Isolation and Characterization of a Cone Snail Protease with Homology to CRISP Proteins of the Pathogenesis-related Protein Superfamily*

Trudy J. Milne‡§, Giovanni Abbenante‡, Joel D. A. Tyndall‡, Judy Halliday‡‡, and Richard J. Lewis‡
From the Institute for Molecular Bioscience, The University of Queensland, Queensland 4072, Australia and the Centre for Molecular Biotechnology, School of Life Sciences, Queensland University of Technology, Queensland 4001, Australia

The pathogenesis-related (PR) protein superfamily is widely distributed in the animal, plant, and fungal kingdoms and is implicated in human brain tumor growth and plant pathogenesis. The precise biological activity of PR proteins, however, has remained elusive. Here we report the characterization, cloning and structural homology modeling of Tex31 from the venom duct of Conus textile. Tex31 was isolated to >95% purity by activity-guided fractionation using a para-nitroanilide substrate based on the putative cleavage site residues found in the propeptide precursor of conotoxin TxVIA. Tex31 requires four residues including a leucine N-terminal of the cleavage site for efficient substrate processing. The sequence of Tex31 was determined using two degenerate PCR primers designed from N-terminal and tryptic digest Edman sequences. A BLAST search revealed that Tex31 was a member of the PR protein superfamily and most closely related to the CRISP family of mammalian proteins that have a cysteine-rich C-terminal tail. A homology model constructed from two PR proteins revealed that the likely catalytic residues in Tex31 fall within a structurally conserved domain found in PR proteins. Thus, it is possible that other PR proteins may also be substrate-specific proteases.

Of the genomes that have been completely sequenced, as many as 2% of the gene products encode known proteases, many of which regulate physiological processes such as blood coagulation, fibrinolysis, the complement system, and the processing of protein hormone precursors by specific convertases. Prohormones and neuropeptides (3–40 amino acids), important coordinators of cellular function in the endocrine and nervous systems, are often synthesized as propeptides. The propeptide is subsequently processed by substrate-specific proteases to yield the mature bioactive form (1). As many of the substrate-specific proteases remain unidentified, predicting new bioactive peptides from cDNA sequences is presently difficult, if not impossible (2, 3). Linderström-Lang coined the term “limited proteolysis” to describe enzymes with restricted specificity (4). Enzyme activity can be restricted by substrate structure, enzyme processing state, endogenous protease inhibitor levels or a combination of these factors (5). Peptides derived from prohormones are typically flanked by a pair of dibasic residues (Lys-Arg, Arg-Arg, Lys-Lys, or Arg-Lys) and are cleaved by proteases found in secretory vesicles (1). However, many precursor peptides contain multiple sets of basic residues, suggesting that highly substrate specific or differentially expressed proteases can determine processing outcomes.

The venom of cone snails (predatory marine molluscs of the genus Conus) has yielded a rich source of novel neuroactive peptides (“conotoxins”) (6). However, the enzymes responsible for cleaving the conotoxins from their propeptide precursors have not been isolated. Here we report the purification, cloning, and initial characterization of Tex31, a protease from venom Conus textile. Interestingly, Tex31 sequence was found to be homologous to members of the PR protein superfamily.

MATERIALS AND METHODS

Substrate Design and Activity Assay—Based on the cdNA sequence of the propeptide and the mature amino acid sequence of the TxVIA peptide, a β-conotoxin contained in the venom of C. textile, propeptide cleavage site substrate analogues were designed. The peptic para-nitroanilide (pNA) substrate analogues were synthesized by solid phase synthesis (7). A typical reaction mixture consisted of 100 mM Bistris, pH 8.5, 100 μM pNA substrate (acetyl-KLNRKWabuKQGG-NH2), and initial characterization of Tex31, a protease from venom Conus textile. Interestingly, Tex31 sequence was found to be homologous to members of the PR protein superfamily.

Received for publication, May 8, 2003
Published, JBC Papers in Press, May 20, 2003, DOI 10.1074/jbc.M304843200

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.jbc.org
 tileSize in chilled extraction buffer (100 mM Bistris propane, pH 7, containing 0.1% Triton X-100). Extracted material was collected and centrifuged at 10,000 × g for 30 min (4 °C). The supernatant was concentrated with an Ultrafree 15 Biomax 5k cutoff centrifugal filter (Millipore) in preparation for the purification of Tex31. The Tex31 isolated for sequencing was first purified on a reversed-phase C4 semi-preparative column eluted with a 1% gradient of 0–80% acetonitrile containing 0.1% (v/v) trifluoroacetic acid. Active fractions were pooled, concentrated, and applied to a Superdex 75 column (Amersham Biosciences) eluted with 50 mM Tris-HCl, pH 8, containing 150 mM NaCl. Active fractions contained a major protein with an apparent molecular weight of ~28,000 and of >95% purity as assessed by SDS-PAGE (Fig. 1A, lane 4). These fractions were pooled, concentrated, and applied to an analytical reversed-phase C4 HPLC column eluted with a 0.5% gradient of 40–70% acetonitrile containing 0.1% (v/v) trifluoroacetic acid. A single major peak (C4) eluting at 52% acetonitrile, with an apparent molecular weight of ~28,000 and purity >99% as assessed by SDS-PAGE, was collected and sequenced (Fig. 1A, lane 5).

To confirm that activity was associated with the ~28,000-Da band and to purify sufficient material for biochemical analysis, additional Tex31 was purified using a modified method. First, an ion-exchange DEAE-Fast Flow Hi-Trap column (Amersham Biosciences) was equilibrated, and crude venom duct extract was eluted with 20 mM Bistris propane, pH 7. Tex31 was eluted with a linear 0–1 mM NaCl gradient in 20 mM Bistris propane, pH 7, and fractions were assayed in 50 mM Tris-HCl, pH 8.5. Pooled active fractions eluting at ~150 mM NaCl were concentrated by ultrafiltration (as described above) and applied to a Biosep-SEC-S 2000 column (Phenomenex) equilibrated in 20 mM Bistris propane, pH 7, containing 150 mM NaCl. The eluted active fractions of >95% purity as assessed by SDS-PAGE were pooled (SEC I) (Fig. 1B, lane 3) and used to characterize Tex31 proteolytic activity. This pooled fraction was reapplied to the SEC column eluted at 0.3 ml/min to obtain an active fraction that gave a single band on SDS-PAGE (SEC II) ~28,000-Da protein (Fig. 1B, lane 4), confirming that proteolytic activity corresponded to the sequenced material.

Mass Spectrometry and Sequencing—Electrospray quadrupole-time of flight mass spectra of Tex31, SEC II, and C4 were acquired on an Applied Biosystems QSTAR pulsar mass spectrometer (Applied Biosystems). Purified C4 was digested with trypsin (1:1-tonsilamido-2-phenylthiochromomethyl ketone-treated, Sigma) for 6 h at 37 °C in 200 mM N-ethylmorpholine, pH 8.5. Digested peptides were purified on a reversed-phase C18 column (Zorbax C18 300SB). Amino acid sequencing was performed on both the intact Tex31 and tryptic peptides by Edman degradation on an Applied Biosystems 477A protein sequencer.

Cloning—TRizol (Invitrogen) was used to extract total RNA from a single C. textile venom duct, previously stored in RNAlater (Ambion) at −20 °C, according to the manufacturer’s protocol. Single-stranded cDNA was 3’ end-synthesized from total RNA templates using an oligoDT18 primer (PE Biosystems) extended with Superscript II reverse transcriptase (Invitrogen). Degenerate PCR oligonucleotide primers were designed based on the N-terminal protein sequence (5’-TAY-TATTTGCTYNACNCCNGCAYATG-3’) and the internal tryptic peptide (5’-CARTTTNGCNGCTNGTNCRCNGC-3’), designated Pr1S and Pr2R, respectively. Touchdown PCR (64 °C to 50 °C) was carried out on C. textile cDNA using Red Hot DNA polymerase (Advanced Biotechnologies), 3 mM MgCl₂, 200 μM dideoxynucleotides, primers Pr1S and Pr2R, producing a 140-bp cDNA fragment. This fragment was ligated into a cloning vector (TA cloning vector, Invitrogen) and sequenced as described above. The primers Pr5S (5’-CGgACTACACAGTTCGGC-3’) and Pr6R (5’-CGgACgCCTTACgCgGgC-3’) were used to amplify full-length Tex31 cDNA.

Biochemical Characterization—The optimal pH for protease activity was determined over a pH range of 7.2–9.4 in 100 mM Bistris propane buffer. pNA substrate analogues based on the TxVIA conotoxin propeptide cleavage site along with three control substrates were assayed at 100 μM in 100 mM Bistris propane, pH 8.5 at 37 °C. Protease inhibitors were preincubated with Tex31 for 30 min at 37 °C before the addition of 100 μM pNA and the reaction incubated for a further 1 h. RVKR-CMK (decanoyl-Arg-Val-Arg-Lys-chloromethylketone) was assayed at 100 μM, Kunitz-soy bean trypsin inhibitor, 100 μM; leupeptin, 100 μM; EGTA, 1 mM; EDTA, 1 mM; AEBSP (4-(2-aminoethyl)benzenesulfonyl fluoride) 100 μM, Nα-tosyl-Lys-chloromethylketone, 100 μM;

![Image](https://example.com/image.png)

**Fig. 1.** Purification and characterization of Tex31. A, Coomassie Blue-stained 12% SDS-PAGE gel showing molecular mass markers (LMW, Amersham Biosciences) (lane 1), crude C. textile venom (lane 2), preparative C4 fraction (lane 3), Superdex 75 fraction (lane 4), C4 purified Tex31 used for Edman sequencing (lane 5). B, Coomassie Blue-stained 12% SDS-PAGE gel showing crude extract (lane 1), DEAE active fraction (lane 2), SEC I active fraction (lane 3), SEC II active fraction (puriﬁed active Tex31) (lane 4), molecular mass markers (Seeblue, Novex) (lane 5). Tex31 is seen as a band at ~28,000 Da in lanes A and B. C, reconstructed mass spectra of Tex31 (see lane 5 of A). Tex31 was identiﬁed at 30,825 Da together with a number of related masses. SEC II-puriﬁed Tex31 (lane 4 of B) gave a similar mass spectrum.

peptatin A, 2 μM; E-64, 20 μM; and benzamidine, at 1 mM. Tex31 was preincubated for 30 min at 37 °C in buffers containing 2 mM metal ion, 1 mM EDTA, 100 mM Bistris propane, pH 8.5, before the addition of 100 μM Ac-LNKRP-pNA. Tex31 in 100 mM Bistris propane, pH 8.5, buffer was included as a control. Initial velocities (v) were determined with 0–150 μM Ac-KLNKR-pNA, Tex31 SEC I pool (200 pmol) in 100 mM Bistris propane, 1 mM CaCl₂, pH 8.5, at 37 °C. Kinetic parameters were calculated from nonlinear regression of initial velocities as a function of seven substrate concentrations [S] using GraphPad Prism software. The kₚ/Kᵥ values were calculated assuming Michaelis-Menten kinetics. Tex31 SEC I pool (200 pmol) and 100 μM Ac-LNKRP-pNA was used unless stated. Tex31 protein concentration was determined using the BCA protein assay kit (Pierce).

**Homology Modeling**—The PR protein, Ves v 5, and P14a structures were overlaid within the InsightII (9) environment to ascertain structural homology. The structurally conserved regions were then formally superimposed using InsightII. The structural alignment and homology modeling of Tex31 was carried out using the homology module with subsequent refinement using Discover 3. The validity of the model was tested using WHATCHECK and PROCHECK (10). Electrostatic potentials were calculated using GRASP (11).

**RESULTS AND DISCUSSION**

Using an Ac-LNKRP-pNA substrate based on the propeptide cleavage site of conotoxin TxVIA (12) in a simple colorimetric
assay, we isolated and purified to homogeneity a major protease Tex31 from the
C. textile venom duct (Fig. 1). The venom duct contents contained the majority of Ac-LNKR-pNA-specific
cleavage activity, suggesting that Tex31 is secreted from the venom duct tissue. Tex31 appeared not to be specifically asso-
ciated with the insoluble granules found in the crude venom

FIG. 2. Nucelotide and deduced amino acid sequence of Tex31 cDNA (GenBank™/EBI accession number AJ4913118). SignalP analysis (13) of the translated amino acid sequence predicts a 21-amino acid signal peptide with a cleavage site on the carboxyl side of Gly21 (indicated here by the downward arrow). The amino acid sequence also contains the sequence RKR (boxed), a classic proprotein cleavage site motif for proteolytic processing by prohormone convertase-like proteases (1). As the N-terminal amino acid of Tex31 is a histidine, Tex31 must be secreted after a post-translational cleavage between the arginine and histidine residues (indicated here with an upward arrow). The polyadenylation signal is underlined twice.

FIG. 3. Sequence alignment of representative pathogenesis-related proteins. Tex31; Ves v 5, antigen 5 from V. vilgaris (GenBank™/EBI accession number Q05110); P14a, tomato PR protein (GenBank™/EBI accession number P04284); Tpx-1, human testis-specific protein (GenBank™/EBI accession number P16562) (also known as CRISP-2); Heth, helothermine (Mexican beaded lizard) (GenBank™/EBI accession number Q91055); and GliPR, human glioma PR protein (GenBank™/EBI accession number P48060). Sequences were aligned using ClustalX (31). Tpx-1 and helothermine are members of the CRISP protein family. PR protein family signatures (32) 1 and 2 residues are highlighted as indicated. Conserved residues have been highlighted in green and semiconserved residues in blue. Putative active site residues are highlighted in red.
Under reducing conditions, Tex31 migrated with an apparent molecular weight of 28,000 on 12% SDS-PAGE gel (Fig. 1A, lanes 3–5, and B, lanes 2–4). Mass spectral analysis of Tex31 showed a mass of 30,825 Da, as well as a number of slightly heavier masses (Fig. 1C) that may represent modified forms of the enzyme. Similar profiles of related masses were seen in batches of Tex31 prepared by two separate purification methods (SEC II and C4; data not shown). Edman degradation of C4 purified native Tex31 yielded a single N-terminal sequence, HHXDSKYYELTPAHTM. Based on this protein sequence, and a tryptic digest fragment ADVTDAANMLK, we designed two degenerate primers to PCR-amplify a 140-bp fragment of cDNA encoding Tex31. Utilizing this sequence, we designed specific primers to isolate the 5'- and 3'-untranslated regions using SMART rapid amplification of cDNA ends. The final assembled cDNA sequence encoded a 300-amino acid protein (GenBank™/EBI data base: accession number AJ4913118) (Fig. 2). The predicted post-translationally processed secreted protein of 276 residues contains both the N-terminal and tryptic digest sequences and has a calculated molecular weight of 30,875. This result is consistent with SDS-PAGE (~28,000 Da) and mass spectrometry (30,853 Da).

![Fig. 4. Proteolytic activity of Tex31. A, propeptide substrate cleavage. HPLC trace of the products of cleavage of the Ac-KLNKRWA-buKQSG-NH₂ propeptide substrate after 3-h incubation at 37 °C in the presence of 1.5 µM Tex31. Cleavage products were separated with a 2% gradient of 0–40% acetonitrile containing 0.1% (v/v) trifluoroacetic acid on a Vydac C18 reversed-phase column. B, substrate specificity. Each pNA substrate analogue was assayed at 100 µM. Proteolytic activity toward Ac-KLNKR-pNA defines 100% activity. Three control substrates are included (shown in light gray).](image)

![Fig. 5. Biochemical characterization of Tex31. A, pH profile for Tex31. Activity was determined over a pH range in 100 mM Bistris propane. B, metal ion dependence. All divalent chloride salt buffers (2 mM) were EDTA (1 mM) buffered. Proteolytic activity in the presence of CaCl₂ (1 mM) is taken as 100%. C, inhibitor profile. Inhibitors of serine, cysteine, aspartyl, and metalloproteases are represented. D and E, Kᵦ and Vₘₐₓ were determined using the Ac-KLNKR-pNA substrate in the presence of 1 mM CaCl₂. All data represent mean ± S.E. of duplicate results from two separate experiments (n = 4).](image)
Sequence homology to the plant PR-1 family have now been identified in the animal (invertebrates and vertebrates, including humans) and fungal kingdoms. Proteins include the human brain tumor PR proteins GluPR (17) and RTVP-1 (18), mammalian Tpx-1, testis-specific protein (19), helothermine, from lizard venom (20), antigen 5 from Vespuca vilgaris venom (21), and proteins from fungi (Saccharomyces, Schizopyllam) (22). Sequence alignment of PR proteins reveals a highly conserved core of amino acids (Fig. 3), suggesting that they may share a similar function. This conservation has been interpreted as evidence that PR proteins arose from a common ancestor and evolved into a large “PR protein superfamily” (23). Among the PR proteins, Tex31 displays the highest sequence identity (36%) to a subfamily of PR proteins found in mammals and known as the cysteine-rich secretory proteins (CRISPs). Like the CRISP subfamily (24), which includes Tpx-1 (19) and helothermine (20), Tex31 possess a characteristic cysteine-rich C-terminal region, with 19 of its 22 cysteine residues found in the C terminus.

Biochemical characterization confirms that Tex31 is proteolytic. Tex31 cleaved the conotoxin propeptide substrate Ac-KLNKRWAbuKQSG-NH\(_2\) at the predicted propeptide cleavage site (Fig. 4A). The generation of the two resulting peptide fragments, Ac-KLNKR-OH and H-WAbuKQSG-NH\(_2\), was confirmed by mass spectrometry. A number of pNA substrate analogues were used to assess the substrate specificity of Tex31 (Fig. 4B). Tex31 cleaved single basic amino acid pNA substrates inefficiently. However, the incremental addition of a second, third, and fourth residue increased cleavage activity by 13-, 2.4-, and 4.4-fold, respectively. A glycine substituted for the leucine residue at position P4 resulted in a 11-fold loss in activity, demonstrating that the substrate specificity of Tex31 preferred a leucine residue at position P4. Proteolytic activity was little altered with the addition of a lysine residue found in the native propeptide substrate. In contrast, the isostructural, but uncharged, norleucine caused ~50% loss of activity. The substantial influence on cleavage activity of the P4 and P5 residues demonstrates that the specificity of Tex31 is influenced by residues far removed from the cleavage site. To investigate the selectivity toward other conotoxin family propeptides of Tex31, the substrate Ac-LEKR-pNA based on the cleavage site of the \(\mu\)-conotoxin family propeptides was synthesized. Tex31 also cleaved this substrate efficiently, again demonstrating its preference for substrates with a leucine at P4 and suggesting that a related protease could be used by mollusc hunting cone snails to cleave \(\mu\)-conotoxins.

The ability of Tex31 to cleave the native substrate Ac-KLNKR-pNA was 4–33-fold greater than for control substrates Ac-EVKKQR-pNA, Ac-TTSTRR-pNA, and Ac-RTSKKR-pNA, which are the preferred substrates for the dengue 2 virus NS3 protease (25), and 166-fold greater than for the Ac-R-pNA substrate, which is efficiently cleaved by trypsin (26).

Tex31 is active over a broad pH range with maximal activity at pH 8.5 (Fig. 5A) near the pH of C. textile crude venom (pH ~8). Tex31 cleaves the peptidyl Ac-KLNKR-pNA substrate efficiently, with an apparent affinity (\(K_m\)) of 80 \(\mu\)M and a proteolytic efficiency (\(k_{cat}/K_m\)) of \(8.7 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}\) (Fig. 5B). To further characterize Tex31 activity, the effect of metal ions and serine, cysteine, aspartyl, and metalloprotease inhibitors were assessed (Fig. 5, C and D). The irreversible serine-protease inhibitors decanoyl-RVKR-chloromethylketone and Kunitz-soy bean trypsin inhibitor fully inhibited Tex31. Leupeptin, a reversible inhibitor of serine and cysteine proteases, inhibited 75% of the activity of Tex31, while the irreversible cysteine-protease inhibitor E-64 had no significant effect. In contrast, the irreversible serine inhibitor AEBSF caused significant in-
Isolation and Characterization of a Cone Snail Protease

The identification of catalytic residues modified by irreversible inhibitors using mass spectral peptide fragment analysis was made impossible due to the protease resistant nature of Tex31. The addition of the metal chelators EDTA and EGTA (1 mM) reduced cleavage activity by 58 and 65%, respectively, whereas the addition of 1 mM Ca\(^{2+}\) caused a 5-fold increase in substrate cleavage (Fig. 5D). This result indicates that calcium ions may contribute to the structural stability (27) of Tex31. Neither Mg\(^{2+}\) nor Zn\(^{2+}\) were found to effectively substitute for Cu\(^{2+}\) (Fig. 5D). Due to the nonselective nature of many protease inhibitors, the question of to which protease class Tex31 belongs may be more accurately answered by site-directed mutagenesis studies.

To explore the structure of Tex31, we constructed a homology model for two PR protein structures. The highly conserved core amino acids from two PR protein structures, namely the crystal structure of antigen 5 from V. vilgaris venom (Ves v 5) (28) and the NMR structure of the plant protein P14a (29), were aligned with Tex31 and a model constructed in InsightII (Fig. 6A). Statistically, enzyme active sites are located in the largest cleft of an enzyme (30). On viewing the homology model, we identified His\(^{130}\) in close hydrogen bonding distance to Glu\(^{115}\) (Fig. 6B) within a large, well defined electronegative cleft (Fig. 6C). Given their proximity and placement, we propose that His\(^{130}\) and Glu\(^{115}\) are two members of a catalytic triad that cleaves basic peptides. A conserved serine nucleophile, typically the third element of the catalytic triad for serine proteases, which is predicted from the AEBSF inhibition studies (Fig. 4C), is more easily recognized from the homology model. However, the relatively conserved Ser\(^{205}\) (Fig. 3) is a possible candidate, since it is located on a loop that could position it in close proximity to His\(^{130}\) and Glu\(^{115}\).

Tex31 is the first member of the PR protein superfamily with a defined function. Substrates comprising at least four residues N-terminal of the cleavage site, and including a hydrophobic leucine at P4, characterize proteolysis by Tex31. A possible basic substrate recognition site, which includes the likely catalytic residues His\(^{130}\) and Glu\(^{115}\), was identified in a Tex31 homology model. Given that the proposed catalytic residues in Tex31 fall within structurally conserved regions of the PR proteins, we suggest that PR proteins may include presently uncharacterized substrate-specific proteases. Lack of knowledge of the specific substrate requirements for PR proteins may have confounded attempts to identify their proteolytic activity. Tex31 may provide the first insight into the role played by PR proteins in the cell in response to disease and stress.

Acknowledgements—We thank A. Jones for mass spectrometry, D. Leung for control pNA substrates, D. Adams for peptide synthesis, I. Sharpe for assistance with manuscript preparation, and T. Walsh (Queensland University of Technology) for support during this project.

REFERENCES

Isolation and Characterization of a Cone Snail Protease with Homology to CRISP Proteins of the Pathogenesis-related Protein Superfamily
Trudy J. Milne, Giovanni Abbenante, Joel D. A. Tyndall, Judy Halliday and Richard J. Lewis

doi: 10.1074/jbc.M304843200 originally published online May 20, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M304843200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 30 references, 8 of which can be accessed free at
http://www.jbc.org/content/278/33/31105.full.html#ref-list-1