cDNA Cloning and Deduced Amino Acid Sequence of Prothrombin Activator (Ecarin) from Kenyan *Echis carinatus* Venom[†,‡]

Shinji Nishida,[§] Taizo Fujita,[§] Noriatsu Kohno,[∥] Hideko Atoda,[∥] Takashi Morita,[∥] Hiroyuki Takeya,[§,⊥] Isao Kido,[§] Mark J. I. Paine,[▽] Shun-ichiro Kawabata,[§] and Sadaaki Iwanaga*[,§]

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan, Department of Biochemistry, Meiji College of Pharmacy, Tanashi, Tokyo 188, Japan, and Wolfson Unit of Molecular Genetics, Liverpool School of Tropical Medicine, Liverpool, Pembrock Place, Liverpool L3 5QA, United Kingdom

Received July 26, 1994; Revised Manuscript Received November 21, 1994[⊗]

ABSTRACT: The complete amino acid sequence of ecarin is deduced from the nucleotide sequence of a cDNA clone isolated by screening a venomous gland cDNA library of Kenyan *Echis carinatus*. The cDNA sequence with 2379 base pairs encodes an open reading frame of 616 amino acids with a remarkable sequence homology to the putative precursor protein of trigramin from *Trimeresurus gramineus* venom (61% identity) and a large hemorrhagin, jararhagin, from the pit viper *Bothrops jararaca* venom (62% identity). Thus, ecarin, as well as jararhagin and trigramin, is translated as a precursor protein, which may be processed posttranslationally. The ecarin proprotein has a "cysteine switch" motif (-Pro-Lys-Met-Cys-Gly-Val-) similar to that involved in the activation of matrix metalloproteinase zymogens. The processed mature protein consists of 426 amino acid residues (residues 191−616), showing the strongest sequence similarity with that of Russell's viper venom factor X activator (RVV-X) heavy chain (64% identity). Like RVV-X heavy chain, ecarin contains metalloproteinase, disintegrin, and cysteine-rich domains. The metalloproteinase domain has a typical zinc-chelating sequence (-His-Glu-Xaa-Xaa-His-Xaa-Xaa-Gly-Xaa-Xaa-His-), as found in crayfish astacin. In the disintegrin domain of ecarin, the Arg-Gly-Asp sequence is replaced by Arg-Asp-Asp, as found in the disintegrin domains of RVV-X heavy chain (Arg-Asp-Glu) and a guinea pig sperm fusion protein, PH-30β (Thr-Asp-Glu). These findings show that while there are structural and evolutionary relationships among these proteins, each has a unique functional activity.

Snake venom contains several metalloproteinases with various pathological effects on blood coagulation, fibrinolysis, and complement systems (Ouyang *et al.*, 1992; Tu, 1988). In addition, the venom contains hemorrhagic metalloendopeptidases which disrupt collagenous basement membrane of the vascular endothelium, resulting in bleeding (Ohsaka, 1979). The complete amino acid sequences for several venom metalloproteinases have been determined including hemorrhagic factors and procoagulants from the genera *Crotalus*, *Trimeresurus*, and *Lachesis* of the Crotalidae family (Iwanaga & Takeya, 1993). It would thus appear that these venom procoagulants and hemorrhagic factors are newly identified members of the metalloproteinase subfamily and are mosaic proteins consisting of an amino-terminal metalloproteinase domain, a large noncatalytic middle segment with a disintegrin domain, and a unique carboxyl-terminal cysteine-rich domain (Takeya *et al.*, 1990a; Paine *et al.*, 1992; Hite *et al.*, 1992).

In previous papers, we reported the primary structures of two types of procoagulants, factor V activator (RVV-V)[1] and factor X activator (RVV-X), isolated from Russell's viper venom, which catalyze respectively the activation of factor V to factor Va and factor X to factor Xa in mammalian plasma (Tokunaga *et al.*, 1988; Takeya *et al.*, 1992). The prothrombin activator ecarin, which is contained in the venom of *Echis carinatus,* is also a procoagulant (Kornalik, 1963). Ecarin catalyzes specifically the conversion of prothrombin to α-thrombin through formation of a meizothrombin intermediate (Morita *et al.*, 1976). It is also able to generate α-thrombin from abnormal prothrombin produced by treating mammals with vitamin K antagonists (Nelsestuen & Suttie, 1972). Thus, ecarin is being used for selective determination to quantitate acarboxyprothrombin in patients with hepatic diseases or those being treated with vitamin K antagonists (Rosing & Tans, 1992).

Ecarin has been characterized as a single-chain glycoprotein with a molecular weight of 56 000 that has metalloproteinase activity (Morita *et al.*, 1976; Kornalik & Blombäck, 1975). Our interest in the structure of ecarin arises from its

[†] This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

[‡] The nucleotide sequences reported in this paper have been submitted to DNA Data Bank of Japan, European Molecular Biology Laboratory, and National Center for Biotechnology Information under accession number D32212.

* Author to whom correspondence should be addressed: Department of Biology, Faculty of Science, Kyushu University 33, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812, Japan. Telephone: 092-641-1101 (ext. 4428). Fax: 092-632-2742.

[§] Kyushu University.

[∥] Meiji College of Pharmacy.

[⊥] Present address: Department of Molecular Biology, Mie University School of Medicine, Tsu, Mie 514, Japan.

[▽] Liverpool School of Tropical Medicine.

[⊗] Abstract published in *Advance ACS Abstracts,* January 15, 1995.

[1] Abbreviations: *E., Echis*; RVV-V, factor V activating enzyme from Russell's viper venom; RVV-X, factor X activating enzyme from Russell's viper venom; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; Boc, *tert*-butoxycarbonyl group; MCA, 4-methylcoumaryl-7-amide.

0006-2960/95/0434-1771$09.00/0

	oligonucleotide sequence	position	base
Sense Primer			
	Lys Phe Glu Lys Lys Phe Ile 5'- GC GAATTC AAA TTT GAA AAA AAA TTT AT-3' *Eco*RI G C G G G C	197-203	28
Antisense Primer			
	Glu Ile Gly Tyr Tyr Gly 3'- CTT TAA CCA ATA ATA CC CTTAAG CG -5' C G G G G *Eco*RI T T C	545-550	25

FIGURE 1: Structures of the oligonucleotides used for PCR primers. The partial amino acid sequences chosen for syntheses of the oligonucleotides and the numbers of bases are indicated. The third bases of the triplet codons for the carboxyl-terminal amino acids were removed.

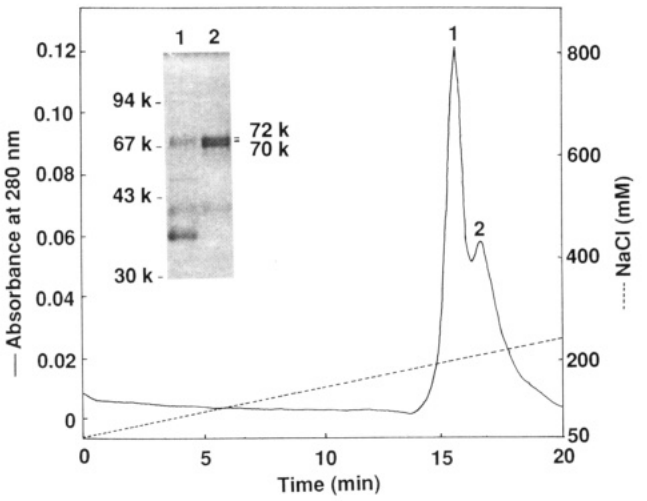

FIGURE 2: Elution profile for ecarin from a Mono S column. The sample obtained from the fourth stage of the purification procedure for ecarin (Rhee *et al.*, 1982) was applied to the Mono S column connected to an FPLC apparatus. The elution was performed at a flow rate of 1 mL/min with a linear gradient of 0.05−0.24 M NaCl containing 20 mM Tris-HCl (pH 8.0), at room temperature. In the inset, SDS−PAGEs of purified ecarin (peak 2) and a sample from peak 1 are shown.

unique metalloproteinase properties with respect to strict substrate specificity toward prothrombin and centers in particular on how this glycoprotein specifically activates only prothrombin. For the elucidation of structure−function relationships, knowledge of the primary structure is a prerequisite for a better understanding of related functions. It is also of interest to compare the covalent structures of RVV-X and ecarin, and we determined the complete cDNA sequence and the translated protein sequence of ecarin.

EXPERIMENTAL PROCEDURES

Materials. Bovine prothrombin was purified as described (Hashimoto *et al.*, 1985). All DNA modifying enzymes were purchased from Nippon Gene Co., Toyama, Japan; Toyobo Co., LTD., Osaka; Takara Shuzo Co., Kyoto; or Bethesda Research Laboratories Life Technologies, Inc., Gaithersburg, MD. [α-^{32}P]dCTP was obtained from Amersham-Japan, Tokyo. Lysyl endopeptidase was purchased from Wako Pure Chemical Industries, Ltd., Osaka. Sephadex G-150 fine, DEAE-Sephadex A-25, wheat germ lectin 6MB, and DEAE-Sephacel were from Pharmacia, Uppsala. The phenyl-5PW RP column was from Tosoh, Osaka; the Cosmosil 5C4-300 column was from Nacalai Tesque, Inc., Kyoto; and the μBondasphere C8 column was from Nihon Waters Ltd., Tokyo. Boc-Val-Pro-Arg-MCA was the product of the Protein Research Foundation, Minoh, Osaka. All other chemicals were of the highest quality commercially available.

Purification of Ecarin and Determination of Its Partial Amino Acid Sequence. Ecarin was purified from Kenyan *E. carinatus* venom collected by us and venom from Sigma Chemical Co., St. Louis, MO (lot V-8250), as described (Morita & Iwanaga, 1978; Rhee *et al.*, 1982). The ecarin-derived peptides suitable for the synthesis of DNA primers for polymerase chain reaction (PCR) were obtained from the *S*-pyridylethylated ecarin (Takeya *et al.*, 1990b). The sample was digested with lysyl endopeptidase (E/S = 1/50, by weight) in 50 mM Tris-HCl, pH 8.0, at 37 °C for 20 h. The peptides were purified by reversed-phase high-performance liquid chromatography (HPLC), as described (Takeya *et al.*, 1990b). The isolated peptides were analyzed on a gas-phase amino acid sequencer, an Applied Biosystems 473A or 477A protein sequencer. Amino acid analysis was done on a PICO-TAG system (Waters, Millipore Corp., Milfold, MA) or a Hitachi L-8500 automatic analyzer.

Sodium Dodecyl Sulfate−Polyacrylamide Gel Electrophoresis (SDS−PAGE). SDS−PAGE was carried out by the method of Laemmli (1970). The reference protein kit for molecular weight estimates was obtained from Pharmacia Fine Chemicals, Uppsala.

Fluorometric Assay of Prothrombin Activation. Ten microliters of bovine prothrombin (1.9 mg/mL) and 80 μL of 50 mM Tris-HCl buffer, pH 8.0, containing 100 mM NaCl and 0.1 mg/mL BSA were mixed with 10 μL of a 1×10^{-3} sample appropriately diluted with the buffer, and the mixture was incubated at 37 °C for 5 min. Twenty microliters of 0.1 mM EDTA was added to halt the reaction. The solution was then mixed with 800 μL of the same buffer and 10 μL of 5 mM Boc-Val-Pro-Arg-MCA, and the mixture was incubated further at 37 °C for 5 min. Then, 200 μL of 10% acetic acid was added to stop the reaction. The resulting fluorescence was measured in a fluorometer with excitation at 380 nm and emission at 440 nm.

Construction of Venomous Gland cDNA Library. Total RNA was extracted from the venomous gland of Kenyan *E. carinatus*, and poly(A)$^+$ RNA was purified using oligo(dT)−cellulose chromatography (Boehringer Mannheim). Double-stranded cDNA was synthesized from the poly(A)$^+$ RNA

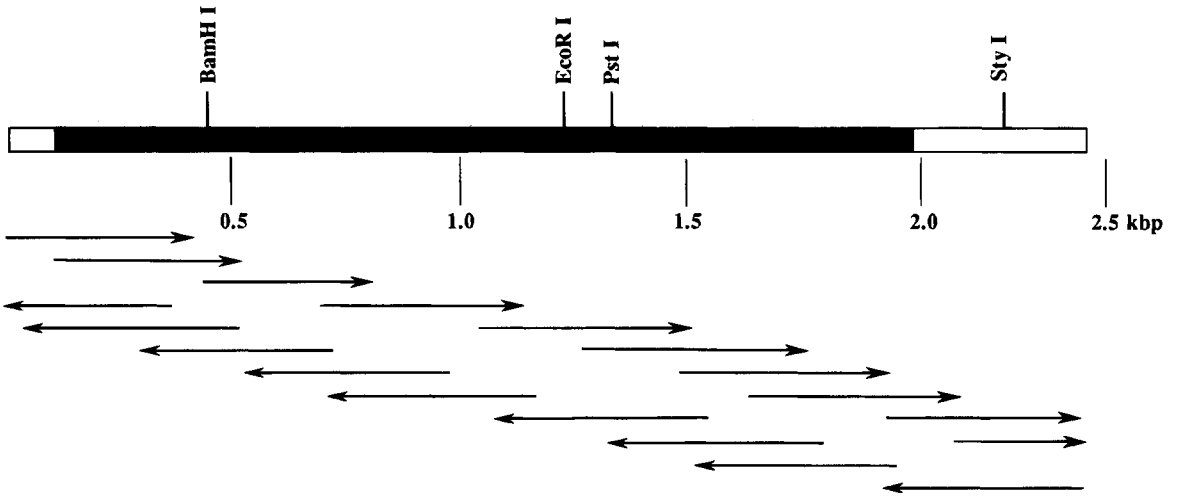

FIGURE 3: Restriction map and sequence strategy. A black rectangle shows the open reading frame. The directions and extents of sequence determinations are shown by horizontal arrows.

and then cloned into λZAPII, using cDNA synthesis and cDNA cloning kits (Stratagene, CA), respectively (Lipman & Pearson, 1985).

Oligonucleotide Synthesis. Oligonucleotide primers for PCR were prepared on a DNA synthesizer (Applied Biosystems, Model 380B).

Polymerase Chain Reaction (PCR). DNAs were amplified by PCR, using the total cDNA library as a template and two oligonucleotide primers corresponding to two peptide sequences of ecarin (Figure 1). Reactions for PCR were carried out in a Perkin-Elmer Cetus thermal cycler for 35 cycles of denaturation (90 °C, 1 min), annealing (40 °C, 1 min), and extension (60 °C, 2 min); a 1.06-kb fragment was thus amplified. After subcloning of the PCR products into the pUC118 vector, the amplified DNAs were confirmed by partial sequencing. An *Eco*RI fragment (0.53 kb) of the PCR product was used to screen a cDNA library after being radiolabeled by the use of a DNA labeling kit (Nippon gene).

Screening of the cDNA Library. To screen cDNAs, about 1.3×10^6 plaques from a λZAPII cDNA library were made using the PCR-derived probe. Hybridization was carried out overnight at 65 °C (7% polyethylene glycol, 10% SDS, and 100 mg/mL tRNA). The filters were washed in a change of 6× SSC/0.1% SDS (1× SSC = 15 mM sodium citrate and 150 mM NaCl, pH 7.0) and in two changes of 2× SSC/0.1% SDS at 65 °C. After tertiary screening, plaque-purified positive clones were then rescued into pBlueScript II phagemids using R408 helper phage (Sambrook *et al.*, 1989). One of the largest clones (λE24) was used for DNA sequence analysis.

DNA Sequencing. Clone λE24 was sequenced using sequential exonuclease digestion and a deletion kit (Takara Shuzo Co., Kyoto). The dideoxy termination method was used to sequence double-stranded DNA, using Taq dye primer cycle sequencing kits (Applied Biosystems) and nucleotide sequences were determined using an Applied Biosystems Model 370A DNA sequencer.

Homology Search. The amino acid and DNA sequences were compared with sequences in the GenBank (release 73.1, January, 1993) and in the NBRF protein Database (release 34.0, January, 1993) with the GENETYX system (version 20, Software Development Co., Ltd., Tokyo), based on the algorithm of the computer program FASTA (Gavel & von Heijne, 1990).

RESULTS AND DISCUSSION

Purification of Ecarin. Ecarin was highly purified from crude venom purchased from Sigma Chemical Co. and from the crude venom of Kenyan *E. carinatus*, according to previously published methods (Morita & Iwanaga, 1978; Rhee *et al.*, 1982). Information on the partial amino acid sequence of the former was used for oligonucleotide primers of PCR, and the latter was used for confirmation of the sequence of the λE24 clone. Figure 2 shows the fast protein liquid chromatographic (FPLC) pattern for the partially purified Kenyan ecarin obtained from the fourth step of the purification procedure (Rhee *et al.*, 1982). Of the two peaks, the second one contained ecarin, as judged from the amino acid composition and the prothrombin activating activity (Morita & Iwanaga, 1978). The purified ecarin thus obtained gave two protein bands in close proximity (72 and 70 kDa on SDS−PAGE), as shown in the inset of Figure 2. These proteins could not be further separated on HPLC, thereby suggesting the existence in the ecarin preparation of an isoform with a different carbohydrate composition or a variant. Previously, we and others (Morita & Iwanaga, 1978; Rhee *et al.*, 1982) have published the purification procedures for ecarin, using commercial *E. carinatus* venom (Sigma and Miami Serpentarium), which resulted in the purification of ecarin to homogeneity on SDS−PAGE with apparent molecular masses of 56 and 55 kDa, respectively. In the present study, however, a 70−72-kDa protein has been purified from the crude venom of Kenyan *E. carinatus*. The 70−72-kDa protein is obviously a processed active form of the precursor of ecarin, because the amino-terminal sequence starts at Val-191, as shown in Figure 4. Although we have no definite information to explain the discrepancy between the molecular masses of the present protein and the protein previously reported, the latter may represent a further proteolytic processed form or a different glycosylation form of the protein, probably due to a difference of subspecies of *Echis carinatus*.

Protein Sequencing of Ecarin. The amino-terminal amino acid sequence of the purified Kenyan ecarin was determined up to 38 residues by Edman degradation and peptides

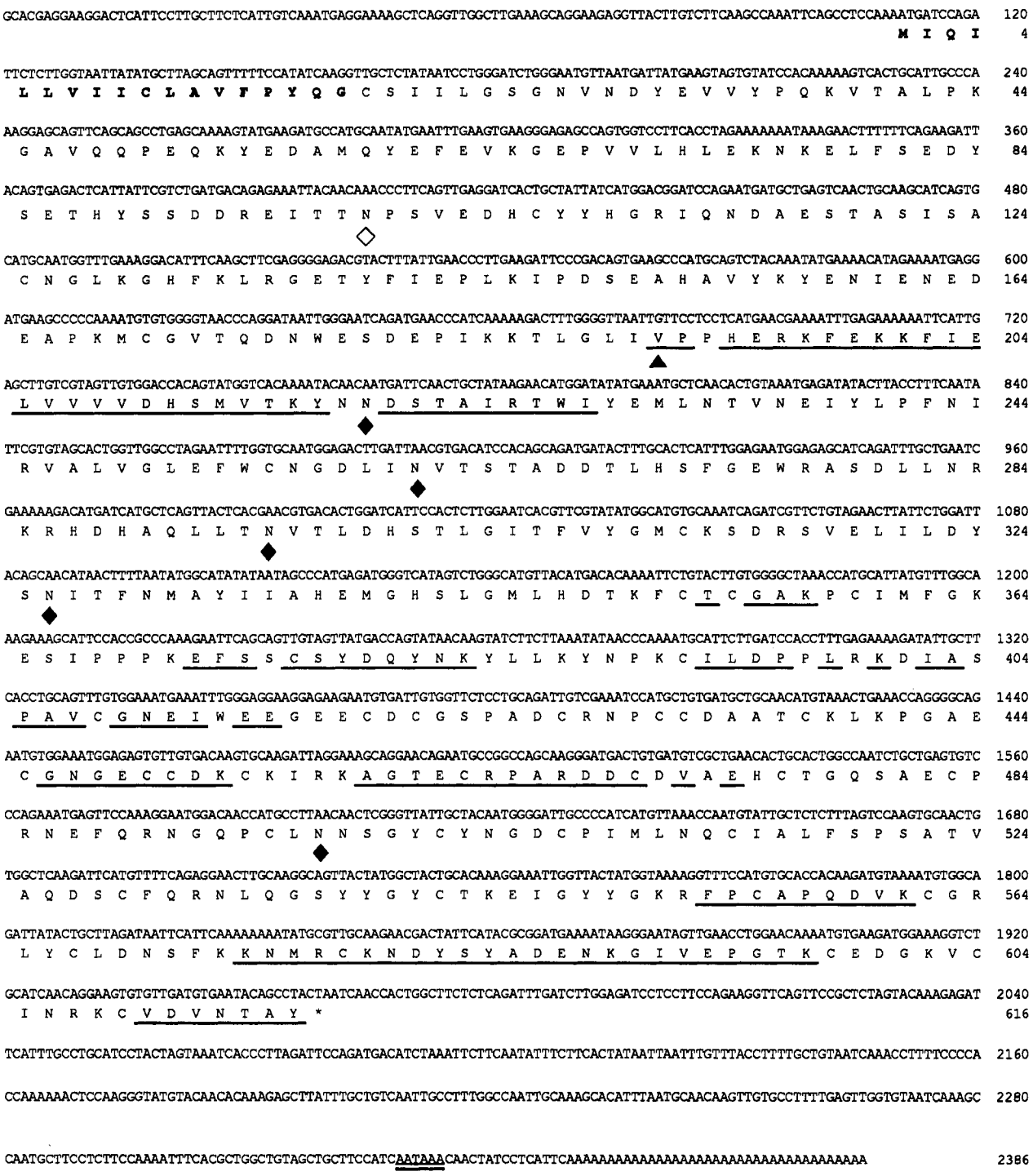

FIGURE 4: cDNA and deduced amino acid sequences of ecarin clone λE24. The putative signal sequence is denoted in boldface, and the mature protein starts from the closed triangle. A polyadenylation signal (AATAAA) is double-underlined. The amino acid residues confirmed by sequencing the purified peptides are underlined. The potential carbohydrate attachment sites for Asn-Xaa-Ser/Thr are indicated by closed diamonds, and a site (Asn-Pro-Ser) which may not be glycosylated due to the presence of Pro is indicated by an open diamond.

digested with lysyl endopeptidase were also sequenced. These results gave amino acid sequences covering a total of 120 residues (underlined in Figure 4), which comprise 26.3% of the mature protein. These sequences showed good agreement with those deduced from the cDNA sequence.

cDNA Cloning and the Predicted Amino Acid Sequence of Ecarin. In the 13 positive clones selected, the longest one, λE24, was used for sequencing and consisted of 2386 bp. The restriction map and the strategy used to analyze the entire insert of λE24 and its sequence with the deduced amino acid sequence of ecarin are shown in Figures 3 and 4. There were two candidates for an initiation Met at nucleotide positions 39 and 111, and the latter one with the upstream sequence TCCAAA agreed well with the proposed GCCACC motif optimal for the initiation of translation (Kozak, 1986). The amino-terminal 18 residues possessed a hydrophobic core with a typical cleavage motif for signal peptidase at a Gly18−Cys19 bond (von Heijne, 1990). Cleavage at Ile190−Val191 yielded a mature protein with an amino-terminal sequence identical with that of the purified ecarin. The cDNA, therefore, included an open reading frame of 1848 nucleotides which encoded a pre-proecarin

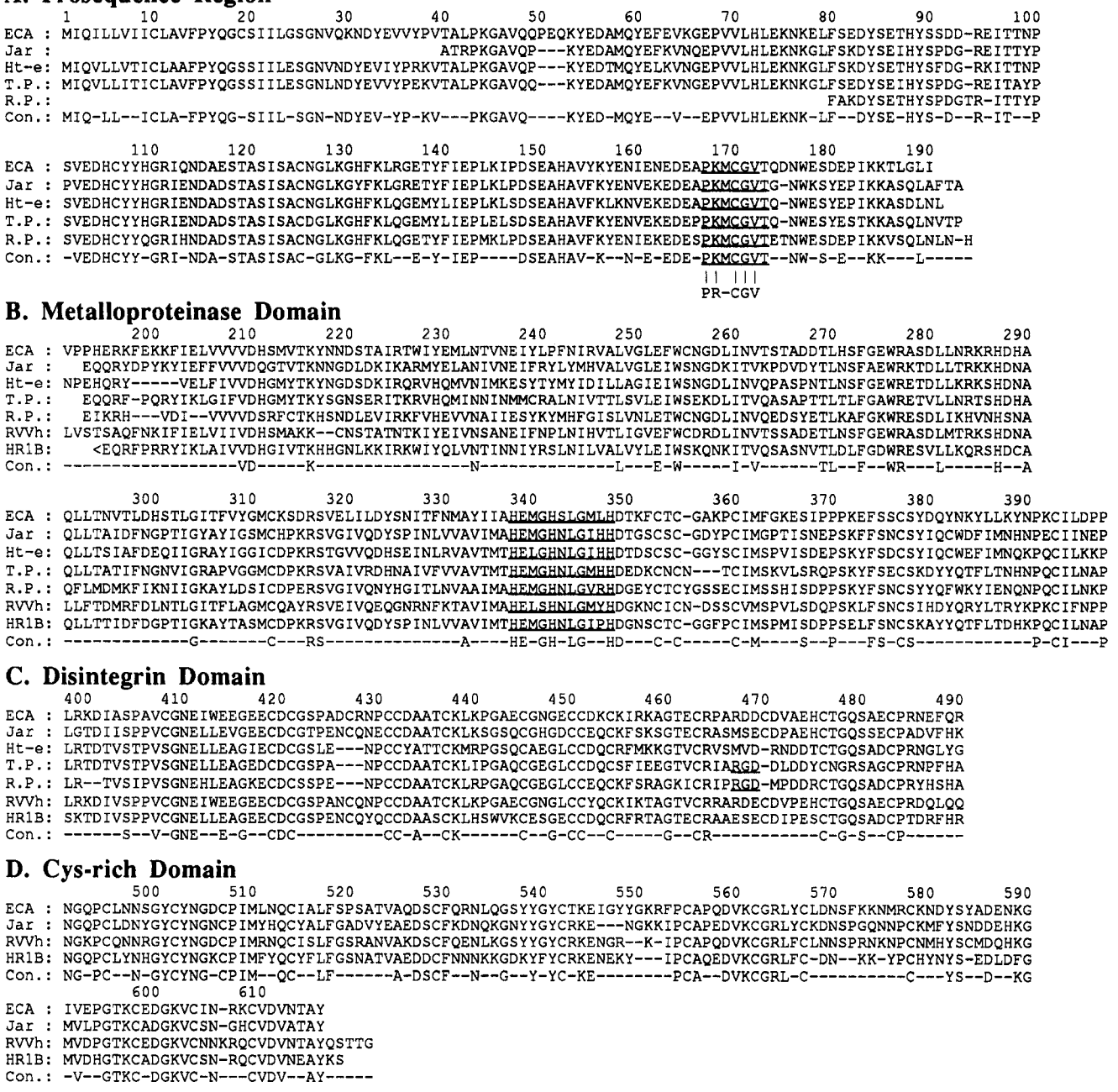

FIGURE 5: Comparison of the predicted amino acid sequence of ecarin with members of the snake venom metalloproteinase and disintegrin families. ECA denotes ecarin. Jararhagin (Jar) from *Bothrops jararaca* (Paine *et al.*, 1992) and HR1B from *Trimeresrus flavoviridis* (Takeya *et al.*, 1990a) are high molecular mass hemorrhagic proteins. Ht-e is a low molecular mass hemorrhagic protein from *Crotalus atrox* (Hite *et al.*, 1992). Protrigramin (T.P.) from *Trimeresurus gramineus* (Neeper & Jacobson, 1990) and prorhodostomin (R.P.) from *Calloselasma rhodostoma* (Au *et al.*, 1991) are disintegrins in the venoms. RVVh is a heavy chain of the blood coagulation factor X activating enzyme isolated from Russell's viper venom (Takeya *et al.*, 1992). Residues conserved in all proteins are shown at the bottom (Con.). Jararhagin, Ht-e, protrigramin, and prorhodostomin are identified by cDNA. Alignments are divided into four functional regions on the basis of sequence similarity; A, prosequence region; B, metalloproteinase domain; C, disintegrin domain; and D, Cys-rich domain. The "cysteine switch" in (A), the putative zinc ligands and active site in (B), and the RGD sequences in (C) are underlined.

of 616 residues consisting of a signal peptide, a propeptide, and a mature protein.

The molecular mass of the mature protein without carbohydrate chains was calculated to be 47 921 Da. There are five potential *N*-glycosylation sites with the sequence Asn-Xaa-Ser/Thr in the mature protein and one site in the prosequence (Figure 4), but the latter potential site (Asn99) may not be glycosylated because of the sequence Asn-Pro-Ser (Gavel & von Heijne, 1990). The molecular mass value of the purified ecarin estimated on SDS−PAGE was higher than that calculated from the entire sequence. Although it is difficult to explain this disagreement, one possible reason may be the glycoprotein nature of ecarin, since most glycoproteins do show abnormal behavior on SDS−PAGE.

Sequence Similarities of Ecarin to Other Proteins. Table 1 shows sequence identities among procoagulant, hemorrhagic factors, and disintegrins isolated from various venoms. When the nucleotide sequence of ecarin clone was compared with the GenBank database, several significant similarities were found in the zinc metalloproteinases, hemorrhagic toxin e (Ht-e) from *Crotalus atrox* venom (Hite *et al.*, 1992), protrigramin from *Trimeresurus gramineus* venom (Neeper

Table 1: Sequence Similarities (%) of Ecarin-Derived Domains with the Venom Metalloproteinase and Disintegrin Families[a]

snake venom counterparts	prosequence region	metalloproteinase domain	disintegrin domain	Cys-rich domain	whole (mature)
jararhagin/ecarin	76	48	63	65	62 (56)
Ht-e/ecarin	80	45	52		59
protrigramin/ecarin	78	43	60		61
prorhodostomin/ecarin	79	43	55		56
RVV-X heavy chain/ecarin		54	82	68	(64)
HR1B/ecarin		45	60	59	(52)

[a] Each domain of ecarin is compared with the prosequence region (residues 1−190 in ecarin), the metalloproteinase (residues 191−397 in ecarin), disintegrin (residues 398−490 in ecarin), and Cys-rich (residues 491−616 in ecarin) domains, and the entire and/or mature amino acid sequence of jararhagin, Ht-e, protrigramin, prorhodostomin, RVV-X heavy chain, and HR1B. The degree of conservation of each pair is expressed as the percentage of identical residues. In the prosequence regions of jararhagin and prorhodostomin, values are shown in parentheses, since the nucleotide sequences followed by initiation Met codon have not been determined. HT-e, protrigramin and prorhodostomin consist of prosequence region, metalloproteinase and disintegrin domains. In comparison with the entire sequence, the identities in mature proteins are shown in parentheses.

```
           1       10        20        30        40        50        60        70
Ecarin     : MIQILLVIICLAVFPYQGCSIILGSG-NVNDYEVVYPQKVTALPKGAVQQPEQKYEDAMQYEFEVKGEPVV
Cyritestin:      MAAALFLVLSYLGQVIAAGKDVETPLLQITVPEKIDT----NIQDAKEA-ETQVTYVVTIEGKAYT

             80        90       100       110       120       130       140
Ecarin     : LHLEKNKELFSEDYSETHYSSDDREITTNPSVEDHCYYHGRIQNDAESTASISACNGLKGHFKLRGETYFI
Cyritestin: LQLEK-QSFLHPLFGTYLRDKLGTLQPYFSLVKTHCFYQGHAAEIPVSTVTLSTCSGLRGLLQLENITYGI
                                              *                 *

                  150       160       170       180       190       200       210
Ecarin     : EPLKIPDSEAHAVYKYENIENEDEAPKMCGVTQDNWESDEPIKKTLGLIVPPHERKFEKKFIELVVVVDHS
Cyritestin: EPLESSATFEHILY--EIKNNKIDYSPLKENFANS-EQESQSYRILVKPEKGSNSTLTKRILRIKIIMDKA

                  220       230       240       250       260       270       280
Ecarin     : MVTKYNNDSTAIRTWIYEMLNTVNEIYLPFNIRVALVGLEFWCNGDLINVTSTADDTLHSFGEWRASDLLN
Cyritestin: MFDHMGSEVGVATQKVVHIFGLINTMFSQLKMTVMLNSLEIWSEQDKIETNGDADEVLQRFLLWKSKEISQ

                  290       300       310       320       330       340       350
Ecarin     : RKRHDHAQLLTNVTLDHSTLGITFVYGM-CKSDRSVELILDYSNITFN-MAYIIAHEMGHSLGMLHDTKF-
Cyritestin: -KAQDITYLLLYKD-HPDYVGATYH-GMACNPNFTAGIALHPKTLAVEGFAIVLSQLLGINLGLAYDDVYN
                                        *

                  360       370       380       390       400       410
Ecarin     : CTCGAKPCIMFGKESIPPP-KEFSSCSYDQYNKYLLKYNPKCILDPP---LRKDIASPAVCGNEIWEEGEE
Cyritestin: CFCPGSTCIMNPSAIRSQGIKVFSSCSVDEFKQLASQPELDCLRNTSETEFVVQPQGGSYCGNHLLEVPEQ
PH-30  β   :                                                   SNPVCGNNRVEQGED
             * *     *                 *                 *                 *

            420       430       440       450       460       470       480
Ecarin     : CDCGSPADC-RNPCCDAATCKLKPGAECGNGECCDK--CKIRKAGTECRPARDDCDVAEHCTGQSAECPRN
Cyritestin: CDCGPPETCTHKKCCNPKDCTLIDAAQCGTGPCCDKRTCTIAERGRLCRKSKDQCDFPEFCNGETEGCAPD
PH-30  β   : CDCGSQEEC-QDTCCDAATCRLKSTSRCAQGPCCNQ--CEFKTKGEVCRESTDECDLPEYCNGSSGACQED
PH-30  α   :                                                             YCTGQSGKCPLD
             * *     *    **    *       *    **    *       *       *     *      *

             490       500       510       520       530       540           550
Ecarin     : EFQRNGQPCLNNSGYCYNGDCPIMLNQCIALFSPSATVAQDSCFQR-NLQGSYYGYCTKEIGY---YGKRF
Cyritestin: TKAADLEPCNNETAYCFGGVCRDPDRQCTDLFGKYAKGPNYVCAQEVNLQNDKFGNCHGRCNYSAIFCGKA
PH-30  β   : LYVINGHRCANEEWICMNGRCLSGKAQCQETFGTEMEMGSVDCFEQLNTKNDITGNCGILSPG--NY-K-A
PH-30  α   : TYKQDGTPC-NEGFFCVSKGCTDPGIQCATYFGHGARSAPDACYTTLNSIGNIFGNCGQSGNPT-TY---V
                     *      *    *      *               *               *

                  560                       570           580       590       600
Ecarin     : PCA--PQDVKCGRLYC--------------LDNSFKKNMRC---KNDYSYADENK-GIVEPGTKCEDGKVC
Cyritestin: VCYWNFAEVIQTEKYD--------------VQYTYLGGQVCVSAHLRSQTGTRDD-TYVHDGTVCGSGQVC
PH-30  β   : -CG--ASNWKCGKLICSYDKSEILRNKEGMTIYANISGHICVSIEYPPGHAKSAL-MWVRDGTVCGPSEVC
PH-30  α   : GCS--GDSTKCGKLICTGISSIPPIRALFAAIQIPHGDDWCWSISNFGDPASSPTEGAVSAGTSCASGKAC
              *                                       *                       *     *

             610
Ecarin     : INRKCVDVNTAY
Cyritestin: FRGDCLRVHVLR
PH-30  β   : RQQQCVSSSYLG
PH-30  α   : VNAQCSTFTLDT
              *
```

FIGURE 6: Alignment of the deduced amino acid sequence of ecarin and those of cyritestin, PH-30 β, and PH-30 α. Cyritestin, PH-30 β, and PH-30 α encode 823, 353, and 289 residues, respectively. The alignment with residues 1−619 in cyritestin, residues 1−229 in PH-30 β, and residues 1−159 in PH-30 α is shown. Conserved cysteine residues are indicated by asterisks, and the predicted cysteine switch motif (residues 167−171) and the zinc-chelating motif (residues 337−347) in the ecarin sequence are underlined.

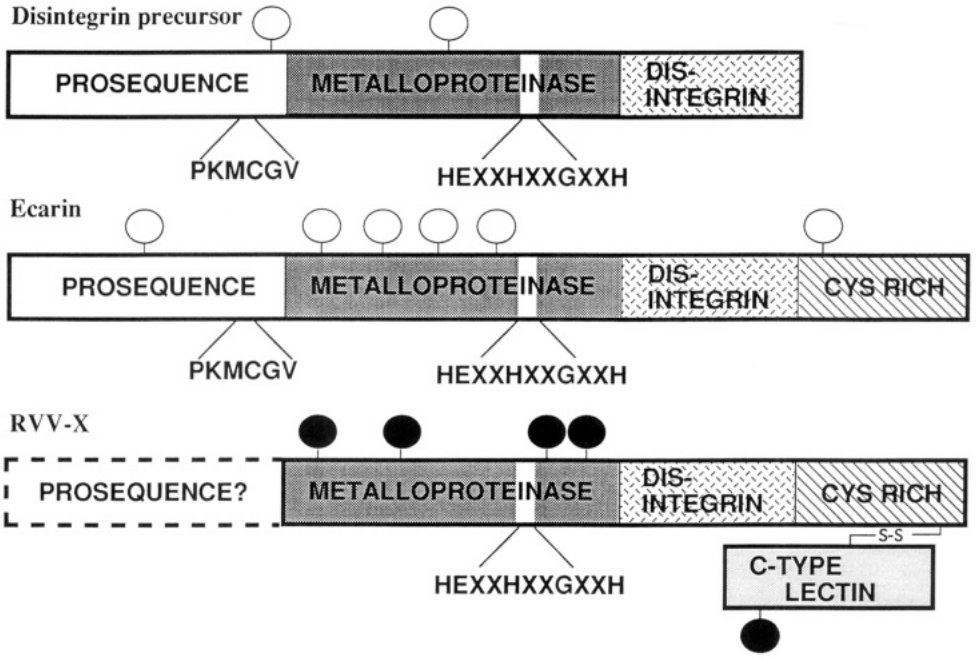

FIGURE 7: Gross structures of snake venom metalloproteinases: (A) precursor protein of trigramin deduced from its cDNA sequence; (B) ecarin; (C) RVV-X, factor X activator. Identified and potential *N*-linked sugar chains are shown by closed and open ovals. The locations of the conserved HEXXHXXGXXH sequence are indicated.

& Jacobson, 1990), and prorhodostomin from *Calloselasma rhodostoma* venom (Au *et al.*, 1991). Sequence comparison of the translated open reading frame of these three clones with that of the ecarin clone showed 59%, 61%, and 56% identities, respectively. Ecarin and RVV-X heavy chain show the closest similarity, indicating a 64% identity. Furthermore, ecarin has the same domain structure as that of a high molecular mass hemorrhagic protein, jararhagin (Paine *et al.*, 1992). Both contain a prosequence region (residues 1−190), a metalloproteinase region (residues 191−397), a disintegrin region (residues 398−490), and a Cys-rich region (residues 491−616).

As shown in Figure 5A, an underlined sequence, -Pro-Lys-Met-Cys-Gly-Val-, is located in the carboxyl-terminal portion of the prosequence in ecarin, which shows a sequence similar to the -Pro-Arg-Cys-Gly-Val- sequence found in the prosequence of matrix metalloproteinases (MMP) (Valle & Auld, 1990; van Wart & Birkedal-Hansen, 1990; Woessner, 1991; Nagase *et al.*, 1990). This sequence is closely associated with activation of these proenzymes, the so-called "cysteine switch" mechanism. In the latent form of the proenzyme, a free sulfhydryl group in that sequence is coordinated to the catalytic zinc ion in the active site. On cleavage and dissociation of the propeptide from the latent form of the proenzyme, it provides a free ligand of zinc, resulting in the activation of the proenzyme (Valle & Auld, 1990). The prosequence region consisting of 172 residues in ecarin has a striking homology to those found in protrigramin, prorhodostomin, Ht-e, and jararhagin (up to 80% identity; Table 1). These findings suggest that the venom metalloproteinases are synthesized in a similar fashion and are activated by proteolysis of the propeptide region like MMP. However, except for the sequence of the cysteine switch located in the ecarin propeptide region, there are no similarities to MMPs.

Figure 5B summarizes the sequence of the metalloproteinase region. As expected, the sequence of ecarin is similar to those of the venom procoagulant and hemorrhagic factors, in addition to those of low molecular mass metalloproteinases HR2a (Miyata *et al.*, 1989), HT-2 (Takeya *et al.*, 1990b), Ht-d (Shannon *et al.*, 1989), and H_2-proteinase (Takeya *et al.*, 1989), with 41−48 percent identities (data not shown). Bode *et al.* (1992) reported the X-ray crystal structure of astacin, a zinc metalloproteinase isolated from the crayfish *Astacus astacus* L. Astacin, which contains the consensus sequence His-Glu-Xaa-Xaa-His-Xaa-Xaa-Gly-Xaa-Xaa-His, and which is also found in many vertebrate collagenases; bacterial thermolysin lacks Gly and the third His in the consensus sequence (Priestle, 1988). Similarly, all the venom metalloproteinases contain this consensus sequence, thereby indicating the presence of an analogous zinc-chelating site corresponding to that of astacin. Furthermore, the positioning of the three histidine residues and the conservation of most of the intervening residues in the consensus sequence indicate a similar spatial arrangement of these residues around the active site zinc ion in these proteins. Except for this consensus sequence, however, not one of the venom metalloproteinase domains has a significant sequence homology with other metalloproteinases.

Ecarin also contains a domain (residues 398−490; Figure 5C and Table 1) with sequence homology to disintegrins, a family of Arg-Gly-Asp (RGD) sequence containing snake venom polypeptides which inhibit strongly platelet aggregation (Ouyang *et al.*, 1992). This region of ecarin shows the closest similarity with those of the venom metalloproteinase and disintegrin families, particularly with that of RVV-X heavy chain (82% identity; Figure 5B). Although there is sequence similarity between the ecarin disintegrin family and the snake venom disintegrin family, such as echistatin (Gan *et al.*, 1988), bitistatin (Shebuski *et al.*, 1989), and barbourin (Scarborough *et al.*, 1991) (42−61% identities), the ecarin-derived disintegrin, as well as jararhagin, Ht-e (Hite, *et al.*, 1992), RVV-X (Takeya *et al.*, 1992) and HR1B (Takeya *et al.*, 1990a), lacks the RGD sequence but contains the Arg-Asp-Asp sequence in the corresponding region (Figure 5C). Interestingly, ecarin and echistatin, the latter of which is one

of the RGD sequence containing disintegrins, are derived from the same venom, *E. carinatus*, but their disintegrin domain sequences differ (42% identity).

Ecarin, in addition to jararhagin, RVV-X heavy chain, and HR1B, contains a Cys-rich domain, which also shows a high sequence similarity (59−68% identities; Figure 5D). The Cys-rich domain has been detected also in the guinea pig sperm fusion proteins PH-30 α and PH-30 β, which participate in cell fusion of guinea pig sperm cells to the egg surface during fertilization (Blobel *et al.*, 1992). Although disintegrin binds to the platelet GP IIb-IIIa integrin target via the RGD sequence (Ouyang *et al.*, 1992), the integrin specificities of the Cys-rich domain remain to be determined. Another surprising observation is that the amino-terminal region of cyritestin (residues 1−619), which is known as pre- and postmeiotic germ cell specific antigen of TAZ 83 (U. A. O. Heinlein, EMBL accession no. X644277), shares a sequence similarity to the carboxyl-terminal region of ecarin (24% identity). As shown in Figure 6, almost all the cysteine residues in the carboxyl-terminal region of ecarin are aligned with the same positions found in cyritestin, although cyritestin does not contain a cysteine switch motif and a zinc-chelating motif. The sequence similarities between ecarin, cyritestin, PH-30 α, and PH-30 β indicate a divergent evolution from a common ancestry. These domains of snake venom disintegrin and the Cys-rich region have also been found in tumor suppressor genes for breast and ovarian cancers (Emi *et al.*, 1993). The gene (MDC) encodes a 524 amino acid metalloproteinase-like domain, a non-RGD disintegrin, and a Cys-rich domain with sequence similarity to the snake venom metalloproteinase subfamily.

In Figure 7, the gross structure of ecarin is illustrated with those of other known members of the venom metalloproteinases, including trigramin precursor and RVV-X. While the mosaic structures of these proteins are similar, there is a significant difference in structural domains: the trigramin precursor and rhodostomin do not contain the Cys-rich domain, but ecarin contains the Cys-rich domain in the carboxyl-terminal portion, like RVV-X heavy chain. The whole RVV-X has an additional C-type lectin domain in the light chain, which might participate in recognizing some portion of the zymogen factor X (Takeya *et al.*, 1992). Although these proteins have diverse functional activities, these findings do show structural and evolutionary relationships among the venom metalloproteinase. The structural information described here may provide new insights into the substrate specificity of ecarin.

ACKNOWLEDGMENT

We thank Dave Theakston for the gift of the crude venom of Kenyan *E. carinatus*, Hiroko Hashimoto-Komatsu for skillful analyses of the amino acids, Itsuko Edamitsu and Satomi Matsumura for expert secretarial assistance, and M. Ohara for helpful comments.

REFERENCES

Au, L. C., Huang, Y. B., Huang, T. F., Teh, G. W., Lin, H. H., & Choo, K. B. (1991) *Biochem. Biophys. Res. Commun. 181*, 585−593.

Blobel, C. P., Wolfsberg, T. G., Turk, C. W., Myles, D. G., Primakoff, P., & White, J. M. (1992) *Nature 356*, 248−252.

Bode, W., Gomis-Ruth, F. X., Huber, R., Zwilling, R., & Stocker, W. (1992) *Nature 358*, 164−167.

Emi, M., Katagiri, T., Harada, Y., Saito, H., Inazawa, J., Ito, I., Kasumi, F., & Nakamura, Y. (1993) *Nat. Genet. 5*, 151−157.

Gan, Z.-R., Gould, R. J., Jacobs, J. W., Friedman, P. A., & Polokoff, M. A. (1988) *J. Biol. Chem. 263*, 19827−19832.

Gavel, Y., & von Heijne, G. (1990) *Protein Eng. 3*, 433−442.

Hashimoto, N., Morita, T., & Iwanaga, S. (1985) *J. Biochem. (Tokyo) 97*, 1347−1355.

Hite, L. A., Shannon, J. D., Bjarnason, J. B., & Fox, J. W. (1992) *Biochemistry 31*, 6203−6211.

Huang, T. F., Holt, J. C., Lukasiewicz, H., & Niewiarowski, S. (1987) *J. Biol. Chem. 262*, 16157−16163.

Iwanaga, S., & Takeya, H. (1993) in *Methods in Protein Sequence Analysis* (Imahori, K., & Sakiyama, F., Eds.) pp 107−115, Plenum Press, London.

Kornalik, F. (1963) *Folia Haematol. 80*, 73−78.

Kornalik, F., & Blombäck, B. (1975) *Thromb. Res. 6*, 53−63.

Kozak, M. (1986) *Cell 44*, 283−292.

Laemmli, U. K. (1970) *Nature 227*, 680−685.

Lipman, D. J., & Pearson, W. R. (1985) *Science 227*, 1435−1441.

Miyata, T., Takeya, H., Ozeki, Y., Arakawa, M., Tokunaga, F., Iwanaga, S., & Omori-Satoh, T. (1989) *J. Biochem. (Tokyo) 105*, 847−853.

Morita, T., & Iwanaga, S. (1978) *J. Biochem. (Tokyo) 83*, 559−570.

Morita, T., Iwanaga, S., & Suzuki, T. (1976) *J. Biochem. (Tokyo) 79*, 1089−1108.

Nagase, H., Enghild, J. J., Suzuki, K., & Salvesen, G. (1990) *Biochemistry 29*, 5783−5789.

Neeper, M. P., & Jacobson, M. A. (1990) *Nucleic Acids Res. 18*, 4255.

Nelsestuen, G. L., & Suttie, J. W. (1972) *J. Biol. Chem. 247*, 8176−8182.

Ohsaka, A. (1979) in *Handbook of Experimental Pharmacology, Volume 52, Snake Venoms* (Lee, C.-Y., Ed.) pp 450−546, Springer-Verlag, New York.

Ouyang, C., Teng, C.-M., & Huang, T.-F. (1992) *Toxicon 30*, 945−966.

Paine, M. J. I., Desmond, H. P., Theakston, R. D. G., & Crampton, J. M. (1992) *J. Biol. Chem. 267*, 22869−22876.

Priestle, J. P. (1988) *J. Appl. Crystallogr. 21*, 572−576.

Rhee, M.-J., Morris, S., & Kosow, D. P. (1982) *Biochemistry 21*, 3437−3443

Rosing, J., & Tans, G. (1992) *Toxicon 30*, 1515−1527.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.

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*, 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., Bencen, G. H., & Polokoff, M. A. (1989) *J. Biol. Chem. 264*, 21550−21556.

Takeya, H., Arakawa, M., Miyata, T., Iwanaga, S., & Omori-Satoh, T. (1989) *J. Biochem. (Tokyo) 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. (Tokyo) 108*, 711−719.

Takeya, H., Nishida, S., Miyata, T., Kawada, S., Saisaka, Y., Morita, T., & Iwanaga, S. (1992) *J. Biol. Chem. 267*, 14109−14117.

Tokunaga, F., Nagasawa, K., Tamura, S., Miyata, T., Iwanaga, S., & Kisiel, W. (1988) *J. Biol. Chem. 263*, 17471−17481.

Tu, A. T. (1988) in *Hemostasis and Animal Venoms* (Pirkle, H., & Markland, F. S., Jr., Eds.) pp 445−455, Marcel Dekker, New York.

Valle, B. L., & Auld, D. S. (1990) *Biochemistry 29*, 5647−5659.

Van Wart, H. E., & Birkedal-Hansen, H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5578−5582.

Von Heijne, G. (1990) *J. Membr. Biol. 115*, 195−201.

Woessner, J. F. (1991) *FASEB J. 5*, 2145−2154.

BI941703U