Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 5312–5316, July 1990
Biochemistry

Characterization of the gene encoding the hemocyanin subunit *e* from the tarantula *Eurypelma californicum*

(oxygen carrier/copper proteins/intron/exon junctions/protein domains)

WOLFGANG VOLL*† AND RENATE VOIT

Zoologisches Institut, Universität München, Luisenstr. 14, D-8000 Munich 2, Federál Republic of Germany

Communicated by Kensal E. van Holde, April 16, 1990 (received for review November 14, 1989)

ABSTRACT **The gene for the hemocyanin subunit *e* of the tarantula *Eurypelma californicum* has been isolated from a genomic phage library by using a corresponding cDNA clone as a probe. The transcriptional unit spans a chromosomal region of about 55 kilobase pairs (kbp). The gene consists of nine exons that are separated by large introns. The intron/exon boundaries were determined by direct comparison of genomic and cDNA sequences. A putative promoter region ("TATA" box, reversed "CAAT" box) 100 bp 5′ to the translational initiation codon strongly suggests the presence of a functional gene. The 3′ flanking region carries the polyadenylylation signal (AATAAA) and several conserved structures for the 3′ splicing of the pre-mRNA. A comparison of the gene architecture of the subunit *e* gene with the three-dimensional structure of the arthropod hemocyanin subunit shows a good correspondence with the division of the subunit into three domains (two exons coding for the first, three coding for the second, and four coding for the third domain). The relationship to molluscan hemocyanins, different tyrosinases, and the larval serum proteins is discussed.**

Hemocyanins are the predominant proteins in the hemolymph of many arthropods and molluscs (1, 2). Common features of the hemocyanins in both phyla are a binuclear copper(I) active site, the ability to transport oxygen, and a bluish color when oxygenated. Arthropod and molluscan hemocyanins are substantially different in their subunit composition and molecular architecture. Molluscan hemocyanins are cylindrical molecules that consist of 10 or 20 polypeptides with a M_r of 450,000. Each of those polypeptides (= subunits) contains seven or eight oxygen-binding units of M_r 55,000. In contrast, arthropod hemocyanins are built up of hexamers of about M_r 75,000, each bearing one oxygen-binding site. *In vivo* arthropod hemocyanins are found as 1 × 6, 2 × 6, 4 × 6, 6 × 6, and 8 × 6 aggregates, depending upon the species (3).

The primary structure of several subunits of arthropod hemocyanins and two functional units of molluscan hemocyanins has been elucidated (4–10). Arthropod hemocyanin subunits reveal an overall homology of about 30% (the positions with isofunctional residues included). Highly conserved regions are clustered among the binuclear copper-binding center with six histidine residues that have been found to be the copper ligands, and a common tertiary structure for all arthropod hemocyanins has been deduced from comparison to the x-ray crystallography of a lobster hemocyanin subunit (11, 12).

The polypeptide chain is clearly divided into three domains. The central domain bears the active center with the copper-binding sites A and B (CuA and CuB). Each copper ion is bound by three histidine residues, two of them located in one α-helix with a characteristic sequence -His-Xaa-Xaa-Xaa-His-, and the third histidine residue located about 30 amino acids further toward the carboxyl terminus in a second α-helix.

The hemocyanin of the North American tarantula *Eurypelma californicum*, a 4 × 6-mer comprising seven distinct subunit types, termed *a* through *g*, has been extensively analyzed. Details of the primary structure, quaternary structure, and oxygen-binding function are known (see refs. 3, 12–14). It has been shown by immunohistochemistry (15) and *in vitro* translation of isolated mRNA (16) that the hemocyanin of *Eurypelma* is synthesized in cells proliferating from the inner heart wall. A cDNA library was established from RNA prepared from the total spider heart and screened with a 17-mer oligonucleotide probe corresponding to the amino acid sequences of the CuA oxygen-binding site. The first cDNA clone for an arthropod hemocyanin, termed pHC4, was found to encode the subunit *e* of *Eurypelma* hemocyanin (16).

In this paper the isolation and characterization of the corresponding gene are reported.‡ The structural features of the gene are discussed in relation to the three-dimensional (3-D) structure of arthropod hemocyanins and are compared with the structure of a mouse tyrosinase gene (17, 18) and insect larval serum proteins (19, 20).

MATERIALS AND METHODS

Animals. The tarantula *E. californicum* was purchased from Carolina Biological Supply. Animals were kept under standard conditions as described (21).

Construction and Screening of Genomic Libraries. Genomic DNA was isolated from the heart, lung, and ovary of one tarantula (22), partially digested with *Sau*3AI, and fractionated on a CsCl gradient (23). Purified and 5′-dephosphorylated target DNA [12–20 kilobase pairs (kbp)] was ligated to the arms of the replacement phage λ vector λEMBL3 (24). The ligated DNA was packaged *in vitro* by using extracts that had been prepared by a modified method of Hohn and Hohn (25). Genetic selection for recombinant phages was carried out by using the Spi phenotype system with the bacterial strain Q358/Q359 as the host system (26). Five libraries with 4.3×10^6 plaque-forming units (pfu) were established and amplified. Approximately 6×10^5 pfu from the largest one (1.7×10^6 pfu) were screened at a density of 20,000 plaques per 150-mm plate (27). Filters were hybridized with ^{32}P-labeled nick-translated pHC4 clone DNA as described (28).

Characterization of Genomic Clones. Phage DNA was isolated from purified positive plaques (29), digested with var-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CuA and CuB: copper binding site A and B.

*To whom reprint requests should be addressed.
†Present address: Bayerisches Landeskriminalamt, Maillinger Str. 15, D-8000 Munich 19, F.R.G.

‡The sequences reported here have been deposited in the GenBank data base (accession nos. X16650 to X16657).

ious restriction endonucleases, separated by agarose gel electrophoresis, and transferred to nitrocellulose membranes (30). The filters were hybridized either with the ^{32}P-labeled, nick-translated pHC4 clone DNA or with the appropriate subprobes and with two different [γ-^{32}P]ATP-labeled 17-mer synthetic oligonucleotide mixtures, respectively (31): one probe (16) was derived from the highly conserved CuA region, and the second probe was derived from a specific *e* peptide (-$^{46}_{Asp}$-Phe-His-Glu-Glu-$^{51}_{Asp}$-) (5). Hybridization conditions were as follows. Filters were prehybridized in 0.9 M NaCl/0.09 M sodium citrate, pH 7/0.1% Ficoll/0.1% bovine serum albumin/0.1% polyvinylpyrrolidone/0.5% sodium dodecyl sulfate/100 μg of denatured salmon DNA per ml/0.05% sodium pyrophosphate at 37°C for at least 2 hr. Hybridization was performed with the same solution including 1–2 × 10^6 cpm of end-labeled oligonucleotides per ml. Hybridization temperature and washing conditions were as described by Szostak *et al.* (32). Selected genomic fragments carrying exon sequences and 5′ and 3′ flanking genomic fragments were subcloned into pUC8, pUC9 (33), or pUC19 (34) and were sequenced by using the dideoxynucleotide chain-termination method in both orientations (35, 36). Directed deletions were obtained with exonuclease III and VII (34).

Computer Analysis. DNA sequences were analyzed on a Micro VAX computer (Digital Equipment) with the GAP and the FASTN programs of the University of Wisconsin Genetics Computer Group program package version 4 and the FASTP program of the Protein Identification Resource program package using the EMBL data bank as data source.

RESULTS

Isolation and Characterization of the Subunit *e* Gene from a Genomic Library. Independent phages (6 × 10^5) of a *Eurypelma* genomic library constructed in λEMBL3 were screened by using the nick-translated DNA from the cDNA clone pHC4 as a probe. Eight of 15 overlapping different clones (Hc1, -2, -4, -6, -10, -11, -12, and -15) containing exon segments were isolated and characterized by complete restriction enzyme analysis with *Bam*HI, *Eco*RI, *Hin*dIII, and *Sal* I and by hybridization with total pHC4 cDNA, subclones of the cDNA probe, or specific oligonucleotides (exon 1 by oligonucleotide e and exon 3 by oligonucleotide CuA). All of these recombinant clones could be arranged in a 5′ → 3′ series by overlaps in the restriction pattern (Fig. 1*A*), indicating the presence of a single copy gene for the hemocyanin subunit *e*.

Several genomic DNA fragments that had hybridized with cDNA probes (subclones b–g) or with specific oligonucleotides (subclones a and c) were subcloned into the plasmid vectors pUC8, pUC9, or pUC19 and further analyzed with restriction enzyme digests and Southern blotting followed by hybridization with pHC4, derived subprobes, or specific oligonucleotides (Fig. 1*B*). The subsequent sequencing of suitable fragments revealed the entire nucleotide sequence, consisting of nine exons and the adjacent exon position boundaries, and included the 5′ and 3′ untranslated regions of the gene (Table 1).

Structural Features of the 5′ End of the Subunit *e* Gene. The 5′ noncoding region with the first exon of the gene encoding subunit *e* could be detected on the subclone HclH/E0.6 by using oligonucleotide e as a probe. The nucleotide sequence of this clone shows several putative regulatory elements (Fig. 2).

A TATA-like (37) sequence was found −100 nucleotides upstream from the initiation codon, whereas a degenerate form (CAAT) of the canonical CCAAT box sequences (5′GGYCAATCT3′, . . .) (38) was located at position −142 on the antisense strand. Simian virus 40-like (39) G+C-rich motifs occurred in the forward orientation at position −110 and in the reverse orientation at position −134. The best match for the point of transcription initiation could be identified at position −76 (40). A good correspondence with the nucleotide sequences around the AUG initiation codon, postulated for the best translation initiation by Kozak *et al.* (41) (CCACCATGG), was found. No leader sequence for a signal peptide was detected between the AUG initiation codon and the first amino acid of the hemocyanin subunit *e*, as determined by protein chemical methods (5). This fact supports the idea that hemocyanin is released by holocrine secretion (42).

Structural Features of the 3′ End of the Subunit *e* Gene. The 3′ untranslated region of the subunit *e* gene was detected on subclone Hc6Hi 0.7 and contains 286 nucleotides counting from the UAG stop codon to the guanosine before the putative first adenosine of the poly(A) tail (Fig. 3). This is in accord with the data obtained by cDNA analysis.

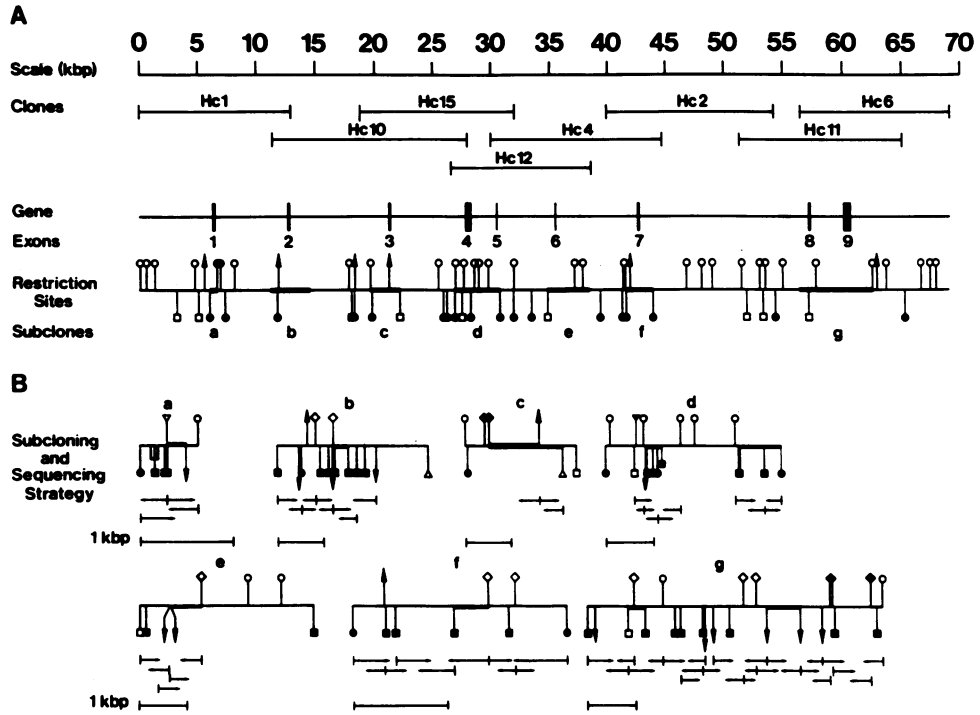

FIG. 1. (*A*) Physical map of the subunit *e* gene. Inserts from eight different phage λ clones were aligned based upon restriction mapping. The positions of the exons are indicated by filled boxes. The size of subclones a–g is indicated by bold lines. (*B*) Subcloning and sequencing strategy of subclones a–g. Horizontal arrows indicate the direction and extent of DNA regions sequenced. ▬, Region hybridizing with oligonucleotide e (subclone a) or with oligonucleotide CuA (subclone c); ▬, region hybridizing with cDNA clone pHC4. ↑ *Bam*HI; ♀ *Eco*RI; ↓ *Hin*dIII; ⏚ *Sal* I; ♀ *Pst* I; ♈ *Xho* I; ⅄ *Xba* I; ↓ *Sau*3AI; ↑ *Bgl* II; ↓ *Hin*cII.

Table 1. Intron/exon organization of the hemocyanin subunit *e* gene of *E. californicum*

Exon	Position in open reading frame	Size of exon, bp	Sequence of intron/exon junctions: 5′-splice donor. . . . . intron . . . . . .3′-splice acceptor	intron in kbp	
1	1–190	190*	CTT TAC A gtaagtctca.	6.1	. . . . . ttttttttag CT GCA AAA
			Leu Tyr T		hr Ala Lys
2	191–408	218	GAT CAA gtaagtacat.	8.6	. . . . . tttttatag GAC ATT TCT
			Asp Gln		Asp Ile Ser
3	409–614	206	GCG AG gtaagtccaa.	6.3	. . . . . gtaactctag G TAC GAC
			Ala Ar		g Tyr Asp
4	615–1011	397	TTC AAT gtaagtatgt.	2.1	. . . . . tatttttag GAA AAC CCA
			Phe Asn		Glu Asn Pro
5	1012–1140	129	GAA GAT gtaagtagtc.	5.0	. . . . . tcatccgcag TTG ACA TGT
			Glu Asp		Leu Thr Cys
6	1141–1275	135	GTG AAG gtatttctgc.	6.9	. . . . . catttacag GTA CTG TAC
			Val Lys		Val Ile Tyr
7	1276–1462	187	GCC ACT C gtatgttaag.	14.3	. . . . . tttattttag TC GAC CCT
			Ala Thr L		eu Asp Pro
8	1463–1677	215	GAT GCT gtaagtgcta.	2.7	. . . . . cttctttcag GTC AAC GGA
			Asp Ala		Val Asn Gly
9	1678–2153	476†	TCGACAG‡		

Nucleotide sequence of intron/exon junctions was determined. Exon sequences are in capital letters; intron sequences are in lowercase letters. The size of the introns I_1–I_7 is estimated. I_8 was sequenced in its entirety.
*Plus 70–80 nucleotides of 5′ untranslated region; the transcription start site is still to be determined.
†This exon contains 284 nucleotides of the 3′ untranslated region.
‡Presumptive last nucleotide before the poly(A) tail.

Thirteen nucleotides 5′ to the deduced polyadenylylation site is the common polyadenylylation signal AATAAA (43). Additional less-conserved recognition elements are a G+T cluster (44), a T-rich sequence, and, at a more downstream position, the CACTG-motif that occurs three times in short succession (45).

DISCUSSION

In this paper the structure and organization of a hemocyanin subunit gene is presented. There are several striking features. The total size of the transcriptional unit is large (≈54 kbp) in relation to the 2.3-kbp cDNA. As in most genes, the introns show a wide range of size (2.1–14.3 kbp). There are five introns with a minimum length of 6 kbp, one of them (between exons 7 and 8) comprises >25% of the total gene size. The size of the exons—except exon 4 with 397 bp—lies in the usual range of 100–200 bp (46).

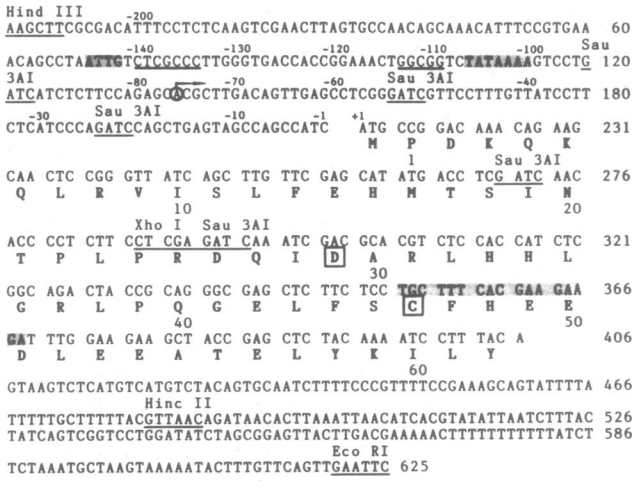

FIG. 2. Nucleotide sequence of the 5′ terminus and of the first exon of the tarantula hemocyanin subunit *e* gene. The first base of the ATG initiation codon is designated 1. TATA box and reversed CAAT box are shaded as well as the binding site for the oligonucleotide e mixed probe. Sites for factor SP1 binding corresponding to the core consensus sequence are underlined. The arrow indicates the putative transcription start site. Selected restriction enzyme sites are indicated. Amino acids are given in single-letter code. Differences from the amino acids determined by protein chemistry are boxed. Numbering of the nucleotides is according to the Hc1H/E0.6 subclone.

The nucleotide sequences of the coding region could be compared with data obtained by analysis of the subunit *e* full-length cDNA clone λM1. The comparison of the amino acid sequence deduced from genomic DNA with data obtained by protein chemical sequencing (5) should be restricted to exon 1 (see Fig. 2), since the remaining differences have been comprehensively analyzed by Voit and Schneider (16). By sequencing genomic DNA of the subunit *e* gene, we found arginine replaced by aspartate at position 29 and aspartate by cysteine at position 41 in the first exon. These differences may be due to ambiguous identification of amino acid phenylhydantoin derivatives.

Since all arthropod hemocyanins are homologous (12), one may compare the data from the *Eurypelma* subunit *e* gene

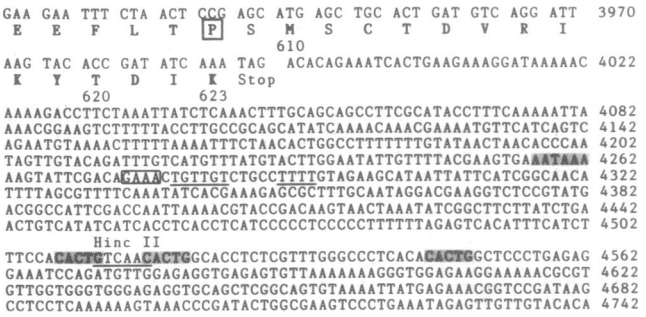

FIG. 3. Nucleotide sequence of the 3′ end of the tarantula subunit *e* gene-containing part of the last exon and of the 3′ untranslated region. Poly(A) addition occurs at one of the three boxed nucleotides. The polyadenylylation signal sequence AATAAA and the sequence CACTG are shaded. Other putative 3′ termination signal sequences are underlined. Amino acids are given in single-letter code. Differences from the amino acids determined by protein chemistry are boxed. Numbering of the nucleotides is according to the subclones Hc6E/S0.6 and Hc6E4.7.

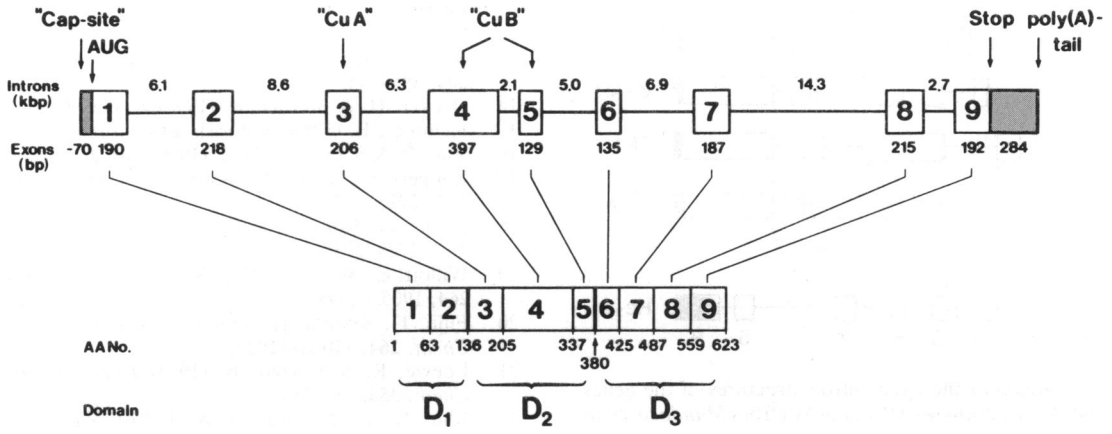

FIG. 4. Exon/intron architecture of the hemocyanin subunit *e* gene in comparison with the proposed division of one subunit in domains. The 5′ and 3′ untranslated regions are shaded.

with the division of the hemocyanin subunit into domains, established for *Panulirus* hemocyanin by Gaykema *et al.* (11). The gene structure appears to support the hypothesis that the position of exon boundaries correlates with the functional or structural domains (47–49). Domain 1 is encoded by two exons; domain 2 containing CuA and CuB, by three exons; and domain 3, by four exons (Fig. 4). However, it should be mentioned that the boundaries are slightly different from those presented in the paper by Linzen *et al.* (12). According to the data obtained from genomic DNA analysis, the transition from domain 1 to domain 2 is in the β-strand 1B of domain 1, and the exon boundary separating domains 2 and 3 is in α-helix 3.1 of domain 3.

Surprisingly, CuB is encoded by two exons, whereas only one exon codes for CuA. Assuming that CuB is the ancestral copper-binding site (for reasons, see later in *Discussion*) we must postulate that an intron was lost during the evolution of CuA from a copy of CuB.

The exon-shuffling model suggests that intron/exon junctions usually map outside the hydrophobic core and near the surface of the proteins (50) to avoid disturbing the three-dimensional structure of the protein after a shuffling event. A stereo view of the splice junctions supports these ideas (Fig. 5). All intervening sequences, indicated by I_1–I_8 "cut" the polypeptide chain at amino acids lying at or near the protein surface.

The descriptions of primary structures of one functional unit of the mollusc *Helix pomatia* (9) as well as several tyrosinases (51, 52) have suggested the possible homology of the oxygen-binding site CuB of all hemocyanins to that in tyrosinases. It has been presumed that hemocyanins and tyrosinases are derivatives of a common ancestral mononuclear CuB polypeptide (53). When we compare the gene architecture of the hemocyanin subunit *e* of *Eurypelma* with the published data of a mouse tyrosinase gene structure (17, 18), it turns out that (*i*) the structure of the tyrosinase gene does not harmonize in any way with the present hemocyanin gene, even if one allows intron sliding and (*ii*) CuB in the mouse tyrosinase is encoded only by one exon with a completely different size than exons 4 and 5 of the hemocyanin gene.

New insight into the evolution of hemocyanins at the level of DNA comes from the structural analysis of two insect larval serum proteins (19, 20). Based on the alignment of the amino acid sequences, arthropod hemocyanins and larval serum proteins show a significant sequence identity of 27% (19). This is the same degree of identity found between crustacean and chelicerate hemocyanins (12). In addition, both classes of proteins are hexamers exhibiting similar quaternary structures, and a monoclonal antibody against the tarantula hemocyanin was found to cross-react with the larval serum protein calliphorin (54). Now we are able to compare both types of proteins as to their gene structures (Fig. 6). Surprisingly we find one exon/intron boundary to be conserved among the hemocyanin *e* gene and the genes encoding the larval storage proteins (19, 20). This conserved splice site

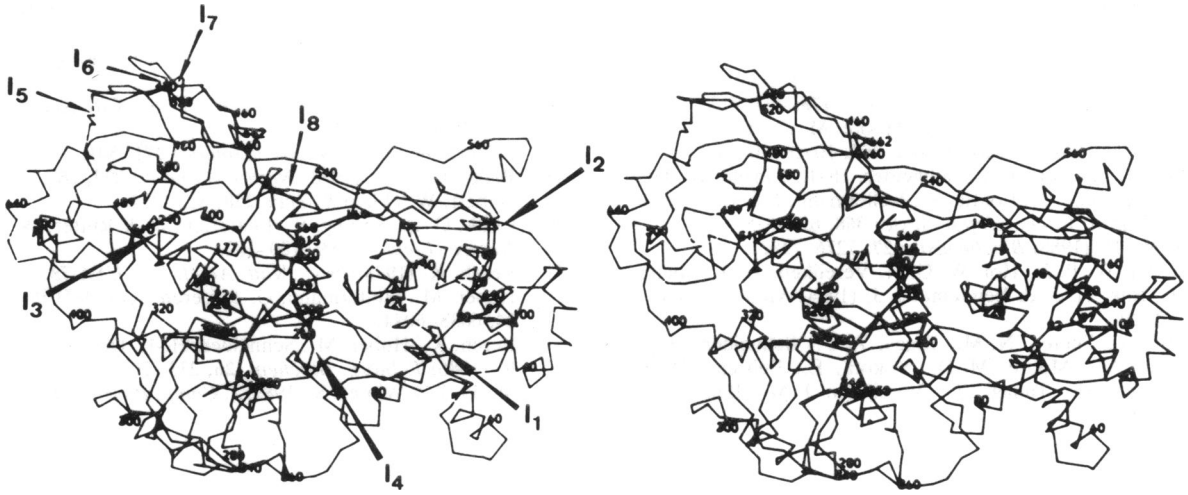

FIG. 5. Stereo view of the α-carbon chain tracing of the *Panulirus interruptus* subunit a hemocyanin showing the location of the splice junctions (I_1–I_8) mapped on the protein sequence.

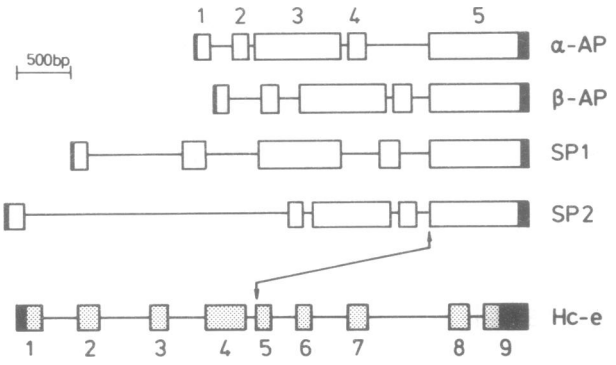

FIG. 6. Comparison of the exon/intron structures of the genes encoding α- and β-arylphorin (α-AP and β-AP) from *Manduca sexta* (ref. 19), larval serum proteins from *Bombyx mori* (SP1 and SP2; ref. 20) and hemocyanin subunit *e* from *E. californicum*. Open boxes represent exons, and filled boxes indicate the 5′ and 3′ untranslated regions. The introns of the hemocyanin gene are drawn in reduced scale.

is located within the CuB binding region of arthropod hemocyanins. Thus, we can conclude from our present knowledge that there is a strong relationship between arthropod hemocyanins and larval storage proteins. On the other hand, molluscan hemocyanins are related to tyrosinases rather than to arthropod hemocyanins, suggesting that arthropod hemocyanins and molluscan hemocyanins have most likely evolved from different ancestors.

This paper is dedicated to Prof. Dr. Bernt Linzen who initiated and supported this work; he suddenly died during its completion. We thank Dr. R. Mertz (Genzentrum, München) for synthesizing the oligonucleotide probes; Dr. W. Schartau, Prof. Dr. J. Markl, and Prof. Dr. H. MacWilliams for critical reading of the manuscript; and Mrs. I. Krella for typing the manuscript. Financial support was received from the Deutsche Forschungsgemeinschaft (Li 107/26-2 and Scha 317/4-1).

1. Van Holde, K. E. & Miller, K. I. (1982) *Q. Rev. Biophys.* **43,** 1–129.
2. Préaux, G. & Gielens, C. (1984) in *Copper Proteins and Copper Enzymes*, ed. Lontie, R. (CRC Press, Boca Raton, FL), Vol. 2, pp. 159–205.
3. Markl, J. (1986) *Biol. Bull.* **171,** 90–115.
4. Schartau, W., Eyerle, F., Reisinger, P., Geisert, H., Storz, H. & Linzen, B. (1983) *Hoppe-Seylers Z. Physiol. Chem.* **364,** 1383–1409.
5. Schneider, H.-J., Drexel, R., Feldmaier, G., Linzen, B., Lottspeich, F. & Henschen, A. (1983) *Hoppe-Seylers Z. Physiol. Chem.* **364,** 1357–1381.
6. Bak, H. J. & Beintema, J. J. (1987) *Eur. J. Biochem.* **169,** 333–348.
7. Jekel, P. H., Bak, H. J., Soeter, N. M., Vereijken, J. M. & Beintema, J. J. (1988) *Eur. J. Biochem.* **178,** 403–412.
8. Nakashima, H., Behrens, P. Q., Moore, M. D., Yokota, E. & Riggs, A. F. (1986) *J. Biol. Chem.* **261,** 10526–10533.
9. Drexel, R., Siegmund, S., Schneider, H.-J., Linzen, B., Gielens, C., Préaux, G., Lontie, R., Kellermann, J. & Lottspeich, F. (1987) *Biol. Chem. Hoppe-Seyler* **368,** 617–635.
10. Lang, W. H. (1988) *Biochemistry* **27,** 7276–7282.
11. 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.
12. Linzen, B., Soeter, N. M., Riggs, A. F., Schneider, H.-J., Schartau, W., Moore, M. D., Yokota, E., Behrens, P. Q., Nakashima, H., Takagi, T., Vereijken, J. M., Bak, H. J., Beintema, J. J., Volbeda, A., Gaykema, W. P. J. & Hol, W. G. J. (1985) *Science* **229,** 519–524.
13. Savel-Niemann, A., Markl, J. & Linzen, B. (1988) *J. Mol. Biol.* **204,** 385–395.
14. Decker, H. & Sterner, R. (1990) *J. Mol. Biol.* **211,** 281–293.
15. Kempter, B. (1983) *Naturwissenschaften* **70,** 255–256.
16. Voit, R. & Schneider, H.-J. (1986) *Eur. J. Biochem.* **159,** 23–29.
17. Ruppert, S., Müller, G., Kwon, B. & Schütz, G. (1988) *EMBO J.* **7,** 2715–2722.
18. Müller, G., Ruppert, S., Schmid, E. & Schütz, G. (1988) *EMBO J.* **7,** 2723–2730.
19. Willott, E., Wang, X.-Y. & Wells, M. A. (1989) *J. Biol. Chem.* **264,** 19052–19059.
20. Fujii, T., Sakurai, H., Izumi, S. & Tomino, S. (1989) *J. Biol. Chem.* **264,** 11020–11025.
21. Loewe, R. & Linzen, B. (1973) *Hoppe-Seylers Z. Physiol. Chem.* **354,** 182–184.
22. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303–2308.
23. Seed, B., Parker, R. C. & Davidson, N. (1982) *Gene* **19,** 201–209.
24. Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842.
25. Hohn, B. & Hohn T. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 2372–2376.
26. Karn, J., Brenner, S., Barnett, L. & Cesarini, G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5172–5176.
27. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
28. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
29. Leder, P., Tiemeier, D. & Enquist, C. (1977) *Science* **196,** 175–177.
30. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
31. Woods, D. (1984) *BRL Focus* **6,** (3), 1–3.
32. Szostak, J. W., Stiles, J. I., Tye, B.-K., Chiu, P., Sherman, F. & Wa, R. (1979) *Methods Enzymol.* **68,** 419–429.
33. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
34. Yanish-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
35. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
36. Chen, E. J. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
37. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
38. Benoist, C. & Chambon, P. (1981) *Nature (London)* **290,** 304–310.
39. Jones, S. M., Yamamoto, K. R. & Tjian, R. (1985) *Cell* **42,** 559–572.
40. Hentschel, C. C. & Birnstiel, M. L. (1981) *Cell* **25,** 301–313.
41. Kozak, M. (1983) *Microbiol. Rev.* **47,** 1–45.
42. Sminia, T. (1977) in *Structure and Function of Hemocyanins*, ed. Bannister, J. (Springer, Heidelberg), pp. 279–288.
43. Proudfoot, N. J. & Brownlee, C. G. (1976) *Nature (London)* **263,** 211–214.
44. McLauchlan, J., Gaffney, D., Whitton, J. C. & Clements, J. B. (1985) *Nucleic Acids Res.* **13,** 1347–1368.
45. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41,** 349–359.
46. Naora, H. & Deacon, N. J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6196–6200.
47. Gilbert, W. (1978) *Nature (London)* **271,** 501.
48. Blake, C. C. F. (1978) *Nature (London)* **273,** 267.
49. Gilbert, W. (1985) *Science* **228,** 823–824.
50. Craik, C. S., Sprang, S., Fletterick, R. & Rutter, W. J. (1982) *Nature (London)* **299,** 180–182.
51. Lerch, K. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3635–3639.
52. Huber, M., Hintermann, G. & Lerch, K. (1985) *Biochemistry* **24,** 6038–6044.
53. Lerch, K., Huber, M., Schneider, H.-J., Drexel, R. & Linzen, B. (1986) *J. Inorg. Biochem.* **26,** 213–217.
54. Markl, J. & Winter, S. (1989) *J. Comp. Physiol.* **B159,** 139–151.