Protein Science (1999), *8*:242–248. Cambridge University Press. Printed in the USA.

FOR THE RECORD

The insect immune protein scolexin is a novel serine proteinase homolog

CASEY M. FINNERTY,[1,2] P. ANDREW KARPLUS,[2,3] AND ROBERT R. GRANADOS[1]

[1]Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York 14853
[2]Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

(RECEIVED June 5, 1998; ACCEPTED September 2, 1998)

Abstract: Scolexin is a coagulation-provoking plasma protein induced in response to bacterial or viral infection of larval *Manduca sexta*, a large lepidopterous insect. Here we report the isolation and sequencing of two cDNA clones that code for scolexin isoforms sharing 80% sequence identity. The scolexin sequences have low but recognizable sequence similarity to members of the chymotrypsin family and represent a new subfamily of chymotrypsin-like serine proteinases. Comparison with known structures reveals the conservation of key catalytic residues and a possible specificity for small nonpolar residues. Most remarkable is the absence of a canonical activation peptide cleavage site. This suggests that the regulation of scolexin activity will involve a novel activation mechanism.

Keywords: alignment; cDNA; sequence; serine proteinase; zymogen activation

The insect immune system comprises a coordinated set of cellular and humoral responses to infection that can often prevent the insect from succumbing to disease. Larger invaders of the hemocoel, such as parasites or fungi, may become entrapped by attaching insect hemocytes that secret enzymes to form a melanized capsule. Bacteria entering the hemocoel may become phagocytosed, sealed off in hemocytic nodules or digested by secreted lysozyme and eventually lysed by peptides secreted by the insect fat body into the hemocoel (Hultmark, 1993).

One insect protein that has been associated with response to pathogenic challenge is scolexin, found in larvae of *Manduca sexta*, also known as the tobacco hornworm. Scolexin was originally identified as a 36 kD glycoprotein induced in the plasma of *M. sexta* larvae injected with bacteria (Hughes et al., 1983) and was subsequently shown to be induced by other immune challenges, including yeast and lipopolysaccharide injection as well as baculovirus infection (Finnerty et al., 1994; Finnerty & Granados, 1997), making the it the first baculovirus-induced insect protein to be characterized.

The induction of scolexin following such diverse challenges strongly suggests that it occupies a central role in the immune response of *M. sexta*, but the precise function of the protein in vivo remains unknown. Unlike the inducible lysozyme and antibacterial peptides found in insect plasma (Hultmark et al., 1980), scolexin does not appear to possess bacteriolytic activity (Hurlbert et al., 1985). However, scolexin was subsequently shown to possess a potent coagulation-inducing activity on insect hemolymph in vitro (Minnick et al., 1986) and to localize in vivo within a coagulum that was associated with hemocytic nodules formed in response to bacteria injection (Kyriakides et al., 1993). The combination of infection-inducibility and coagulation activity suggests that scolexin may represent the first identified member of an insect-blood clotting and defense system, analogous to the clotting cascade that performs these functions in the horseshoe crab, *Limulus* (Iwanaga, 1993).

Here we report the cloning, sequencing, and analysis of cDNAs coding for the scolexin previously isolated from insect hemolymph, as well as another scolexin isoform previously undescribed. Using sequence comparisons, we assign both scolexins to a new subfamily of chymotrypsin-like serine proteinases, analyze the substrate-binding pocket specificity, and evaluate potential mechanisms for zymogen activation.

Results: *Cloning of scolexins A and B:* Screening of the day one, fifth instar epidermis library (E-5-1) with polyclonal antiserum yielded five immunopositive clones. No positive clones were obtained from the epidermis or fat body derived libraries. Four of the five positive clones from this screening expressed fusion proteins that reacted with the antiserum and had SDS-PAGE apparent molecular weights similar to scolexin's expected molecular weight of 36 kD. Sequencing the 5′ ends of the sense strands in these four clones indicated that they were identical, and one of them was

Reprint requests to: Casey M. Finnerty, Section of Biochemistry, Molecular and Cell Biology, Cornell University 221 Biotechnology Building, Ithaca, New York 14853; e-mail: cmf5@cornell.edu.

[3]Present address: Department of Biochemistry and Biophysics, 2011 Agricultural and Life Sciences Building, Oregon State University, Corvallis, Oregon 97331.

Abbreviations: SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; EDTA, ethylenediaminetetra-acetic acid; PMSF, phenylmethylsulfonyl fluoride; DIG, digoxigenin-deoxyuracil triphosphate; GCG, Genetics Computer Group Sequence Analysis Software Package; ScA, scolexin A; ScB, scolexin B; SCR, structurally conserved region.

chosen for nested deletion construction and sequencing. The fifth clone expressed a truncated form of scolexin. All five clones also expressed a smaller, cross-reacting protein (~17–20 kD). This protein was not further characterized, but it may represent a proteolysis product of the scolexin fusion protein or a product of internal transcription or translation initiation.

The correct reading frame of this first clone was identified by comparison with the N-terminal sequence of scolexin obtained from protein sequencing (Fig. 1) (Finnerty & Granados, 1997) (Kyriakides et al., 1995). The upstream cDNA sequence did not contain a translation initiation codon, indicating that this clone was incomplete. To isolate a full-length cDNA, the E-5-1 library was rescreened with a DNA probe prepared from the first clone, and three larger clones were isolated. Restriction analysis showed two distinct patterns, and the largest clone of each was chosen for sequencing. The first of these two had a sequence identical to the original clone, except that it contained 35 additional base pairs at the 5′ end. The additional sequence corresponded to a reasonable sequence for a signal peptide, but it lacked a translation initiation site. The sequence of the second clone was 76% identical to the first, and it also lacked a codon for translation initiation. The second clone codes for a protein that matches 13 of the 15 N-terminal residues determined by direct peptide sequencing of scolexin (Kyriakides et al., 1995). The identification of scolexin's N-terminal sequence in both clones confirms that they code for scolexin isoforms, and we have named them scolexin A (or ScA) and scolexin B (or ScB).

The putative signal peptides for ScA and ScB are 22 and 24 amino acids long, respectively, and begin with two to four polar residues. Since signal peptides rarely contain more than 29 amino acids and usually start with a few polar residues followed by a large cluster of hydrophobic residues (Lewin, 1987), we suspect that the clones are missing only a few nucleotides at their 5′ ends. Sequencing of the 5′ ends of the ScA and ScB transcripts was attempted using inverse RT-PCR from internal primers, but this was unsuccessful due to 5′ sequence termination at the same location as in clones isolated from the cDNA library (results not shown). This result is consistent with premature termination due to a highly stable secondary structure at the 5′ end of the scolexin transcript.

```
Scolexin-A       CGGCAGTCGGTTGTGTTGGCAGTGGCGGCGGTGCTCTTCGGGTGCGCGTGCGCAG 55
                 R  Q  S  V  V  L  A  V  A  A  V  L  F  G  C  A  C  A
               S  K  Q  S  V  V  L  A  V  A  A  A  L  V  A  C  A  C  A
Scolexin-B  CGTCGAAGCAGTCGGTTGTGTTGGCAGTGGCGGCGGCGCTCGTCGCGTGCGCGTGCGCAG 60

Scolexin-A  CGCCCAATCCTGGCGCCAACGACATACAACTTAATCAAAAATTAAGTATCGAAGCTAAGG 85
            A  P  N  P  G  A  N  D  I  Q  L  N  Q  K  L  S  I  E  A  K
            A  P  D  P  G  A  N  D  I  Q  L  N  Q  K  L  S  V  D  A  K
Scolexin-B  CGCCCGACCCCGGCGCCAACGATATACAACTTAATCAAAAATTAAGTGTTGATGCCAAGG 90

Scolexin-A  GGGCAAAGCAGCCAATTGATACGAGGGCAGTGAAGGAACGGTATCCATACGCAGTTCGGA 115
            G  A  K  Q  P  I  D  T  R  A  V  K  E  R  Y  P  Y  A  V  R
            G  A  K  Q  P  I  D  T  R  A  V  N  E  R  Y  P  H  A  V
Scolexin-B  GGGCAAAGCAGCCAATTGATACGAGGGCAGTCAACGAACGGTATCCACATGCAGTT---C 117

Scolexin-A  GTTTCGGAGGCTTCTGCGGAGGAACCATTATCAGTCCCACCTGGATCCTGACCGCCGGCC 145
            S  F  G  G  F  C  G  G  T  I  I  S  P  T  W  I  L  T  A  G
            L  F  G  G  T  C  G  G  T  I  I  S  P  T  W  I  L  T  A  G
Scolexin-B  TATTCGGAGGCACCTGCGGAGGAACCATTATCAGTCCCACCTGGATCCTGACCGCCGGTC 147

Scolexin-A  ACTGCTCGATACTCTATGCGGGGAGCGGCCTACCGGCCGGCACCAACATTACCGAGGTAT 175
            H  C  S  I  L  Y  A  G  S  G  L  P  A  G  T  N• I  T  E  V
            H  C  T  L  F  N  D  G  R  G  V  L  A  G  T  N• N  S  D  V
Scolexin-B  ACTGCACACTATTCAATGACGGGCGCGGCGTCCTGGCCGGCACCAACAACAGCGACGTGT 177

Scolexin-A  CTAGCTTGTACCGCTTCCCCAAGCGGCTCGTCATACACCCGCTCTTCTCCATAGGACCCG 205
            S  S  L  Y  R  F  P  K  R  L  V  I  H  P  L  F  S  I  G  P
            S  G  V  Y  R  F  T  K  R  L  I  I  H  P  L  F  S  V  G  P
Scolexin-B  CTGGCGTGTACCGCTTCACCAAGCGGCTCATCATACATCCGCTCTTCTCCGTAGGACCAT 207

Scolexin-A  TCTGGCTCAACGCTACGGAGTTCAACCTCAAACAGGCGGCTGCACGATGGACTTCTTGT 235
            V  W  L  N• A  T  E  F  N  L  K  Q  A  A  A  R  W  D  F  L
            Y  W  L  N  A  E  E  F  N  L  K  Q  V  A  A  R  W  D  F  L
Scolexin-B  ACTGGCTCAACGCCGAAGAGTTCAACCTCAAACAGGTGGCTGCACGATGGACTTCTTGT 237

Scolexin-A  TGATAGAACTGGAGGAACCGCTGCCGTTGGACGGCAAGATCCTGGCGGCTGCGAAGCTCG 265
            L  I  E  L  E  E  P  L  P  L  D  G  K  I  L  A  A  A  K  L
            L  A  E  L  E  E  P  L  P  L  D  G  K  I  M  A  A  A  K  L
Scolexin-B  TGGCGGAACTGGAGGAACCGCTGCCGTTGGACGGCAAGATCATGGCGGCTGCGAAGCTCG 267

Scolexin-A  ACGACCAGCCCGACCTCCCCGCAGGCCTCGACGTGGGCTATCCGAGCTACAGCACCGACA 295
            D  D  Q  P  D  L  P  A  G  L  D  V  G  Y  P  S  Y  S  T  D
            D  D  Q  P  D  L  P  A  G  L  D  V  G  Y  A  G  Y  G  T  D
Scolexin-B  ACGACCAGCCCGACCTCCCCGCAGGCCTCGACGTGGGCTATGCGGGCTACGGCACCGACC 297

Scolexin-A  CCTACGAGGCTAAGATACAAAGCGAGATGCACGGAAAGAAGCTTTCGGTTCAATCTAACG 325
            T  Y  E  A  K  I  Q  S  E  M  H  G  K  K  L  S  V  Q  S  N
            H  H  G  G  T  M  R  S  E  M  H  A  M  E  L  S  V  Q  S  N
Scolexin-B  ACCATGGGGGCACGATGCGAAGCGAGATGCATGCAATGGAGCTTTCGGTTCAATCTAACG 327

Scolexin-A  AGGTGTGCTCGAAGCTAGAGCAGTTCAAGGCGGAGGACATGTTGTGCGCCAAGGGACGTC 355
            E  V  C  S  K  L  E  Q  F  K  A  E  D  M  L  C  A  K  G  R
            E  V  C  S  K  L  E  Q  F  E  A  K  D  M  L  C  A  K  G  R
Scolexin-B  AAGTGTGCTCGAAGCTAGAGCAGTTCGAGGCGAAGGACATGTTGTGCGCCAAGGGACGTC 357

Scolexin-A  CACCGCGATACGACTTCGTCTGCTTCAGCGACAGTGGCAGTGGGCTAGTAGACAACAATG 385
            P  P  R  Y  D  F  V  C  F  S  D  S  G  S  G  L  V  D  N  N
            P  P  R  Y  D  S  A  C  N  G  D  S  G  S  G  L  V  D  N  N
Scolexin-B  CGCCACGATACGACTCCGCCTGTAACGGCGACAGTGGCAGTGGGCTAGTAGACAACAATG 387

Scolexin-A  GTCGCCTAGTCGGCGTGGTGTCGTGGGCCGAGAACAACGCTTTCGAGTGCCGCAACGGCA 415
            G  R  L  V  G  V  V  S  W  A  E  N  N  A  F  E  C  R  N  G
            G  R  L  V  G  V  A  S  W  V  E  N  D  A  F  E  C  R  N  G
Scolexin-B  GTCGCCTAGTCGGTGTGGCGTCGTGGGTAGAGAACGACGCTTTCGAGTGCCGCAACGGCA 417

Scolexin-A  ACCTGGCGGTCTTCTCGCGAGTGTCCAGCGTACGCGAGTGGATCCGACAAGTCACCAACA 445
            N  L  A  V  F  S  R  V  S  S  V  R  E  W  I  R  Q  V  T  N
            N  L  V  V  F  S  R  V  S  S  V  R  E  W  I  R  Q  V  T  N
Scolexin-B  ACCTGGTGGTCTTCTCGCGAGTGTCCAGCGTACGCGAGTGGATCCGACAAGTCACCAACA 447

Scolexin-A  TATAATAGACCATACCGTTGTCATTTTGGGCTGTAGTGTATAGATAATAAATATAGACGC 475
            I  *
            I  *
Scolexin-B  TATAATGTACCATCGTCATTTTGGGCGATCGATGTTAACTAGTGGGCGAACGCAGATACC 477

Scolexin-A  GTGTACTGGTGTGACGTACGGAAGTGGAGAGTTGGGAGCGACAGCTCACCGCTCACTCCA 505
Scolexin-B  AAAATCTATTTATTGTTTTAGAAGTAAATGAGATCAGTTTCGTTTCGCCCACGTTGTTTA 507

Scolexin-A  CTCCCGGCCGCCGCGCGAGTAGCAGTGTCAGTGATTGCAAAAAAAAAAAAAAAAAAAAAA 535
Scolexin-B  CCATTACTAGTCATCTCATCACTTCACAAGCAATACTTTGTTTTCCTTTAAGTTTAGTAA 537

Scolexin-A  AAAAAAAAAAAAAAA 550
Scolexin-B  CATTTATTGATTTCTCGTCAGTGTTGTTACGTACTTAACAATGTCAGTAATTTATAGACA 567

Scolexin-B  ATTTGTATATAAAAGTACAAATATATTTTTAGTGTTAATACGGTTGTGATAAAATGTGTA 597

Scolexin-B  CAAACAGAACAGAAAGAACGTAGTATCATAAACCATTCAAAATCAAAATGAATAAAAAAT 627

Scolexin-B  GAAATAAAAAAAAAAAAAAAAAAAAAAAAAAA 657
```

Fig. 1. cDNA sequences of scolexins A and B. Polyadenylation signals are shown in bold italics. The translated amino acid sequences are shown in bold, and the N-terminal sequence determined from direct protein sequencing of gel-purified scolexin is shaded. The residues constituting the catalytic triad for serine proteinases are boxed, and the potential sites for N-linked glycosylation are followed by "●."

Predicted biochemical properties for the scolexin A and B proteins: The ScA and ScB clones code for nascent proteins with 80% identity and predicted molecular weights of 30,261 and 30,024, respectively. These sizes are close to the 30–31 kD polypeptide produced by in vitro translation of *M. sexta* epidermal transcripts followed by precipitation with antiscolexin antiserum (Spence et al., 1992), supporting our contention that the clones are nearly full length. Translation of ScA and ScB from the N-terminus identified for mature scolexin yields polypeptide molecular weights of 28,220 and 27,830. The calculated isoelectric points of these proteins are 6.6 and 5.4, respectively.

Scolexin has been demonstrated to be glycosylated (Hughes et al., 1983), and its carbohydrate content may account for the 8 kD needed to make up the apparent molecular weight of 36 kD for the secreted protein (Spence et al., 1992). Previous analysis of the carbohydrate content of scolexin indicates that its glycosylation is of the complex type (Kyriakides et al., 1995). Depending on the type of oligosaccharide side chain(s) attached to scolexin, the 8 kD estimate would require that two to five residues be glycosylated. The scolexin A sequence contains two potential N-linked glycosylation sites at N71 and N99, and the scolexin B sequence contains one at N70 (Fig. 1).

Database searching: A nonredundant database BLASTP v2.0.3 search using the predicted amino acid sequences of either ScA or ScB yielded >75 significant hits (i.e., having E-values less than 10^{-4}), and all of these appeared to be serine proteinases of the chymotrypsin S1 family (Rawlings & Barrett, 1994). Both scolexins A and B also conserve the catalytic triad found in this family (cH57-cD102-cS195, where "c" denotes chymotrypsinogen numbering) (Rawlings & Barrett, 1994). The residues surrounding cH57 and cS195 tend to be highly conserved, and the scolexin A and B sequences around the catalytic histidine match the consensus sequence from the PROSITE database (Bairoch et al., 1997). Neither scolexin possesses the S195-flanking consensus pattern from PROSITE, but this property is shared by at least 18 other members of the chymotrypsin family. The low similarities shared by either scolexin with their nearest neighbors (ScA vs. DER3_DERPT-1 = 19.9%, ScB vs. TRY3_AEDAE-1 = 22%) suggest that scolexins A and B are novel chymotrypsin-like serine proteinases belonging to a new subfamily. This conclusion is supported by a cladogram comparing scolexin to protein sequences representing diverse subfamilies within the chymotrypsin family (Fig. 2).

Interestingly, both scolexins utilize AGT codons for S195. The two major groups of chymotrypsin-like proteinases, i.e., the plasma proteinases used to regulate complex enzymatic cascades such as coagulation and the pancreatic proteinases used for digestion, tend to segregate based on their codon usage (Brenner, 1988). The AGY-type codon used by ScA and ScB for S195 is shared by the regulatory serine proteinases, which is consistent with the observed coagulation-provoking activity of scolexin (Minnick et al., 1986).

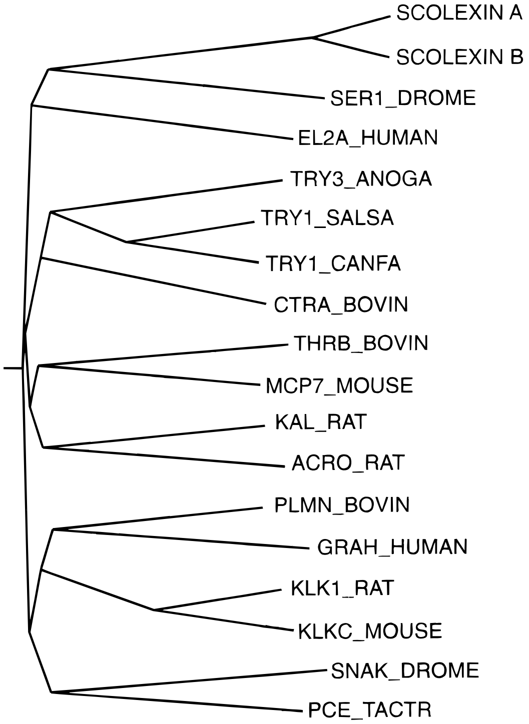


Fig. 2. Cladogram comparing the scolexin A and B sequences to diverse members of the chymotrypsin family. Proteinase sequences were selected from the major branch groups of a phylogenetic tree of 170 sequences belonging to the chymotrypsin family. This tree is based on a BLOCKS alignment (Henikoff & Henikoff, 1994) and is available on the World Wide Web at http://www.blocks.fhcrc.org/blocks/. Scolexin A and B were aligned to the selected sequences using the CLUSTALW algorithm (Thompson et al., 1994). The resulting output was converted to a cladogram using the Drawgram program of the PHYLIP v3.57c package (Felsenstein, 1989). Nonscolexin sequences are labeled with their SWISSPROT identifiers. *SER1_DROME*, *Drosophila melanogaster* serine proteinase; *EL2A_HUMAN*, human pancreatic elastase IIA; *TRY3_ANOGA*, *Anopheles gambiae* trypsin 3; *TRY1_SALSA*, salmon trypsinogen 1; *TRY1_CANFA*, dog trypsinogen 1; *CTRA_BOVIN*, bovine chymotrypsinogen A; *THRB_BOVIN*, bovine prothrombin; *MCP7_MOUSE*, mouse mast cell protease 7; *KAL_RAT*, rat plasma kallikrein; *ACRO_RAT*, rat acrosin; *PLMN_BOVIN*, bovine plasminogen; *GRAH_HUMAN*, human granzyme H; *KLK1_RAT*, rat glandular kallikrein 1; *KLKC_MOUSE*, mouse glandular kallikrein; *SNAK_DROME*, *Drosophila melanogaster* snake protease; *PCE_TACTR*, *Tachypleus tridentatus* proclotting enzyme.

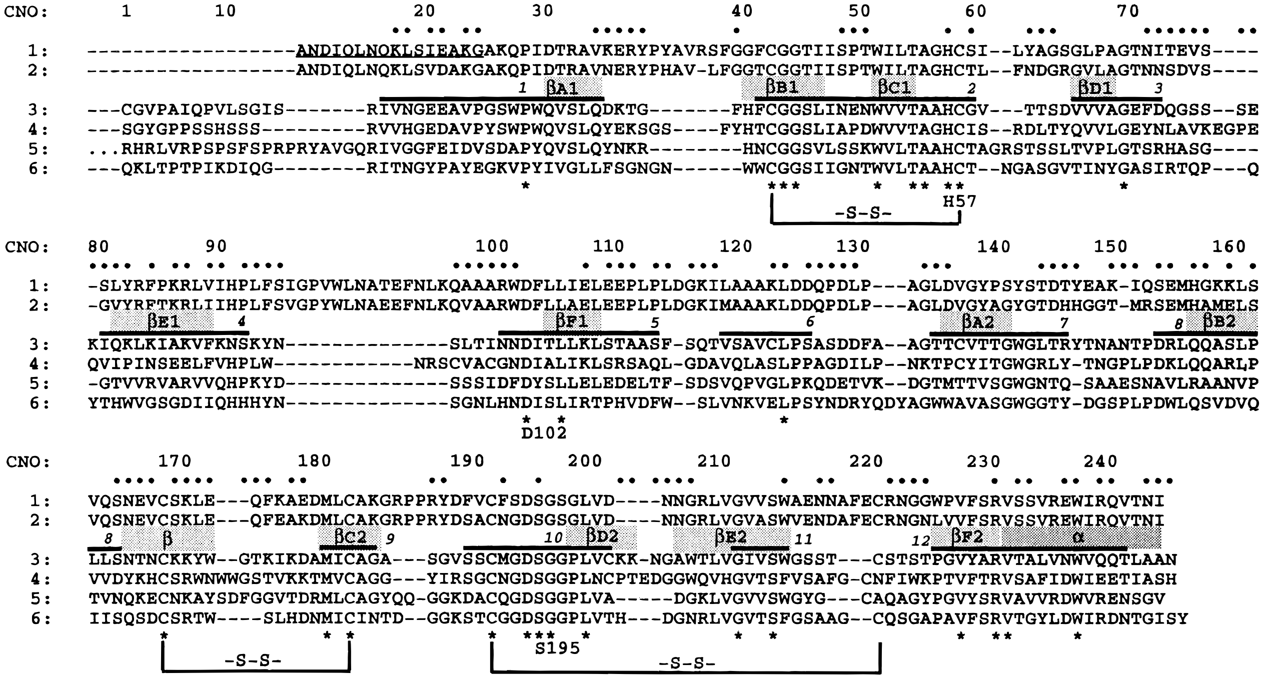
Structural alignment of the scolexin A and B sequences: To make inferences about scolexin structure/function relations, we constructed a structure-based sequence alignment of the ScA and ScB sequences. The three-dimensional structures of eight chymotrypsin homologs were used to define 12 structurally conserved regions (SCRs) (Greer, 1981) in which no gaps would be allowed in the alignment. The resulting definition of the SCRs and the alignment of scolexins A and B with two vertebrate and two invertebrate serine proteinases is shown in Figure 3.

Examination of the alignment strongly supports the assignment of ScA and ScB to the chymotrypsin family. In addition to the chymotrypsin catalytic triad, the six Cys residues found in the secreted forms of ScA and ScB align with the location of the three disulfide bridges (c42–c58, c168–c182, and c191–c220) that are conserved in bacterial and invertebrate trypsins (Zwilling & Neurath, 1981). Detailed examination of the SCRs also indicates that both scolexins will adopt the overall fold of a chymotrypsin homolog. Four of the 12 SCRs have high degrees of amino acid identity (38–42%) among all six sequences in the alignment. Furthermore, residues in the remaining SCRs corresponding to buried positions in chymotrypsin are generally well conserved, and the few nonconservative substitutions that occur in ScA or ScB can be rationalized by inspection of the local environment for the corresponding residue in chymotrypsin (Fig. 3).

The most remarkable feature of the aligned scolexin sequences is their lack of the highly conserved activation peptide cleavage site I-V-N-G in SCR1. This feature makes the alignment in this region uncertain, but we have chosen the alignment shown because it contains some residues (i.e., E21, A22, and especially P28) that are conserved in the chymotrypsin family. If our alignment is correct and scolexin is proteolytically activated in a manner analogous to chymotrypsin, cleavage may occur at the scolexin site QLN/QKL. The absence of any conserved residues at this site emphasizes, however, that such a prediction is highly speculative (see below).

Discussion: The two cDNA clones isolated and sequenced in this study code for isoforms of scolexin that we have named scolexin A and B. The scolexin A clone matches the first 15 residues of the previously published N-terminal sequence (Kyriakides et al., 1995; Finnerty & Granados, 1997), whereas the scolexin B sequence differs at two positions. These results indicate that the scolexin isolated from hemolymph, after induction with either bacteria (Kyriakides et al., 1995) or baculovirus (Finnerty & Granados, 1997), corresponds to the scolexin A clone reported here. An N-terminal sequence matching that predicted for our scolexin B clone has not yet been published, so the properties and function of this new isoform of scolexin remain unknown. However, the presence of scolexin B clones in the cDNA library indicates that at least its transcript is synthesized in *M. sexta* epidermis.

Structural and functional inferences for scolexins A and B: Sequence comparison shows that scolexins A and B belong to a



```
CNO:      1        10                 20        30           40        50        60        70
1:  ------------------ANDIQLNQKLSIEAKGAKQPIDTRAVKERYPYAVRSFGGFCGGTIISPTWILTAGHCSI--LYAGSGLPAGTNITEVS----
2:  ------------------ANDIQLNQKLSVDAKGAKQPIDTRAVNERYPHAV-LFGGTCGGTIISPTWILTAGHCTL--FNDGRGVLAGTNNSDVS----
3:  ---CGVPAIQPVLSGIS-------RIVNGEEAVPGSWPWQVSLQDKTG-------FHFCGGSLINENWVVTAAHCGV---TTSDVVVAGEFDQGSS--SE
4:  ---SGYGPPSSHSSS---------RVVHGEDAVPYSWPWQVSLQYEKSGS----FYHTCGGSLIAPDWVVTAGHCIS--RDLTYQVVLGEYNLAVKEGPE
5:  ...RHRLVRPSPSFSPRPRYAVGQRIVGGFEIDVSDAPYQVSLQYNKR---------HNCGGSVLSSKWVLTAAHCTAGRSTSSLTVPLGTSRHASG----
6:  ---QKLTPTPIKDIQG--------RITNGYPAYEGKVPYIVGLLFSGNGN------WWCGGSIIGNTWVLTAAHCT--NGASGVTINYGASIRTQP---Q

CNO:  80        90                   100       110        120        130        140        150       160
1:  -SLYRFPKRLVIHPLFSIGPVWLNATEFNLKQAAARWDFLLIELEEPLPLDGKILAAAKLDDQPDLP---AGLDVGYPSYSTDTYEAK-IQSEMHGKKLS
2:  -GVYRFTKRLIIHPLFSVGPYWLNAEEFNLKQVAARWDFLLAELEEPLPLDGKIMAAAKLDDQPDLP---AGLDVGYAGYGTDHHGGT-MRSEMHAMELS
3:  KIQKLKIAKVFKNSKYN--------------SLTINNDITLLKLSTAASF-SQTVSAVCLPSASDDFA--AGTTCVTTGWGLTRYTNANTPDRLQQASLP
4:  QVIPINSEELFVHPLW------------NRSCVACGNDIALIKLSRSAQL-GDAVQLASLPPAGDILP--NKTPCYITGWGRLY-TNGPLPDKLQQARLP
5:  -GTVVRVARVVQHPKYD--------------SSSIDFDYSLLELEDELTF-SDSVQPVGLPKQDETVK--DGTMTTVSGWGNTQ-SAAESNAVLRAANVP
6:  YTHWVGSGDIIQHHHYN--------------SGNLHNDISLIRTPHVDFW--SLVNKVELPSYNDRYQDYAGWWAVASGWGGTY-DGSPLPDWLQSVDVQ

CNO:       170        180          190       200        210          220        230       240
1:  VQSNEVCSKLE---QFKAEDMLCAKGRPPRYDFVCFSDSGSGLVD----NNGRLVGVVSWAENNAFECRNGGWPVFSRVSSVREWIRQVTNI
2:  VQSNEVCSKLE---QFEAKDMLCAKGRPPRYDSACNGDSGSGLVD----NNGRLVGVASWVENDAFECRNGNLVVFSRVSSVREWIRQVTNI
3:  LLSNTNCKKYW--GTKIKDAMICAGA---SGVSSCMGDSGGPLVCKK-NGAWTLVGIVSWGSST---CSTSTPGVYARVTALVNWVQQTLAAN
4:  VVDYKHCSRWNWWGSTVKKTMVCAGG---YIRSGCNGDSGGPLNCPTEDGGWQVHGVTSFVSAFG-CNFIWKPTVFTRVSAFIDWIEETIASH
5:  TVNQKECNKAYSDFGGVTDRMLCAGYQQ-GGKDACQGDSGGPLVA-----DGKLVGVVSWGYG---CAQAGYPGVYSRVAVVRDWVRENSGV
6:  IISQSDCSRTW----SLHDNMICINTD--GGKSTCGGDSGGPLVTH---DGNRLVGVTSFGSAAG--CQSGAPAVFSRVTGYLDWIRDNTGISY
```

Fig. 3. Alignment of translated scolexin A and B sequences with four other chymotrypsin-like serine proteinase sequences. The nonscolexin sequences are included for comparison only and should not be construed as being the most similar proteins to scolexin. *CNO*: chymotrypsinogen numbering; *Row 1*: scolexin A; *Row 2*: scolexin B; *Row 3*: bovine alpha-chymotrypsinogen; *Row 4*: human pancreatic elastase; *Row 5*: *Anopheles* trypsin; *Row 6*: *Drosophila* serine protease. "*" denotes residues that are identical in all six sequences. The residues of the serine proteinase catalytic triad (H57, D102, S195) are indicated. The N-terminus of scolexin determined from direct protein sequencing is underlined in the scolexin A sequence. The filled dots (●) above the alignment designate residues in chymotrypsin with side chains having >20 Å accessible surface area (Greer, 1981). The numbered black lines above row 3 indicate SCRs where no gap was allowed in the alignment. The stippled boxes indicate the location of secondary structures found in chymotrypsin. "-S-S-" denotes the location of conserved disulfide bridges. SCRs 2, 9, 10, and 11 possess high degrees of amino acid identity (38–42%) among all six aligned sequences. Within SCRs 3, 5, 6, and 12, ScA and ScB have conservative substitutions at residues predicted by analogy to chymotrypsin to have buried side chains. The local environment of the buried chymotrypsin residues in SCRS 4, 7, and 8 makes the alignment of the other sequences tenable in these regions, despite the presence of nonconservative substitutions.

new subfamily of the chymotrypsin family of serine proteinases. Chymotrypsin-like serine proteinases share a highly conserved structure consisting of two domains formed by antiparallel six-stranded β-barrels (Branden & Tooze, 1991). The active and specificity site residues are located on conserved loops located in the crevice between the two β-barrels.

One of the major determinants of serine proteinase specificity is the pocket that interacts with the side chain of the P1 residue in the substrate (Schecter & Berger, 1967; Branden & Tooze, 1991). Based on our alignment, this specificity pocket is predicted to be lined by cF189-cA216-cP226 in scolexin A and cS189-cV216-cV226 in scolexin B. The absence of charged residues at these positions in both scolexins suggests that they will cleave after uncharged residues in the substrate. In addition, the Ala, Pro, and Val residues at positions c216 and c226 would probably restrict or block the specificity pocket entrance, just as the cV216 and cT226 residues of human elastase block the pocket of that enzyme (Navia et al., 1989). This conformation would further restrict the specificity of scolexin to substrates having small to moderately sized side chains at the P1 position.

Is scolexin proteolytically active?: All clotting systems examined to date involve serine proteinase cascades, so our identification of scolexin as a serine proteinase homolog is consistent with previous reports of coagulation-provoking activity for this molecule (Minnick et al., 1986). However, attempts in our lab to detect proteolytic activity of either plasma-derived or recombinant scolexin, using albumin, casein or gelatin substrates and a variety of assay formats, have produced negative results (data not shown). Several possibilities could explain the lack of detectable proteolytic activity in scolexin: (1) scolexin is a pseudoproteinase, i.e., a proteinase homolog that lacks proteolytic activity, (2) scolexin is highly specific for its proteolytic substrate, or (3) the scolexin tested was a zymogen that needs activation before its proteolytic activity can be detected.

Pseudoproteinases are found in the chymotrypsin family, but invariably these molecules have a clear defect that would eliminate their proteolytic activity, such as mutation of the catalytic residues (Rawlings & Barrett, 1994) or severe truncation at the N-terminus (e.g., cattle procarboxypeptidase subunit III (Venot et al., 1986)). In contrast, scolexin is not missing any of the active site residues, and, though highly divergent, the N-terminal region does not appear to be truncated. The conservation of the catalytic residues and absence of any obvious truncation suggest to us that scolexin is a true proteinase.

The regulatory role implied by scolexin's coagulation-provoking activity (Minnick et al., 1986) and its catalytic serine codon usage

(Brenner, 1988) suggests that it will possess high substrate specificity (i.e., low proteolytic activity against exogenous substrates). However, most regulatory proteinases will exhibit a low level of detectable proteolytic activity against exogenous substrates. The complete lack of proteolytic activity observed for scolexin suggests to us that an even more potent factor may be preventing activity.

We feel that the most likely explanation for the lack of detectable proteolytic activity in scolexin is that the protein we examine is in the zymogen state, and the mode of its activation remains unknown. Serine proteinases of the chymotrypsin family are typically secreted as zymogens, and the catalytic domain of these proteinases is a conserved set of approximately 220 amino acid residues found at the C-terminal portion of the sequence. With very few exceptions (see below), these proteinases are activated via the proteolytic cleavage of an amino terminal portion termed the activation peptide (Rawlings & Barrett, 1994). Both scolexin sequences are very unusual in lacking any sequence resembling the consensus for this activation peptide cleavage site.

Possible mechanisms of zymogen activation: To the best of our knowledge, the only other active serine proteinases in the chymotrypsin family that do not possess this conserved cleavage site are complement factor C2 (e.g., CO2_HUMAN) (Bentley, 1986), of the classical complement activation pathway, its functional analog, complement factor B, of the alternative pathway (e.g., CFAB_HUMAN) (Mole et al., 1984) and the salivary tissue plasminogen activators from the vampire bat (e.g., URT1_DESRO) (Gardell et al., 1989). These proteinases and a few others suggest two alternative mechanisms for regulating serine proteinase activity that may apply to scolexin: (1) activation via proteolytic cleavage at a noncanonical site or (2) regulation of proteolytic activity via interaction with substrate or a cofactor.

In the standard activation mechanism exemplified by chymotrypsin, proteolysis at R/IVNG creates a new amino terminus, cI16, which forms a salt bridge with cD194 to induce the active conformation, with the cI16 side chain sitting in a hydrophobic pocket (Freer et al., 1970). In contrast, the noncanonical proteolytic activation sites of human factors C2 and B are R/KIQI and R/KIVL, and these sites are located far (ca. 250 residues) upstream of the catalytic domain (Mole et al., 1984; Bentley, 1986). Interestingly, cD194 is conserved in both mouse and human factor B, whereas it has been mutated to the nearly equivalent Glu in complement factor 2 from both species. Assuming that the lysine residues immediately following the cleavage sites of factors C2 and B form salt bridges with cD/E194 following proteolytic activation, it demonstrates that a hydrophobic side chain is not a strict requirement for the new N-terminal residue. Thus, cQ16 and the polar and charged residues surrounding it in both scolexins may not be precluded a priori from participating in a salt bridge with cD194, which is conserved in both ScA and ScB.

An alternate possibility, that scolexin is activated not through proteolysis but via substrate or cofactor binding, is exemplified by at least three other chymotrypsin homologs: complement factor D, plasminogen, and vampire bat salivary tissue plasminogen activator (Bat-PA). Complement factor D circulates in the plasma in the cleaved (i.e., active) form, and it is thought to be regulated via substrate binding (Volanakis & Narayana, 1996). Plasminogen can be proteolytically activated to plasmin by forming a 1:1 complex with the bacterial protein streptokinase, which itself is a proteolytically inactive serine proteinase homolog. Catalytic amounts of the plasmin: streptokinase complex are capable of activating other plasminogen molecules, though neither member of the complex possesses this activity by itself (for a review, see Castellino, 1979). Unlike factor D or plasminogen, Bat-PA has a severely mutated activation cleavage site that is presumed to be nonfunctional. Remarkably, Bat-PA is stimulated ca. 50,000-fold in the presence of fibrin, and assembly on fibrin is the suggested mechanism for Bat-PA regulation (Gardell et al., 1989). These precedents and the lack of a recognizable activation peptide cleavage site in scolexin make it plausible that a similar mechanism of cofactor interaction might be involved in the regulation of scolexin activity.

Association of scolexin with lectin activity: Scolexin has been reported to have a hemagglutinating or lectin activity, and inhibition of this hemagglutinating activity with glucose or antiscolexin antiserum also inhibited scolexin's coagulation-provoking activity (Minnick et al., 1986). Interestingly, the *Limulus* coagulation cascade is initiated by a molecule having both lectin and serine proteinase domains that are functional but separate (Muta et al., 1991). The hemagglutinating activity attributed to scolexin was resistant to both boiling and trypsin treatment, however, while the coagulation-provoking activity was abolished by both these treatments (Minnick et al., 1986). These results may indicate that a nonproteinaceous component was responsible for the reported hemagglutinating activity. There is no discernable lectin domain in the scolexin sequence, and to the best of our knowledge there is no example of a serine proteinase domain having lectin activity. These observations, combined with the fact that we have not been able to duplicate the hemagglutinating activity using purified scolexin (data not shown), suggest that the hemagglutination previously attributed to scolexin may be caused by a hemolymph component copurifying with scolexin.

If a protein is responsible for the hemagglutination previously reported, a lectin may form a complex with scolexin, and as a result the two molecules may copurify under certain conditions. Such a lectin/serine proteinase complex has been observed between human mannose-binding protein and at least two associated serine proteinases that activate complement (Sato et al., 1994; Thiel et al., 1997). The serine proteinases of this complex are activated by the lectin component after a sugar-binding event. This example provides some precedence for lectin-mediated activation of an immunity-related serine proteinase, and scolexin may participate in an insect analog to these systems. In any case, should scolexin prove to be proteolytically active, it is clear that such a novel serine proteinase will reveal new mechanisms of proteolytic activation and regulation. Additional study of scolexin should also illuminate important aspects of the insect immune response.

Materials and methods: *Library screening:* Lambda ZAP II cDNA expression libraries (Stratagene) prepared from epidermis of day one (E-5-1) and day three (E-5-3), fifth instar *M. sexta* larvae and from fat body of day one, fifth instar larvae (FB-5-1) were generated as described previously (Horodyski et al., 1989; Lerro & Prestwich, 1990; Li & Riddiford, 1992). Screening of these libraries was performed according to the manufacturer's protocol. Inserts were expressed as β-galactosidase fusion proteins. Approximately 10^5 plaque-forming units from each library were screened with a rabbit polyclonal antiserum raised against scolexin (Finnerty & Granados, 1997).

Positive plaques were taken through two rounds of plaque purification and then amplified in *Escherichia coli* XL1-Blue. The amplified phage clones were immunoscreened to confirm their purity, and excised as Bluescript SK plasmids using the ExAssist helper phage/*E. coli* SOLR host cell system (Stratagene).

Production of fusion protein and analysis by Western blot: E. coli SOLR cells containing plasmid were grown in LB broth at 37 °C to midlog phase ($OD_{600} = 0.2$). IPTG was added to 1 mM final concentration, and the cultures grown until just reaching stationary phase ($OD_{600} = 1.0$). Cells were pelleted at 1,600 × *g* for 15 min and the supernatant decanted. Cells were resuspended in four volumes of lysis buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 1 mM PMSF, and 10% sucrose). Lysozyme was added to the resuspended cells to 1 mg/mL final concentration and the cells incubated on ice for 10 min. Triton X-100 was added to 0.1% final concentration, and the mixture was incubated on ice for 10 min. The lysed cell mixtures were centrifuged at 13,000 × *g* for 1 h, after which an aliquot of the supernatant was removed for Western blotting with antiscolexin antiserum. SDS-PAGE, electro-transfer of protein, and Western blotting were performed as previously described (Finnerty & Granados, 1997).

Restriction endonuclease analysis: The clones were restriction mapped with the following enzymes: *Bam*HI, *Eco*RI, *Hind*III, *Kpn*I, *Pst*I, *Sac*I, *Sac*II, *Sal*I, *Xba*I, and *Xho*I. All enzymes and enzyme buffers were purchased from Promega Corp.

Double-stranded DNA sequencing: Cloned plasmids were electroporated into *E. coli* JM101 *rec A*$^-$ and amplified. After isolating plasmid DNA on CsCl gradients, nested deletions were prepared with the Erase-A-Base kit (Promega). Plasmid DNA was sequenced using the USB Sequenase 7-deaza-dGTP DNA sequencing kit according to the manufacturer's protocol. Radiolabel was γ-^{35}S-dATP from DuPont NEN. Sequencing reactions were separated on 6% denaturing polyacrylamide gels. Both strands were sequenced for each cDNA.

Preparation of DIG-labeled DNA probe and rescreening of cDNA library: The entire insert of one of the positive clones from the immunoscreen was excised from the vector using *Eco*RI digestion followed by gel purification. The DNA was labeled with digoxigenin-dUTP (DIG) using the Genius DNA labeling kit with the random priming protocol, as described by the manufacturer (Boehringer Mannheim). The E-5-1 library was rescreened twice using the DIG-labeled probe. Sixteen positive plaques were taken through two rounds of plaque purification and amplification. Plasmids were excised as described above and electroporated into JM101. Plasmid minipreps were cut with EcoRI to excise their inserts, and these were separated on a 1% agarose, 1× TAE gel, transferred to Magnagraph nylon membrane (Micron Separations, Inc.) (Zhou et al., 1994), and hybridized overnight with the DIG-labeled DNA probe according to the Genius kit protocol. Those clones detected with the probe and found to be larger than the clone obtained through antibody screening were restriction mapped to determine whether they were similar to the original clone. In brief, clones were digested with combinations of *Bam*HI, *Eco*RI, *Hind*III, and *Xho*I and then separated, blotted, and hybridized with the DIG-labeled probe.

Sequence analysis and comparison: Sequence analysis was performed using the Lasergene software package (DNA Star Inc., Madison, Wisconsin) and the Sequence Analysis Software Package version 7.1 by Genetics Computer Group, Inc. (GCG). Nucleotide sequences were submitted to the National Center for Biotechnology Information server for sequence comparisons using the BLASTN, version 1.3.13MP algorithm (Altschul et al., 1990). Protein sequence comparison utilized gapped alignments constructed by BLASTP, v2.0.3 (Altschul et al., 1997).

Multiple sequence alignment: Pairwise sequence overlays of eight chymotrypsin-like serine proteinases for which there is three-dimensional structural information (bovine alpha-chymotrypsin (5CHA), rat mast cell proteinase II (3RP2), porcine kallikrein (2PKA), rat submaxillary gland tonin (1TON), *Streptomyces griseus* trypsin (1SGT), porcine pancreatic elastase (3EST), bovine pancreatic trypsin (2PTN), and human neutrophil elastase (1HNE)) were made according to the method of Chothia and Lesk (1986), using a 3 Å cutoff for alpha-carbon distances (K. Field & P.A. Karplus., unpubl. results). This information was pooled as in Rozwarski et al. (1994) to yield a common core structure. The segments of this core have a high probability of being structurally conserved among members of the chymotrypsin family. The scolexin cDNA sequences, along with two representative mammalian serine proteinases (bovine chymotrypsinogen A (CTRA_BOVIN) and human pancreatic elastase IIIA (EL3A_HUMAN)) and two other insect serine proteinase sequences (*Anopheles* trypsin (TRY1_ANOGA) and *Drosophila* serine proteinase (SER1_DROME)) identified as similar by BLASTP searching, were aligned to the eight sequences used to construct the core sequence, with no gaps allowed within the core segments and no penalty for gaps between the core segments. The computer program GPAlign (K. Clark & G. Reeck, © 1989, Kansas State University Research Foundation) with the MDM78 scoring matrix (Schwartz & Dayhoff, 1978) was used for this alignment. The automated alignment was manually adjusted. Local environments for individual residues in bovine chymotrypsinogen were examined using the molecular modeling program CHAIN and the PDB coordinate file 5cha.

Note added in proof: Both scolexin sequences reported in this paper have been submitted to Genbank. The accession number of ScA is AF087004, and the accession number for ScB is AF087005.

Acknowledgments: The authors wish to thank Dr. Lynn Riddiford for her gift of the *M. sexta* cDNA libraries and Kenneth Field for his work on identifying the serine proteinase core sequence used in the multiple sequence alignment. C.M.F. was supported by a Cornell Biotechnology Fellowship, a Boyce Thompson Institute Graduate Assistantship, and a James Palmer Graduate Assistantship for parts of this study, and the support of these sponsors is gratefully acknowledged. The authors also wish to thank Dr. Gary Blissard for his helpful comments on the manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl Acids Res 25*:3389–3402.

Appel LF, Prout M, Abu-Shumays R, Hammonds AG, Fristrom D, Fristrom J. 1993. The *Drosophila* stubble-stubbloid gene encodes an apparent transmembrane serine protease required for epithelial morphogenesis. *Proc Natl Acad Sci USA 90*:4937–4941.

Bairoch A, Bucher P, Hofmann K. 1997. The PROSITE database, its status in 1997. *Nucl Acids Res 25*:217–221.

Bentley DR. 1986. Primary structure of human complement component C2. *Biochem J 239*:339–345.

Branden C, Tooze J. 1991. *Introduction to protein structure*. New York: Garland Publishing, Inc.

Brenner S. 1988. The molecular evolution of genes and proteins: A tale of two serines. *Nature 334*:528–530.

Castellino FJ. 1979. A unique enzyme-protein substrate modifier reaction: Plasmin/streptokinase interaction. *TIBS 4*:1–5.

Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J 5*:823–826.

Felsenstein J. 1989. PHYLIP: Phylogeny inference package (Version 3.2). *Cladistics 5*:164–166.

Finnerty CM, Granados RR, Hughes PR, Bellotti AC. 1994. Bioassay of several baculoviruses for virus-induced mortality in *Manduca sexta* larvae and induction of infection-specific protein. *J Invert Path 63*:140–144.

Finnerty CM, Granados RR. 1997. The plasma protein scolexin from *Manduca sexta* is induced by baculovirus infection and other immune challenges. *Insect Biochem Molec Biol 27*:1–7.

Freer ST, Kraut J, Robertus JD, Wright HT, Xuong NH. 1970. Chymotrypsinogen: 2.5-Å crystal structure, comparison with a-chymotrypsin, and implications for zymogen activation. *Biochemistry 9*:1997–2009.

Gardell SJ, Duong LT, Diehl RE, York JD, Hare TR, Register RB, Jacobs JW, Dixon RAF, Friedman PA. 1989. Isolation, characterization, and cDNA cloning of a vampire bat salivary plasminogen activator. *J Biol Chem 264*:17947–17952.

Greer J. 1981. Comparative model-building of the mammalian serine proteases. *J Mol Biol 153*:1027–1042.

Henikoff S, Henikoff JG. 1994. Protein family classification based on searching a database of blocks. *Genomics 19*:97–107.

Horodyski FM, Riddiford LM, Truman JW. 1989. Isolation and expression of the eclosion hormone gene from the tobacco hornworm, *Manduca sexta*. *Proc Natl Acad Sci USA 86*:8123–8127.

Hughes JA, Hurlbert RE, Rupp RA, Spence KD. 1983. Bacteria-induced haemolymph proteins of *Manduca sexta* pupae and larvae. *J Insect Physiol 29*:625–632.

Hultmark D. 1993. Immune reactions in *Drosophila* and other insects: A model for innate immunity. *TIG 9*:178–183.

Hultmark D, Steiner H, Rasmuson T, Boman HG. 1980. Insect immunity. Purification and properties of three inducible bactericidal proteins from hemolymph of immunized pupae of *Hyalophora cecropia*. *Eur J Biochem 106*:7–16.

Hurlbert RE, Karlinsey JE, Spence KD. 1985. Differential synthesis of bacteria-induced proteins of *Manduca sexta* larvae and pupae. *J Insect Physiol 31*:205–215.

Iwanaga S. 1993. The limulus clotting reaction. *Curr Opinion Immunol 5*:74–82.

Kyriakides TR, Bedoyan JK, Patil CS, Spence KD. 1993. *In vivo* distribution of immune protein scolexin in bacteria-injected *Manduca sexta* larvae. *Tiss Cell 25*:423–434.

Kyriakides TR, McKillip JL, Spence JD. 1995. Biochemical characterization, developmental expression, and induction of the immune protein scolexin from *Manduca sexta*. *Arch Insect Biochem Physiol 29*:269–280.

Lerro KA, Prestwich GD. 1990. Cloning and sequencing of a cDNA for the hemolymph juvenile hormone binding protein of larval *Manduca sexta*. *J Biol Chem 265*:19800–19806.

Lewin B. 1987. *Genes*. New York: John Wiley and Sons.

Li W, Riddiford LM. 1992. Two distinct genes encode two major isoelectric forms of insecticyanin in the tobacco hornworm, *Manduca sexta*. *Eur J Biochem 205*:491–499.

Minnick MF, Rupp RA, Spence KD. 1986. A bacterial-induced lectin which triggers hemocyte coagulation in *Manduca sexta*. *Biochem Biophys Res Comm 137*:729–735.

Mole JE, Anderson JK, Davidson EA, Woods DE. 1984. Complete primary structure for the zymogen of human complement factor B. *J Biol Chem 259*:3407–3412.

Muta T, Miyata T, Misumi Y, Tokunaga F, Nakamura T, Toh Y, Ikehara Y, Iwanaga S. 1991. Limulus factor C: An endotoxin-sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like, and lectin-like domains. *J Biol Chem 266*:6554–6561.

Navia MA, McKeever BM, Springer JP, Lin T, Williams HR, Fluder EM, Dorn CP, Hoogsteen K. 1989. Structure of human neutrophil elastase in complex with a peptide chloromethyl ketone inhibitor at 1.84-Å resolution. *Proc Natl Acad Sci USA 86*:7–11.

Rawlings ND, Barrett AJ. 1994. Families of serine peptidases. *Methods Enzymol 244*:19–61.

Rozwarski DA, Gronenborn AM, Clore GM, Bazan JF, Bohm A, Wlodawer A, Hatada M, Karplus PA. 1994. Structural comparisons among the short-chain helical cytokines. *Structure 2*:159–173.

Sato T, Endo Y, Matsushita M, Fujita T. 1994. Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein. *Internat Immunol 6*:665–669.

Schecter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Comm 27*:157–162.

Schwartz RM, Dayhoff MO. 1978. Matrices for detecting distant relationships. In: Dayhoff MO, ed. *Atlas of protein sequence and structure*. Washington, D.C.: National Biomedical Research Foundation. pp 353–358.

Sidén-Kiamos I, Skavdis G, Rubio J, Papagiannakis G, Louis C. 1996. Isolation and characterization of three serine protease genes in the mosquito *Anopheles gambiae*. *Insect Molec Biol 5*:61–71.

Smith CL, DeLotto R. 1992. A common domain within the proenzyme regions of the *Drosophila* snake and easter proteins and *Tachypleus* proclotting enzyme defines a new subfamily of serine proteases. *Protein Sci 1*:1225–1226.

Spence KD, Karlinsey JE, Kyriakides TR, Patil CS, Minnick MF. 1992. Regulation and synthesis of selected bacteria-induced proteins in *Manduca sexta*. *Insect Biochem Molec Biol 22*:321–331.

Thiel S, Vorup-Jensen T, Stover CM, Schwaeble W, Laursen SB, Poulsen K, Willis AC, Eggleton P, Hansen S, Holmskov U, Reid KBM, Jensenius JC. 1997. A second serine protease associated with mannan-binding lectin that activates complement. *Nature 386*:506–510.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl Acids Res 22*:4673–4680.

Venot N, Sciaky M, Puigserver A, Desnuelle P, Laurent G. 1986. Amino acid sequence and disulfide bridges of subunit III, a defective endopeptidase present in the bovine pancreatic 6 S procarboxypeptidase A complex. *Eur J Biochem 157*:91–99.

Volanakis JE, Narayana SVL. 1996. Complement factor D, a novel serine protease. *Protein Sci 5*:553–564.

Zhou MY, Xue D, Gomez-Sanchez EP, Gomez-Sanchez CE. 1994. Improved downward capillary transfer for blotting of DNA and RNA. *Biotechniques 16*:58–60.

Zwilling R, Neurath H. 1981. Invertebrate proteases. *Methods Enzymol 80*:633–664.