

Identifying exoskeleton proteins in the blue crab from an expressed sequence tag (EST) library

Thomas H. Shafer¹, Michael A. McCartney, Lindsay M. Faircloth

Department of Biology and Marine Biology, University of North Carolina–Wilmington, 601 South College Road, Wilmington, NC 28403-5915, USA

Synopsis A blue crab (*Callinectes sapidus*) expressed sequence tag project was designed for multiple purposes including discovery of genes for cuticular (exoskeletal) proteins, some of which may regulate mineralization. One of the expression libraries sequenced was from the hypodermis (the epithelium depositing the cuticle). RNAs used for cDNA synthesis were pooled from arthroal and mid-dorsal hypodermis at both pre-ecdysis and post-ecdysis. This ensured representation from both calcifying and non-calcifying regions and from layers of cuticle deposited both before and after ecdysis. The EST database was mined for cuticular protein sequences in three ways. First, we searched for sequences coding for known cuticle-specific motifs like the Rebers-Riddiford chitin-binding sequence and a motif known only from proteins extracted from mineralized exoskeletons of other decapods. Second, we checked the associated annotations in the EST project for similarity to known cuticular proteins, often from insects. Third, BLAST was used to search the EST data for significant homology to published cuticular protein sequences from other crustaceans. In all, the database contains at least 73 contigs or singlets representing transcripts of cuticular proteins. Forty-five of these distribute among ten clusters of very similar transcripts, possibly representing alternative splicing or recent gene duplications. The rest share less similarity. We have obtained complete sequences for 25 of the transcripts, have produced phylogenetics trees comparing them with similar proteins from insects and other crustaceans, and have determined expression patterns across the molt in calcifying versus non-calcifying cuticle. The combination of homology analysis and gene expression analysis allows us to infer putative functions in cuticle synthesis and calcification.

Introduction

Expressed sequence tags (ESTs) are partial sequences of cDNAs typically produced by rapid single-pass sequencing of randomly chosen clones from a cDNA library. An EST project is a large dataset of such partial sequences from one or more libraries and is thus a representation of the transcriptome of a particular tissue or tissues at a particular stage of development and physiological condition. For organisms without genome projects and thus having as yet poorly described genetic backgrounds, the ease and relative economy of partial sequencing of cDNA libraries can be a highly productive way to generate a great deal of information about functioning genes (Boguski and others 1993; Rudd 2003).

Typically, EST datasets either lead to or derive from the production of microarrays, and so their annotations are useful for characterizing and understanding multi-gene expression patterns (for example Whitfield and others 2002; papers in this Symposium including

McClintock and others 2006, and Stillman and Teranishi 2006). In a more targeted approach, however, datasets produced in EST projects can be mined for the presence of transcripts with similarity to known protein sequences that might indicate a relationship to a particular physiological or developmental question (Blackshear and others 2001). This is the approach we report here, where the database produced in a blue crab (*Callinectes sapidus*) EST project (Coblentz and others 2006), though designed for multiple purposes, has been used to obtain sequence information about proteins deposited in the cuticle. Some of these proteins may regulate exoskeleton mineralization.

To date, 73 different cuticle transcripts have been discovered. We have obtained complete sequences and gene expression patterns for 25 of these. After reviewing the blue crab as a model system for analysis of the control of mineralization and the known cuticle protein sequences from other decapod crustaceans, we summarize the relevant facts about our *C. sapidus*

From the symposium "Genomic and Proteomic Approaches in Crustacean Biology" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 4–8, 2006, at Orlando, Florida.

¹ E-mail: shafert@uncw.edu

Integrative and Comparative Biology, volume 46, number 6, pp. 978–990
doi:10.1093/icb/icl022

Advance Access publication July 20, 2006

© The Author 2006. Published by Oxford University Press on behalf of the Society for Integrative and Comparative Biology. All rights reserved. For permissions please email: journals.permissions@oxfordjournals.org.

EST project (Coblentz and others 2006). Next, we describe the approaches used to mine the database for cuticular transcripts. Finally, the general results obtained so far are reviewed in an attempt to illustrate the approach and construct a synthesis of our findings. The actual sequences and the details of the gene expression data are not given either because they are already published (Wynn and Shafer 2005), have been submitted for publication (Faircloth and Shafer in review), or are being prepared for publication (Kennedy and others manuscript in preparation), or because they are preliminary and need additional replication.

Crustacean exoskeleton as a model system to study matrix deposition and mineralization

Biomineralization, the deposition of insoluble inorganic material, is a widespread phenomenon, occurring in organisms as diverse as coccolithophores, sponges, corals, mollusks, crustaceans, echinoderms, and vertebrates (Lowenstam and Weiner 1989; Mann 2002). The minerals formed are used for support, protection, mastication, gravity perception, and various other functions. The mineral found most commonly in the hard structures of invertebrates is calcium carbonate (CaCO_3), which may exist either in amorphous form or as one of the crystal morphs calcite, aragonite, or vaterite (Weiner and Dove 2003). In all cases, it is deposited in conjunction with an organic matrix. The macromolecules of that matrix, especially the proteins, glycoproteins, and proteoglycans, are believed to play significant roles in the regulation of calcification. They probably control the nucleation, orientation, and cessation of crystal growth (Addadi and Weiner 1985). It is not surprising, therefore, that the matrices in mineralized tissues are complex. The crustacean cuticle, or exoskeleton, is no exception. It is a composite material composed of a variety of proteins and glycoproteins intimately associated with chitin fibrils. While many of these components are undoubtedly shared across the Arthropoda, the crustaceans have evolved the ability to use CaCO_3 to augment cuticle hardening and so might be expected to have calcification-regulating proteins not found in insects, where sclerotization is the only hardening mechanism.

Crustaceans, especially large decapods, are ideal model systems in which to study matrix formation and calcification. Growth in these animals requires that they undergo cyclic shedding of the exoskeleton. Unlike other mineralizing organisms, they must periodically deposit and calcify an entirely new chitin-protein matrix at each molt. The tissue responsible for cuticle synthesis is the hypodermis, an epithelium

of only 1 cell layer. The process of exoskeleton calcification is tightly regulated both temporally and anatomically. The temporal regulation is exemplified in the molt cycle, where calcification must closely follow the emergence of the animal from its old exoskeleton (Roer and Dillaman 1984, 1993). Preparation for ecdysis begins with apolysis, the separation of the underlying hypodermis from the old exoskeleton. Other events of pre-ecdysis, comprising stages D_0 , D_1 , D_2 , D_3 , and D_4 , include partial resorption of the organic and mineral components of the old cuticle and generation of a new epicuticle and exocuticle (the pre-exuvial layers) beneath the old exoskeleton (Drach and Tchernigovtzeff 1967). The fully formed pre-exuvial layers remain completely uncalcified at this time. Ecdysis is referred to as stage E, after which the animal enters post-ecdysis, referred to as stages A_1 , A_2 , B_1 , B_2 , C_1 , C_2 , and C_3 . The pre-exuvial layers mineralize after expansion by water uptake and the endocuticle is deposited during post-ecdysis. The endocuticular lamellae mineralize as they appear or very soon thereafter. The deposition of the thin, uncalcified membranous layer signals entry into an ecdysis or intermolt, stage C_4 (Drach and Tchernigovtzeff 1967; Roer and Dillaman 1984, 1993). Obviously, the timing of cuticular synthesis and the control of the initial CaCO_3 accumulation in each layer is critical.

The need for anatomical regulation of exoskeleton calcification by proteins secreted by the hypodermis is evident in the differential calcification of specific body parts. For example, the dorsal carapace and the chelipeds are heavily calcified and become extremely rigid while a thin suture line calcifies less and in a pattern that promotes its systematic rupture at the next molt (Priester and others 2005). At the other extreme, the cuticle of the joints, gills, and walls of the branchial chamber are uncalcified and remain flexible. Of particular interest to us are the arthroal membranes, cuticle at each joint that remains completely uncalcified and allows for movement. Arthroal membrane is similar to hard cuticle in the timing of deposition, in overall thickness of pre-exuvial and post-exuvial layers and in the morphology of the lamellae (Williams and others 2003). The composition of its organic matrix, however, is considerably different from calcified cuticle (Hepburn and Chandler 1976), and a comparison of the proteins in the 2 locations can be useful for understanding cuticular mineralization.

The blue crab, *C. sapidus*, is our organism of choice because of its economic importance to our region and because we are able to obtain integument material at precise time points both pre-ecdysis and post-ecdysis from a soft-shell crab "shedding" operation. The histology, histochemistry, and ultrastructure of cuticular

deposition have been studied and key patterns of initial and subsequent mineralization have been described (Elliott and Dillaman 1999; Williams and others 2003; Dillaman and others 2005; Priester and others 2005). For example, post-ecdysial deposition of calcium carbonate into the exocuticle is now known to proceed from both distal and proximal aspects along septa between polygonal columns (interprismatic septa) (Dillaman and others 2005). The resulting honeycomb pattern produces a rigid cuticle with a minimum of mineral. Only later do the prisms in-fill with CaCO_3 . Furthermore, evidence has been obtained indicating that mineral deposited first (in 3–5 h) at the epi-exocuticular boundary and within the interprismatic septa is amorphous calcium carbonate, which only later (8–12 h) is transformed to calcite (Dillaman and others 2005). Presumably all these fine-scale differences in the location and type of mineral deposited must be under the control of cuticular proteins.

In a previous study (Shafer and others 1995) we used isolated crab cuticle explants free of cellular material to show that the control of CaCO_3 deposition after the molt resides in the organic matrix itself rather than in, for example, the ion-pumping activity of the underlying hypodermis. Dramatic post-ecdysial changes in glycoproteins synchronous with initial mineralization in the exocuticle were documented histochemically (Marlowe and others 1994) and by lectin blotting following SDS-PAGE (Shafer and others 1994, 1995). Subsequently, cuticle proteins were extracted with various solvents at critical times at and after the molt and were tested for their ability to affect calcite nucleation (Coblentz and others 1998). The molecular weights of the proteins that associated with calcite crystals as they formed *in vitro* were determined by electrophoresis. The results were used to formulate a model wherein one or more acidic proteins form nucleating sites in the cuticle and glycoproteins present only in pre-ecdysis and for the first few hours post-ecdysis shield these nucleators from the calcium and carbonate ions or otherwise inhibit their function (Coblentz and others 1998). Recently, the purification of a specific cuticular glycoprotein from ecdysial cuticle provided strong support for this hypothesis (Tweedie and others 2004). Immunoblot analysis using an antibody produced against the N-terminal amino acid sequence suggested that several changes occur in the glycosylation pattern of this protein, including the loss of the highest molecular mass glycoform, during the first few hours after the molt. Immunohistochemistry showed that the antigen decreases in the exocuticle during early post-ecdysis and, most importantly, that the decrease occurs first in the interprismatic septa beginning as

early as 2 h after the molt. Thus, we produced temporal and fine-scale spatial correlations between the loss of a possible inhibitor and the initiation of CaCO_3 deposition. This protein, at least in its most heavily glycosylated form, does not exist in the arthroal membrane either pre-ecdysis or post-ecdysis (T.H. Shafer unpublished observations). We also have identified a glycosidase that appears in the cuticle during the early post-ecdysial hours (Roer and others 2001) and have produced evidence that N-acetylhexosaminidase activity can alter cuticular glycans and increase cuticle-explant calcifying potential in a manner that mimics *in vivo* post-ecdysis processes (Pierce and others 2001).

Crustacean cuticular proteins directly sequenced or translated from cDNA

Since a major objective of what follows is to characterize *C. sapidus* cuticular proteins based on their similarity with sequenced proteins from other species, it is important to review work on cuticular proteins in other decapods. Andersen and his associates have extracted and directly sequenced proteins from the insoluble matrix of intermolt crustacean cuticles after decalcification in acetic acid. These include 1 protein from the shrimp *Pandalus borealis* (Jacobsen and others 1994); 17 proteins from the mineralized cuticle of the lobster *Homarus americanus* (Kragh and others 1997; Nousiainen and others 1998); 6 arthroal membrane proteins from the same species (Andersen 1998b); and 16 proteins from the crab *Cancer pagurus* (Andersen 1999), consisting of 11 from hard cuticle only, 4 from arthroal membrane only, and 1 found in both types of cuticle. All sequenced proteins were rather small (4–20 kDa) and tended to fall into similarity groups related to the cuticle of their origin and to recognized or described amino acid motifs. All the arthroal membrane proteins in these studies contained a Rebers-Riddiford (RR) motif. The RR consensus sequence was first described in *Manduca sexta* larval cuticle (Rebers and Riddiford 1988) and has been shown to bind chitin *in vitro* (Rebers and Willis 2001; Togawa and others 2004). It is common to many insect cuticular proteins, where sequences generally fall into 2 subgroups, RR-1 and RR-2, from the flexible cuticle and heavily sclerotized cuticle, respectively (Andersen 1998a). The motifs in the crustacean arthroal proteins are RR-1. Three of the mineralized crustacean cuticular proteins also contain Rebers-Riddiford motifs, but these variants cannot be classified as RR-1 or RR-2. All the other hard-cuticle proteins lack the motif completely. On the other hand, a hydrophobic 18-residue motif is described in 10 of the

17 hard-cuticle proteins from *H. americanus* (Kragh and others 1997; Nousiainen and others 1998) and 8 of the 15 hard-cuticle proteins from *C. pagurus* (Andersen 1999). This motif occurs either 2 or 4 times in each protein and is never found in arthroal membranes of crustaceans or in any cuticular proteins of insects. In other words, it is completely restricted to cuticle that is capable of mineralization. For convenience, we have named this motif "crust-18" (Faircloth and Shafer in review). One additional motif was recognized to be present in two of the *H. americanus* and one of the *C. pagurus* hard-cuticle proteins after it had been thoroughly characterized in post-ecdysial insect cuticle (Andersen 2000). For convenience, we have named this hydrophobic 18-residue motif "postmolt-18" (Faircloth and Shafer in review).

A study designed to discover transcripts expressed predominantly or exclusively in post-ecdysis tail-fan epidermis using the technique of differential display has produced sequence information on several cuticular proteins in the shrimp *Marsupenaeus japonicus*. DD9A and DD9B, 2 proteins with similar deduced amino acid sequences, contain partial RR-1 consensus sequences (Watanabe and others 2000). These transcripts are found in the lateral uncalcified exoskeletal region of the tail blade of shrimp, and not in the calcified medial region. DD5 encodes a cuticular protein unusual for the fact that it consists of tandem repeats of units of ~100 amino acid residues, each with an RR consensus sequence (Ikeya and others 2001). Another differentially expressed *M. japonicus* transcript identified exclusively in calcified cuticle was originally called DD4 (Endo and others 2000) but was renamed crustocalcin when additional open reading frame (ORF) sequence was obtained (Endo and others 2004). The encoded protein contains a Rebers-Riddiford-like motif in close association with a region highly enriched in acidic amino acid residues. The recombinant protein is capable of nucleating CaCO_3 *in vitro* (Endo and others 2004). The only other protein described from the *M. japonicus* differential display study is DD1, which contains no recognized cuticle motif (Watanabe and others 2006).

Two interesting proteins from the calcified cuticle of the crayfish *Procambarus clarkii* have been sequenced both directly (Inoue and others 2001, 2004) and from cDNA (Inoue and others 2003, 2004). They are termed calcification-associated peptide-1 and -2 (CAP-1 and CAP-2). Each contains a RR consensus motif flanked by stretches of acidic amino acid residues. They have been shown experimentally to bind chitin and to inhibit the formation of CaCO_3 in solution, a known property of proteins that could have either positive or negative effects on crystal formation

when associated with an insoluble organic matrix. The CAP-1 and CAP-2 transcripts are only expressed in the *P. clarkii* tail-fan blade during post-ecdysis, the time of exoskeleton calcification, and are described as candidate calcite-nucleating agents.

A summary of the blue crab EST project

We have produced an EST database from blue crab (*C. sapidus*) transcripts designed, in part, to facilitate the discovery of cuticular protein sequences in that species (Coblentz and others 2006). Two cDNA libraries were made from pooled RNAs. One included RNA extracts from both anterior (respiratory) and posterior (ion transporting) gills from animals acclimated to 2 salinities, 35 and 5 ppt. The other, more germane to the study of cuticular synthesis and composition, was constructed from hypodermis RNAs from both the arthroal membrane and the mid-dorsal carapace taken at both premolt stage D₂ and early postmolt (3 h). Each library was normalized in order to increase the relative proportion of less abundant transcripts (Bonaldo and others 1996). After normalization, the average size of the inserts increased significantly in the hypodermis library but not in the gill library. This suggests that the original hypodermis library contained small transcripts present in very high copy number. Indeed, we have since discovered that most of the transcripts for cuticular proteins, expected to be produced in the hypodermis but not in the gill, are small and abundant. Inserts of clones chosen at random from each library were sequenced in one direction using high-throughput one-pass techniques. A total of 11,763 ESTs met our quality criteria for inclusion in the project. About 45% were from the hypodermis library.

All ESTs, whether from gill or hypodermis, were put together in the assembly phase of the project (Coblentz and others 2006). Paracel Transcript Assembler software organized the sequences into putative transcripts in 2 steps. In the first, ESTs with regions that are highly similar were placed in clusters. Ninety percent of our sequences cluster and the rest are singlets, assumed to represent transcripts sequenced only once in the project. In the second step, sequences within clusters were arranged as contiguous putative transcripts. Ideally, one would expect all the ESTs in a cluster to form a single contig. However, sometimes clusters contain sequences with regions of dissimilarity and then more than one contig forms or else singlets form because they do not contig with the other ESTs in the cluster. Such complex clusters are thought to represent cases of either alternative splicing or

transcription from members of recently duplicated gene families. In the end, our data assembled into 883 contigs and 1293 singlets, representing 2176 putative blue crab transcripts. Almost 17% of these are part of complex clusters. There are 363 contigs that contain at least 1 EST from each library, indicating that they are reasonably abundant transcripts in both tissue types. The putative transcripts sequenced uniquely from hypodermal tissue are of the most interest for the discovery of cuticular proteins. There were 578 of these, 216 contigs and 362 singlets.

Several approaches were used to annotate the sequences and make them publicly available (Coblentz and others 2006). The 2176 putative transcripts were compared with the NCBI non-redundant protein database using BLAST, and the descriptions of the 3 most similar proteins, when significant, were retained in the database. GOBlet software (Hennig and others 2003) was used to produce and summarize the Gene Ontology (GO) results associated with the significant BLAST hits. This was possible for 32% of the putative blue crab transcripts. For transcripts with no significant BLAST results, known protein domains were revealed using InterPro (European Bioinformatics Institute). The annotation results, as well as the actual sequences of the ESTs and contigs, have been made available to the public on a website (<http://firedev.bear.uncw.edu:8080/shaferlab/>) that may be searched by sequence or by keyword. Users may also input a sequence to compare with our data using BLAST. The ESTs themselves have all been submitted to NCBI (dbEST) using trace2dbest software (<http://www.nematodes.org/PartiGene/>; A. Anthony, J. Parkinson, and M. Blaxter, unpublished data) that annotates each sequence with its most similar BLAST result at the time of submission. All of our sequences will also be submitted to the Marine Genomics Project (<http://www.marinegenomics.org>) where the BLAST annotations will be periodically updated and where other *C. sapidus* ESTs submitted later can be automatically assembled with these.

Approaches to “mining” the EST data for cuticle-related transcripts

Three strategies have been employed for discovering transcripts for cuticular proteins in the *C. sapidus* EST database. First, we searched the sequences for the invariant residues of known cuticular protein motifs such as the Rebers-Riddiford chitin-binding sequence and crust-18, the mineralized tissue motif described by Nousiainen and others (1998) and Andersen (1999). For example, the nucleotide sequence $\text{Tx}_4\text{GGxCCx}_4\text{GGx}_{13}\text{GxGGx}_{10}\text{CA}$ codes for

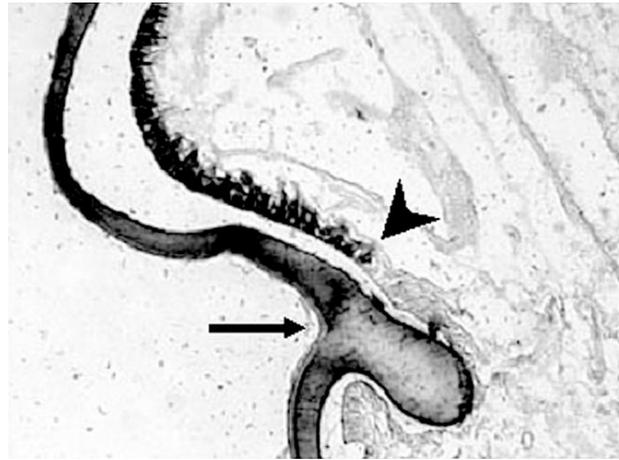


Fig. 1 *In situ* hybridization using an antisense riboprobe to detect the mRNA transcript coding for *Callinectes sapidus* arthroal membrane protein 8.1 (CsAMP8.1). Cross-section of a walking leg of the crab 3 h after molting. Arrow indicates a structural feature marking the boundary between arthroal membrane (above) and cuticle that will soon calcify (below). Staining in both types of cuticle is non-specific because an identical pattern is seen with the sense probe. The arthroal hypodermis (above the arrowhead) reacts specifically, indicating the presence of the mRNA, while the hypodermis underlying cuticle that will harden (below the arrowhead) does not. The separation of the cuticle from the hypodermis in this preparation is a fixation artifact.

$\text{LxGPSGx}_4\text{DGx}_3\text{Q}$, the invariant amino acids within the crust-18 motif, and is present in 10 of the contigs representing putative transcripts. Second, we conducted a word search for “cuticle” in the BLAST and InterPro annotations and compared the results with the GOBlet results for category “cuticle structure” in the GO molecular function hierarchy. Searching these annotations was especially useful for finding transcripts in our library that are more closely related to insect proteins than to any of the known crustacean cuticle proteins. Finally, we searched our own data using BLAST for translated proteins with similarity to previously sequenced crustacean proteins. These known sequences from other species were divided into groups based on the motifs they contain (RR-1, other Rebers-Riddiford variants, crust-18, and postmolt-18), and the putative transcripts in the EST dataset were searched for similarity to at least some members of each group.

These 3 search strategies produced highly overlapping results as might be expected. In all, 73 transcripts likely to code for cuticle proteins were identified. Forty-eight of these fall into 10 complex clusters. This is a very high proportion given that only 16.8% of all transcripts were in complex clusters. It suggests that at least some cuticular proteins are coded by alternatively

spliced genes or members of closely related gene families. The other transcripts for possible cuticular proteins are represented by 18 simple contigs and 7 singlets. Since 941 of all the putative transcripts in the EST database are expressed in the hypodermis (363 hypodermis-gill mixed contigs, 216 hypodermis-only contigs, and 362 hypodermis singlets), we estimate that the 73 mRNAs for cuticular proteins represent ~7.8% of the diversity in the hypodermal transcriptome as we have sequenced it.

So far, we have done additional work on 25 of these potential cuticular protein transcripts. At this time, that effort includes obtaining complete transcript sequences and ascertaining gene expression patterns across the molt cycle in arthroal membrane versus calcifying tissue hypodermis. Some of these data are published (Wynn and Shafer 2005), some submitted for publication (Faircloth and Shafer in review), and some are about to be submitted (Kennedy and others manuscript in preparation). Gene expression data are from combinations of real-time PCR, northern blotting, and *in situ* hybridization (ISH). The 4 published expression patterns (Wynn and Shafer 2005) do not include ISH data, so Figure 1 is included as an example of the use of this technique. The transcript is detected in the hypodermis depositing arthroal membrane, but not in the hypodermis depositing cuticle destined to calcify. It is also not present in muscle or other internal tissues.

ORFs analyzed to date: sequence similarities and expression patterns

Translations of the 25 cuticular transcripts have been analyzed for their relatedness to other known proteins, and those data, when combined with summaries of the patterns of gene expression, are being used to infer possible functions for the encoded proteins. Following the convention used for *C. pagurus* and *H. americanus* (Andersen 1998b, 1999), the names indicate the species (Cs for *C. sapidus*) and the probable location of the mature protein [AMP for arthroal membrane and CP for hard (dorsal) cuticle].

The analysis began by using BLAST to obtain all similar protein sequences from both the NCBI non-redundant protein database and the SwissProt/TrEMBL database for the translation of each ORF. The resulting “hits” with E -values $< 10^{-5}$ always included some subset of the previously sequenced crustacean cuticular proteins and in some cases various insect proteins. Next, *C. sapidus* sequences that had overlapping hits in BLAST were placed together with the other arthropod sequences with which they shared similarity, and each of these sets of presumptive

homologs was aligned using *Clustal X* (Thompson and others 1997), using the default Gonnet series to weight amino acid similarities. Redundant sequences obtained from BLAST were removed, and alignments were trimmed.

Bayesian phylogenetic analysis was performed using *MrBayes 3.1.2* (Ronquist and Huelsenbeck 2003). We chose *MrBayes* in part due to its performance in estimating relationships between proteins over the broad range of evolutionary distances represented in our alignments, due to its modest need for computing time, and also due to the availability of the “protein mixed model” in the software. This is a procedure that efficiently jumps between several alternative empirically derived models of amino acid evolution (each based on large “training sets” of amino acid sequences) to estimate the most appropriate model. We felt this to be superior to selecting a model *a priori* for these previously unstudied proteins. In each of the 3 large alignments we analyzed, *MrBayes* selected the WAG model (Whelan and Goldman 2001) as the best fitting model with a posterior probability of $>95\%$.

To run Bayesian analysis, we allowed rates to be gamma distributed across residues. We ran a sufficient number of generations for an ~24 h run (500,000 for the crust-18 alignment and 250,000 generations for the RR alignments) and set the burn-in at 10% of the trees saved. Each run was examined to ensure that log likelihoods had converged, which they did well in advance of termination of the run.

The first group of *C. sapidus* transcripts consists of the 7 obtained sequences to date that code for RR-1 consensus sequences within their ORFs. When BLAST searches were done on the public protein databases, each of the inferred *C. sapidus* proteins was found to have a high degree of similarity to the other 6, to all RR-1-containing proteins known from other crustacean species, and to a large number of insect protein sequences containing the RR-1 consensus. Since the number of insect “hits” was quite large and since the proteins included were not all the same when each of the *C. sapidus* proteins was BLASTed separately, we decided to include in a single alignment only those crustacean and insect sequences that had E -values $< 10^{-8}$ for at least 6 of our 7 *C. sapidus* sequences. Phylogenetic analysis showed that the crustacean RR-1 sequences and the insect RR-1-containing proteins form separate clades (Fig. 2A), except for 1 insect sequence, *M. sexta* cuticle protein 20 (Suderman and others 2003), which is basal to the crustacean clade. Four blue crab proteins (CsAMP6.0, CsAMP16.5, CsAMP8.1, and CsAMP13.4) showed strongly supported sister relationship to different *C. pagurus* proteins. Of particular interest is the fact

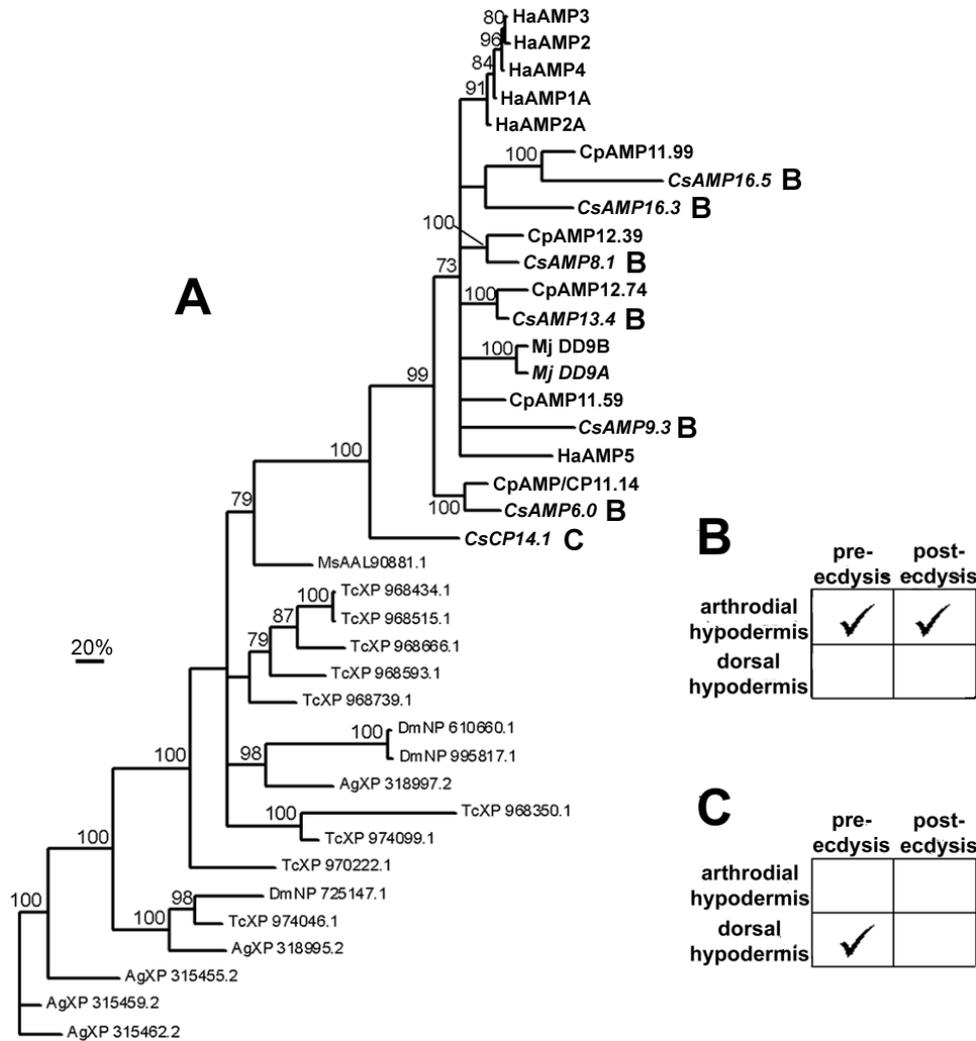


Fig. 2 Phylogenetic relationships and gene expression patterns for 7 *Callinectes sapidus* transcripts coding for proteins containing a RR-1 consensus sequence. **(A)** The Bayesian consensus phylogram is shown with branch lengths drawn to scale of percentage amino acid sequence divergence. Numbers label nodes with Bayesian reliabilities ≥ 70 . Crustacean proteins are in boldface, and the *C. sapidus* sequences are in boldface and italics. Species key: Ag—*Anopheles gambiae*, Cp—*Cancer pagurus*, Cs—*Callinectes sapidus*, Dm—*Drosophila melanogaster*, Ha—*Homarus americanus*, Mj—*Marsupeneus japonicus*, Ms—*Manduca sexta*, Tc—*Tribolium castaneum*. Insect sequences are named using their NCBI protein accession numbers. **(B)** Summary of gene expression results for CsAMP6.0, CsAMP16.3, CsAMP16.5, CsAMP8.1, CsAMP12.74, CsAMP13.4, and CsAMP9.3. **(C)** Summary of gene expression results for CsCP14.1.

that the most basal sequence falling within the crustacean clade is CsCP14.0. This protein is also unique among *C. sapidus* RR-1-containing sequences in being transcribed only pre-ecdysis in calcifying tissue (Fig. 2C) rather than in arthrodial membrane. No other known crustacean RR-1 protein is exclusively from calcifying cuticle, though CpAMP/CP11.4 was isolated from both cuticle types (Andersen 1999). The 6 more typical *C. sapidus* RR-1 proteins found in arthrodial membranes, including CsAMP6.0, which shares greatest similarity with CpAMP/CP11.4, are expressed both pre-ecdysis and post-ecdysis (Fig. 2B). This suggests that these flexible

cuticle proteins are common to both exocuticle and endocuticle. Presumably they are structural proteins important for any matrix that remains uncalcified. They range from very common transcripts in arthrodial hypodermis to one sequence, CsAMP9.3, which appears to be transcribed only in scattered clusters of arthrodial hypodermis cells at low levels (Faircloth and Shafer in review).

In contrast to the RR-1 sequences, a second group of Rebers-Riddiford motif-containing *C. sapidus* cuticular proteins produced BLAST search results revealing very few similar proteins in the public databases. These 3 sequences have moderately significant

	pre-ecdysis	post-ecdysis
A		
arthrodial hypodermis		
dorsal hypodermis	✓	✓
B		
arthrodial hypodermis		
dorsal hypodermis		✓

Fig. 3 Summary of gene expression results for *Callinectes sapidus* cuticular proteins showing the characteristics of CaCO_3 nucleators. **(A)** CsCP8.2 and CsCP8.5. **(B)** CsCP6.1.

“hits” (E -values $< 10^{-4}$) only with each other and with 4 previously sequenced proteins from other crustaceans. CsCP8.2 and CsCP8.5 (Wynn and Shafer 2005) have local similarity to both calcification-associated peptides from *P. clarkii*, though they are considerably more similar to CAP-1 than to CAP-2 (Inoue and others 2003, 2004). CsCP 6.1 (Faircloth and Shafer in review) shows local similarity to CAP-2, crustocalcin from *M. japonicus* (Endo and others 2004), and CpCP5.75 (Andersen 1999). The small size of this group of similar proteins made detailed phylogenetic analysis at this point unnecessary, but neighbor joining trees produced from *Clustal* alignments of these 7 sequences concur with the relationships stated (data not shown).

We suggest that these 3 proteins discovered from the blue crab EST dataset are potential calcium carbonate nucleators for the following reasons. First, they are expressed in the dorsal hypodermis that is making cuticle destined to calcify but not in arthrodial hypodermis (Fig. 3). CsCP8.2 and 8.5 are exclusively pre-ecdysis transcripts and therefore likely to code for exocuticular proteins (Fig. 3A), and CsCP6.1 is exclusively post-ecdysis (Fig. 3B) and therefore likely to be endocuticular. Second, CAP-1 and -2 (Inoue and others 2004) and crustocalcin (Endo and others 2004) have been shown to either bind calcium or affect *in vitro* CaCO_3 formation and have thus been implicated as important components in controlling the initiation of mineralization in their respective species. CpCP5.75 has not been tested for calcification potential, but it is the clear crab homolog of CsCP6.1. Third, like the other proteins, the 3 *C. sapidus* sequences contain not only an RR motif not assignable to any of the variants described in insects but also additional

regions rich in aspartic acid and glutamic acid residues. These portions of the protein could attract or bind calcium ions. Finally, these sequences show no homology with any proteins found in insect cuticles, where mineralization never occurs. If the 3 blue crab proteins do act as nucleators for CaCO_3 , the action of CsCP8.2 and 8.5, the exocuticular transcripts, must somehow be inhibited until after ecdysis.

The third group of *C. sapidus* transcripts discovered by mining the EST data actually have more similarity, as revealed by BLAST, to insect cuticular proteins than to any of the known crustacean cuticular protein sequences. Several other sequences among the 73 putative cuticular transcripts represented in the EST database also appear to have closer BLAST matches with insect sequences than with known crustacean sequences, but so far we have obtained complete ORFs and tentative expression results for only these two. Both of them show unusual expression patterns (unpublished data). CsAMP/CP13.7 is the only gene we have examined to date that is transcribed in arthrodial hypodermis and calcification-associated hypodermis during both pre-ecdysis and post-ecdysis (Fig. 4B). Real-time PCR shows that its expression is 6- to 7-fold higher before the molt. CsAMP23.6 is the only arthrodial transcript analyzed so far that is strictly post-ecdysial (Fig. 4C). Both of these proteins contain reasonably well-conserved Rebers-Riddiford motifs somewhat similar to the RR-2 variant found in insects, yet without the typical RR-2 upstream consensus. CsAMP/CP13.7 shows a degree of similarity to the less-clearly characterized RR-3, which has been recognized as existing in *H. americanus* proteins as well as in some insect sequences (Anderson 2000). CsAMP23.7 codes for the longest ORF we have sequenced so far in blue crab cuticular transcripts. It contains an unusual arrangement of 3 short motifs each repeated 6 times in the C-terminal portion of the sequence. Some insect cuticular proteins show a similar feature, though the motifs themselves are quite different.

BLAST searches with these 2 crustacean proteins returned a number of insect and other arthropod sequences. Figure 4A shows phylogenetic relationships between the members of this diverse alignment, containing CsAMP/CP13.7, CsAMP23.7, the insect sequences that were found similar to one or the other of them by BLAST (E -value $< 10^{-4}$), and HaCP18.8 and HaCP14.2, 2 RR-containing crustacean proteins that also were found similar by BLAST. Interestingly, the crustacean sequences did not form a clade as they did for the RR-1 sequences, suggesting that divergence of these proteins from ancestral forms may have pre-dated the split between insects and crustaceans. In contrast, there is a well-supported chelicerate

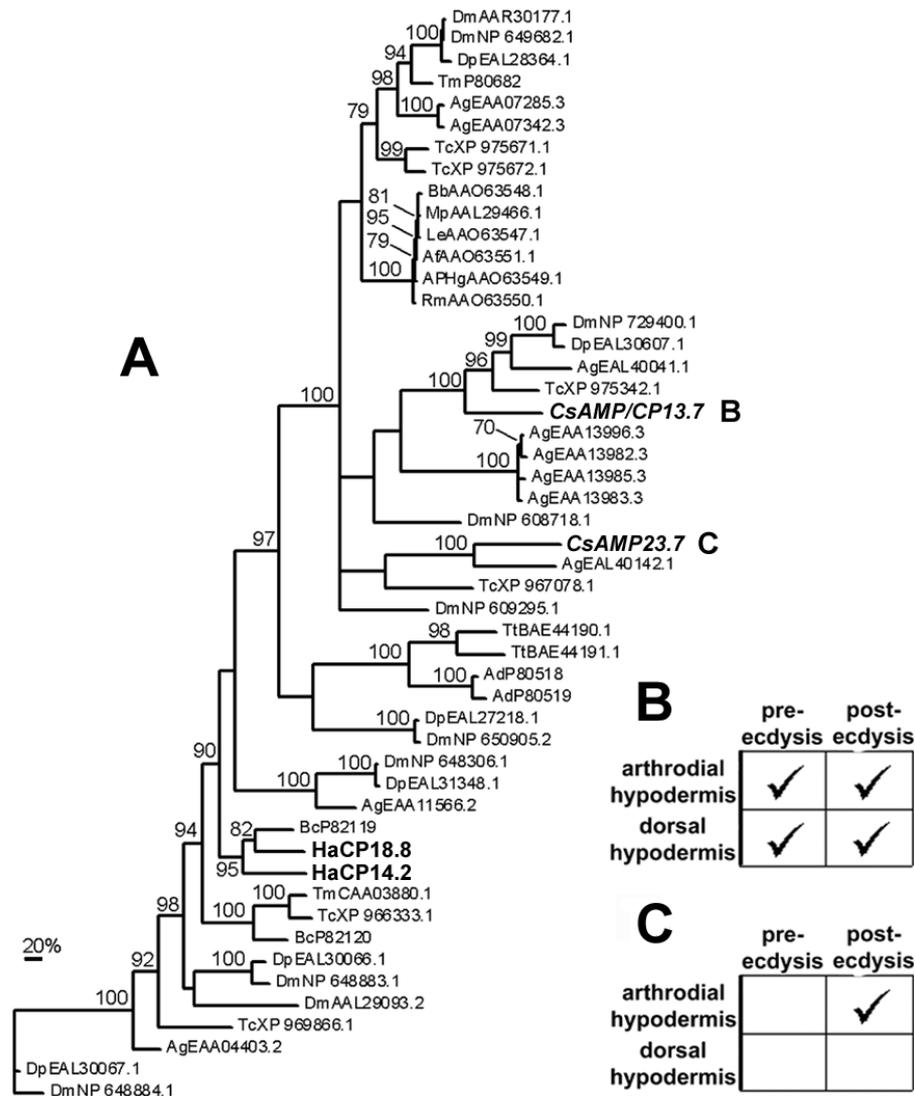


Fig. 4 Phylogeny and gene expression patterns for 2 *Callinectes sapidus* transcripts coding for proteins containing insect-like RR motifs. **(A)** Bayesian consensus phylogram with most closely related cuticular proteins from other arthropods as in Figure 2. Crustacean proteins are in boldface, and the *C. sapidus* sequences are in boldface and italics. Species key as in Figure 2 with the addition of Ad—*Araneus diadematus*, Af—*Aphis fabae*, Apg—*Aphis gossypii*, Bb—*Brevicoryne brassicae*, Bc—*Blaberus craniifer*, Dp—*Drosophila pseudoobscura*, Le—*Lipaphis erysimi*, Mp—*Myzus persicae*, Rm—*Rhopalosiphum maidis*, Tm—*Tenebrio molitor*, Tt—*Tachypleus tridentatus*. **(B)** Summary of gene expression results for CsAMP/CP13.7. **(C)** Summary of gene expression results for CsAMP23.7.

clade containing 2 sequences from the horseshoe crab *Tachypleus tridentatus* (Tt) and 2 sequences from the spider *Araneus diadematus* (Ad). It is possible that a clearer picture of the evolutionary history of the insect and crustacean proteins, and of the motifs they contain, may result when more species can be sampled. The highly supported homology of CsAMP23.7 with a predicted protein from the mosquito, *Anopheles gambiae*, genome that has not yet been annotated is worthy of note.

Another expected natural group of *C. sapidus* sequences is one with ORFs coding for crust-18, the motif found in either 2 or 4 copies in several

hard-cuticle proteins of *H. americanus* and *C. pagurus* (Kragh and others 1997; Nousiainen and others 1998; Andersen 1998b). It is never found in arthrodial membrane proteins or in insect cuticular proteins. When the EST dataset was probed for nucleotides coding for this motif, 10 contigs were found (Kennedy and others manuscript in preparation). Complete transcript sequences were obtained and the translations of the ORFs showed typical signal sequences followed by varying lengths of secreted polypeptides containing between 4 and 14 variant copies of the motif. In each case the crust-18 motifs constituted ~70% of the total length.

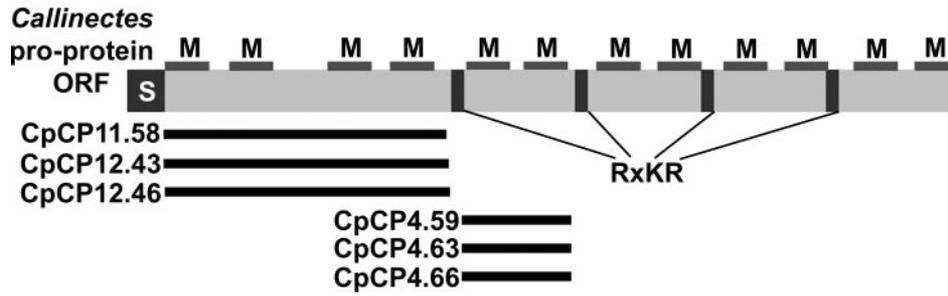


Fig. 5 Diagram of the ORF of one of the *Callinectes sapidus* transcripts coding for multiple variant copies of crust-18, a motif found in calcified cuticle. The initial segment labeled S is the signal peptide. Twelve motifs are labeled M and indicated by dark bars above the lighter segments of the pro-protein. Putative peptide cleavage sites consisting of the sequence RxKR are indicated by the short dark segments separating groups of 2 or 4 motifs. Six known motif-containing *Cancer* cuticle proteins are shown where they align with the *Callinectes* translation with greatest similarity.

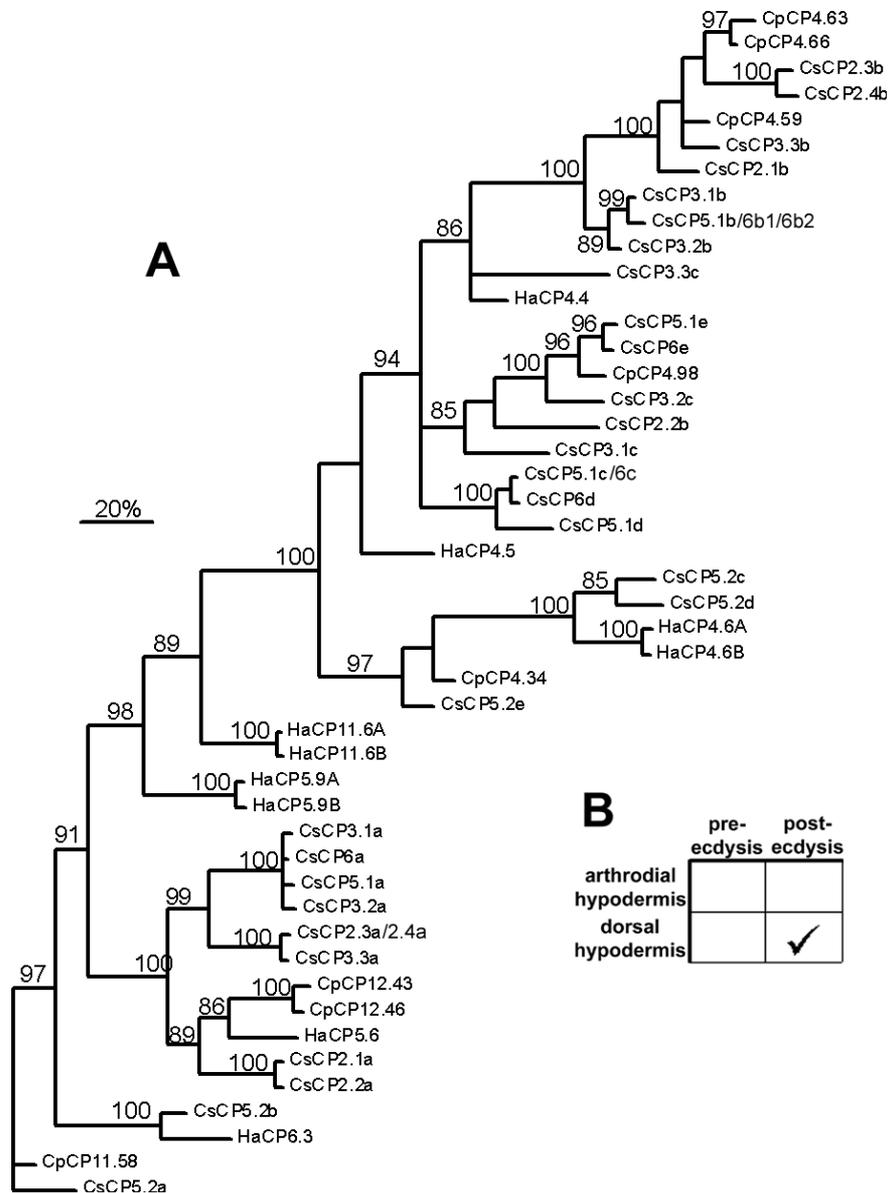


Fig. 6 Phylogenetic relationships (A) and summary of gene expression patterns (B) for the cuticular proteins containing 2 or 4 copies of the crust-18 motif that could result from cleaving pro-proteins coded by 10 *Callinectes sapidus* transcripts at every putative protease site. Consensus phylogram as in Figure 2. Species key as in Figure 2.

In an attempt to understand the differences in motif number among these translations as well as to explain why they differ from the pattern of 2 or 4 motifs found in cuticular proteins extracted from *C. pagurus* and *H. americanus*, we examined the amino acid sequences more closely. Between groups of either 2 or 4 motifs was found the short sequence RxKR. Figure 5 is a diagrammatic example of the ORF from one of the transcripts described by Kennedy and others (manuscript in preparation) illustrating this arrangement. These groupings of the basic amino acid residues arginine and lysine are not found in the proteins previously extracted from decapod calcified cuticle and sequenced as polypeptides. They are rather typical of recognition sites for trypsin-like serine proteases. We suggest (Kennedy and others manuscript in preparation) that the transcripts code for pro-proteins that are cleaved after translation into 2-motif or 4-motif peptides that then become resident in the exoskeleton. As evidence for this suggestion, we show that when attempts are made to align the direct-sequenced *Cancer* proteins with the ORFs of the *Callinectes* transcripts, high degrees of full-length similarity are always found to sequences of amino acid residues between the groups of arginanyl and lysyl residues and never to sequences including 2 or 4 motifs but spanning the putative cleavage sites (Fig. 5). Assuming cleavage at all the basic residue sites in all the pro-proteins, the *C. sapidus* transcripts discovered become 33 mature cuticular proteins, 7 of which contain 4 crust-18 motifs and the rest 2 motifs. Only 29 of these short peptides are actually different sequences since there are 2 identical pairs and 1 identical group of 3.

When each of these virtually cleaved peptides was used in a BLAST search of the public databases, only the known crust-18-containing sequences were returned. In other words, there is never significant similarity to any insect protein. Analysis of the crust-18 alignment (Fig. 6A) revealed a great deal of phylogenetic structure. Apart from CsCP5.2b and HaCP6.3, which are sister sequences, and CsCP5.2a and CpCP11.58, which are basal to the entire tree, there are 2 major clades. Both of these clades contain mixtures of *Callinectes*, *Cancer* and *Homarus* sequences. Crust-18, like the other major classes of cuticle proteins, is a diverse family of proteins with a complex evolutionary history whose diversity of function awaits further study.

Three northern blot probes were designed against either unique or shared parts of several of the 10 crust-18 transcripts. Differences in cDNA lengths allowed us to predict that the probes would detect 1, 2, or 3 bands if all the transcripts were actually present

	pre-ecdysis	post-ecdysis
A		
arthrodial hypodermis		
dorsal hypodermis	✓	
B		
arthrodial hypodermis		
dorsal hypodermis		✓

Fig. 7 Summary of gene expression results for *Callinectes sapidus* cuticular proteins each containing 3 copies of the postmolt-18 motif. **(A)** CsCP15.0. **(B)** CsCP19.0A, and CsCP19.0B.

in an RNA sample (Kennedy and others manuscript in preparation). Each probe detected the predicted number and size of RNA bands, and the expression pattern for each was the same. The subset of crust-18-containing transcripts that could be detected on these northern blots was strictly dorsal (that is, hard cuticle) hypodermis and post-ecdysis (Fig. 6B).

The final 3 of the 25 transcripts discovered through mining of the EST data and for which sequencing and gene expression pattern analysis are complete are CsCP15.0, CsCP19.0A, and CsCP19.0B (Faircloth and Shafer in review). Their ORFs code for proteins containing 3 copies of the postmolt-18 motif. Though this hydrophobic motif was first described in insects, the BLAST search result for each of these proteins found no significantly similar insect proteins in the public databases. The only related sequences known are the crustacean proteins CpCP14.99, CpCP18.79, HaCP20.2A, and HaCP20.2B. A neighbor joining tree (not shown) suggests that CsCP15.0 and CpCP14.99 are clear orthologs while all other relationships among these crustacean sequences are less obvious. The *C. sapidus* transcripts are expressed principally, if not exclusively, in hypodermis that is making calcifying cuticle rather than in the arthrodial hypodermis (Fig. 7). Interestingly, CsCP15.0 (Fig. 7A), but not CsCP19.0A or CsCP19.0B (Fig. 7B), is transcribed both pre- and post-ecdysis. It is the only transcript with this particular expression pattern, so the protein for which it codes is unusual among the presumptive hard-tissue peptides we have analyzed in that it may be present in both exocuticle and endocuticle. It could be a key structural protein of all calcifying parts of the exoskeleton.

Only 6 cuticular proteins known from species other than *C. sapidus* show no apparent sequence similarity to any of the 25 putative transcripts analyzed to date from the EST database. They are the single protein sequenced from *P. borealis* (Jacobsen and others 1994); DD1 (Watanabe and others 2006) and DD5 (Ikeya and others 2001) from *M. japonicus*; and the lobster proteins HaCP9.3, HaCP11.8, and HaCP12.7 (Nousiainen and others 1998). Perhaps *C. sapidus* transcripts coding for proteins with similarity to these await discovery.

Conclusion

We have documented how an expressed sequence tag project from the blue crab (*C. sapidus*) was designed, in part, to contain a maximal number of partial sequences coding for cuticular proteins and how this dataset is being “mined” to discover and analyze these sequences. Though this is still a work in progress, the data presented here clearly demonstrate that this transcriptome approach is working. When the analysis is complete, this species will be the best-understood decapod in terms of the structural and, perhaps, the regulatory proteins of its cuticle. This will greatly enhance future efforts to elucidate the control of biomineralization in this model system.

Acknowledgments

The authors thank D.L. Mykles and D.W. Towle for the invitation to participate in the 2006 Society for Integrative and Comparative Biology symposium on Genomic and Proteomic Approaches to Crustacean Biology. Special thanks go to long-term collaborators F.E. Coblenz, R.M. Dillaman, R.D. Roer, and D.W. Towle for their multiple contributions. This work was supported by National Science Foundation grant IBN-0114597 and by a Genomics Research Initiative Award from the Office of the President, University of North Carolina.

Conflict of interest: None declared.

References

- Addadi L, Weiner S. 1985. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. *Proc Natl Acad Sci USA* 82:4110–4.
- Andersen SO. 1998a. Amino acid sequence studies of endocuticular proteins from the desert locust *Schistocerca gregaria*. *Insect Biochem Mol Biol* 28:421–34.
- Andersen SO. 1998b. Characterization of proteins from arthro-dial membranes of the lobster *Homarus americanus*. *Comp Biochem Physiol A* 121:375–83.
- Andersen SO. 1999. Exoskeletal proteins from the crab, *Cancer pagurus*. *Comp Biochem Physiol A* 123:203–11.
- Andersen SO. 2000. Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*. *Insect Biochem Mol Biol* 30:569–77.
- Blackshear PJ, Lai WS, Thorn JM, Kennington EA, Staffa NG, Moore DT, Bouffard GG, Beckstrom-Sternberg SM, Touchman JW, Bonaldo MF, Soares MB. 2001. The NIEHS *Xenopus* maternal EST project: interim analysis of the first 13,879 ESTs from unfertilized eggs. *Gene* 267:71–87.
- Boguski MS, Lowe TM, Tolstoshev CM. 1993. dbEST—database for “expressed sequence tags”. *Nat Genet* 4:332–3.
- Bonaldo MF, Lennon G, Soares MB. 1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res* 6:791–806.
- Coblenz FE, Shafer TH, Roer RD. 1998. Cuticular proteins from the blue crab alter *in vitro* calcium carbonate mineralization. *Comp Biochem Physiol B* 121:349–60.
- Coblenz FE, Towle DW, Shafer TH. 2006. Expressed sequence tags from normalized cDNA libraries prepared from gill and hypodermis tissues of the blue crab, *Callinectes sapidus*. *Comp Biochem Physiol D* 1:200–8.
- Dillaman RM, Hequembourg S, Gay M. 2005. Early pattern of calcification in the dorsal carapace of the blue crab, *Callinectes sapidus*. *J Morphol* 263:356–74.
- Drach P, Tchernigovtzeff C. 1967. Sur la méthode de détermination des stades d’intermue et son application générale aux crustacés. *Vie Milieu Ser A Biol Mar* 18:595–610.
- Elliott EA, Dillaman RM. 1999. Formation of the inner branchiostegal cuticle of the blue crab, *Callinectes sapidus*. *J Morphol* 240:267–81.
- Endo H, Persson P, Watanabe T. 2000. Molecular cloning of the crustacean DD4 cDNA encoding a Ca²⁺-binding protein. *Biochem Biophys Res Commun* 276:286–91.
- Endo H, Takagi Y, Ozaki N, Kogure T, Watanabe T. 2004. A crustacean Ca²⁺-binding protein with a glutamate-rich sequence promotes CaCO₃ crystallization. *Biochem J* 384:159–67.
- Faircloth LM, Shafer TH. In review. Differential expression of eight transcripts and their roles in the cuticle of the blue crab, *Callinectes sapidus*. *Comp Biochem Physiol B* submitted.
- Hennig S, Groth D, Lehrach H. 2003. Automated gene ontology annotation for anonymous sequence data. *Nucleic Acids Res* 31:3712–5.
- Hepburn HR, Chandler HD. 1976. Material properties of arthropod cuticles: the arthro-dial membrane. *J Comp Physiol* 109:177–8.
- Ikeya T, Persson P, Kono M, Watanabe T. 2001. The DD5 gene of the decapod crustacean *Penaeus japonicus* encodes a putative exoskeletal protein with a novel tandem repeat structure. *Comp Biochem Physiol B* 128:379–88.
- Inoue H, Ozaki N, Nagasawa H. 2001. Purification and structural determination of a phosphorylated peptide with anti-calcification and chitin-binding activities in the exoskeleton of the crayfish, *Procambarus clarkii*. *Biosci Biotechnol Biochem* 65:1840–8.
- Inoue H, Ohira T, Ozaki N, Nagasawa H. 2003. Cloning and expression of a cDNA encoding a matrix peptide associated

- with calcification in the exoskeleton of the crayfish. *Comp Biochem Physiol B* 136:755–65.
- Inoue H, Ohira T, Ozaki N, Nagasawa H. 2004. A novel calcium-binding peptide from the cuticle of the crayfish, *Procambarus clarkii*. *Biochem Biophys Res Commun* 318:649–54.
- Jacobsen SL, Andersen SO, Højrup P. 1994. Amino acid sequence determination of a protein purified from the shell of the shrimp, *Pandalus borealis*. *Comp Biochem Physiol B* 109:209–17.
- Kragh M, Mølbak L, Andersen SO. 1997. Cuticular proteins from the lobster, *Homarus americanus*. *Comp Biochem Physiol B* 118:147–54.
- Lowenstam HA, Weiner S. 1989. *On Biomineralization*. New York: Oxford University Press.
- Mann S. 2002. *Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry*. New York: Oxford University Press.
- Marlowe RL, Dillaman RM, Roer RD. 1994. Lectin binding by the crustacean cuticle: the cuticle of *Callinectes sapidus* throughout the molt cycle, and the intermolt cuticle of *Procambarus clarkii* and *Ocypoe quadrata*. *J Crust Biol* 14:231–46.
- McClintock TS, Derby CD, Ache BW. 2006. Physiological genomics of lobster olfaction. Society for Integrative and Comparative Biology. Orlando, Florida.
- Nousiainen M, Rafn K, Skou L, Roepstorff P, Andersen SO. 1998. Characterization of exoskeletal proteins from the American lobster, *Homarus americanus*. *Comp Biochem Physiol B* 119:189–99.
- Pierce DC, Butler KD, Roer RD. 2001. Effects of exogenous N-acetylhexosaminidase on the structure and mineralization of the post-ecdysial exoskeleton of the blue crab, *Callinectes sapidus*. *Comp Biochem Physiol B* 128:691–700.
- Priester C, Dillaman RM, Gay DM. 2005. Ultrastructure, histochemistry, and mineralization patterns in the ecdysial suture of the blue crab, *Callinectes sapidus*. *Microsc Microanal* 11:479–99.
- Rebers JE, Riddiford LM. 1988. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle proteins. *J Mol Biol* 203:411–23.
- Rebers JE, Willis JH. 2001. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem Mol Biol* 31:1083–93.
- Roer RD, Dillaman RM. 1984. The structure and calcification of the crustacean cuticle. *Am Zool* 24:893–909.
- Roer RD, Dillaman RM. 1993. Molt-related change in integumental structure and function. In: Horst MN, Freeman JA, editors. *Crustacean Integument*. Boca Raton, Florida: CRC Press. p 1–36.
- Roer RD, Halbrook KE, Shafer TH. 2001. Glycosidase activity in the post-ecdysial cuticle of the blue crab, *Callinectes sapidus*. *Comp Biochem Physiol B* 128:683–90.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–4.
- Rudd S. 2003. Expressed sequence tags: alternative or complement to whole genome sequences? *Trends Plant Sci* 8:321–9.
- Shafer TH, Roer RD, Midgette-Luther C, Brookins TA. 1995. Postecdysial cuticle alteration in the blue crab, *Callinectes sapidus*: synchronous changes in glycoproteins and mineral nucleation. *J Exp Zool* 271:171–82.
- Shafer TH, Roer RD, Miller CG, Dillaman RM. 1994. Postecdysial changes in the protein and glycoprotein composition of the cuticle of the blue crab *Callinectes sapidus*. *J Crust Biol* 14:210–9.
- Stillman J, Teranishi K. 2006. Transcriptome changes during thermal acclimation, acclimatization, and stress in porcelain crabs. Orlando, Florida: Society for Integrative and Comparative Biology.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–82.
- Suderman RJ, Andersen SO, Hopkins TL, Kanost MR, Kramer KJ. 2003. Characterization and cloning of three major proteins from pharate pupal cuticle of *Manduca sexta*. *Insect Biochem Mol Biol* 33:331–43.
- Togawa T, Nakato H, Izumi S. 2004. Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* 34:1059–67.
- Tweedie EP, Coblenz FE, Shafer TH. 2004. Purification of a soluble glycoprotein from the uncalcified ecdysial cuticle of the blue crab *Callinectes sapidus* and its possible role in initial mineralization. *J Exp Biol* 207:2589–98.
- Watanabe T, Persson P, Endo H, Kono M. 2000. Molecular analysis of two genes, DD9A and B, which are expressed during the post molt stage in the decapod crustacean *Penaeus japonicus*. *Comp Biochem Physiol B* 125:127–36.
- Watanabe T, Persson P, Endo H, Fukuda I, Furukawa K, Kono M. 2006. Identification of a novel cuticular protein in the kuruma prawn *Penaeus japonicus*. *Fish Sci* 72:452–4.
- Weiner S, Dove PM. 2003. An overview of biomineralization processes and the problem of the vital effect. In: Dove PM, De Yoreo JJ, Weiner S, editors. *Reviews in Mineralogy and Geochemistry, vol. 54*. Washington, DC: The Mineralogical Society of America. p 1–29.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–9.
- Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinias JR, Robertson HM, Soares MW, Robinson GE. 2002. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res* 12:555–66.
- Williams CL, Dillaman RM, Elliott EA, Gay DM. 2003. Formation of the arthroal membrane in the blue crab, *Callinectes sapidus*. *J Morphol* 256:260–9.
- Wynn A, Shafer TH. 2005. Four differentially expressed cDNAs in *Callinectes sapidus* containing the Rebers-Riddiford consensus sequence. *Comp Biochem Physiol B* 141:294–306.