copper, 7440-50-8.

REFERENCES

Avissar, I., Daniel, E., & Daniel, V. (1986) *Biochem. J. 233*, 253–257.

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

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 2251–2266.

Chen, E. Y., & Seeburg, P. H. (1985) *DNA 4*, 165–170.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5249–5299.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Deng, G., & Wu, R. (1981) *Nucleic Acids Res. 9*, 4173–4188.

Drexel, R., Siegmund, S., Schneider, H. J., Linzen, B., Gielens, C., Preaux, G., Kellerman, J., & Lottspeich, F. (1987) *Biol. Chem. Hoppe-Seyler 368*, 617–635.

Ellerton, H. D., Ellerton, N. F., & Robinson, H. A. (1983) *Prog. Biophys. Mol. Biol. 41*, 143–248.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Gaykema, W. P. J., Hol, W. G. J., Vereijken, J. M., Soeter, N. M., Bak, H. J., & Beintema, J. J. (1984) *Nature (London) 309*, 23–29.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–269.

Hanahan, D., & Meselson, M. (1983) *Methods Enzymol. 100*, 333–342.

Heidecker, G., & Messing, J. (1983) *Nucleic Acids Res. 11*, 4891–4907.

Huber, M., & Lerch, K. (1986) in *Invertebrate Oxygen Carriers* (Linzen, B., Ed.) pp 265–276, Springer-Verlag, Heidelberg.

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., & Hol, W. G. J. (1985) *Science (Washington, D.C.) 229*, 519–524.

Lamy, J., Lamy, J. N., Leclerc, M., Compin, S., Miller, K. I., & 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., & van Holde, K. E. (1987) *Biochemistry 26*, 3509–3518.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1983) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Micard, D., Sorbier, M. L., Couderc, J. L., & Dastague, B. (1985) *Anal. Biochem. 148*, 121–126.

Miller, K. I., & van Holde, K. E. (1982) *Comp. Biochem. Physiol., B: Comp. Biochem. 73B*, 1013–1018.

Miller, K. I., van Holde, K. E., Toumadje, A., Johnson, W. C., Jr., & Lamy, J. (1988) *Biochemistry* (following paper in this issue).

Nakahara, A., Shinnichiro, S., & Kino, J. (1983) *Life Chem. Rep., Suppl. Ser. 1*, 319–322.

Nakashima, H., Behrens, P. Q., Moore, M. D., Yokota, E., & Riggs, A. F. (1986) *J. Biol. Chem. 261*, 10526–10533.

Preaux, G., & Gielens, C. (1984) in *Copper Proteins* and *Copper Enzymes* (Lontie, R., Ed.) Vol. 2, pp 159–205, CRC, Boca Raton, FL.

Salvato, B., Jori, G., Piazzese, A., Ghiretti, F., Beltramini, M., & Lerch, K. (1983) *Life Chem. Rep., Suppl. Ser. 1*, 313–318.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Struhl, K. (1985) *BioTechniques 3*, 2–3.
Takagi, T. (1986) in *Invertebrate Oxygen Carriers* (Linzen, B., Ed.) pp 259–262, Springer-Verlag, Heidelberg.
van Holde, K. E., & Miller, K. I. (1982) *Q. Rev. Biophys. 15*, 1–129.
Wood, D. J. (1984) *Focus 4*, 1–3.
Yochelsen, E. L. (1979) in *The Origin of Major Invertebrate Groups* (House, M. R., Ed.) pp 323–358, Academic, New York.

Structure and Function of the Carboxyl-Terminal Oxygen-Binding Domain from the Subunit of *Octopus dofleini* Hemocyanin†

Karen I. Miller,*,‡ K. E. van Holde,‡ Arazdordi Toumadje,‡ W. C. Johnson, Jr.,‡ and Josette Lamy§

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331, and Laboratoire de Biochimie, Faculté de Pharmacie, Université Francois Rabelais, 37032 Tours Cedex, France

Received February 22, 1988; Revised Manuscript Received May 17, 1988

ABSTRACT: The C-terminal domain, Od-1, of the 7-domain subunit of *Octopus dofleini* hemocyanin has been prepared by partial trypsinolysis followed by ion-exchange chromatography. It binds oxygen reversibly and is homogeneous in molecular weight. Its physical properties have been compared with those of the subunit. The domain molecular weight is found by sedimentation equilibrium to be 4.7×10^4, in excellent agreement with the result recently obtained in our laboratory from cDNA sequencing of this domain [Lang, W. H. (1988) *Biochemistry* (preceding paper in this issue)]. It has a sedimentation coefficient of 3.8 S. Both the molecular weight and sedimentation coefficient are consistent with the domain constituting approximately one-seventh of the M_r 3.5×10^5 subunit. Its amino acid composition and carbohydrate content differ significantly from that of the whole subunit, confirming the heterogeneity in domains previously established on an immunological basis. Circular dichroism predicts similar secondary structure for the domain and subunit. The domain does not self-associate in the presence of Mg^{2+} but does bind to the whole molecule in a ratio of approximately 1 domain/subunit. The oxygen affinity of this domain is quite low. It shows intrinsic magnesium and Bohr effects similar to those of the whole molecule but of greatly reduced magnitude.

Hemocyanins are high molecular weight copper proteins responsible for oxygen transport in arthropods and molluscs. They are composed, in the case of arthropods, of ~70 000-dalton subunits each binding one oxygen, arranged into hexamers and then into higher order structures with molecular weights as high as 3 million. Although early studies on the properties of the subunits were complicated by problems of subunit heterogeneity, it has been possible to obtain fairly detailed information on structure and function of the hemocyanins of several arthropods. Details of structure now known include amino acid sequences for several subunits as well as precise subunit localization within the whole molecules of *Limulus* and *Androctonus* (Lamy et al., 1981, 1985, 1987) and *Eurypelma* (Markl et al., 1981; Linzen et al., 1985) hemocyanins and 3.4-Å resolution X-ray crystallography on the subunit of *Panulirus* hemocyanin (Gaykema et al., 1984). Detailed functional studies have been performed on *Eurypelma* hemocyanin as well (Savel et al., 1986; Markl et al., 1986).

Molluscan hemocyanins, on the other hand, are composed of much larger subunits, each containing seven or eight oxygen-binding domains of about 50 000 daltons. Here, too, subunit heterogeneity has been a problem, especially with gastropod hemocyanins. *Helix pomatia* hemocyanin, for example, contains two types of subunits, termed α and β, which assemble in stoichiometrically complex ways to form whole molecules (Gielens et al., 1973; Brouwer et al., 1977; Verschueren et al., 1981). Each subunit contains eight immunologically distinct domains, of which one, domain d from the βc subunit, has now been purified and used to determine the amino acid sequence (Drexel et al., 1987). This domain was not obtained intact, however, but in the form of two disulfide-bridged fragments. So although the sequence has been determined, it has not been possible to obtain much purified domain in a functionally intact condition. Thus, the many studies that might relate structure and function of this domain to the parent molecule have not been possible.

The properties of cephalopod hemocyanins have made them better subjects for such investigations. *Octopus dofleini* hemocyanin has significant advantages as a model molluscan hemocyanin. The parent molecule is a 51S decamer composed of *identical* 11S subunits, each containing seven immunologically distinct oxygen-binding domains (Lamy et al., 1987). The dissociation and reassociation of the molecule is fully reversible upon removal and readdition of divalent cations (Miller & van Holde, 1982). The order of the domains is now almost completely known (Lamy et al., 1987). In the accompanying paper (Lang, 1988), cloning of the cDNA corresponding to the *O. dofleini* hemocyanin subunit is described, and the sequence of the C-terminal domain, Od-1 (Lamy et al., 1987), is presented.

† This work was supported by Grant DMB 85 10310 from the National Science Foundation (K.V.H. and K.I.M.), Grant RCP 080816 of Centre National de la Recherche Scientifique (CNRS) (J.L.), and USPHS GM21479 (W.C.J.).
‡ Oregon State University.
§ Université Francois Rabelais.

0006-2960/88/0427-7282$01.50/0