

## The crayfish plasma clotting protein: A vitellogenin-related protein responsible for clot formation in crustacean blood

MARTIN HALL\*<sup>†</sup>, RUIGONG WANG\*<sup>†</sup>, RIK VAN ANTWERPEN<sup>‡</sup>§, LARS SOTTRUP-JENSEN<sup>¶</sup>, AND KENNETH SÖDERHÄLL\*<sup>||</sup>

\*Department of Comparative Physiology, Evolutionary Biology Centre, Uppsala University, Villavägen 6, S-752 36 Uppsala, Sweden; <sup>‡</sup>Department of Biochemistry, University of Arizona, Tucson, AZ 85721; and <sup>¶</sup>Department of Molecular and Structural Biology, University of Aarhus, DK-8000 Århus C, Denmark

Communicated by Russell F. Doolittle, University of California, San Diego, La Jolla, CA, December 29, 1998 (received for review November 13, 1998)

**ABSTRACT** Coagulation in crayfish blood is based on the transglutaminase-mediated crosslinking of a specific plasma clotting protein. Here we report the cloning of the subunit of this clotting protein from a crayfish hepatopancreas cDNA library. The ORF encodes a protein of 1,721 amino acids, including a signal peptide of 15 amino acids. Sequence analysis reveals that the clotting protein is homologous to vitellogenins, which are proteins found in vitellogenic females of egg-laying animals. The clotting protein and vitellogenins are all lipoproteins and share a limited sequence similarity to certain other lipoproteins (e.g., mammalian apolipoprotein B and microsomal triglyceride transfer protein) and contain a stretch with similarity to the D domain of mammalian von Willebrand factor. The crayfish clotting protein is present in both sexes, unlike the female-specific vitellogenins. Electron microscopy was used to visualize individual clotting protein molecules and to study the transglutaminase-mediated clotting reaction. In the presence of an endogenous transglutaminase, the purified clotting protein molecules rapidly assemble into long, flexible chains that occasionally branch.

Vertebrate and invertebrate animals have evolved efficient molecular mechanisms to immediately form clots of blood components. This is necessary to prevent loss of blood in case of injury. All vertebrates have similar coagulation systems based on the proteolytically induced aggregation of fibrinogen into insoluble fibrin (1–3). The fibrin aggregates, which initially are noncovalently associated, are further stabilized by intermolecular covalent crosslinks formed by a proteolytically activated transglutaminase (TGase), factor XIIIa. TGases (EC 2.3.2.13) are Ca<sup>2+</sup>-dependent enzymes capable of forming covalent bonds between the side chains of specific lysine and glutamine residues on certain proteins (4–6). Among the invertebrates, which is a more diverse group, different coagulation mechanisms seem to have evolved, and detailed information on the coagulatory mechanisms at the molecular level is lacking in most groups. One exception is the hemocyte (blood cell)-derived clotting cascade in horseshoe crabs, which has been characterized in detail (7). This horseshoe crab (*Limulus*) clotting system is activated by microbial lipopolysaccharides or  $\beta$ -1,3-glucans, and it has some resemblance to the vertebrate coagulation system, as it is based on a proteolytic cascade that leads to the conversion of a soluble protein (coagulogen) into an insoluble aggregate (coagulin). However, the proteins participating in the *Limulus* clotting system are all from the hemocytes and are not homologous to the vertebrate plasma coagulation proteins. A TGase has been characterized and cloned from *Limulus* hemocytes, but it does not appear to

recognize the coagulogen as a substrate (8, 9), and its role during clotting is unclear.

The clotting reaction also is characterized at the molecular level in crustaceans, where it depends on the TGase-mediated crosslinking of a specific plasma clotting protein (CP) (10–13). The crayfish CP, cloning of which is reported in this article, has previously been biochemically and functionally characterized (12, 14). It is a very high density lipoprotein (VHDL) (14) consisting of two identical 210-kDa subunits held together by disulfide bonds (12). Each one of the 210-kDa subunits has both lysine and glutamine sidechains, which are recognized and become covalently linked to each other by TGases (12). Clotting is induced when a TGase is released from hemocytes or tissue, becomes activated by the Ca<sup>2+</sup>-content in plasma, and starts crosslinking the plasma CP molecules into large aggregates. The hemocytes also contain components of the so-called prophenoloxidase activating system (proPO system), which constitutes an important part of the immediate immune response in crustaceans (15, 16). Components of the proPO system cause degranulation and lysis of hemocytes, and as a result more proPO components and TGase are released (15, 16). In this way, the proPO system could affect the clotting reaction by causing the release of TGase activity. However, the proPO system and the clotting reaction do not appear to share a common activation pathway, as the proPO system is activated by a proteolytic cascade [triggered by microbial polysaccharides (15, 16)], and the initiation of the clotting reaction requires no proteolytic processing [only Ca<sup>2+</sup>, which activates the TGase (11, 12)].

The N terminus of the lobster fibrinogen (the CP homologue in lobster) was reported to have sequence similarity to vitellogenins (VTGs) (17), which are proteins expressed only in females of egg-laying animals [vertebrates as well as invertebrates (18–20)]. Besides having similar functions, the CP does not appear to share any characteristics with fibrinogen or coagulogen, the proteins forming clots in vertebrate animals and horseshoe crabs, respectively. This indicates that the crayfish CP (12, 14), and its homologues in other crustaceans (10, 11, 13, 17, 21) constitute a separate group of blood CPs.

To establish the evolutionary origin of the crustacean CPs, we have cloned the crayfish CP. Our results show that the crustacean CPs constitute a distinct group of proteins unre-

Abbreviations: CP, clotting protein; TGase, transglutaminase; UTR, untranslated region; VHDL, very high density lipoprotein; VTG, vitellogenin; vWF, von Willebrand factor; proPO, prophenol oxidase; HLS, hemocyte lysate supernatant.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF102268).

<sup>†</sup>M.H. and R.W. contributed equally to this work.

<sup>§</sup>Present address: Department of Biochemistry and Molecular Biophysics, Virginia Commonwealth University, Richmond, VA 23298-0614.

<sup>||</sup>To whom reprint requests should be addressed. e-mail: Kenneth.Soderhall@fysbot.uu.se.

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.

PNAS is available online at www.pnas.org.

lated to proteins involved in clotting reactions in other groups of animals.

## MATERIALS AND METHODS

**Animals.** Freshwater crayfish, *Pacifastacus leniusculus*, were kept in tanks in aerated tap water. Only intermolt animals were used in the experiments.

**Protein Purification and Sequencing.** For protein sequencing, CP from crayfish hemolymph was isolated by precipitation at low ionic strength according to Kopacek *et al.* (12). Purified CP (20 mg) was reduced, alkylated, and digested with trypsin. Tryptic peptides were isolated by anion exchange chromatography on a DEAE-Sephacel column followed by reversed phase HPLC on a Nucleosil C<sub>18</sub> column. The N-terminal amino acid sequences of the intact protein and internal tryptic peptides were determined on an Applied Biosystem 470A automated gas-phase sequencer.

For electron microscopy study (see below), CP was isolated by repeated KBr density gradient ultracentrifugation using a different rotor and different KBr gradients compared with the previously described procedure (14). Ultracentrifugation tubes (13 × 51 mm polyallomer Quickseal tubes, Beckman) were filled with crayfish KBr-plasma (0.445 g of KBr per ml of plasma) and centrifuged in a vertical rotor (VTi 65.1, Beckman) at 250,000 × *g* at 10°C. After 22 h, the rotor was allowed to stop without braking (to minimize the disturbance of the formed gradients). The CP is easily identified in the KBr density gradient as it is the only VHDL in crayfish plasma and has an orange color from its carotenoid content (14). The VHDL/CP fraction was collected from the gradient, and its KBr content was adjusted to 0.445 g/ml. A 2.5-ml portion of this VHDL/CP sample was added to ultracentrifuge tubes and carefully overlaid with 10 mM sodium phosphate/0.15 M NaCl/1 mM EDTA, pH 7.0, containing 0.3 g/ml KBr. A second ultracentrifugation was performed as described above, except that the centrifuge was stopped after 5 h. Again, the VHDL/CP fraction was collected from the formed density gradient, and the KBr content was adjusted to 0.445 g/ml. The procedure of overlaying 2.5 ml of the CP sample with 10 mM sodium phosphate/0.15 M NaCl, pH 7.0 containing 0.3 g/ml KBr and subjecting it to a 5-h ultracentrifugation was repeated. After this third ultracentrifugation, the collected VHDL/CP sample was dialyzed against 10 mM sodium phosphate/0.15 M NaCl/1 mM EDTA, pH 7.0, and it contained pure CP, as judged by using SDS/PAGE. Centricon 30 concentrators (Amicon) were used for concentrating the sample. The CP sample (≈35 mg in 7 ml) was passed through a 0.2 μm disposable sterile filter (Schleicher & Schuell) and applied to a Sephacryl S-300 HR (Amersham Pharmacia) gel filtration column (26 × 600 mm, 80 ml/hr, 2 ml per fraction). The gel filtration was performed in 50 mM Tris/0.15 M NaCl/1 mM EDTA, pH 8.0. Fractions which in 7.5% SDS/PAGE under reducing conditions (22) were found to contain mainly the 210-kDa subunit and very little TGase-crosslinked aggregates that resisted reduction were used in the crosslinking assay for electron microscopy.

**Crosslinking Assay.** A hemocyte lysate supernatant (HLS) containing endogenous TGase activity was prepared by homogenizing the pelleted hemocytes from 15 crayfish in 1 ml of 10 mM sodium cacodylate/10 mM EDTA, pH 7.0 as described in Kopacek *et al.* (12). CP (435 μl, 0.7 mg/ml) in 50 mM Tris/0.15 M NaCl/1 mM EDTA, pH 8.0, and 5 μl of HLS (2 mg/ml protein) were mixed in plastic Eppendorf tubes. The crosslinking reaction was initiated by activating the Ca<sup>2+</sup>-dependent TGase in HLS by adding 10 μl of 1 M CaCl<sub>2</sub> and carefully mixing the samples by pipetting. As a control with no TGase activity, 10 μl of H<sub>2</sub>O was added to a sample instead of the CaCl<sub>2</sub>. The Ca<sup>2+</sup>-dependent crosslinking reaction was stopped by the addition of 50 μl of 0.5 M EDTA, pH 8.0, after

15 s, 30 s, 45 s, 1 min, or 2 min at room temperature. The samples were analyzed by using electron microscopy. The ability of CP-containing samples to form stabilized clots was examined in a similar way by mixing CP, HLS, and CaCl<sub>2</sub> (or H<sub>2</sub>O in a control) as described above. After 1 h at room temperature, the presence of stabilized clots was examined by shaking and tilting the tubes.

**Protein Determination.** Protein concentrations were determined according to Bradford (23), using BSA as standard.

**Electron Microscopy.** The samples were analyzed by negative staining, essentially according to the method of Valentine *et al.* (24). A thin film of carbon was evaporated onto a piece of freshly cleaved mica. The carbon-coated mica was inserted in a solution of clotting protein (diluted to ≈50 μg/ml in PBS) in such a way that the carbon film partly detached from the mica and floated on the solution surface. After 30 s, the carbon film was retracted from the solution of clotting protein, floated for a few seconds on deionized water, and transferred to a 2% solution of methylamine tungstate in deionized water (pH 7.5). Subsequently, the carbon film was picked up from the tungstate solution with a 300 mesh copper grid coated with lacey substrate (Ted Pella, Redding, CA). Excess methylamine tungstate solution was removed with filter paper, after which the preparation was allowed to air dry at room temperature. Negatively stained preparations were analyzed with a Philips EM 420 electron microscope at an acceleration voltage of 80 kV.

**cDNA Cloning by Using PCR.** Two pairs of nested, degenerate oligonucleotide primers were designed by reverse translation of the italicized amino acids in one of the tryptic peptides, ISNTMVGEYHDELECDTRAPLPIEYTP. The first pair of primers was used to amplify genomic DNA from crayfish by using PCR. The resulting PCR products were used as DNA template in a second PCR with the second pair of nested primers. This PCR reaction resulted in a short PCR product, which was subcloned into a T vector (25) and sequenced. Beside the primer sequences, the PCR product encoded for the anticipated amino acids HDELECDTRAPL. In a similar way, by using two pairs of nested, degenerate primers designed from the italicized sequences from a second tryptic peptide, NLSYHIDDIIQHELYSVTTYDHTAINPEH, a second PCR product was obtained that contained the primers and the nucleotide sequence encoding HELYSVTTYDHT. From these two PCR products, CP-specific oligonucleotide primers were selected from both the plus and minus strands. A λgt 11 cDNA library from the hepatopancreas of male crayfish (26) was amplified by PCR, using the different CP-specific primers in combination with λgt 11-specific primers. Several PCR products were subcloned, sequenced, found to be CP-specific, i.e., having overlapping nucleotide sequences with the initial two short PCR products from genomic DNA and containing sequences encoding other CP tryptic peptides. A cDNA containing the coding region of the CP was established by so-called PCR walking, i.e., repeating the procedure of selecting new CP-specific primers from the most recently obtained PCR products, amplifying the hepatopancreas cDNA library by PCR with the new primers in combination with vector-specific primers, and sequencing the PCR products. This PCR cDNA consisted of nucleotides 115-6265 (Fig. 1). The sequences of the PCR products were determined by manual DNA sequencing using the T7 sequencing kit (Amersham Pharmacia) or by using automated DNA sequencing on an ABI Prism 377 DNA sequencer (Applied Biosystems).

**Cloning of CP by cDNA Library Screening.** Six different PCR products were amplified from different regions of the CP PCR cDNA, labeled with <sup>32</sup>P by using the Megaprime labeling kit (Amersham Pharmacia), and used as probes (probes I-VI; Fig. 1) to screen the crayfish hepatopancreas cDNA library by plaque hybridization. The λ DNA from the positive plaques

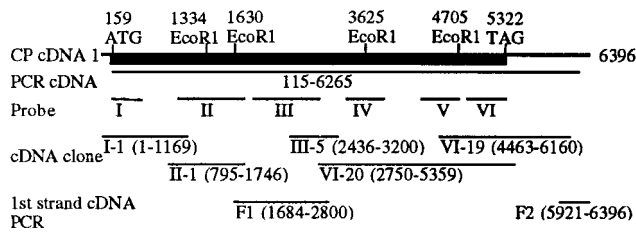


FIG. 1. Cloning strategy, schematic view of the crayfish CP cDNA. The obtained CP cDNA with positions indicated for the ORF (thick, black line), start codon (159 ATG), and stop codon (5322 TAG). Nucleotides 115-6265 were first obtained by repeated PCRs, labeled PCR cDNA in the figure. Radioactively labeled PCR products from different parts of the PCR cDNA (probes I-VI) were used to screen the cDNA library by plaque hybridization. The selected clones from the cDNA library screening that were sequenced are indicated as I-1, II-1, III-5, VI-19, and VI-19. Roman numerals indicate which probe was used to obtain the clones. F1 and F2 are PCR products obtained from first-strand cDNA, covering sequences that are not present in any of the cDNA clones.

were digested with *EcoRI*, and the released cDNA inserts were subcloned into *EcoRI*-digested pBluescript II (SK+) plasmids (Stratagene). The inserts of some clones contained internal *EcoRI* sites present in the CP cDNA and were therefore cleaved into two or three fragments (Fig. 1). The order of these fragments was confirmed by sequencing PCR products from the cDNA library, covering these *EcoRI* sites (not indicated in Fig. 1). The sequences of the inserts were determined by automated DNA sequencing on an ABI Prism 377 DNA sequencer (Applied Biosystems).

**PCR with First-Strand cDNA as Template.** Total RNA was isolated from the hepatopancreas of male crayfish by using the

Trizol LS reagent (GIBCO/BRL) according to the manufacturer's instructions. mRNA was purified from the total RNA by using Quickprep Micro mRNA purification system (Amersham Pharmacia). By using mRNA as template, first-strand cDNA was synthesized with an oligo(dT) primer (22-mer) and Ready-To-Go You-Prime First-strand beads (Amersham Pharmacia). With the first-strand cDNA as template, PCRs were performed to obtain the 3' end with poly(A) tail by using an oligo(dT) primer in combination with a specific primer from the 3' untranslated region (UTR) (F2 in Fig. 1) and to obtain PCR products confirming the sequence of a gap in the cDNA obtained from cDNA clones (F1 in Fig. 1). The PCR products were subcloned into T vectors (25) and sequenced on an ABI Prism 377 DNA sequencer (Applied Biosystems).

**Sequence Analysis.** MACVECTOR 4.1.4 (Kodak) was used to analyze the obtained sequences. The presence of related sequences in the databases was examined by using the BLAST algorithm (27).

**Northern Blot Analysis.** Total RNA was isolated from the hepatopancreas of male crayfish by using the Trizol LS reagent (GIBCO/BRL). Total RNA (30 µg) was separated by electrophoresis on 1% agarose gel containing formaldehyde and transferred to a Hybond N nylon membrane (Amersham Pharmacia), by using standard techniques (28). The nylon membrane was hybridized with a radioactive CP-specific probe, <sup>32</sup>P-labeled using the Megaprime labeling kit (Amersham Pharmacia). Hybridization and washing were performed at 65°C. Presence of radioactivity on the membrane was examined by exposure to x-ray film for 24-48 hr.

RESULTS AND DISCUSSION

A total of 299 amino acids from the CP were determined by amino acid sequencing of the N terminus (20 residues) and

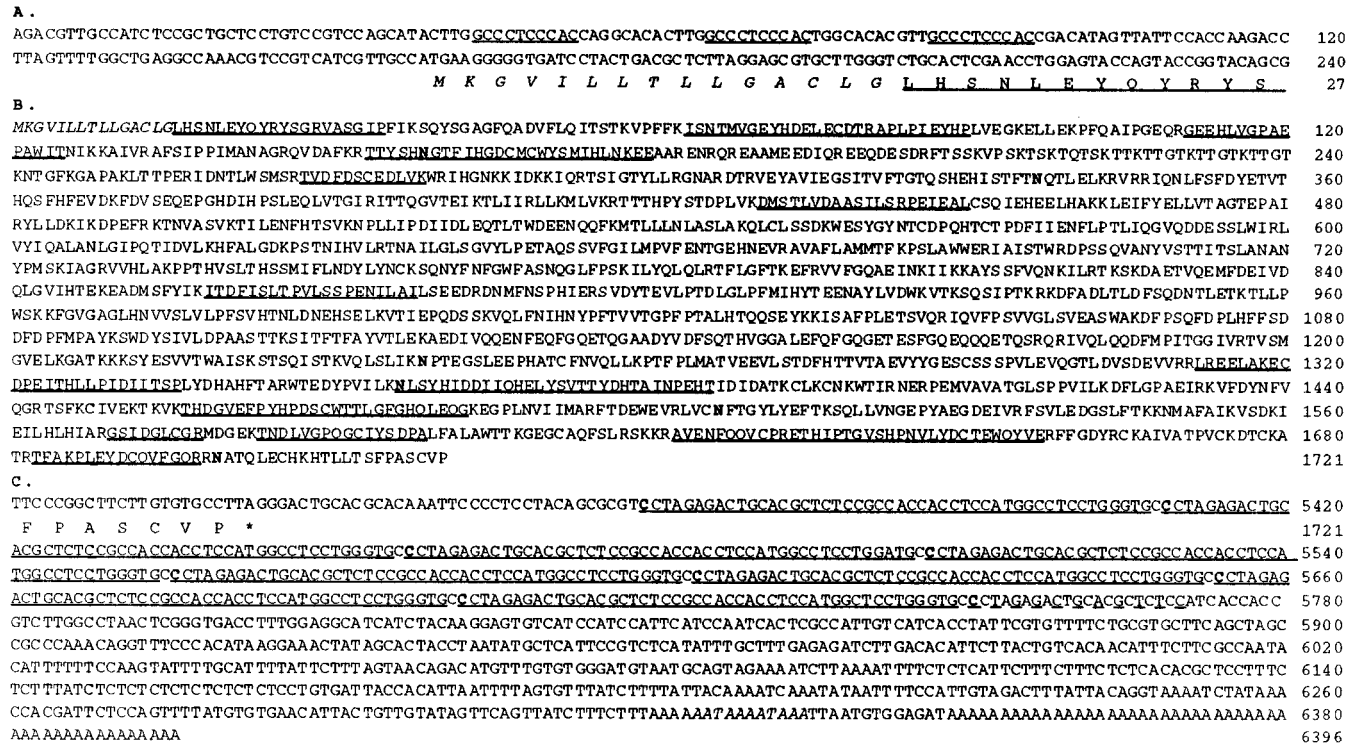


FIG. 2. Deduced amino acid sequence and 5' and 3' UTRs of the crayfish CP cDNA. (A) Nucleotides 1-240 of the CP cDNA, containing the 5' UTR (1-158), and the beginning of the coding region (159-240) with its deduced amino acid sequence (1-27). A repeated sequence motif, 10 nt long, is underlined. (B) The deduced amino acid sequence of the CP. The putative signal sequence is in italics. Sequences confirmed by peptide sequencing are underlined. Putative N-glycosylation sites are in bold. (C) Nucleotides 5301-6396 of the CP cDNA, containing the end of the coding region including the stop codon (5301-5324) with its deduced amino acid sequence (1715-1721), and the 3' UTR (5325-6396). A repeated sequence motif, 49 nt long, is underlined. The repeats start with a C (in bold) and ends with a C that is not underlined. The eight complete repeats (underlined) are identical except for nucleotide 5503 in repeat number 3, which is an A instead of the consensus G, and repeat number 8 is missing one C next to nucleotide 5740. One incomplete repeat is indicated by a dashed line. The two polyadenylation signals are in italics.

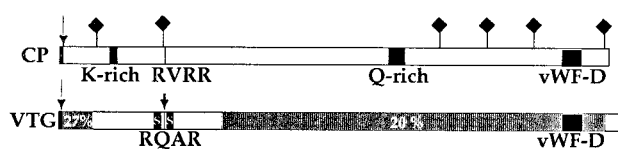


Fig. 3. Schematic view of the crayfish CP. Black areas in the CP indicate, from left to right, the short putative signal peptide with cleavage site (↓), the lysine-rich region (K-rich), the glutamine-rich region (Q-rich), and the region with sequence similarity to the vWF D-domain (vWF-D). Also indicated are putative *N*-glycosylation sites (◆), and the characteristic processing site that is not cleaved in the CP (RVRR). To illustrate the sequence similarity, the schematic view of an insect (*Athalia rosae*) VTG (36) is added below (VTG), where regions with sequence similarity are indicated as two gray areas, one short *N*-terminal region (27% identity to the CP sequence; see also Fig. 4A), and a long region spanning the C-terminal two thirds of the protein (20% identity to the CP sequence; see also Fig. 4B and C). Black areas in the VTG indicate, from left to right, the putative signal peptide with cleavage site (↓), the two polyserine stretches (S) on each side of the proteolytic processing site in the VTG (RQAR, ↓), and the region with sequence similarity to the vWF D-domain (vWF-D).

tryptic peptides (14 peptides) of the protein (Fig. 2). These sequences were used for designing degenerate oligonucleotide primers for the initial PCRs and for confirming the authenticity of the cDNA sequences later obtained.

The cDNA for the crayfish CP was established from the hepatopancreas cDNA library and hepatopancreas first-strand cDNA [synthesized with an oligo(dT) primer] by sequencing both PCR products and cDNA library clones (Fig. 1). Three independent PCR products were sequenced to determine the nucleotide sequence in the coding region (1747–2435) not covered by cDNA library clones (Fig. 1). The 3' end (nucleotides 6161–6396) was determined from two PCR products (Fig. 1).

The cDNA for the CP consists of 6,396 nt which is comparable to an estimated size of approximately 7 kb for the CP mRNA transcript from hepatopancreas in Northern blots (data not shown). The cDNA has a 5' UTR of 158 nt, a 5,166-nt coding region (including the stop codon), and a 1,072-nt 3' UTR (Figs. 1 and 2). Repeated sequence motifs are found in both UTRs—three repeats of 10 nucleotides in the 5' UTR, and eight repeats of 49 nucleotides in the 3' UTR (Fig. 2A and C). The functions of these repeats are not known, but they could be involved in regulating the translation and/or stabilizing the mRNA. The 3' end also contains two polyadenylation signals and a poly(A) tail (Fig. 2C).

The coding region translates into a protein of 1,721 aa, starting with a characteristic signal peptide sequence of 15

amino acids, followed by the N-terminal sequence of the CP subunit determined by amino acid sequencing (Fig. 2B). The mature protein, consisting of 1,706 aa, has an estimated molecular mass of 193 kDa and a pI of 5.55. There are six potential *N*-glycosylation sites (Figs. 2B and 3), which can account for the higher molecular mass of the CP subunit estimated from SDS/PAGE, 210 kDa (12). Two of the potential *N*-glycosylation sites (N159 and N1357) are found in peptides determined by amino acid sequencing (Fig. 2B). No amino acid was detected in these two positions, and it is likely that both these asparagines are glycosylated.

The N terminus of the CP is very similar to the N termini of other crustacean CPs, the lobster fibrinogen (17), and sand crayfish VHDL (13) (Fig. 4A). We therefore assume that conclusions based on the sequence of the cloned crayfish CP are also valid for the crustacean CPs as a group. The deduced amino acid sequence of the CP displays sequence similarity to VTGs, showing highest similarity to insect VTGs (Figs. 3 and 4). Although the overall similarity is relatively low, there are stretches distributed along the entire sequence that can confidently be aligned to the corresponding regions in VTGs (Figs. 3 and 4). Some features, like the position of a dibasic proteolytic processing site RXRR present in insect VTGs and the spacing of the cysteines toward the C terminus (19, 20) are also found in the crayfish CP (Fig. 3). Unlike most insect VTGs, the crayfish CP is not cleaved at the RVRR sequence (residues 344–347, Figs. 2B and 3). The CP also lacks a poly(serine) region present in most VTGs (refs. 19 and 20; Fig. 3). VTGs are reported to display sequence similarity to other lipoproteins such as apolipoprotein B (29, 19, 20), and microsomal triglyceride transfer protein (30), suggesting that these proteins have regions with a common evolutionary origin. Toward the C-terminal end, the CP and VTGs have a cysteine-containing stretch with sequence similarity to the D domain of von Willebrand factor (vWF) in mammals (refs. 31–33; Figs. 3 and 4C). vWF is involved in the coagulation of blood in mammals, and its D domains are important for the formation of multimers of human vWF (34, 35).

The CP has previously been purified and characterized only from male crayfish (12, 14). Because of its sequence similarity to the female-specific VTGs, we purified CP from plasma of female crayfish by precipitation at low ionic strength and KBr gradient ultracentrifugation. CP from males and females are both VHDLs, have the same molecular mass in SDS/PAGE under reducing and nonreducing conditions, and have the same N-terminal sequence (11 residues determined for the female CP). Female CP also functions as a clotting protein, *i.e.* it forms a stabilized clot in the presence of active, endogenous

<b>A.</b>				
P.l. CP	16	LHNSLEYQYRYSGRVASGIPFFIKSQYSGAGFOADVFLQITSTKVVFFKISNTMVGEYHDELECDTRAPLPIEYHPLVEGKE	96	
P.i. fibrinogen	1	LQPKLEYQYKYLIVLSEIPSYKNOFYDAH	30	53 %
I.c. VHDL	1	LQPGLEYQYFYNGRVA	17	71 %
A.r. VTG	21	WKTQGEYFYVVRGRFLAALHQVALQYTSIALKAQKLCQPKGSDALSCNDRPQVAEILHAQPPGGWDSRPPKSKTNYQKFE	100	25 %
<b>B.</b>				
P.l. CP	601	VYIQALANLGIPTIDVLKHFALGDKPSTNIHVLRTNAILGLSGVYLPETAQSSVFGILMPVFENTGEHNEVRAVAFLAMM	681	
A.r. VTG	681	VYIYFAIGNLGHKMLVFEFYLEQDELSSDFQKSMVLSMNLARLMEKQRRAV--FYKIYQVNGEAYQVRTAEVFOIL	757	
<b>C.</b>				
P.l. CP	682	TFKPSLAWWERIALISTWRDPSSQVANYVSTTITSLANANYPMSKIAGRNVHLAKPPHVSLSLTHSSMIFLNDVLYNCKSQNY	762	
A.r. VTG	758	RTNEPPSMLRMAEENQITSKQVNAVAKSITETAFAKLDDAELAHNAFAARDLLSTKNMGQYERG-VHEDYTSSEAMSSCY	837	27 %
<b>C.</b>				
P.l. CP	1534	VRFVLEEDGSLFTKKNMAF---AIKVSDKIEILHLHIARSGSIDLGCGRMDGKTNLGVGQCCIYSDPALFALAWTTKGGG	1611	
A.r. VTG	1646	QFEAYIKPSGVVYIHSEEGVRAVLYDGERIKLILNNEYFDIIRGLCCFTNGEPATLFTAPCNCLKNEHFAASYALTKDQ	1726	23 %
H.s. vWF-D	2050	VRENHGHIFTEPQNE---LQLSHEKTF-----FASKT-YGLCCICENGANPFMLRRCVTVTWKTLVQAEVQRRK	2120	28 %

Fig. 4. Sequence comparisons. Sequences from crayfish (*P. leniusculus*) CP (*P.l.* CP) compared with the corresponding regions of an insect (*A. rosae*) VTG (*A.r.* VTG; ref. 36). The comparison in *A* corresponds to the beginning of the short N-terminal region with sequence similarities shown in Fig. 3. *B* and *C* correspond to the beginning and the vWF-D region, respectively, of the long region with sequence similarity in Fig. 3. Also added in *A* are the N-terminal sequences of lobster (*Panulirus interruptus*) fibrinogen (*P.i.* fibrinogen, 17) and sand crayfish (*Ibacus ciliatus*) VHDL (*I.c.* VHDL, 13) and in *C* part of the human (*Homo sapiens*) vWF D4-domain (*H.s.* vWF-D, 37). The sequences are denoted by the position in published sequences. The percentage of identity to the CP sequence is listed to the right, the identical amino acids being shown in white on black background.

TGase activity from the hemocytes. This shows that the CP is present and has identical function in both sexes.

The shape of the CP molecule and the formation of crosslinked CP aggregates were examined by using electron microscopy. The pictures in Fig. 5 show negatively stained crayfish CP molecules purified by KBr gradient ultracentrifugation followed by gel filtration. CP isolated by precipitation at low ionic strength had similar shape in electron microscopy, but the preparations appeared to be more aggregated. Both triangular (Fig. 5 *E1* and *E2*) and diamond-shaped (Fig. 5 *E3* and *E4*) projections of the CP particle can be recognized. The sides of the triangular projections appear to be of equal length ( $11.7 \pm 0.6$  nm). The different projections most likely represent different views of the particle, suggesting that the overall shape of the CP is tetrahedral. A similar shape was described from electron micrographs of the lobster fibrinogen molecules (11).

The TGase-dependent clotting reaction results in a time-dependent, linear polymerization of CP particles (Fig. 5 *A–D*).

A crayfish hemocyte lysate containing endogenous TGase activity was used to crosslink the CP molecules. Elongation of the polymers is relatively rapid, resulting in chains of 60 units or more within 60 s. During polymerization, the individual CP units seem to maintain their overall shape, as both triangular and diamond-shaped projections of the particles are recognized within the polymer chain. The association of individual units does not involve large particle surfaces. Instead, polymerized particles appear to interact at very localized points at the corners of the tetrahedrons, leaving considerable room for the deposition of stain between individual units of the chain (Fig. 5*F*, arrowheads). The chains appear to be very flexible and bend randomly in different shapes (Fig. 5 *C*, *D*, and *F*). Linear polymers of CP often get entangled during the electron microscope preparation, which may lead to the false impression of bidirectional branching in the electron micrograph. However, certain arrangements of CP molecules clearly suggest unidirectional branching of the polymer chain (Fig. 5 *C*,

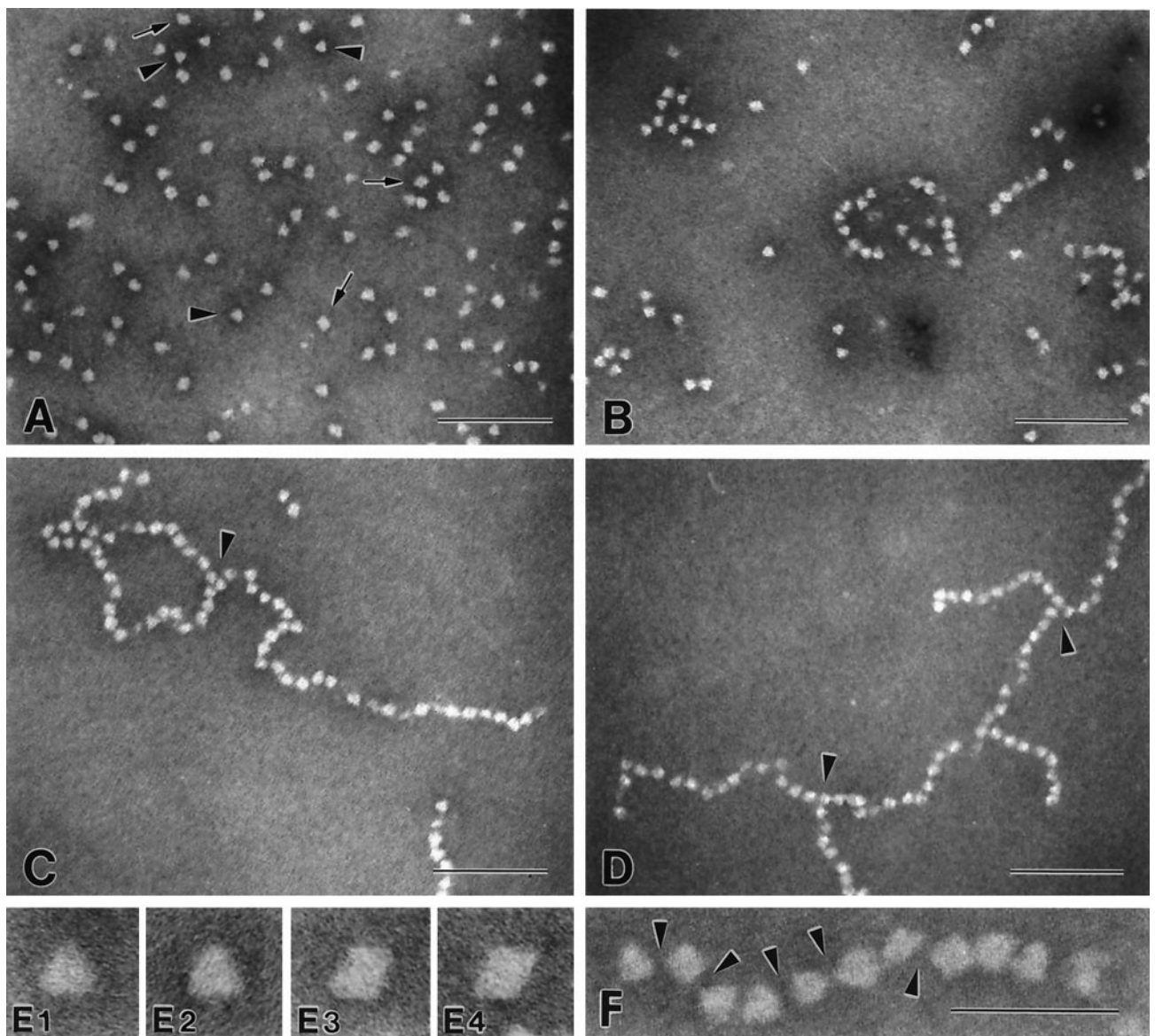


FIG. 5. Electron microscopy of crayfish CP samples. Negatively stained preparations of crayfish CP were analyzed by using electron microscopy. (*A*) CP monomers (CP monomers, HLS, and EDTA mixed); arrowheads indicate triangular projections; arrows indicate diamond-shaped projections. (*B–D*) CP monomers, HLS, and  $Ca^{2+}$  mixed, and the TGase-dependent crosslinking was stopped by addition of EDTA after 15 s (*B*), 45 s (*C*), and 60 s (*D*). Arrowheads in *C* and *D* indicate branchpoints. (*A–D*) Magnification  $\times 172,000$  (Bar = 100 nm.) (*E*) Selected triangular (*E1* and *E2*) and diamond-shaped (*E3* and *E4*) projections from the monomer preparation, magnification  $\times 810,000$ . (*F*) Polymer of crosslinked CP molecules; arrowheads illustrate the deposition of stain between the individual CP molecules, indicating that the CP molecules interact at very localized points and that the interaction does not involve large surfaces of the individual CP molecules. Magnification  $\times 515,000$ . (Bar = 50 nm.)

and D). Because each crayfish CP subunit contains at least one lysine and one glutamine sidechain available for TGase-mediated crosslinking (12) and the CP molecule consists of two identical disulfide-linked subunits, there should be a minimum of four residues—two lysines and two glutamines—available for crosslinking on each CP molecule. This makes branching of the polymerizing CP chains possible. Two stretches, which could be involved in the crosslinking, can be identified in the deduced amino acid sequence. One stretch is rich in lysine residues (amino acids 211–241, 26% lysine, Figs. 2B and 3) and the other stretch is rich in glutamine residues (amino acids 1124–1185, 27% glutamine, Figs. 2B and 3). The lysine-rich stretch has a repeated lysine-containing motif, (S/T)KT(T).

In conclusion, the crustacean CPs are the second type of clot-forming proteins in animal plasma to be characterized and cloned, besides the vertebrate fibrinogens (1–3). The crustacean CPs are evolutionary related to VTGs, but the CPs should not be considered as true VTGs, as they have a completely different function and they are constitutively expressed in both males and females.

We thank Edita Navratilova, University of Arizona, for technical assistance, and Dr Jan J. Enghild, Duke University Medical Center, for determining the N-terminal protein sequence of crayfish CP from females. This work was supported by grants from the Swedish Natural Research Science Research Council (K.S.), the Swedish Research Council for Agriculture and Forestry (K.S.), European Union-Agriculture and Fisheries Programme (FAIR) (PL-97-3660 to K.S.), the Nordic Academy for Advanced Study (M.H.), and the National Institutes of Health (Grant GM50551 to R.v.A.).

- Doolittle, R. F. (1984) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, San Diego), 2nd Ed., Vol. 10, pp. 421–544.
- Furie, B. & Furie, B. C. (1988) *Cell* **53**, 505–518.
- Davie, E. W., Fujikawa, K. & Kisiel, W. (1991) *Biochemistry* **30**, 10363–10370.
- Folk, J. E. (1980) *Annu. Rev. Biochem.* **49**, 517–531.
- Lorand, L. & Conrad, S. M. (1984) *Mol. Cell. Biochem.* **58**, 9–35.
- Greenberg, C. S., Birckbichler, P. J. & Rice, R. H. (1991) *FASEB J.* **5**, 3071–3077.
- Kawabata, S.-i., Muta, T. & Iwanaga, S. (1996) in *New Directions in Invertebrate Immunology*, eds. Söderhäll, K., Iwanaga, S. & Vasta, G. R. (SOS Publications, Fair Haven, NJ), pp. 255–283.
- Tokunaga, F., Yamada, M., Miyata, T., Ding, Y.-L., Hiranaga-Kawabata, M., Muta, T., Iwanaga, S., Ichinose, A. & Davie, E. W. (1993) *J. Biol. Chem.* **268**, 252–261.
- Tokunaga, F., Muta, T., Iwanaga, S., Ichinose, A., Davie, E. W., Kuma, K.-i. & Miyata, T. (1993) *J. Biol. Chem.* **268**, 262–268.
- Fuller, G. M. & Doolittle, R. F. (1971) *Biochemistry* **10**, 1305–1311.
- Fuller, G. M. & Doolittle, R. F. (1971) *Biochemistry* **10**, 1311–1315.
- Kopacek, P., Hall, M. & Söderhäll, K. (1993) *Eur. J. Biochem.* **213**, 591–597.
- Komatsu, M. & Ando, S. (1998) *Biosci. Biotechnol. Biochem.* **62**, 459–463.
- Hall, M., van Heusden, M. C. & Söderhäll, K. (1995) *Biochem. Biophys. Res. Commun.* **216**, 939–946.
- Söderhäll, K., Cerenius, L. & Johansson, M. W. (1996) in *New Directions in Invertebrate Immunology*, eds. Söderhäll, K., Iwanaga, S. & Vasta, G. R. (SOS Publications, Fair Haven, NJ), pp. 229–253.
- Söderhäll, K. & Cerenius, L. (1998) *Curr. Opin. Immunol.* **10**, 23–28.
- Doolittle, R. F. & Riley, M. (1990) *Biochem. Biophys. Res. Commun.* **167**, 16–19.
- Byrne, B. M., Gruber, M. & Ab, G. (1989) *Prog. Biophys. Molec. Biol.* **53**, 33–69.
- Chen, J.-S., Sappington, T. W. & Raikhel, A. S. (1997) *J. Mol. Evol.* **44**, 440–451.
- Sappington, T. W. & Raikhel, A. S. (1998) *Insect Biochem. Mol. Biol.* **28**, 277–300.
- Doolittle, R. F. & Fuller, G. M. (1972) *Biochim. Biophys. Acta* **263**, 805–809.
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254.
- Valentine, R. C., Shapiro, B. M. & Stadtman, E. R. (1968) *Biochemistry* **7**, 2143–2152.
- Marchuk, D., Drumm, M., Saulino, A. & Collins, F. S. (1991) *Nucleic Acids Res.* **19**, 1154.
- Cerenius, L., Liang, Z., Ducic, B., Keyser, P., Hellman, U., Palva, E. P., Iwanaga, S. & Söderhäll, K. (1994) *J. Biol. Chem.* **269**, 29462–29467.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Baker, M. E. (1988) *Biochem. J.* **255**, 1057–1060.
- Shoulders, C. C., Narcisi, T. M. E., Read, J., Chester, S. A., Brett, D. J., Scott, J., Anderson, T. A., Levitt, T. A. & Banaszak, L. J. (1994) *Struct. Biol.* **1**, 285–286.
- Baker, M. E. (1988) *Biochem. J.* **256**, 1059–1063.
- Sadler, J. E. (1991) *J. Biol. Chem.* **266**, 22777–22780.
- Ruggieri, Z. M. & Ware, J. (1993) *FASEB J.* **7**, 309–316.
- Voorberg, J., Fontijn, R., van Mourik, J. A. & Pannekoek, H. (1990) *EMBO J.* **9**, 797–803.
- Mayadas, T. N. & Wagner, D. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3531–3535.
- Nose, Y., Lee, J. M., Ueno, T., Hatakeyama, M. & Oishi, K. (1997) *Insect Biochem. Mol. Biol.* **27**, 1047–1056.
- Bonthron, D., Orr, E. C., Mitscock, L. M., Ginsburg, D., Handin, R. I. & Orkin, S. H. (1986) *Nucleic Acids Res.* **14**, 7125–7127.