

# Sequence of a cDNA Clone Encoding the Zinc Metalloproteinase Hemorrhagic Toxin e from *Crotalus atrox*: Evidence for Signal, Zymogen, and Disintegrin-like Structures<sup>†,‡</sup>

Lawrence A. Hite,<sup>§</sup> John D. Shannon,<sup>§</sup> Jón B. Bjarnason,<sup>||</sup> and Jay W. Fox<sup>\*,§</sup>

Department of Microbiology, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908, and Science Institute, University of Iceland, Reykjavik, Iceland

Received September 17, 1991; Revised Manuscript Received January 17, 1992

**ABSTRACT:** The sequence of two overlapping cDNA clones for the zinc metalloproteinase hemorrhagic toxin e (also known as atrolysin e, EC 3.4.24.44) from the venom gland of *Crotalus atrox*, the Western diamondback rattlesnake, is presented. The assembled cDNA sequence is 1975 nucleotides in length and encodes an open reading frame of 478 amino acids. The mature hemorrhagic toxin e protein as isolated from the crude venom has a molecular weight of approximately 24 000 and thus represents the processed product of this open reading frame. From the deduced amino acid sequence, it can be hypothesized that the enzyme is translated with a signal sequence of 18 amino acids, an amino-terminal propeptide of 169 amino acids, a central hemorrhagic proteinase domain of 202 amino acids, and a carboxy-terminal sequence of 89 amino acids. The propeptide has a short region similar to the region involved in the activation of matrix metalloproteinase zymogens. The proteinase domain is similar to other snake venom metalloproteinases, with over 57% identity to the low molecular weight proteinases HR2a and H<sub>2</sub>-proteinase from the Habu snake *Trimeresurus flavoviridis*. The carboxy-terminal region, which is not observed in the mature protein, strongly resembles the protein sequence immediately following the proteinase domain of HR1B (a high molecular weight hemorrhagic proteinase from the venom of *T. flavoviridis*) and the members of a different family of snake venom polypeptides known for their platelet aggregation inhibitory activity, the disintegrins. The cDNA sequence bears striking similarity to a previously reported sequence for a disintegrin cDNA. This report is evidence that this subfamily of venom metalloproteinases is synthesized in a proenzyme form which must be proteolytically activated. Furthermore, these studies suggest a structural link between the low and high molecular weight hemorrhagic toxins and the disintegrin-like family of polypeptides.

A prominent feature of crotalid snake venoms is that they contain an abundance of proteolytic enzymes. Some of these proteinases are directly capable of causing hemorrhage at the site of envenomation and often disseminate from the site, affecting the entire appendage. In severe cases, systemic hemorrhage has been observed, affecting the heart, lungs, kidneys, and brain. Both types of hemorrhage are the result of the action of zinc metalloproteinases found in the venom [reviewed in Bjarnason and Fox (1989)]. Bjarnason and Tu (1978) first isolated and partially characterized five hemorrhagic metalloproteinases from *Crotalus atrox* venom. The cleavage specificities of these enzymes were initially investigated using insulin B chain as substrate in order to understand more about the mechanism of hemorrhage induction (Bjarnason & Fox, 1983, 1987; Fox et al., 1986; Bjarnason et al., 1988). From both microscopic pathological studies (Ownby et al., 1978) and specificity studies using purified basement membrane components (Baramova et al., 1989, 1990a), it has been well established that these hemorrhagic metalloproteinases degrade the proteins of basement membrane including collagen type IV, nidogen (entactin), and laminin.

Despite these studies, it was still not understood why the sizes of these metalloproteinases are so varied (from 24 to 68

kDa) and why the higher molecular weight toxins are generally more effective at producing hemorrhage than the low molecular weight toxins. One explanation based upon our previous studies (Baramova et al., 1990b) is that the low molecular weight hemorrhagic toxins (Ht-c, -d, and -e)<sup>1</sup> are effectively inhibited by plasma  $\alpha_2$ -macroglobulin while the high molecular weight toxin Ht-a (atrolysin A, EC 3.4.24.1) was not. Partial protein sequence data are now available for many such enzymes from a broad range of crotalid snake venoms, which suggests that these enzymes are evolutionarily related members of a subfamily of zinc metalloproteinases (Shannon et al., 1989; Miyata et al., 1989; Takeya et al., 1989, 1990a,b; Tanizaki et al., 1989; Nakagawa et al., 1989; Sanchez et al., 1991). Indeed, the entire protein sequences of five hemorrhagic proteinases and one nonhemorrhagic proteinase have been completed by traditional Edman protein sequencing. With the elucidation of the protein sequence for HR1B (trimerelysin I, EC 3.4.24.52) from *Trimeresurus flavoviridis*

<sup>†</sup> This work was supported by Grant RO1GM31289 to J.W.F. from the National Institutes of Health and Grant INT-8902965 to J.W.F. from the National Science Foundation.

<sup>‡</sup> The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05372.

\* To whom correspondence should be addressed.

<sup>§</sup> University of Virginia Health Sciences Center.

<sup>||</sup> University of Iceland.

<sup>1</sup> Abbreviations: EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; Ht, hemorrhagic toxin; kDa, kilodalton(s); KGD, tripeptide sequence lysine-glycine-aspartate; M-MuLV, murine Molony leukemia virus; MBIR, Molecular Biology Information Resource; MMP, matrix metalloproteinase; NZY, rich bacteriological broth containing 10 g of NZ amine, 5 g of sodium chloride, 5 g of yeast extract, and 2 g of magnesium sulfate heptahydrate per liter, adjusted to pH 7.5; ORF, open reading frame; RGD, tripeptide sequence arginine-glycine-aspartate; SDS, sodium dodecyl sulfate; SM, phage buffer consisting of 0.1 M sodium chloride, 8 mM magnesium sulfate, 50 mM Tris-HCl, pH 7.5, and 2% gelatin; SSC, standard saline citrate (0.15 M sodium chloride/15 mM sodium citrate, pH 7.0); TAE, Tris-acetate-EDTA buffer (40 mM Tris-acetate/1 mM EDTA); TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane.

(Takeya et al., 1990a), it was shown that this high molecular weight toxin consists of three domains. At the amino-terminal end is a proteinase domain similar to the small molecular weight hemorrhagic metalloproteinases, followed by two additional domains, one of which appears to be related to the disintegrin family of proteins while the other shows weak similarity to von Willebrand factor. Here we present the complete cDNA sequence and the translated protein sequence of a low molecular weight hemorrhagic proteinase, Ht-e, from *Crotalus atrox*, the Western diamondback rattlesnake. On the basis of this work, we can now suggest a new explanation for the structural relationship between the high and low molecular weight snake venom metalloproteinases and the disintegrin-like family of polypeptides.

#### EXPERIMENTAL PROCEDURES

**Protein Preparation.** Hemorrhagic toxins were purified from crude *C. atrox* venom (Miami Serpentarium) using the purification scheme of Bjarnason and Tu (1978). Hemorrhagic toxins were alkylated with iodoacetamide by the method of Hirs (1967), or with 4-vinylpyridine by the method of Friedman et al. (1970), then HPLC-purified away from the reduction/alkylation reagents using a reverse-phase (C-4) column with a starting solvent of 0.1% aqueous TFA, and eluted with 0.1% TFA in acetonitrile at a gradient of 1% increase in eluting solvent per milliliter. Cyanogen bromide digestion of Ht-e by the method of Gross (1967) followed by HPLC (using the same conditions outlined previously) yielded purified CNBr fragments.

**Amino Acid Analysis.** Amino acid analysis of Ht-e was done by a modification (Shannon et al., 1989) of the method of Bidlingmeyer et al. (1984).

**Protein Sequencing.** Protein sequencing was performed on an Applied Biosystems 470A gas-phase sequencer (Applied Biosystems, Foster City, CA) with an on-line Applied Biosystems Model 120A analyzer using the manufacturer's standard protocols.

**Poly(A<sup>+</sup>) RNA Isolation.** RNA was isolated from the venom glands of two freshly sacrificed adult rattlesnakes (Maverick Trading Post, Farmer's Branch, TX) approximately 4 ft in length. Briefly, the snakes were sacrificed by decapitation; the glands were removed and placed directly into a denaturing solution of 4 M guanidinium isothiocyanate and homogenized using a motorized homogenizer. Total RNA was isolated by the method of Chirgwin et al. (1979). Homogenate was layered over ultracentrifuge tubes containing 5.7 M cesium chloride plus 0.1 M EDTA and centrifuged for 18 h at 33 000 rpm at 20 °C. The pelleted RNA was recovered and ethanol-precipitated. Purified total RNA was then passed over an oligo(dT)-cellulose column and eluted as poly(A<sup>+</sup>) RNA. The total yield of poly(A<sup>+</sup>) RNA was about 10 µg from two snakes.

**cDNA Library Construction.** cDNA synthesis was accomplished using the UniZAP XR kit by Stratagene (La Jolla, CA). First-strand synthesis from 1.2 µg of poly(A<sup>+</sup>) RNA was primed by an *Xho*I/oligo(dT) linker/primer and catalyzed by M-MuLV reverse transcriptase in the presence of [ $\alpha$ -<sup>32</sup>P]dATP and 5-methyl-dCTP (to hemimethylate the cDNA at all internal *Xho*I sites). After second-strand synthesis using *Escherichia coli* DNA polymerase the cDNA ends were blunted with T4 DNA polymerase, and *Eco*RI adapters were ligated on. The cDNA was then digested with *Xho*I to remove the adapter from the end containing the *Xho*I linker/primer sequence (internal *Xho*I sites were protected by the hemimethylated state of the cDNA). The resultant cDNA had an *Eco*RI site at the 5' end of the message and an *Xho*I site at the 3' [poly (A)] end. cDNA was then purified from the

excess adapters and cleaved *Xho*I linkers by gel filtration chromatography using Sepharose CL-4B resin. This step was also used to size-select for larger cDNAs. This cDNA was then ligated into Stratagene's UniZAP XR vector (a derivative of phage  $\lambda$ ) and packaged using Stratagene's Gigapack II packaging extract. The packaged DNA was then titrated and found to contain  $2.9 \times 10^7$  recombinants. Of these,  $2 \times 10^6$  were amplified to make the library.

**Design and Synthesis of Oligonucleotide Probes.** The amino-terminal protein sequence of Ht-e was compared to similar regions of other *C. atrox* hemorrhagic toxins (partial protein sequences of which we had previously determined), and a heptameric amino acid sequence was chosen which would be unlikely to cross-hybridize with other hemorrhagic toxin sequences, yet which would have as few as possible degenerate positions. A mixture of oligonucleotide probes of 20 bases in length was synthesized using a Biosearch 8600 synthesizer (MilliGen/Biosearch, Burlington, MA), incorporating the proper mixture of bases at each site of degeneracy. The DNA sequence of the probe was A-T-G-T-A-C/T-A-C-N-A-A-A/G-T-A-C/T-A-A-C/T-G-G, corresponding to the peptide sequence M-Y-T-K-Y-N-G (positions 18–24) from the amino-terminal region of the mature Ht-e proteinase. Prior to hybridization, the probe was labeled using T4 polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]ATP (6000 Ci/mmol; New England Nuclear, Boston, MA) (Ausubel et al., 1987) and purified by spun column chromatography using BioGel P-2 (Bio-Rad, Richmond, CA) and a method adapted from Sambrook et al. (1982).

**cDNA Library Screening.** Since the hemorrhagic toxins were expected to be highly expressed proteins in the venom gland, the primary screening of the library was done at a low plaque density. The library was plated on 150-mm NZY plates and grown 9 h at 37 °C. Plates were then chilled for several hours, and plaques were lifted in duplicate onto nitrocellulose filters and processed as recommended by Stratagene. After vacuum baking at 80 °C, filters were washed in three changes of  $3 \times \text{SSC}/0.1\%$  SDS at room temperature and then prehybridized for 3 h at 42 °C. Labeled probe and hybridization solution were then added, and the hybridization was allowed to proceed overnight at 42 °C. The next day the solution was removed, and the filters were washed in three changes of  $6 \times \text{SSC}/0.1\%$  SDS at room temperature and then in the same buffer at 42 °C. Filters were then sealed in plastic wrap, and Kodak XAR film (Rochester, NY) was placed over them in an autoradiography cassette. Film was exposed overnight at -70 °C. Positive plaques were identified after the film was developed and the autoradiograph was aligned to the plates.

**Characterization of Positive Clones.** Plaques which were both positive and clearly defined from surrounding plaques were removed using a sterile Pasteur pipet and eluted in SM buffer. Clones were then rescued into pBlueScript II phagemids using R408 helper phage. Plasmid DNA was then obtained using a Qiagen plasmid kit (Qiagen Inc., Chatsworth, CA). The plasmid DNA was restriction-digested with *Eco*RI and *Xho*I to remove the insert and analyzed using 1% agarose gels in TAE buffer. Two of the largest clones (HO11 and HO12) were chosen for DNA sequence analysis. Those clones to be sequenced were also isolated in single-stranded form using the in vivo packaging signals in the pBlueScript II vector using the protocols provided by Stratagene.

**DNA Sequencing.** DNA sequencing was performed on a du Pont Genesis 2000 automated DNA sequencer (E. I. du Pont de Nemours & Co., Wilmington, DE). Sanger dideoxy

sequencing reactions (Sanger et al., 1977) were performed on both single- and double-stranded DNA using reagents and protocols furnished by du Pont. Typical reactions gave readable sequences of 250–350 bases after manual editing of the results given by the base-calling software. The single-stranded reactions generally gave superior data. Sequence overlaps were obtained by “walking” across the insert, sequencing in from both ends starting with primers within the vector sequence and making new oligonucleotide primers to the end of each sequence run until each strand was completed. Synthesized primers were 17 or more bases in length. The coding strand was sequenced from single-stranded template and the noncoding strand from double-stranded template. Both strands were sequenced in their entirety at least twice.

**Sequence Assembly.** Individual sequence runs were assembled into overlaps using the computer program SAM (Sequence Assembly Manager) of the MBIR suite of programs (Lawrence et al., 1989). After being edited, the assembled sequence was translated into the coding sequence, and both the protein and DNA sequences were compared to sequences in the GenBank (Bilofsky et al., 1986) and PIR (Protein Identification Resource, National Biomedical Research Foundation, Georgetown University, Washington, D. C.) databases using the *fasta* program of Pearson and Lipman (1988).

## RESULTS

**Protein Sequencing and Amino Acid Analysis of Ht-e.** The amino acid sequence of the amino terminus of the mature Ht-e protein was determined to 42 residues by Edman degradation, and part of this protein sequence was used to design a suitable oligonucleotide probe for screening the cDNA library. Several Ht-e CNBr peptides were sequenced to verify that positive clones isolated from the library encoded the correct protein sequence. The carboxy-terminal CNBr peptide was sequenced from Asn-192 to Pro-202, where the signal ended, suggesting the end of the peptide. Amino acid analysis of this peptide (data not shown) confirmed that Pro-202 was the final residue of the mature proteinase. The amino acid composition of Ht-e was determined. Table I displays the average results of five amino acid analyses compared to the predicted composition from the cDNA sequence.

**Isolation of cDNA Clones.** As stated earlier, it was expected that Ht-e would be highly expressed in the venom gland library. This assumption was borne out by the screening of the library with the Ht-e 20-mer oligonucleotide probe. The average number of clones which tested positive to this probe from the screening of three plates was greater than 1%. Due to the high number of positives, 16 clones were selected for further work. The clone (HO12) which was first selected to be completely sequenced consisted of 1795 base pairs. When it was discovered that this was not sufficient to unambiguously determine the translation initiation codon, a second, overlapping clone (HO11) was sequenced. This clone is slightly larger than the HO12 clone and gave an additional 180 base pairs of 5'-untranslated sequence for a total length of 1975 base pairs.

Figure 1 shows the cDNA sequence of Ht-e and the translated protein sequence from the initiation codon to the termination codon. The deduced amino acid sequence from this open reading frame is 478 amino acids, much larger than would be necessary to encode a 24-kDa protein. Within this sequence, we assume the presence of a signal peptide, a propeptide, the mature protein, and a carboxy-terminal peptide which must be posttranslationally removed.

**Similarity of Ht-e to Other Sequences.** When the nucleotide sequence of the Ht-e clone was compared with the

Table I: Amino Acid Composition of Ht-e<sup>a</sup>

| residue | protein analysis results | deduced from translated cDNA  |                               |
|---------|--------------------------|-------------------------------|-------------------------------|
|         |                          | Asn-1 to Pro-202 <sup>b</sup> | Asn-1 to Gly-291 <sup>c</sup> |
| Asx     | 27                       | 25                            | 36                            |
| Glx     | 23                       | 21                            | 29                            |
| Ser     | 17                       | 17                            | 23                            |
| Gly     | 14                       | 13                            | 22                            |
| His     | 9                        | 9                             | 9                             |
| Thr     | 10                       | 9                             | 17                            |
| Ala     | 7                        | 6                             | 10                            |
| Arg     | 10                       | 8                             | 14                            |
| Pro     | 8                        | 8                             | 12                            |
| Tyr     | 11                       | 10                            | 12                            |
| Val     | 11                       | 11                            | 16                            |
| Met     | 7                        | 7                             | 10                            |
| Ile     | 20                       | 20                            | 21                            |
| Leu     | 14                       | 13                            | 19                            |
| Phe     | 5                        | 5                             | 6                             |
| Lys     | 10                       | 10                            | 13                            |
| Cys     | 6                        | 7                             | 19                            |
| Trp     | ND                       | 3                             | 3                             |

<sup>a</sup> Values for the protein analysis were averaged from five samples of Ht-e, calculated for a 24-kDa protein and rounded to integer values expressed as residues per molecule. Asx = sum of Asn + Asp; Glx = sum of Gln + Glu; ND = not determined. Asn-1 was determined to be the amino terminus of mature Ht-e protein using amino-terminal protein sequence analysis. <sup>b</sup> Pro-202 is the predicted carboxy-terminal residue of mature Ht-e. <sup>c</sup> Gly-291 is the last residue of the open reading frame.

GenBank database, one significant similarity was found, that of the 2017-nucleotide cDNA sequence for trigamin (Neeper & Jacobson, 1990) from *Trimeresurus gramineus*. These two sequences are 85.3% identical over a 1997-nucleotide overlap (which includes the entire Ht-e cDNA sequence plus gaps inserted for optimal alignment). Comparison of the translated open reading frames of these two clones showed a 70.3% identity over a 482 amino acid overlap. Comparison of the deduced protein sequence to the PIR database, as was expected, confirmed a strong similarity to other snake venom metalloproteinases (51–57%) as seen in Figure 2, and also to the disintegrins (up to 75%) as seen in Figure 3.

No other significant similarities were seen, except to the zinc binding site of other metalloproteinases, the consensus sequence (H-E-X-X-H) which was described by Jongeneel et al. (1989), and the extended motif described by Murphy et al. (1991), (H-E-X-X-H)X-X-G-X-X-H.

The proposed signal peptide region of Ht-e is similar in amino acid sequence to the reported trigamin cDNA signal sequence. The sequence structure (a central region rich in hydrophobic residues and a likely cleavage site 18 amino acids from the amino terminus) is consistent with other signal sequences.

Upon close examination, a sequence, P-K-M-C-G-V, was located between the proposed signal sequence and the metalloproteinase domain (residues 164–169) which resembles a region in the propeptide of matrix metalloproteinases (MMPs). This sequence is thought to be involved in activation of these enzymes (Valee & Auld, 1990; Van Wart & Birkedal-Hansen, 1990) and might serve a similar function in the zymogen form of Ht-e.

## DISCUSSION

A comparison of the amino acid sequence of mature Ht-e with the sequences of other known snake venom metalloproteinases is made in Figure 2. As was expected, the protein sequence of Ht-e is very similar to the other known venom metalloproteinases. When all of the proteinases shown in the

```

1          21          41          61          81
CTCAGATTGGCTTGAAGCGGGGAAGAGATTGCCTGTCTTCCAGCCAAATCCAGCCTCCAAAATGATCCAAAGTTCTCTTGGTGACTATATGCTTAGCAGCT
          M I Q V L L V T I C L A A
101        121        141        161        181
TTTCCTTATCAAGGGAGCTCTATAATCCTGGAATCTGGGAAGCTCAATGATTATGAAGTAATTATCCACGAAAGTCACTGCATTGCCCAAAGGAGCAG
F P Y Q G S S I I L E S G N V N D Y E V I Y P R K V T A L P K G A
201        221        241        261        281
TTCAGCCAAAGTATGAAGACACCATGCAATATGAATGAAAGTGAATGGAGAGCCAGTGGTCTTCATCTGGAATAAAAGGACTTTTTCAAAAGA
V Q P K Y E D T M Q Y E L K V N G E P V V L H L E K N K G L F S K D
301        321        341        361        381
TTACAGTGAGACTCATTATTCCTTTGATGGCAGAAAAATTACAACAAACCCCTTCAGTTGAGGATCAGTCTATTATCATGGACGCATCGAGAATGATGCT
Y S E T H Y S F D G R K I T T N P S V E D H C Y Y H G R I E N D A
401        421        441        461        481
GACTCAACTGCAAGCATCAGTGCATGCAACGGTTTGAAGGACATTTCAAGCTTCAAGGGGAGATGTACCTTATTGAACCCCTTGAAGCTTTCCGACAGTG
D S T A S I S A C N G L K G H F K L Q G E M Y L I E P L K L S D S
501        521        541        561        581
AAGCCCATGCAGTCTTCAAATGAAAAATGTAGAAAAAGAGGATGAGGCCCAAAATGTGTGGGTAACCCAGAATTTGGGAATCATATGAGCCCATCAP
E A H A V F K L K N V E K E D E A P K M C G V T Q N W E S Y E P I K
601        621        641        661        681
AAAGGCCTCTGATTAAATCTTAATCCTGAACATCAAAGATATGTTGAGCTTTTCATTGTGTGGACCATGGAATGTACAGAAAAATATAATGGCGATTCA
K A S D L N L N P E H Q R Y V E L F I V V D H G M Y T K Y N G D S
701        721        741        761        781
GATAAGATAAGACAACGGGTACATCAATGGTCAACATTATGAAGGAGAGTTACACTTATATGTACATCGATATATTACTGGCTGGTATAGAAATTTGGT
D K I R Q R V H Q M V N I M K E S Y T Y M Y I D I L L A G I E I W
801        821        841        861        881
CCAACGGAGATTGATTAACGTGCAGCCAGCATCGCCTAATACTTTGAAGTCAATTTGGAGAATGGAGAGACAGATTGTGTAAGCGCAAAAGTCATGA
S N G D L I N V Q P A S P N T L N S F G E W R E T D L L K R K S H D
901        921        941        961        981
TAATGCTCAGTTACTCAGTCCATTGCCTTCGATGAACAAATATAGGACGGGCTTATATAGGTGGCATATGCGACCCAAAGCGTTCTACAGGAGTTGTC
N A Q L L T S I A F D E Q I I G R A Y I G G I C D P K R S T G V V
1001       1021       1041       1061       1081
CAGGATCATAGCGAAATAAATCTTAGGGTTGCAGTTACAATGACCCATGAGCTGGGTCATAATCTGGGCATTTCATCATGACACAGATTCTCTTCTTGGC
Q D H S E I N L R V A V T M T H E L G H N L G I H H D T D S C S C
1101       1121       1141       1161       1181
GTGGTTACTCATGCATTATGTCTCTGTGATAAGCGATGAACCTTCCAATATTTACAGCGATTGTAGTTACATCCAATGTTGGGAATTTATTATGAATCA
G G Y S C I M S P V I S D E P S K Y F S D C S Y I Q C W E F I M N Q
1201       1221       1241       1261       1281
GAAGCCACAATGCATTCTCAAGAAACCCCTTGAGAACAGATACTGTTCAACTCCAGTTTCTGGAATGAACCTTTGGAGGCGGGAATAGAAATGTGACTGT
K P Q C I L K K P L R T D T V S T P V S G N E L L E A G I E C D C
1301       1321       1341       1361       1381
GGCTCTCTTGAAAAATCCGTGCTGTATGCTACAACCTGTAAAATGAGACCAGGGTCACAGTGTGCAGAAGGACTGTGTGTGACCAGTGCAGATTTATGA
G S L E N P C C Y A T T C K M R P G S Q C A E G L C D Q C R F M
1401       1421       1441       1461       1481
AAAAAGGAACAGTATGCCGGGTATCAATGSGTTGATAGGAATGATGATACCTGCACTGGCCAATCTGCTGACTGTCCAGAAATGGCCTCTATGGCTAAAC
K K G T V C R V S M V D R N D D T C T G Q S A D C P R N G L Y G
1501       1521       1541       1561       1581
AACAATGSAGATGGAAGGTCTGCAGCAACAGGCAGTGTGTGATGTGACTACAGCCTACTAATCAACCTCTGGCTTCTCTCAGATTTGATCTTGGAGAT
1601       1621       1641       1661       1681
CCTTCTTTCAGAAAGTTCAACTTCCCTTTAGTCCAAGAGACCCACCTGCCTGCATCCTACTAGTAAATCACTCTTACCTTTCATACGGAATCTAAATTC
1701       1721       1741       1761       1781
TGCAATATTTCTTCTCCATATTTAATCTGTTTACCTTTTGTGTAATCAAACTTTTCCCAACCAAGCTCTATGGGCATATACAACACCAAGGGCTT
1801       1821       1841       1861       1881
ATTGCTGTCAAGAAAAACAATGGCCATTTTACCGTTTGCCAATTGCAAGCACATTTAATGCAACAAGTTCTGCCTTTTGAGCTGGTATATTGCAAAATC
1901       1921       1941       1961
AATGCTTCTCTTCCAAATTTTCATGCTGCCTTTCCAAGATGTAGCTGCTTCCATCAATAAACTAATATTCTCA

```

FIGURE 1: cDNA sequence of Ht-e clones HO11 and HO12, with the translated open reading frame from the start ATG codon to the termination codon. The signal sequence and mature protein are denoted in boldface.

alignment in Figure 2 are compared, there is an overall shared identity of 25% (50 residues). It has the highest identity with HR2a and H<sub>2</sub>-proteinase (57.3% for each), followed next by a hemorrhagic proteinase from *Lachesis muta muta* (the bushmaster), LHF-II (57.1%). If only the proteinase domain (mature protein) of Ht-e is compared with the high molecular weight proteinase HR1B, a 56.1% identity is obtained, whereas there is a 55.7% identity if both the proteinase and disintegrin-like domains are compared. Surprisingly, of the other snake venom metalloproteinases whose protein sequences have been determined, Ht-e is least similar to a hemorrhagic proteinase, Ht-d, from the same species snake (54.8%).

Another surprising observation is that the mature Ht-e protein sequence is 56.6% identical to the upstream region of the translated open reading frame of the cDNA for the disintegrin trigramin reported by Neeper and Jacobson (1990). The translated open reading frames of these two clones are 70.3% identical overall. The structures of these translated clones parallel each other (signal sequence, zymogen, met-

alloproteinase, disintegrin); indeed, even the 5'- and 3'-untranslated regions of the nucleotide sequence bear striking similarity.

Although there has not been a disintegrin isolated from *C. atrox* venom, it is likely that the processed disintegrin-like peptide (Figure 2, residues 203–295) is a component of the venom. This peptide would lack the RGD sequence found in many of the disintegrins, and its biological activity is not known. This sequence shows a high degree of similarity to the disintegrins, including an identical cysteine pattern (Figure 3). For instance, there is a 75.3% identity between this region of Ht-e and elegantin (55 out of 73 amino acids). The disintegrins are a family of anticoagulant peptides which have a high cysteine content [reviewed by Gould et al. (1990)]. These peptides inhibit platelet aggregation by competing with fibrinogen for binding to the platelet's glycoprotein IIb–IIIa complex, a necessary event for thrombus formation. The RGD sequence found in most disintegrins has been suggested to be important in the binding of integrin proteins to their receptors

|                            | 1   | 10  | 20  | 30  | 40  | 50 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|----------------------------|-----|-----|-----|-----|-----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht-e                       | N   | P   | E   | H   | Q   | R  | Y | V | E | L | F | I | V | V | D | H | G | M | Y | T | K | Y | N | G | D | S | D | K | I | R | Q | R | V | H | Q | M | V | N | I | M | K | E | S | Y | T | Y | M | I | D |   |   |   |
| <i>T. gramineus</i> ORF    | Q   | R   | F   | P   | Q   | R  | Y | I | K | L | G | I | F | V | D | H | G | M | Y | T | K | Y | S | G | N | S | E | R | I | T | K | R | V | H | Q | M | I | N | N | I | M | M | C | R | A | L | N | I | V |   |   |   |
| HR1B                       | <   | E   | Q   | R   | F   | P  | R | R | Y | I | K | L | A | I | V | V | D | H | G | I | V | T | K | H | G | N | L | K | K | I | R | K | W | I | Y | Q | L | V | N | T | I | N | N | I | Y | R | S | L | N | I | L |   |
| HR2a                       | <   | E   | Q   | R   | F   | P  | Q | R | Y | I | E | L | A | I | V | V | D | H | G | M | Y | T | K | Y | S | N | F | K | K | I | R | K | R | V | H | Q | M | V | N | N | I | N | E | M | Y | R | P | L | N | I | A |   |
| H <sub>2</sub> -Proteinase | <   | E   | R   | F   | P   | Q  | R | Y | I | E | L | A | I | V | V | D | H | G | M | Y | K | K | Y | N | Q | N | S | D | K | I | K | V | R | V | H | Q | M | V | N | H | I | N | E | M | Y | R | P | L | N | I | A |   |
| Ht-d                       | <   | E   | Q   | N   | L   | P  | Q | R | Y | I | E | L | V | V | V | A | D | H | R | V | F | M | K | Y | N | S | D | L | N | T | I | R | T | R | V | H | E | I | V | N | F | I | N | G | F | Y | R | S | L | N | I | H |
| LHF-II                     |     | F   | S   | Q   | K   | Y  | I | E | L | V | V | V | A | D | H | G | M | F | T | K | Y | N | G | N | L | N | T | I | R | T | R | V | H | E | I | V | N | T | L | N | G | F | Y | R | S | L | N | I | L |   |   |   |
| HT-2                       | <   | E   | N   | L   | P   | Q  | S | Y | I | E | L | V | V | V | A | D | H | R | M | F | M | K | Y | N | S | D | L | N | T | I | R | T | R | V | H | E | I | V | N | F | I | N | E | F | Y | R | S | L | N | I | R |   |
| Ac-1                       | S   | T   | E   | F   | Q   | R  | Y | M | E | I | V | I | V | V | D | H | S | M | V | K | K | Y | N | G | D | S | X | K | I | K | A | W | V | Y | E | M | I | N | T |   |   |   |   |   |   |   |   |   |   |   |   |   |
| J Protease                 | T   | P   | E   | H   | Q   | R  | Y | I | E | L | F | L | V | V | D | S | G | M | F | M | K | Y | N | G | N | S | D | K | I |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                            | 60  | 70  | 80  | 90  | 100 |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Ht-e                       | I   | L   | L   | A   | G   | I  | E | I | W | S | N | G | D | L | I | N | V | Q | A | S | P | N | T | L | N | S | F | G | E | W | R | E | T | D | L | L | K | R | K | S | H | D | N | A | Q | L | L | T | S |   |   |   |
| <i>T. gramineus</i> ORF    | T   | T   | L   | S   | V   | L  | E | I | W | S | E | K | D | L | I | T | V | Q | A | S | A | P | T | T | L | T | L | F | G | A | W | R | E | T | V | L | L | N | R | T | S | H | D | A | Q | L | L | T | A |   |   |   |
| HR1B                       | V   | A   | L   | V   | Y   | L  | E | I | W | S | K | Q | N | K | I | T | V | Q | S | A | S | N | V | T | L | D | L | F | G | D | W | R | E | S | V | L | L | K | Q | R | S | H | D | C | A | Q | L | L | T | T |   |   |
| HR2a                       | I   | T   | L   | S   | L   | L  | D | V | W | S | E | K | D | L | I | T | M | Q | A | V | A | P | T | T | A | R | L | F | G | D | W | R | E | T | V | L | L | K | Q | K | D | H | D | A | Q | L | L | T | D |   |   |   |
| H <sub>2</sub> -Proteinase | I   | S   | L   | N   | R   | L  | Q | I | W | S | K | K | D | L | I | T | V | Q | S | A | S | N | V | T | L | E | S | F | G | N | W | R | E | T | V | L | L | K | Q | N | N | D | C | A | H | L | L | T | A |   |   |   |
| Ht-d                       | V   | S   | L   | T   | D   | L  | E | I | W | S | N | E | D | Q | I | N | I | Q | S | A | S | S | D | T | L | N | A | F | A | E | W | R | E | T | D | L | L | N | R | K | S | H | D | N | A | Q | L | L | T | A |   |   |
| LHF-II                     | I   | S   | L   | T   | D   | L  | E | I | W | S | N | Q | D | L | I | N | V | Q | S | A | A | N | D | T | L | K | T | F | G | E | W | R | E | R | V | L | L | N | R | I | S | H | D | N | A | Q | L | L | T | A |   |   |
| HT-2                       | V   | S   | L   | T   | D   | L  | E | I | W | S | D | Q | D | F | I | T | V | Q | S | A | K | N | T | L | H | S | F | G | E | W | R | K | S | V | L | L | N | R | K | R | H | D | N | A | Q | L | L | T | A |   |   |   |
|                            | 110 | 120 | 130 | 140 | 150 |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Ht-e                       | I   | A   | F   | D   | E   | Q  | I | G | R | A | Y | I | G | G | I | C | D | P | K | R | S | T | G | V | V | Q | D | H | S | E | I | N | L | R | V | A | V | T | M | T | H | E | L | G | H | N | L | G | I |   |   |   |
| <i>T. gramineus</i> ORF    | T   | I   | F   | N   | G   | N  | V | I | G | R | A | P | V | G | G | M | C | D | P | K | R | S | V | A | I | V | R | D | H | N | A | I | V | F | V | V | A | V | T | M | T | H | E | M | G | H | N | L | G | M |   |   |
| HR1B                       | I   | D   | F   | D   | G   | P  | T | I | G | K | A | Y | T | A | S | M | C | D | P | K | R | S | V | G | I | V | Q | D | Y | S | P | I | N | L | V | V | A | V | I | M | T | H | E | M | G | H | N | L | G | I |   |   |
| HR2a                       | I   | N   | F   | T   | G   | N  | T | I | G | W | A | Y | M | G | G | M | C | N | A | K | N | S | V | G | I | V | K | D | H | S | S | N | V | F | M | V | A | V | T | M | T | H | E | I | G | H | N | L | G | M |   |   |
| H <sub>2</sub> -Proteinase | T   | N   | L   | N   | D   | N  | T | I | G | L | A | Y | K | K | G | M | C | N | P | K | L | S | V | G | L | V | Q | D | Y | S | P | N | V | F | M | V | A | V | T | M | T | H | E | L | G | H | N | L | G | M |   |   |
| Ht-d                       | I   | E   | L   | D   | E   | E  | T | L | G | L | A | P | L | G | T | M | C | D | P | K | L | S | I | G | I | V | Q | D | H | S | P | I | N | L | L | M | G | V | T | M | A | H | E | L | G | H | N | L | G | M |   |   |
| LHF-II                     | I   | D   | L   | A   | D   | N  | T | I | G | I | A | Y | T | G | G | M | C | Y | P | K | N | S | V | G | I | V | Q | D | H | S | P | K | T | L | I | A | V | T | M | A | H | E | L | G | H | N | L | G | M |   |   |   |
| HT-2                       | I   | V   | L   | D   | D   | Y  | T | L | G | L | A | Y | L | N | S | M | C | H | P | R | N | S | V | G | L | I | Q | D | H | S | P | I | N | L | L | M | G | V | T | M | A | H | E | L | G | H | N | L | G | M |   |   |
|                            | 160 | 170 | 180 | 190 | 200 |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Ht-e                       | H   | H   | D   | T   | D   | S  | C | S | C | G | G | Y | S | C | I | M | S | P | V | I | S | D | E | P | S | K | Y | F | S | D | C | S | Y | I | Q | C | W | E | F | I | M | N | Q | K | P | Q | C | I | L | K |   |   |
| <i>T. gramineus</i> ORF    | H   | H   | D   | E   | D   | K  | C | N | C | N | T | - | - | C | I | M | S | K | V | L | S | R | Q | P | S | K | Y | F | S | E | C | S | K | D | Y | Y | Q | T | F | L | T | N | H | N | P | Q | C | I | L | N |   |   |
| HR1B                       | P   | H   | D   | G   | N   | S  | C | T | C | G | G | F | P | C | I | M | S | P | M | I | S | D | P | P | S | E | L | F | S | N | C | S | K | A | Y | Y | Q | T | F | L | T | D | H | K | P | Q | C | I | L | N |   |   |
| HR2a                       | E   | H   | D   | D   | K   | D  | K | C | K | C | E | A | - | C | I | M | S | A | V | I | S | D | K | P | S | K | L | F | S | D | C | S | K | D | Y | Y | Q | T | F | L | T | N | S | K | P | Q | C | I | I | N |   |   |
| H <sub>2</sub> -Proteinase | E   | H   | D   | D   | K   | D  | K | C | K | C | E | A | - | C | I | M | S | D | V | I | S | D | K | P | S | K | L | F | S | D | C | S | K | N | D | Y | Y | Q | T | F | L | T | K | Y | N | P | Q | C | I | L | N |   |
| Ht-d                       | E   | H   | D   | G   | K   | D  | C | L | R | G | A | S | L | C | I | M | R | P | G | L | T | K | G | R | S | Y | E | F | S | D | S | M | H | Y | Y | E | R | F | L | K | Q | Y | K | P | Q | C | I | L | N |   |   |   |
| LHF-II                     | K   | H   | D   | E   | N   | H  | C | H | C | S | A | S | F | C | I | M | P | P | S | I | S | E | G | P | S | Y | E | F | S | D | C | S | K | D | Y | Y | Q | M | F | L | T | K | R | K | P | Q | C | I | L | N |   |   |
| HT-2                       | E   | H   | D   | G   | K   | D  | C | L | R | G | A | S | L | C | I | M | R | P | G | L | T | P | G | R | S | Y | E | F | S | D | A | S | M | R | Y | Y | Q | K | F | L | D | Q | Y | K | P | Q | C | I | L | N |   |   |
|                            | 210 | 220 | 230 | 240 | 250 |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Ht-e                       | K   | P   | L   | R   | T   | D  | T | V | S | T | P | V | S | G | N | E | L | L | E | A | G | I | E | C | D | C | G | S | L | E | N | - | - | - | P | C | C | Y | A | T | T | C | K | M | R | P | G | S | Q | C |   |   |
| <i>T. gramineus</i> ORF    | A   | P   | L   | R   | T   | D  | T | V | S | T | P | V | S | G | N | E | L | L | E | A | G | E | D | C | D | C | G | S | P | A | N | - | - | - | P | C | C | D | A | A | T | C | K | L | I | P | G | A | Q | C |   |   |
| HR1B                       | A   | P   | S   | K   | T   | D  | I | V | S | P | P | V | C | G | N | E | L | L | E | A | G | E | E | C | D | C | G | S | P | E | N | C | Q | Y | Q | C | D | A | A | S | C | K | L | H | S | V | W | K |   |   |   |   |
| HR2a                       | A   | P   |     |     |     |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| H <sub>2</sub> -Proteinase | A   | P   |     |     |     |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Ht-d                       | K   | P   |     |     |     |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LHF-II                     | K   | P   |     |     |     |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| HT-2                       | K   | P   |     |     |     |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                            | 260 | 270 | 280 | 290 | 300 |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Ht-e                       | A   | E   | G   | L   | C   | C  | D | Q | C | R | F | M | K | K | G | T | V | C | R | V | S | M | V | D | R | N | - | D | D | T | C | T | G | Q | S | A | D | C | P | R | N | G | L | Y | G |   |   |   |   |   |   |   |
| <i>T. gramineus</i> ORF    | G   | E   | G   | L   | C   | C  | D | Q | C | S | F | I | E | E | G | T | V | C | R | I | A | R | G | D | D | L | D | D | Y | C | N | G | R | S | A | G | C | P | R | N | P | F | H | A |   |   |   |   |   |   |   |   |
| HR1B                       | E   | S   | G   | E   | C   | C  | D | Q | C | R | F | R | T | A | G | T | E | C | R | A | E | S | E | C | D | I | P | E | S | C | T | G | Q | S | A | D | C | P | T | D | R | F | H | R | N | G | Q | P | C |   |   |   |
|                            | 310 | 320 | 330 | 340 | 350 |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| HR1B                       | L   | Y   | N   | H   | G   | Y  | C | Y | N | G | K | C | P | I | M | F | Y | Q | C | Y | F | L | F | G | S | N | A | T | V | A | E |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

```

E      LRTDVTSTPVSIGNELLEAGIECDGSLN---PCCYATTCMKRPGSQCAEGLCCDQCRFMKKGTVCRVSMVD-RNDDTCTGQSADCPRNGLYG
B      SKTDIVSPFPVCGNELLEAGEECDGSPENCQYQCCDAASCKLHSHWVKESGECCDQCRFRTAGTECRAAESECDIPESCTGQSADCPTRDFHRNG
1      VSPFPVCGNKILEQGEDCDGSPANCQDQCCNAATCKLTGSGQCNHGECCDQCKFKKARTVCRIARGD-WNDDYCTGKSSDCPNWH
2      EAGEECDGSPEN---PCCDAATCKLRPGAQCADGLCCDQCRFMKKGTVCRVAKGD-WNDDTCTGQSADCPRNGLYG
3      EAGEDCDGSPAN---PCCDAATCKLLPGAQCGEGLCCDQCSFMKKGTCRRARGD-DLDDYCNIGISAGCPRNPLHA
4      EAGEECDGSPEN---PCCDAATCKLRPGAQCADGLCCDQCRFKKRTICRRARGD-NPDDRCTGQSADCPRNGLYS
5      EAGEECDGSPEN---PCCDAATCKLRPGAQCAEGLCCDQCKFMKEGTVCR-ARGD-DVNDYCNIGISAGCPRNPFH
6      EAGEDCDGSPAN---PCCDAATCKLIPGAQCGEGLCCDQCSFIEEGTVCRARGD-DLDDYCNIGISAGCPRNPFH
7      EAGEECDGTPEN---PCCDAATCKLRPGAQCAEGLCCDQCRFKGAGKICRRARGD-NPDDRCTGQSADCPNRNF
8      GEECDGSPSN---PCCDAATCKLRPGAQCADGLCCDQCRFKKKTGICRIARGD-FPDDRCTGLSNDPCRWNDL
9      GKECDGSSPEN---PCCDAATCKLRPGAQCGEGLCCDQCKFSRAGKICRIPRGD-MPDDRCTGQSADCPRYH
10     ZRQEEFCATGPCRRCKFKRAGKVCRAVARGD-WNDDYCTGKSCDCPRNPWG
11     ECESGPCRRNCKFLKEGTICKRARGD-DMDDYCNIGTCDPCRNPHKGPAT

```

FIGURE 3: Comparison of the Ht-e disintegrin-like domain, the HR1B disintegrin-like domain, and the disintegrins. The cysteine residues are shown in boldface to emphasize their conserved positions. E, Ht-e disintegrin domain from *Crotalus atrox* (Western diamondback rattlesnake). B, HR1B disintegrin domain from *Trimeresurus flavoviridis* (Habu) (Takeya et al., 1990a). 1, Bitistatin from *Bitis arietans* (Puff adder) (Shebuski et al., 1989). 2, Barbourin from *Sistrurus m. barbouri* (Southeastern pigmy rattlesnake) (Scarborough et al., 1991). 3, Albolabrin from *Trimeresurus albolabris* (white-lipped tree viper) (Williams et al., 1990). 4, Elegantin from *Trimeresurus elegans* (Sakishima habu) (Williams et al., 1990). 5, Applaggin from *Agkistrodon piscivorus piscivorus* (Eastern cottonmouth) (Chao et al., 1989). 6, Trigramin from *Trimeresurus gramineus* (Indian green tree viper) (Huang et al., 1989). 7, Batroxostatin from *Bothrops atrox* (Fer-de-lance) (Rucinski et al., 1990). 8, Flavoridin from *Trimeresurus flavoviridis* (Habu) (Musial et al., 1990). 9, Kistrin (Rhodostomin) from *Agkistrodon rhodostoma* (Malayan pit viper) (Dennis et al., 1989). 10, Eristicophin from *Eristicophis macmahoni* (Asian sand viper) (Scarborough et al., 1991). 11, Echistatin from *Echis carinatus* (Saw-scaled viper) (Gan et al., 1988).

as part of its precursor protein. Thus, it appears that the disintegrin family of polypeptides and at least some of the venom metalloproteinases are intricately linked, sharing common precursor proteins.

This observation was also suggested by Takeya et al. (1991), on the basis of autolysis studies of HR1B protein. HR1B (from *T. flavoviridis*) is a member of the family of large venom metalloproteinases, as is Ht-a (from *C. atrox*). Both are greater than 60 kDa. Takeya et al. stated that HR1B protein, when incubated in the absence of calcium, could be autolyzed to yield a fragment of approximately 30 kDa which contained the disintegrin-like domain. We have seen similar results with Ht-a: when incubated for 1 h at 37 °C in the presence of 0.1% SDS, we get a similar autolysis product (data not shown), the amino-terminal sequence of which corresponds with the disintegrin-like domain of Ht-a.

The difference between the large and small venom metalloproteinases is 2-fold: the large proteinases retain their disintegrin-like domains, while the smaller proteinases do not, and the large proteinases have an extra carboxy-terminal domain not found on the small proteinases. The autolysis studies on Ht-a and HR1B both rely on slight perturbation of protein structure in order to promote cleavage at the proteinase/disintegrin junction. It would appear that the presence of the third carboxy-terminal domain on the large proteinases stabilizes and protects this cleavage site. It is possible that the collection and purification of the venom cause the disintegrin domain to be autolyzed from the precursor (Takeya et al., 1991). It is also possible that the presence of the carboxy-terminal domain of the large proteinases helps to stabilize or mask the processing site between the proteinase and disintegrin domains of the large proteinases, while the disintegrin-like domain is normally autolytically removed from the small proteinases. cDNA expression studies could be helpful in determining the stabilities of these precursor proteins.

In their discussion of the amino acid sequence of the hemorrhagic metalloproteinase HT-2, Takeya et al. (1990b) denote six amino acids (Gln-68, His-92, Gln-96, Ile-101, Tyr-186, and Lys-194 using the numbering system for mature Ht-e) which are conserved in all of the hemorrhagic proteinases they compared but not in the nonhemorrhagic H<sub>2</sub>-proteinase and suggested that these residues might be important in hemorrhage production. Of these, Ht-e also has all but the Tyr-186, which is substituted by a glutamine. Our preliminary sequence

data from the hemorrhagic proteinase Ht-a indicate that the Gln-68 and Lys-194 residues are also not consistently conserved (Hite, unpublished data). This leaves the region containing His-92, Gln-96, and Ile-101 as the region possibly crucial to the ability of these proteinases to generate hemorrhage. The clustering of these 3 residues within 10 amino acids of each other suggests that they might be important in the recognition of and interaction with relevant substrates, the cleavage of which results in hemorrhage.

The translated open reading frame of the Ht-e clone illustrates several features of the snake venom metalloproteinase family of enzymes which were not previously known. As seen in Figure 1, the amino and carboxy termini of the mature protein reflect posttranslational processing events, in that the termini do not correspond to either the start or the end of translation. The first residue of the mature protein is known from amino-terminal sequence analysis. From the molecular weight of the mature protein and the sequencing and amino acid analysis of the carboxy-terminal cyanogen bromide peptide, the carboxy terminus of the mature protein is Pro-202. The amino acid analysis results presented in Table I match the predicted composition from the translated cDNA sequence within the standard 10% error.

As would be expected for a secreted protein, Ht-e has an amino-terminal signal sequence. On the basis of the prediction method of von Heijne (1983), the signal sequence is cleaved between Gly-18 and Ser-19. The resultant signal peptide is 18 amino acids long and conforms to the (-3, -1)-rule. The length of this signal sequence is within the normal limits of eukaryotic signal peptides (15-35 amino acids). It is very similar to the signal sequence found in the trigramin clone (88.9% identical).

Amino acids 19-187 comprise the 169-residue propeptide. The fact that Ht-e is synthesized as a zymogen is not surprising; however, no previous evidence has substantiated this possibility. Given this finding, it is likely that the other venom metalloproteinases are synthesized in a similar fashion and are activated by proteolysis of the propeptide from the mature protein. Analysis of the propeptide allows one to hypothesize that the snake venom metalloproteinases might be activated in a similar manner to the matrix metalloproteinase (MMP) family of metalloproteinases. A "Velcro" mechanism of activation (Valee & Auld, 1990) and a "cysteine switch" mechanism (Van Wart & Birkedal-Hansen, 1990) were re-



|                             |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                    |     |   |   |   |   |   |   |   |   |   |  |  |  |  |
|-----------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--------------------|-----|---|---|---|---|---|---|---|---|---|--|--|--|--|
| Neutral Proteinases         |   |   |   |   |   |   |   |   |   |   |   |   |   |   | Metalloproteinases |     |   |   |   |   |   |   |   |   |   |  |  |  |  |
| Thermolysin                 | V | V | A | H | E | L | T | H | A | V | T | D | Y | T | A                  | G   | A | I | N | E | A | I | S | D |   |  |  |  |  |
| <i>B. cereus</i> Proteinase | V | I | G | H | E | L | T | H | A | V | T | E | N | S | S                  | G   | A | L | N | E | A | I | S | D |   |  |  |  |  |
|                             |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                    |     |   |   |   |   |   |   |   |   |   |  |  |  |  |
| <i>Erwinia</i> Proteinase B | S | F | T | H | E | I | G | H | A | L | G | L | S | H | P                  | 174 | A | A | Y | A | E | D | S | R | Q |  |  |  |  |
| <i>Serratia</i> Proteinase  | T | F | T | H | E | I | G | H | A | L | G | L | S | H | P                  | 178 | V | T | Y | A | E | D | T | R | Q |  |  |  |  |
|                             |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                    |     |   |   |   |   |   |   |   |   |   |  |  |  |  |
| Matrix Metalloproteinases   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | Metalloproteinases |     |   |   |   |   |   |   |   |   |   |  |  |  |  |
| Human Collagenase           | V | A | A | H | E | L | G | H | S | L | G | L | S | H | S                  | 238 | P | S | Y | T | F | S | G | D | V |  |  |  |  |
| Rat Transin                 | V | A | A | H | E | L | G | H | S | L | G | L | F | H | S                  | 242 | P | V | Y | K | S | S | T | D | L |  |  |  |  |
|                             |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                    |     |   |   |   |   |   |   |   |   |   |  |  |  |  |
| Venom Metalloproteinases    |   |   |   |   |   |   |   |   |   |   |   |   |   |   | Metalloproteinases |     |   |   |   |   |   |   |   |   |   |  |  |  |  |
| Ht-e                        | T | M | T | H | E | L | G | H | N | L | G | I | H | H | D                  | 162 | Y | S | C | I | M | S | P | V | I |  |  |  |  |

In summary, the cDNA sequence for a snake venom hemorrhagic metalloproteinase was determined. The translated protein sequence of the mature Ht-e protein shows similarity to other snake venom metalloproteinases. The continuation of the open reading frame beyond the end of the observed carboxy terminus of the mature protein suggests a carboxy-terminal domain, which is processed away from the proteinase. The high degree of similarity between this clone and a cDNA clone for the disintegrin trigramin strengthens this hypothesis. Furthermore, it helps to bridge our understanding of the interrelatedness of the large and small molecular weight venom metalloproteinases to one another and to the disintegrin family of peptides. It also predicts the existence of this disintegrin-like peptide in *C. atrox* venom, the function of which remains to be established. Efforts to isolate this peptide are underway. A cluster of amino acids from residues 92 to 101 was noted to be conserved in the hemorrhagic metalloproteinases but not

conserved in the only nonhemorrhagic snake venom metalloproteinase whose protein sequence has been determined. This could be evidence that this region of the proteinase is crucial to its ability to hydrolyze substrates in a manner which results in hemorrhage. Since this region is now defined to a single, 10 amino acid sequence and a cDNA clone is available, it should be possible to perform site-directed mutagenesis studies in order to test this hypothesis. Evidence exists for a signal sequence and prosequence for Ht-e. The propeptide shows some similarity to the propeptide of MMPs and may have a similar mechanism of activation. This region of the cDNA clone is also a good candidate for site-directed mutagenesis studies. The zinc binding region of Ht-e more closely resembles that of the MMPs than of the thermolysin family of zinc metalloproteinases, which is not surprising since the substrates of Ht-e are more similar to those of the MMPs. More studies are needed to better define the structure and function of this family of metalloproteinases.

#### ACKNOWLEDGMENTS

We thank Ms. Lolita L. Bland for her assistance in DNA sequencing.

**Registry No.** DNA (*Crotalus atrox* atrolysin e mRNA-complementary and flanking region), 141436-80-8; DNA (*Crotalus atrox* atrolysin e mRNA-complementary), 141436-79-5; preproatrolysin e (*Crotalus atrox* reduced), 141436-82-0; atrolysin e (*Crotalus atrox* reduced), 141436-81-9; atrolysine, 81669-70-7; proatrolysin e (*Crotalus atrox* reduced), 141436-83-1.

#### REFERENCES

- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K., Eds. (1987) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.
- Baramova, E. N., Shannon, J. D., Bjarnason, J. B., & Fox, J. W. (1989) *Arch. Biochem. Biophys.* **275**, 63-71.
- Baramova, E. N., Shannon, J. D., Bjarnason, J. B., & Fox, J. W. (1990a) *Matrix* **10**, 91-97.
- Baramova, E. N., Shannon, J. D., Bjarnason, J. B., Gonias, S. L., & Fox, J. W. (1990b) *Biochemistry* **29**, 1069-1074.
- Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr.* **336**, 93-104.
- Bilofsky, H. S., Burks, C., Fickett, J. W., Goad, W. B., Lewitter, F. I., Rindone, W. P., Swindell, C. D., & Tung, C.-S. (1986) *Nucleic Acids Res.* **14**, 1-4.
- Bjarnason, J. B., & Tu, A. T. (1978) *Biochemistry* **17**, 3395-3404.
- Bjarnason, J. B., & Fox, J. W. (1983) *Biochemistry* **22**, 3770-3778.
- Bjarnason, J. B., & Fox, J. W. (1987) *Biochim. Biophys. Acta* **911**, 356-363.
- Bjarnason, J. B., & Fox, J. W. (1989) *J. Toxicol. Toxin Rev.* **7**, 121-209.
- Bjarnason, J. B., Hamilton, D., & Fox, J. W. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 121-129.
- Chao, B. H., Jakubowski, J. A., Savage, B., Ping Chow, E., Marzec, U. M., Harker, L. A., & Maraganore, J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8050-8054.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry* **18**, 5294-5299.
- Deleplaire, P., & Wandersman, C. (1989) *J. Biol. Chem.* **264**, 9083-9089.
- Dennis, M. S., Henzel, W. J., Pitti, R. M., Lipari, M. T., Napier, M. A., Deisher, T. A., Bunting, S., & Lazarus, R. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2471-2475.
- Fox, J. W., Campbell, R., Beggerly, L., & Bjarnason, J. B. (1986) *Eur. J. Biochem.* **156**, 65-72.
- Friedman, M., Krull, L. H., & Cavins, J. F. (1970) *J. Biol. Chem.* **245**, 3868-3871.
- Gan, Z.-R., Gould, R. J., Jacobs, J. W., Friedman, P. A., & Polokoff, M. A. (1988) *J. Biol. Chem.* **263**, 19827-19832.
- Gould, R. J., Polokoff, M. A., Friedman, P. A., Huang, T.-F., Holt, J. C., Cook, J. J., & Niewiarowski, S. (1990) *Proc. Soc. Exp. Biol. Med.* **195**, 168-171.
- Gross, E. (1967) *Methods Enzymol.* **11**, 238-255.
- Hirs, C. H. W. (1967) *Methods Enzymol.* **11**, 199-203.
- Huang, T.-F., Holt, J. C., Kirby, E. P., & Niewiarowski, S. (1989) *Biochemistry* **28**, 661-666.
- Jongeneel, C. V., Bouvier, J., & Bairoch, A. (1989) *FEBS Lett.* **242**, 211-214.
- Lawrence, C. B., Shalom, T., & Honda, S. (1989) *SAM: A Software Package for Sequence Assembly Management for UNIX Systems*, Molecular Biology Information Resource, Baylor College of Medicine, Houston, TX.
- Matthews, B. W., Weaver, L. H., & Kester, W. R. (1974) *J. Biol. Chem.* **249**, 8030-8044.
- Miyata, T., Takeya, H., Ozeki, Y., Arakawa, M., Tokunaga, F., Iwanaga, S., & Omori-Satoh, T. (1989) *J. Biochem.* **105**, 847-853.
- Murphy, G. J. P., Murphy, G., & Reynolds, J. J. (1991) *FEBS Lett.* **289**, 4-7.
- Musial, J., Niewiarowski, S., Rucinski, B., Stewart, G. J., Cook, J. J., Williams, J. A., & Edmunds, L. H., Jr. (1990) *Circulation* **82**, 261-273.
- Nagase, H., Englund, J. J., Suzuki, K., & Salvesen, G. (1990) *Biochemistry* **29**, 5783-5789.
- Nakagawa, H., Miller, R. A., Tu, A. T., Nikai, T., & Sugi-hara, H. (1989) *Biochem. Arch.* **5**, 243-247.
- Nakahama, K., Yoshimura, K., Marumoto, R., Kikuchi, M., Lee, I. S., Hase, T., & Matsubara, H. (1986) *Nucleic Acids Res.* **14**, 5843-5855.
- Neeper, M. P., & Jacobson, M. A. (1990) *Nucleic Acids Res.* **18**, 4255.
- Ownby, C. L., Bjarnason, J. B., & Tu, A. T. (1978) *Am. J. Pathol.* **93**, 201-218.
- Pearson, W. R., & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444-2448.
- Rucinski, B., Niewiarowski, S., Holt, J. C., Soszka, T., & Knudsen, K. A. (1990) *Biochim. Biophys. Acta* **1054**, 257-262.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1982) in *Molecular Cloning: A Laboratory Manual*, pp 466-467, Cold Spring Harbor Laboratory Press, Salem, MA.
- Sanchez, E. F., Diniz, C. R., & Richardson, M. (1991) *FEBS Lett.* **282**, 178-182.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.
- Scarborough, R. M., Rose, J. W., Hsu, M. A., Phillips, D. R., Fried, V. A., Campbell, A. M., Nannizzi, L., & Charo, I. F. (1991) *J. Biol. Chem.* **266**(15), 9359-9362.
- Shannon, J. D., Baramova, E. N., Bjarnason, J. B., & Fox, J. W. (1989) *J. Biol. Chem.* **264**, 11575-11583.
- Shebuski, R. J., Ramjit, D. R., Bensen, G. H., & Polokoff, M. A. (1989) *J. Biol. Chem.* **264**, 21550-21556.
- Sidler, W., Niederer, E., Suter, F., & Zuber, H. (1986) *Biol. Chem. Hoppe-Seyler* **367**, 643-657.
- Takeya, H., Arakawa, M., Iwanaga, S., & Omori-Satoh, T. (1989) *J. Biochem.* **106**, 151-157.
- Takeya, H., Oda, K., Miyata, T., Omori-Satoh, T., & Iwanaga, S. (1990a) *J. Biol. Chem.* **265**, 16068-16073.



- Takeya, H., Onikura, A., Nikai, T., Sugihara, H., & Iwanaga, S. (1990b) *J. Biochem.* 108, 711-719.
- Takeya, H., Kubo, J., Omori-Satoh, T., & Iwanaga, S. (1991) *Thromb. Haemostasis* 65, 842.
- Tanizaki, M. M., Zingali, R. B., Kawazaki, H., Imajoh, S., Yamazaki, S., & Suzuki, K. (1989) *Toxicon* 27, 747-755.
- Titani, K., Hermodson, M. A., Ericsson, L. H., Walsh, K. A., & Neurath, H. (1972) *Nature (London), New Biol.* 238, 35-37.
- Valee, B. L., & Auld, D. S. (1990) *Biochemistry* 29, 5647-5659.
- Valee, B. L., & Auld, D. S. (1991) in *Methods in Protein Sequence Analysis* (Hörmvall, H., Höög, J.-O., & Gustavsson, A.-M., Eds.) pp 363-372, Birkhäuser Verlag, Basel.
- Van Wart, H. E., & Birkedal-Hansen, H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 5578-5582.
- von Heijne, G. (1983) *Eur. J. Biochem.* 133, 17-21.
- Whitham, S. E., Murphy, G., Angel, P., Rahmsdorf, H. J., Smith, B. J., Lyons, A., Harris, T. J. R., Reynolds, J. J., Herrlich, P., & Docherty, A. J. P. (1986) *Biochem. J.* 240, 913-916.
- Williams, J., Rucinski, B., Holt, J. C., & Niewiarowski, S. (1990) *Biochim. Biophys. Acta* 1039, 81-89.
- Woessner, J. F. (1991) *FASEB J.* 5, 2145-2154.
- Yamanaka, M., Matsuda, M., Isobe, J., Ohsaka, A., Takahashi, T., & Omori-Satoh, T. (1974) *Platelets, Thrombosis, and Inhibitors* (Didisheim, P., Ed.) pp 335-344, F. K. Schattauer-Verlag, New York.

## Identification of Five Sites of Acetylation in Alfalfa Histone H4<sup>†</sup>

Jakob H. Waterborg

Division of Cell Biology and Biophysics, School of Basic Life Sciences, University of Missouri—Kansas City, Kansas City, Missouri 64110-2499

Received January 16, 1992; Revised Manuscript Received April 7, 1992

**ABSTRACT:** Radioactive acetylation in vivo of plant histone H4 of alfalfa, *Arabidopsis*, tobacco, and carrot revealed five distinct forms of radioactive, acetylated histone. In histone H4 of eukaryotes ranging from fungi to man, acetylation is restricted to four lysines (residues 5, 8, 12, and 16) possibly caused by a quantitative methylation of lysine-20. Chemical and proteolytic fragmentation of the amino terminally blocked alfalfa H4 protein, dynamically acetylated by radioactive acetate in vivo, allowed protein sequencing and identification of selected peptides. Peptide identification was facilitated by analyzing fully characterized calf histone H4 in parallel. Acetylation in vivo of alfalfa histone H4 was restricted to the lysines in the amino-terminal domain of the protein, residues 1-23. Lysine-20 was shown to be free of methylation, as in pea histone H4. This apparently makes lysine-20 accessible as a novel target for histone acetylation. The in vivo pattern of lysine acetylation ( $16 > 12 > 8 \geq 5 = 20$ ) revealed a preference for lysines-16 and -12 without an apparent strict sequential specificity of acetylation.

Core histone acetylation has been suggested to allow transient displacement of histones from DNA by polymerases during gene transcription and DNA duplication and permanent displacement by protamines during spermatogenesis [for recent reviews, see Van Holde (1989), Grunstein (1990), and Csordas (1990)]. Multiacetylated core histones H3 and H4 appear the functional forms for these processes (Van Holde, 1989; Delcuve & Davie, 1989; Boffa et al., 1990; Walker et al., 1990). In both histones, the amino acid sequence of the histone amino-terminal protein domain with the multiple lysine residues which can be acetylated in vivo is highly conserved (Matthews & Waterborg, 1985). Experiments in yeast (Megee et al., 1990; Grunstein, 1990; Durrin et al., 1991), *Physarum* (Pesis & Matthews, 1986), *Tetrahymena* (Chicoine et al., 1986), *Drosophila* (Munks et al., 1991), and animal cells (Couppez et al., 1987; Turner & Fellows, 1989; Thorne et al., 1990) have revealed functional differences between lysine residues, a nonrandom order of lysine acetylation, and a nonrandom distribution of steady-state acetylation patterns in histone H4. Nonrandom distributions of lysine acetylation

and methylation have recently been described for two distinct histone H3 variants in alfalfa (Waterborg, 1990). This paper presents the pattern of acetylation of histone H4 in alfalfa.

The highest level of histone H4 acetylation observed in animals and lower eukaryotes is four, restricted to lysines-5, -8, -12, and -16 in the amino-terminal domain of histone H4 (Matthews & Waterborg, 1985; Van Holde, 1989; Thorne et al., 1990). Methylation of lysine-20 in lower eukaryotes like *Physarum* (Waterborg et al., 1983) and in animals ranging from sea urchins and fishes to birds and mammals (Van Holde, 1989; Duerre & Buttz, 1990) precludes modification by acetylation. In contrast, when alfalfa cells were labeled in vivo with acetate, five radioactive, charge-modified forms of histone H4 were observed when histone H4 was fractionated in acid/urea/Triton (AUT)<sup>1</sup> gels (Waterborg et al., 1989, 1990). In vivo treatment of alfalfa cells with [<sup>32</sup>P]phosphate failed to label histone H4 (Waterborg et al., 1989), suggesting that phosphorylation (Ruiz-Carrillo et al., 1975) is not involved. This paper describes experiments that identify lysine-20 of

<sup>†</sup> This work was supported by an NIH Biological Research Support Grant to the School of Basic Life Sciences, University of Missouri—Kansas City, and by National Science Foundation Grants DCB-8896292 and DCB-9118999.

<sup>1</sup> AUT, acid/urea/Triton; BAP, 6-benzylaminopurine; BSA, bovine serum albumin fraction V; PTH, phenylthiohydantoin; TEMED, *N,N,N',N'*-tetramethylethylenediamine; TFA, trifluoroacetic acid; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; 2,4D, 2,4-dichlorophenoxyacetic acid.