Crotalase, a Fibrinogen-Clotting Snake Venom Enzyme: Primary Structure and Evidence for a Fibrinogen Recognition Exosite Different from Thrombin

Agnes H. Henschen-Edman¹, Ida Theodor^{1,2}, Brian F.P. Edwards³, Hubert Pirkle²

From the Departments of Molecular Biology and Biochemistry¹ and Pathology², University of California, Irvine, CA, USA and the ³Department of Biochemistry and Molecular Biology, Wayne State University, Detroit, MI, USA

Summary

Crotalase, a fibrinogen-clotting enzyme isolated from the venom of Crotalus adamanteus, and its overlapping fragments were subjected to Edman degradation. The resulting amino acid sequence, VIGGDEC NINEHRFLVALYDYWSQLFLCGGTLINNEWVLTAAHCDRTHI LIYVGVHDRSVQFDKEQRRFPKEKYFFDCSNNFTKWDKDIM LIRLNKPVSYSEHIAPLSLPSSPPIVGSVCRAMGWGQTTSPQET LPDVPHCANINLLDYEVCRTAHPQFRLPATSRTLCAGVLEG GIDTCNRDSGGPLICNGOFOGIVFWGPDPCAOPDKPGLYTK VFDHLDWIQSIIAGEKTVNCP, is characteristic of a serine proteinase. Comparison with thrombin, the physiological fibrinogen-clotting enzyme, showed that thrombin's fibrinogen-recognition exosite (FRE) is poorly represented in crotalase. Hirudin, a FRE-dependent inhibitor, had no effect on crotalase. Spatial modeling of crotalase yielded a possible alternative fibrinogen-recognition site comprised of Arg 60F, Lys 85, Lys 87, and Arg 107 (underlined in the sequence above). Crotalase also lacks thrombin's YPPW loop, as well as its functionally important ETW 146-148, and its heparin-binding site. The enzyme contains a single asparagine-linked glycosylation site, NFT, bearing neutral and amino sugars that account for 8.3% of the enzyme's total molecular weight of 29,027. The calculated absorbance of crotalase at 280 nm, 1%, cm⁻¹ is 15.2.

Introduction

Thrombin is a serine proteinase which catalyzes the last in the series of narrowly specific proteolytic reactions that lead to the clotting of the blood protein fibrinogen. Crotalase (EC 3.4.21.30) is an enzyme contained in the venom of the Eastern diamondback rattlesnake *Crotalus adamanteus* (1) which, like thrombin, triggers the clotting of fibrinogen by hydrolytic release of its fibrinopeptide A (FPA) (2). Venom enzymes with this thrombin-like action are widely distributed, mainly within several pit viper genera (3), and afford the opportunity for comparative examination of the molecular structural requirements for thrombin's narrow proteolytic action on fibrinogen.

In the present study, Edman degradation of the amino terminus and overlapping fragments of crotalase yielded a complete amino acid sequence with features characteristic of a serine proteinase. The fibrinogen recognition exosite (FRE) of thrombin is poorly represented in the crotalase sequence and this fact, along with the finding here that crotalase is not inactivated by hirudin (a FRE-dependent inhibitor), prompted a search for an alternative fibrinogen recognition site. The resulting spatial modeling of the crotalase molecule in the present work and in Massova et al. (4) reveals a possible alternative FRE comprised of Arg 60F, Lys 85, Lys 87, and Arg 107 [chymotrypsinogen numbering (15)].

Experimental Procedures

Materials

Lyophilized venom of the Eastern diamondback rattlesnake (*Crotalus adamanteus*) was purchased from Biotoxins, Inc. (St. Cloud, FL). The following materials were obtained from the sources given in parentheses: trypsin-TPCK (Worthington Biochemical Corp., Freehold NJ), lysyl endopeptidase (Waco Pure Chemicals Industries, Tokyo), Sephadex G-50 and G-100 and SP Sephadex C-50 (Pharmacia), DEAE-cellulose, DE-52 (Whatman), Sepharose 4B aminocaproyl-*p*-aminobenzamidine, *o*-iodosobenzoic acid, and *Staphylococcus aureus* protease V8 (Pierce Chemical Co., Rockford, IL), radiolabeled iodoacetic acid (New England Nuclear Co., Boston), N-Glycanase (Genzyme Corp.), hirudin (Boehringer Mannheim), human α -thrombin (gift from J.W. Fenton, II). Other reagents were the best available commercially.

Purification of Enzyme

Crotalase was purified from *Crotalus adamanteus* venom according to Bajwa and Markland (5). This procedure entails gel filtration on Sephadex G-100 followed by ion exchange chromatography on DEAE-cellulose, affinity chromatography on Sepharose 4B aminocaproyl-*p*-aminobenzamidine, and finally ion exchange chromatography on SP-Sephadex C-50. The product had a fibrinogen-clotting activity of 577 thrombin-equivalent NIH units/mg of protein using a previously described method (6).

Carbohydrate Removal

Crotalase (5 mg/ml in 0.2 M phosphate buffer, 0.01 M EDTA, pH 8.6) was incubated with N-Glycanase (2 units/mg protein) at 37°C for 48 h and the deglycosylated protein was isolated by gel filtration on Sephadex G-50.

Protein Modification

Crotalase was denatured and reduced in 0.1 M Tris, pH 8.5 containing 6 M guanidine hydrochloride and 5% final concentration of 2-mercaptoethanol for 4 h under nitrogen at 37°C in the dark. Sulfhydryl groups were carboxy-methylated with [³H] iodoacetic acid (7) or were pyridylethylated with 4-vinylpyridine in a 1:1 molar ratio to all sulfhydryl groups. Free sulfhydryl groups were determined using 5,5'-dithiobis(2-nitrobenzoic acid) (8). Succinylation of free amino groups was carried out according to Habeeb et al. (9).

Correspondence to: Dr. Hubert Pirkle, Department of Pathology, Medical Sciences I, University of California, Irvine, California 92697 USA – FAX Number +1 949 824 2160; Tel.: +1 949 824 6575; E-mail: hcpirkle@uci.edu

Chemical Cleavage

Methionyl bonds of reduced and alkylated crotalase were cleaved in 70% formic acid containing 10% (w/v) cyanogen bromide for 2 h at room temperature. Crotalase was cleaved at tryptophanyl bonds using *o*-iodosobenzoic acid (10), a 2-fold (w/w) excess of which was added to the reduced and alkylated protein dissolved in 80% acetic acid, 4 M guanidine hydrochloride.

Enzymatic Cleavage

Tryptic cleavage of reduced and alkylated crotalase was performed with or without prior succinylation in 0.1 M ammonium bicarbonate pH 8.2 with a 1:100 (w/w) enzyme to substrate ratio for 18 h at 37°C. Cyanogen bromide fragments of crotalase were digested with Lys-C endoproteinase at a 1:40 (w/w) enzyme to substrate ratio for 8 h at 37°C in 0.1 M Tris pH 8.6. Cyanogen bromide fragments were cleaved at Glu residues using *Staphylococcus aureus* V8 protease at a 1:40 (w/w) enzyme to substrate ratio for 3 h at 37°C in 0.1 M ammonium bicarbonate pH 8.0.

Fig. 1 Amino acid sequence of crotalase. Horizontal lines correspond to continuous sequences determined by automated Edman degradation. Short vertical lines indicate ends of sequenced peptides. Arrows indicate the end of satisfactory identification. Absence of any marking at either end of a horizontal line indicates continuity with the next or the preceding line of the diagram. NT, NH₂-terminal sequence of intact S-alkylated molecule; CB, cyanogen bromide-cleaved peptide; TR, trypsin-digested peptide; IB, *o*-iodosobenzoic acid-cleaved peptide; CB/LE, lysyl endoproteinase-digested fragment of a cyanogen bromide-cleaved peptide; CB/V8, *Staphylococcus aureus* protease V8-digested fragment of a cyanogen bromide-cleaved peptide. (N) is a glycosylation site. Polymorphic sites are shown as one residue symbol above another. Chymotryp-sinogen numbering (15)

The neutral and amino sugars of crotalase were determined by Dr. Adriana Manzi, Glycobiology Core, University of California, San Diego using high performance anion exchange chromatography with pulsed amperometric detection (GlycoStation) after trifluoroacetic acid hydrolysis.

Peptide Isolation

Peptides generated by the chemical and enzymatic cleavage procedures were separated by open column chromatography on Sephadex G-50, eluted with a 0.2 M ammonium bicarbonate buffer pH 8.0, and on DE-52 using a linear NaCl gradient (0 to 0.2 M) in Tris buffer pH 7.8 or by high performance liquid chromatography using a Beckman model 332 instrument with a Vydac C18 RP column. The gradient was linear from a starting eluant of 0.1% trifluoroacetic acid to a limiting eluant of 0.1% trifluoroacetic acid in 80% or 50% acetonitrile.

Amino Acid Sequence Analysis

Automated Edman degradation was performed with an Applied Biosystems Model 477A sequencer equipped with an on-line Model 120A phenylthiohydantoin amino acid analyzer or, earlier, with a Beckman 890B sequencer, in which case the PTH amino acids were identified by high performance liquid chromatography and by thin layer chromatography. The analyzer of the Applied Biosystems sequencer allowed direct identification of pyridylethylated PTH cysteine; radiolabeled, carboxymethylated PTH cysteine in the Beckman sequencer fractions was detected in a scintillation counter.

Molecular Modeling

The space-filling and secondary structure images of the crotalase molecule were generated by INSIGHT II (Kabsch-Sander algorithm for secondary structure; Molecular Simulations Inc.), using the coordinates of the model reported by Massova et al. (4).

Hirudin Inhibition

Crotalase (6.3 NIH thrombin-equivalent units/ml) was incubated with hirudin (30 antithrombin units/ml) in 0.15 M NaCl 0.02 M Tris pH 7.4 at 22°C. Hirudin was in 1.4-fold molar excess over crotalase. The mixture was periodically tested for fibrinogen-clotting activity (6) for 30 min. A comparable mixture of thrombin and hirudin showed immediate loss of fibrinogen-clotting activity.

Results and Discussion

Amino Acid Sequence Determination

The strategy for Edman degradation of crotalase and its overlapping fragments and the resulting amino acid sequence of the enzyme is given in Fig. 1. Here, the sequence analysis relied both on peptides generated by enzymatic digestion and on chemical cleavage of fragments produced by cyanogen bromide cleavage of the pyridylethylated or carboxymethylated protein.

Primary Structural Features of Crotalase

Comparison of the amino acid sequence of crotalase to those of chymotrypsin, kallikrein, trypsin, and thrombin [Fig. 2, chymotrypsinogen numbering (15)] reveals typical characteristics of a serine proteinase. The residues of the catalytic triad, His 57, Asp 102, and Ser 195, are in their expected positions, flanked by sequences that are

	16	30	36A		50	57 60		60I	68	70
CHY	IVNGEEAVPGSW	PWQVSLQI	OKT-GF	HFCGGSLI	NENWVV	TAAHCGVT	TSD		VVVA	GEF
TAL TDV	VVGGYNCEMNSQ	PWQVAVY: DVOUCTN	(FGE	ATCCCATT	DPSWVI	TAAHCATD	NYQVWL	G-RNNL	YEDEPF	AQH
THR	IVGGYTCGANTV IVEGQDAEVGLS	DMUANT LI	SGI	TICCASIT	NSQWVV	TAAHCIKS	GIQ		-VRL	GED
	VIGGDECNINEH									
	VIGGDECNINEH									
Mut	VIGGDECNINEH	RFLVALYI	OGLSGT	FLCGGTLT	NOEWVL	TAOHC	NRS	I	-MNTYL	GMH
Bot	VIGGDECDINEH	PFLAFMY-	YSPQ	YFCGMTLI	NOEWVL	TAAHC	DKT	- Y 	-MRIYI	GTH
Bat	VIGGDECDINEH	PFLAFMY-	YSPR	YFCGMTLI	NQEWVL	ТААНС	NRR	E	-MRIHL	GKV
Fla	VIGGDECNINEH	PFLVALYI	DAWSGR	FLCGGTLI	NPEWVL	таанс	DSK	NFK	-MKL	GAH
Bil	IIGGDECNINEH	RFLVALYI	DVWSGS	FLCGGTLI	NQEWVL	ТААНС	NMS	N	-IYIYL	GMH
Cal	VIGGDECNINEH	RFLVALYN	ISRSRT	LFCGGTLI	NQEWVL	таанс	ERN	NF	RIKL	GIH
			• •							
01117	77A 81		90 777777	97A 1		110		20	129	
	DQGSSS-EKIQK									
	RLVSQSFPHPGF NINVVEGNEQFI									
	SRTRYERKVEKI									
	DRSVQFDKEQRR									
Anc	RKSEKFDDEQER	YPKKRYFI	RCNKT	RTSWDE	DTMT.T-	RENKRVNN	SEHTAP	LSTPSNI	DDT	V
Mut	NKNVKFDDEQRR	YPKKÄYFE	RCNKN	FTKWDE	DI	RENREVRE	SAHIFP	LSLPSNI	 	E
	TRSVANDDEVIR									
	AGSVANYDEVVR									
	SQKVLNEDEQIR									
	NQSVQFDDEERR									
Cal	SKKVPNEDEQTR	VPKEKE	FCLSS	KNYTLWDK	DIMLI-	RLDSPVSN	SEHIAP	LSLPSS	PS	V
	100 140	1 4 6			~ ~					
CUV	133 140	031 0464	A 150	10171	60	170		other second	184 18	
	GTTCVTTGWGLT	RYTNANTE	DRLQQ	ASLPLLSN'	INCKK-	YWG		-AMICA	3A	SGV
KAL	GTTCVTTGWGLT GSTCLASGWGSI	RYTNANTE PPDGLELS	DRLQQ	ASLPLLSN' DDLQCVNI	INCKK- DLLSNE	YWG КСVЕАНКЕ	-EVTD-	-AMICA	3A SEMD	SGV GGK
KAL TRY	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT	RYTNANTE FPDGLELS KSSG- TS -	PDRLQQ SYP	ASLPLLSN' DDLQCVNI DVLKCLKA	INCKK- DLLSNE PILSDS	YWG KCVEAHKE SCKSAYPG	-EVTD- -QITSN	-AMICA -LMLCA MFCA	3A SEMD SYLE	SGV GGK DDK
KAL TRY THB	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR	RYTNANTE FPDGLELS KSSG- TS - RETWTTSV	PDRLQQ SYP VAEVQP	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL	INCKK- DLLSNE PILSDS PLVERP	YWG KCVEAHKE SCKSAYPG VCKASTRI	-EVTD- -QITSN -RITDN	-AMICA -LMLCA MFCA MFCA	5A EMD YLE SYKPGE	SGV GGK DDK GKR
KAL TRY THB Cro	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ-	RYTNANTE IPDGLELS KSSG-TS- RETWTTSV TTSE	PDRLQQ SYP VAEVQP PQETLP	ASLPLLSN' DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII	INCKK- DLLSNE PILSDS PLVERP NLLDYE	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ	-EVTD- -QITSN -RITDN FRLPAT	-AMICA -LMLCA MFCA MFCA SRTLCA	JA SEMD SYLE SYKPGE SVLE	SGV GGK DDK GKR GG-
KAL TRY THB Cro Anc	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR	RYTNANTE IPDGLELS KSSG-TS- RETWTTSV TTSE INRR-	PDRLQQ YP VAEVQP PQETLP -IDVLS	ASLPLLSN' DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANI DEPRCANI	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFT	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK	-EVTD- -QITSN -RITDN FRLPAT MPKKG-	-AMICA -LMLCA -MFCA -MFCA SRTLCA -RVLCA	SA SEMD SYLE SYKPGE SVLE SDLR	SGV GGK DDK GKR GG- GR-
KAL TRY THB Cro Anc Mut	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSDCRVMGWGS-	RYTNANTE FPDGLELS KSSG-TS- RETWTTSV TTSE INRR- ITSP-	PDRLQQ YP VAEVQP PQETLP -IDVLS -PETLP	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANI DEPRCANI DVPHCANI	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFT NLFNYT	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCRGAYP-	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T	-AMICA -LMLCA MFCA MFCA SRTLCA -RVLCA -RVLCA	3A SEMD SYLE SYKPGE SVLE SDLR SVLE	SGV GGK DDK GKR GG- GR- GG-
KAL TRY THB Cro Anc Mut Bot Bat	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSDCRVMGWGS- DSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA-	RYTNANTE FPDGLELS KSSG-TS- RETWTTSV TTSE INRR- ITSP- ITTS- ITTS-	PDRLQQ YP VAEVQP QETLP -IDVLS -PETLP -EDTYP -EDTYP	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANI DEPRCANI DVPHCANI DVPHCANI DVPHCANI	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFT NLFNYT NLFNNT NLFNNT	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRKI VCRGAYP- VCREAYNG VCREAYNG	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA-	-AMICA -IMLCA -MFCA -MFCA SRTLCA -RVLCA -KVLCA -KVLCA -KTLCA	SA SEMD SYLE SYKPGE SVLE SVLE SVLQ SVLQ	SGV GGK DDK GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bot Bat Fla	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSDCRVMGWGQ- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGS-	RYTNANTE IPDGLELS KSSG-TS- RETWTTSV ITSE ITSP- ITSSP- ITTS ITTS ITTS ITTS	PDRLQQ YP VAEVQP QETLP -IDVLS -PETLP -EDTYP -EDTYP -ETTP	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFT NLFNYT NLFNNT NLFNNT NLFNNT	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCRGAYP- VCREAYNG VCREAYNG ECKPGYPE	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- L-LPEY	-AMICA -LMLCA -MFCA SRTLCA -RVLCA -RVLCA -KVLCA -KTLCA -KTLCA	SA SEMD SYLE SYKPGE SULE SULE SVLQ SVLQ SVLQ	SGV GGK DDK GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bot Bat Fla Bil	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSDCRVMGWGG- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGS- GSVCRIMGWGG-	RYTNANTE IPDGLELS KSSG-TS- RETWTTSV ITSE ITSP- ITSSP- ITTS ITTS ITTS ITSP-	PORLQQ YP VAEVQP PQETLP IDVLS -PETLP EDTYP EDTYP -EETFP NETLP	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFT NLFNYT NLFNNT NLFNNT NLLDDV NLFNYT	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG VCREAYNG ECKPGYPE VCRGVFP-	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- L-LPEY -RLPER	-AMICA -IMLCA -MFCA SRTLCA -RVLCA -RVLCA -KVLCA -KTLCA -KTLCA SRILCA	5A 5EMD 5YLE 5VLE 5VLE 5VLQ 5VLQ 5VLQ 5VLQ 5VLE	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bot Bat Fla Bil	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSDCRVMGWGQ- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGS-	RYTNANTE IPDGLELS KSSG-TS- RETWTTSV ITSE ITSP- ITSSP- ITTS ITTS ITTS ITSP-	PORLQQ YP VAEVQP PQETLP IDVLS -PETLP EDTYP EDTYP -EETFP NETLP	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFT NLFNYT NLFNNT NLFNNT NLLDDV NLFNYT	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG VCREAYNG ECKPGYPE VCRGVFP-	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- L-LPEY -RLPER	-AMICA -IMLCA -MFCA SRTLCA -RVLCA -RVLCA -KVLCA -KTLCA -KTLCA SRILCA	5A 5EMD 5YLE 5VLE 5VLE 5VLQ 5VLQ 5VLQ 5VLQ 5VLE	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bot Bat Fla Bil	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSVCRVMGWGQ- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGG- GSVCRIMGWGG- GSVCRIMGWGR-	RYTNANTF TPDGLELS KSSG-TS RETWTTSV TTSF ITSP ITSS ITTS ITSS ITSS ITSS ITSS ITSS ITSS ISST	PDRLQQ PDRLQQ PDRLP PQETLP PDTLP PDTLP EDTYP EDTYP EETFP NETLP KETYP	ASIPLISN DDIQCVNI DVIKCLKA SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFTI NLFNNT NLFNNT NLLDDV NLFNYT NLLEYE	YWG KCVEAHKE SCKSAYPG VCRTAHPQ MCHGLFRK VCRGAYP- VCREAYNG VCREAYNG ECKPGYPE VCRGVFP- MCRAPYPE	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- LPA- L-LPEY -RLPER FGLPAT	-AMICA -LMLCA -MFCA SRTLCA SRTLCA -RVLCA -KVLCA -KTLCA -KTLCA SRTLCA SRTLCA	5A 5EMD 5YLE 5VLE 5VLE 5VLQ 5VLQ 5VLQ 5VLQ 5VLE	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bat Fla Bil Cal	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNT GSVCRVMGWGQ- GSVCRVMGWGQ- GSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGS- GSVCRVMGWGT- GSVCRIMGWGR- 189 195	RYTNANTF TPDGLELS KSSG-TS- RETWTTSV TTSP- ITSP- ITTS- ITTS- ITTS- ITTS- ITSP- 200 204	PORLQQ PORLQQ VAEVQP QETLP IDVLS PETLP EDTYP EDTYP EDTYP NETLP KETYP	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DEPRCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFTI NLFNNT NLFNNT NLLDDV NLFNYT NLLBYE	YWG KCVEAHKE SCKSAYPG VCKTAHPQ WCRGLFRKI VCRGAYP- VCREAYNG ECKPGYPE VCREAYNG ECKPGYPE VCREAYNG ECKPGYPE MCRAPYPE 1A 226 2	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- L-LPEY -RLPER FGLPAT	-AMICA -IMLCA -MFCA -MFCA SRTLCA -RVLCA -KVLCA -KTLCA -KTLCA SRTLCA SRTLCA 240	3A 3EMD 3YLE 3YKPGE 3VLE 3VLE 3VLQ 3VLQ 3VLE 3ILE	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bat Fla Bil Cal	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNT GSVCRAMGWGQ- GSVCRVMGWGQ- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGR- GSVCRIMGWGR- 189 195 -SSCMGDSGGPL	RYTNANTF TPDGLELS KSSG-TS- RETWTTSV TTSP- ITSP- ITTS- ITTS- ITTS- ITSP- 200 204 VCKKNC	PDRLQQ PDRLQQ PQETLP IDVLS PETLP EDTYP EDTYP EDTYP ETTP NETLP KETYP IA 21 GAWTLV	ASIPLLSN DDIQCYNI DVIKCLKA SVLQVVNLI DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII 0 0 -GIVSWGS:	INCKK- DLLSNE PILSDS PLVERP NLLDYE NLHNFT NLFNYT NLFNNT NLLDV NLLDV NLLEYE 22 STCS-T.	YWG KCVEAHKE SCKSAYPG VCKTAHTQ WCRGLFRKI VCRGAYP- VCREAYNG ECKPGYPE VCRGVFP- MCRAPYPE 1A 226 2 STPGVYAR	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- L-LPEY -RLPER FGLPAT 30 VTALVN	-AMICA -IMLCA -MFCA -MFCA SRTLCA -RVLCA -KVLCA -KTLCA -KTLCA SRTLCA SRTLCA 240 WVQQTL	3A SEMD SYLE SYKPGE SULE SULQ SVLQ SVLQ SVLE SILE AAN	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bot Fla Bil Cal CHY KAL	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNT GSVCRAMGWGQ- GSVCRVMGWGS- DSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGG- GSVCRIMGWGR- 189 195 -SSCMGDSGCPL - DTCKGDSGGPL	RYTNANTF IPDGLELS KSSG-TS- TTSF ITSF ITSF ITSF ITSF ITSF ISFT 200 204 VCKKNC IC	PORLQQ 	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII 0 -GIVSWGS: -GITSWGFI	PNCKK- DILSNE PILSDS PLVERP NILDYE NILDYE NILFNYT NLFNNT NLFNNT NLLDV. NLFNYT NLLDV. 22 STCS-T. NPCGEP	YWG KCVEAHKE SCKSAYPG VCRASTRI VCRTAHPQ MCHGLFRK VCRGAYP- VCREAYNG VCREAYNG VCREAYNG VCREAYNG VCRGVFP- MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- L-LPA- L-LPA- RLPER FGLPAT 30 VTALVNI LIKFTP	-AMICA -IMLCA -MFCA -MFCAS SRTLCA -RVLCA -KVLCA -KTLCA SRTLCA SRTLCA SRTLCA 240 WVQQTLL WIKEVM	A SEMD SYLE SYLE SULE SULQ SULQ SULQ SULQ SULE SULE SULE SULE	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bat Fla Bil Cal CHY KAL TRY	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGT- GSVCRIMGWGT- GSVCRIMGWGR- 189 195 -SSCMGDSGCPL -DTCKGDSGGPL -DSCQGDSGGPV	RYTNANTF IPDGLELS KSSG-TS- RETWTTSV ITSF ITSP- ITSP- ITSP- ITSP- ITSP- ITSP- 200 204 VCKKNC ICS	PORLQQ SYP VAEVQP PQETLP -IDVLS PETLP -EDTYP -EDTYP -EETFP -KETYP IA 21 GAWTLV IGVLQ- -SGKLQ-	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII OVPHCANII 0 -GIVSWGS -GITSWGF	INCKK- DILSNE PILSDS PILVERP NULDYE NULDYE NULFNYT NULFNNT NULFNYT NULFNYT NULEYE 22 STCS-T NPCGEP SGCAQK	YWG KCVEAHKE SCKSAYPG VCRASTRI VCRTAHPQ MCHGLFRK VCRGAYP- VCREAYNG VCREAYNG VCREAYNG VCRGVFP- MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK	-EVTD- -QITSN -RITDN FRLPAT. MPKKG- -RMP-T -LPA- -LPA- L-LPEY- RLPER FGLPAT: 30 VTALVNI LIKFTPI VCNYVS	-AMICA -IMLCA -MFCA -MFCAS SRTLCA -RVLCA -KVLCA -KTLCA -KTLCA SRTLCA SRTLCA SRTLCA 240 WVQQTLL WIKEVMI	A EMD SYLE SYLE SVLE SVLQ VLQ VLQ VLE SILE AN KENP ASN	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG-
KAL TRY THB Cro Anc Mut Bat Fla Bil Cal CHY KAL TRY THB	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNT GSVCRAMGWGQ- GSVCRVMGWGS- DSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGG- GSVCRIMGWGR- 189 195 -SSCMGDSGCPL - DTCKGDSGGPL	RYTNANTF IPDGLELS KSSG-TS- RETWTTSV ITSF ITSP- ITSP- ITSP- ITSP- ITSP- 200 204 VCKKNC VCKNC VCS VMKSPYN	DRLQQ DRLQQ YP AEVQP QETLP IDVLS PPTLP EDTYP EDTYP ETTP NETLP KETYP IA 21 SAWTLV IGVLQ- IGVLQ- IRWYQM	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII 0 -GIVSWGS: -GIVSWG-1 -GIVSWG-1	INCKK- DILSNE PILSDS PIVERP NILDYE NILDYE NILFNNT NILFNNT NILFNNT NILEYE 22 STCS-T. NPCGEP SSCAQK EGCDRD	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG VCREAYNG VCREAYNG VCRGVFP- MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK NKPGVYTK GKYGFYTH	-EVTD- -QITSN -RITDN FRIPAT. MPKKG- -LPA- -LPA- L-LPEY- RLPER. FGLPAT: 30 VTALVN LIKFTPI VCNYVSI VFRLKKI	-AMICA -IMICA -MFCA SRTLCA SRTLCA -RVLCA -KVLCA -KTLCA SRTLCA SRTLCA 240 WVQQTLL WIKEVMI WIKEVMI WIKEVII	ANN CENP	SGV GGK DDK GGG GGG GGG GGG GGG GGG GGG GGG
KAL TRY THB Cro Anc Mut Bat Fla Bil Cal CHY KAL TRY THB Cro	GTTCVTTGWGLT GSTCLASGWGSI GTCCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGF- GSVCRIMGWGF- H89 195 -SSCMGDSGGPL -DTCKGDSGGPL -DSCGGDSGGPF	RYTNANTH TPDGLELS KSSG-TS- RETWTTS- TTSF ITSP- ITTS- ITTS- ITTS- ITTS- ITSP- 200 204 VCKKNC VCKKNC VCKN VCS VMKSPYNN IC	DRLQQ DRLQQ QETLP IDVLS PETLP EDTYP EDTYP EDTYP KETYP KETYP IA 21 IAWTLV IGUC- IGKLQ- IGKLQ- IGKLQ-	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID OPPHCANID 0 -GIVSWGS: -GIVSWG- -GIVSWG- -GIVSWG-	INCKK- DILSNE PILSDS PLVERP NILDYE NILFNYT NIFNNT NIFNYT NIFNYT NIFNYT NIFNYT NIFNYT NIFNYT SCCAQK SCCAQK DCCAQP	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG ECKPGYPE MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK NKPGVYTK GKYGFYTH DKPGLYTK	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- -LPA- -RLPER FGLPAT 30 VTALVNI LIKFTPI VCNYVSI VVRLKK VFDHLD	-AMICA -IMLCA -MFCA -MFCA SRTLCA -RVLCA -KVLCA -KTLCA KTLCA SRTLCA SRTLCA 240 WVQCTLA WIKEVMI WIKEVIA WIKEVIA WIKEVIA WICKUIA	ANN CANNERS	SGV GGK DDK GG GG GG GG GG GG GG GG GG GG GG GG GG
KAL TRY THB Cro Anc Mut Bat Fla Bil Cal CHY KAL TRY THB Cro Anc	GTTCVTTGWGLT GSTCLASGWGSI GTCCLISGWGNT GFKGRVTGWGNT GSVCRAMGWGQ- GSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGG- GSVCRIMGWGG- GSVCRIMGWGG- 189 195 -SSGMGDSGGPL -DTCKGDSGGPL -DSCOGDSGGPV GDACEGDSGGPL IDTCNRDSGGPL	RYTNANTF TPDGLELS KSSG-TS- RETWTTSV TTSP- ITSP- ITSP- ITSP- 200 204 YCKKN-C IC VC VCKN-C IC IC IC IC	DRLQQ DRLQQ CALVAN VAEVQP QETLP DVDS PETLP EDTYP EDTYP EDTYP EDTYP NETLP NETLP NETLP NETLP NETLP NETLP NETLP IGVLQ- GRUQ- IRWYQM IGQFQ- IEELH-	ASLPLLSN DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DEPRCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII OVPHCANII 0 -GIVSWGS -GITSWGF -GIVSWG- -GIVFWGP -GIVFWGP	INCKK- DILSNE PILSDS PLVERP NILDYE NILDYE NILFNYT NILFNNT NILFNNT NILFNYT NILFNYT NILFYYT STCS-T SGCAQK EGCDRD DPCAQP NPCAQP	YWG KCVEAHKE SCKSAYPG VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG ECKPGYPE VCREAYNG ECKPGYPE 1A 226 2 STPGVYAR KKPGIYTK NKPGIYTK NKPGLYTK NKPALYTS	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- -LPA- -LPA- -LPA- SO VILLER FGLPAT 30 VTALVN LIKFTP VCNVS VFRLKK VEDHLD VEDHLD	-AMICA -IMLCA -IMLCA -MFCA SRTLCA -KVLCA KVLCA -KTLCA -KTLCA SRTLCA SRTLCA 240 WQQTLA WIKEVMA WIKUTA WIKUTA WIQKIII WIQNII WINVIA	A EMD SYLE SYL-E SVLE SVLE SVLQ VLQ VLQ VLQ VLQ VLQ VLQ VLQ S	SGV GGK DDK GG GG GG GG GG GG GG GG GG GG GG GG GG
KALL TRY THB Cro Anc Mut Bat Bat Bill Cal CHY KAL TRY KAL TRY Cro Anc Chy KAL TRY Bot	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFKGRVTGWGNR GSVCRAMGWGQ- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWG- GSVCRIMGWG- GSVCRIMGWG- GSVCRIMGWG- DSCQCDSGCPL -DTCKGDSGCPL DTCKGDSGCPL IDTCNRDSGCPL IDTCRRDSGCPL IDTCGCDSGCPL	RYTNANTE IPDGLELS KSSG-TS RETWTTSV ITSI ITSP- ISPT- 200 204 VCKKN-CO IC VCCKSPUND IC IC IC IC IC IC	DRLQQ DRLQQ YP VAEVQP 'IDVLS PETTP EDTYP EDTYP EETYP KETYP IA 21 SAWTLV IGVLQ- IGVLQ- IRWYQM IGQLQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ-	ASLPLLSN DDLQCVNI DDLQCVNI DVLKCLKA SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII O -GIVSWGS -GIVSWG- -GIVSWG- -GIVFWGP -GIVFWGP -GILSWGS	INCKK- DILSNE PILSDS PILVERP NILDYE NILDYE NILFNYT NILFNYT NILFNYT NILFNYT NILFNYT SGCAQK EGCDRD DPCAQP DPCAQP DPCAPP	YWG KCVEAHKE SCKSAYPG VCRASTRI VCRTAHPQ MCHGLFRK VCRGAYP- VCREAYNG VCREAYNG VCREAYNG VCREYP- MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK NKPGIYTK NKPGLYTK NKPALYTS DKPGVYTK RKPAFYTK	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- L-LPA- L-LPEY- FGLPAT 30 VTALVNI LIKFTPI VCNYVSI VFRLKKI VFRLKKI VFDYLDI VFDYLDI	-AMICA -IMICA -IMICA -MFCA SRTLCA -KVLCA -KVLCA -KTLCA -KTLCA SRTLCA SRTLCA SRTLCA VQQTLL WIKEVM WIQQTLL WIQSII WIQSII WIQSII WIQSII	ANN CANNERS ANN CANNE CANNERS ANN CANNE CANNE CANNE CANNE CANNE CANNE CANNE CANNE CANNE CANNE CANNE CA	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- CCP CCP CCP TCPP
KAL TRY THB Croc Mut Bot Fla Bil Cal CHY KAL TRY THB Croc Ancc Bot Bot	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFXGRVTGWGNR GSVCRAMGWGQ- GSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- DSCCRIMGWGA- DSCCGDSGPL DTCKGDSGGPL IDTCRDSGGPL IDTCGGDSGGPL IDTCGGDSGGPL	RYTNANTE IPDGLELS KSSG-TS- RETWTTSV ITSI ITSP- ITSP- ITSP ITSP- ITSP- ITSP- ITSP- ITSP- ITSP- ITSP- ITSP- 200 204 VCKKNC VCKKNC VCKKNS VMKSPYNN IC	PORLQQ PORLQQ PORTLP	ASLPLLSN DDLQCVNI DDLQCVNI SVLQVVNL SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII OVPHCANII OVPHCANII GVFKGP -GIVSWGS -GIVSWGP -GIVFWGP -GIVARGP -GIVFWGP -GILSWGS	INCKK- DILSNE PILSDS PIVERP NILDYE NILDYE NILFNNT NILFNNT NILFNNT NILFNNT NILDV. 22 STCS-T. NPCGEP SGCAQK EGCDRD DPCAQP DPCAQP DPCAQP DPCAAP	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG ECKPGYPE MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK NKPGVYTK GKYGFYTH DKPGLYTK NKPALYTS DKPGVYTK RKPAFYTK	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T LPA- -LPA- -LPA- -LPA- -RLPER FGLPAT 30 VTALVNI UTATVNI VFRLKK VFDLLD VFDYLD VFDYLP VFDYLP	-AMICA -IMICA -IMICA -MFCA SRTLCA SRTLCA -RVLCA -KVLCA -KTLCA SRTLCA SRTLCA 240 WVQQTL WIQCTL WIQVI WIQVI WIQVI WIQVI WIQVI WIQVI WIQVI WIQVI WIQVI WIQVI	AAN CENP AAN CAN CENP AAN CAN CENP AAN CAN CENP AAN CAN CENP AAN CAN CAN CENP AAN CAN CAN CENP AAN CAN CAN CAN CAN CAN CAN CAN CAN CAN	SGV GGK DDK GKR- GG- GG- GG- GG- CSP -CSP TCPP
KAL TRY THB Croo Anc Bot Fla Bil Cal CHY KAL TRY THB Croo Anc Bot Bat Fla	GTTCVTTGWGLT GSTCLASGWGSI GTCCLISGWGNT GFKGRVTGWGNR GSVCRIMGWGQ- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGG- GSVCRIMGWGG- BSVCRIMGWGG- GSVCRIMGWGG- DSCCRIMGWGG- GSVCRIMGWGG- DSCCRIMGWGGG - DSCCRIMGWGGG - DSCCRIMGWGGG - DSCCRIMGWGGG - DSCCRIMGWGGG - DSCCRIMGWGGG - DSCCRIMGWGGG - DSCCRIMGWGGG - DSCCRIMGWGG - DSCCRIMGWGG - DSCCRIMGWGG - DSCCRIMGWGG - DSCCGDSGGPL - DTCCGDSGGPL IDTCNRDSGGPL IDTCNRDSGGPL IDTCNRDSGGPL IDTCCGDSGGPL IDTCCGDSGGPL IDTCCGDSGGPL IDTCCGDSGGPL IDTCCGDSGGPL	RYTNANTE IPDGLELS KSSG-TS- RETWTTSV ITTSI ITTSP ITSP ITSP VCKKNC VCKKNC VCCS VMKSPYNN ICN ICN <	PDRLQQ PDRLQQ PDRLQQ VAEVQP VAEVQP VDVLP IDVLS PETLP EDTYP EDTYP EDTYP NETLP NETLP NETLP NETLP IGQTQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ-	ASLPLLSN DDLQCVNI DDLQCVNI SVLQVVNL SVLQVVNL DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII DVPHCANII OUPRCVNII DVPHCANII GVFKGS -GIVSWG- GIVSWG- GIVSWG- GIVSWG- 	INCKK- DILSNE PILSDS PIVERP NILDYE NILDYE NILFNYT NILFNNT NILFNYT NILFNYT NILFNYT NILFNYT NILFNYT NILFYE STCS-T NICSCAQK EGCDRD DPCAQP DPCAQP DPCAP DPCAP DPCAP HPCGQS	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG VCREAYNG VCREAYNG VCRGVFP- MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK NKPGVYTK NKPGLYTK NKPGLYTK RKPAFYTK RKPAFYTK RKPAFYTK	-EVTD- -QITSN -RITDN FRLPAT MPKKG- -RMP-T -LPA- -LPA- -LPA- -LPA- RLPER FGLPAT 30 VTALVNI LIKFTP VFDYLD VFDYLD VFDYLD VFDYLP VFDYLP VFDYLP	-AMICA -IMICA -IMICA -MFCA SRTLCA -KVLCA -KVLCA -KTLCA -KTLCA SRTLCA SRTLCA SRTLCA 240 WVQQTLL WIQCTLL WIQSVI WIQSVI WIQSVI WIQSVI WIQSVI WIQSII WIQSII WIQSII WIQSII	AGNKTA AGNKTA AGNKTA AGNKTA AGNKTA	SGV GGK DDK GKR- GG- GG- GG- GG- GG- SG- SG- SG- CSP TCPP TCCPP TCCPP
KAL TRY THB Croo Bat Fla Bil Cal CHY KAL TRY Croo Anc Croy Anc Sat Fla Bil Bil Bil Bil Bat	GTTCVTTGWGLT GSTCLASGWGSI GTQCLISGWGNT GFXGRVTGWGNR GSVCRAMGWGQ- GSVCRVMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- GSVCRIMGWGA- DSCCRIMGWGA- DSCCGDSGPL DTCKGDSGGPL IDTCRDSGGPL IDTCGGDSGGPL IDTCGGDSGGPL	RYTNANTE IPDGLELS KSSG-TS- RETWTTS- ITTSI ITSI ITTSI 200 204 VCKKNC VCCS VKSPYNN ICN ICN ICN ICN ICN ICN ICN ICN ICN	DRLQQ DRLQQ CALLQ QETLP IDVLS PETLP EDTYP EDTYP EDTYP KETYP KETYP IA 21 IGQTQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ- IGQFQ-	ASLPLLSN DDLQCVNI DDLQCVNI SVLQVVNL DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID DVPHCANID GIVSWGS -GIVSWG-	INCKK- DILSNE PILSDS PIVERP NILDYE NILDYE NILFNYT NILFNNT NILFNYT NILFNYT NILFNYT NILFNYT NILFNYT NILFYE SCCAQK EGCDRD DPCAQP DPCAQP DPCAQP HPCQQS KRCAQP	YWG KCVEAHKE SCKSAYPG VCKASTRI VCRTAHPQ MCHGLFRK VCREAYNG VCREAYNG VCREAYNG ECKPGYPE MCRAPYPE 1A 226 2 STPGVYAR KKPGIYTK NKPGVYTK NKPGLYTK NKPALYTS DKPGCYTK RKPAFYTK RKPAFYTK RKPAFYTK RKPALYSK	-EVTD- -QITSN -RIPAT MPKKG- -RMP-T -LPA- -LPA- -LPA- -LPA- -RLPER FGLPAT 30 VTALVNI LIKFTPI VCNYVSI VCNYVSI VCNYVSI VFDLLDI VFDYLDI VFDYLDI VFDYLDI VFDYLDI VFDYLDI VFDYLDI	-AMICA -IMLCA -IMLCA -MFCA SRTLCA -KVLCA -KVLCA -KTLCA -KTLCA SRTLCA SRTLCA SRTLCA 240 WVQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL WIQULL	AGNKTA AGNKTA AGNKTV	SGV GGK DDK GG- GG- GG- GG- GG- GG- GG- GG- CSP -CSP TCCP TCCP TCCP TCCP TCCP

Fig. 2 Comparison of amino acid sequences of thrombin-like venom enzymes with bovine thrombin and with other serine proteinases. Dashes indicate gaps introduced to optimize the alignment. All residue positions correspond to the first digit of the residue number. CHY, bovine chymotrypsin (35); KAL, rat pancreatic kallikrein (36); TRY, bovine trypsin (37); THB, bovine thrombin B-chain (38); Cro, crotalase (39); Anc, ancrod (40); Mut, *Lachesis muta muta* (41); Bot, bothrombin (42); Bat, batroxobin (12); Fla, flavoxobin (43), Bil, bilineobin (13), Cal, calobin (44). Chymotrypsinogen numbering (15)

largely conserved. The only nonconserved neighboring residue is Arg 193, a position which turns out to be variable also in other thrombinlike enzymes (TLEs) (Fig. 2). As in all other serine proteinases position 194 of crotalase is occupied by an Asp residue which makes an important contribution to stabilizing the active center in a catalytically active configuration. In general this is accomplished by forming a salt bridge with the enzyme's N-terminal alpha amino group (11) and such a salt bridge is a feature of the modeled three-dimensional structure of crotalase (4).

Since crotalase has no free sulfhydryl groups (1) it is likely that all of its cysteines are disulfide bonded to one another. Based on homology with trypsin (12) and with another thrombin-like venom enzyme (13), the disulfide bridges of crotalase are predicted to link the following sequence positions: 22-157, 42-58, 91-250, 136-201, 168-182, and 191-220.

The overall rates of sequence identity between crotalase and the mammalian serine proteinases in Fig. 2 are kallikrein 33%, trypsin 32%, thrombin 30%, and chymotrypsin 24%. As might be expected, the

TLEs exhibit a much higher degree of similarity among themselves with a mean rate of identity of 64%. The primary structures of crotalase and the other TLE are also strikingly similar to venom serine proteinases that do not clot fibrinogen (62% mean rate of identity) (14).

Modeling of the spatial structure of the crotalase molecule shows its core to be comprised, like thrombin (15), of two six-stranded β -barrels between which lies the active center (Fig. 3). Other secondary structural features are three α -helices formed from segments Tyr 165-Ala-171, Val 231-Leu 235, and Trp 237-Ala 243, homologous to three of thrombin's five B-chain helices.

A single consensus site for asparagine-linked glycosylation, Asn 94, was first inferred from a sequencing blank two steps before Thr 96 and later confirmed after enzymatic deglycosylation of the crotalase molecule. The total neutral and amino sugar content of crotalase is 8.34% with molar ratios of mannose 3, galactose 6.1, *N*-acetylglucos-amine 10.5, *N*-acetylglalactosamine 0.6, and fucose 1.5, a composition consistent with a complex type of asparagine-linked chain (16) and not with *O*-linkage. Analysis for sialic acid was not done since earlier



Fig. 3 Secondary structure diagram of the model of crotalase. The program INSIGHT II (Kabsch-Sander algorithm) was used to represent beta sheets as arrows (ribbons), helices as cylinders, and turns as sleeves

analyses yielded values that varied from one crotalase preparation to another (1, 5). Although *N*-acetylgalactosamine, as found here, does not ordinarily occur in asparagine-linked carbohydrate chains, it has now been demonstrated as subterminal to sialic acid residues of batroxobin (17, 18) and ancrod (19), partially or completely replacing galactose in that structural role. The unevenness of molar ratios among the individual sugars of crotalase most likely reflects the heterogeneity of carbohydrate chain structure often exhibited by glycoproteins.

From the amino acid sequence of crotalase its molecular weight can be computed as 26,606 without carbohydrate or 29,027 adjusted for the sugar content just given. From the enzyme's tryptophan, tyrosine, and cystine content its molar absorption coefficient, ϵ (280)(M⁻¹ cm⁻¹), can be calculated as 44,180 (20, 21). Then, assuming a molecular weight of 29,027, the absorbance of crotalase at 280 nm, 1%, cm⁻¹ is 15.2.

Comparison of Crotalase Sequence with Thrombin

An important role in fibrinogen recognition and cleavage specificity has been attributed to the 60A-60D (YPPW) insertion loop of thrombin which forms part of the rim of its active site cleft (22). This feature is absent in crotalase, leaving a more open entrance to the active site (4). This feature is also absent in the mutant des-PPW thrombin (23). But both crotalase and the mutant thrombin exhibit, nonetheless, a narrow specificity for fibrinogen as a coagulant substrate. By contrast, a component of the opposite rim of the active site cleft of thrombin, ETW 146-148, also absent in crotalase, has been shown in mutant forms of thrombin to be critically important to its fibrinogen-clotting activity (24, 25). Thus, whatever the role of ETW 146-148 in the fibrinogenclotting activity of thrombin, the absence of this motif in clotting-active crotalase suggests a different mode of action for the venom enzyme.

Crotalase shares with thrombin and with other enzymes that cleave basic peptide bonds (e.g. kallikrein and trypsin, but not chymotrypsin, Fig. 2) a residue, Asp 189, located at the bottom of a "specificity pocket" near the active site (4, 26). This residue forms a salt bridge with the basic residue of the scissile bond, whose cleavage in the case of fibrinogen releases fibrinopeptide A (FPA) from the parent molecule. Specificity for the release of FPA by thrombin is also served by a hydrophobic binding site near the catalytic center comprised of Tyr 60A, Trp 60D, Leu 99, Ile 174, and Trp 215 which interacts, in particular, with Phe 8 of the fibrinopeptide (26). As crotalase lacks any part of the 60 insertion loop, this hydrophobic site is formed instead by Trp 97A, Phe 174A, Leu 176, and Trp 215 and is favorably configured to accommodate Phe 8 of FPA (4). Other residues of crotalase that are homologous to thrombin's FPA-interacting amino acids are Lys 97 and Gly 216 (26).

Of the seven basic residues of thrombin that form its heparinbinding site (exosite 2) (15), only one is represented in crotalase, this in keeping with the insensitivity of crotalase to heparin (27).

Fibrinogen Recognition Exosite (FRE)

The remarkably narrow specificity of thrombin for fibrinogen and other macromolecular substrates is thought to reside largely in a surface groove that extends "southeast" from the active site cleft, using the conventional orientation of the molecule (28). The interaction of this fibrinogen recognition site (FRE) with a complementary site on the fibrinogen molecule has been, to a major extent, attributed to several positively charged residues at or immediately subjacent to this groove. Eleven basic residues of thrombin [R35, K36, R67, K70, R73, R75, R77A, K81, K109, R110 (human and bovine), K149E (human)] have been implicated in one way or another in this electrostatic interaction with fibrinogen (29, 30). However, of these 11 basic residues of the thrombin molecule, crotalase possesses only three and none of the three is conserved among the other TLEs (Fig. 2).

The absence in the TLEs of a cationic region homologous to the FRE of thrombin raises the question of whether the corresponding region of the crotalase might contain a similar, positively charged site made up of nonhomologous residues. But, in contrast to thrombin's net positive charge of 5 in the critical center of the FRE (positions 67-82), the same crotalase sequence has a positive charge of only 1 and most of the other TLEs have a negative net charge, usually of 1 (Fig. 2).

In addition to the foregoing structural grounds for doubting that crotalase recognizes fibrinogen through an interaction at the same region of the molecule used by thrombin, the following three function-

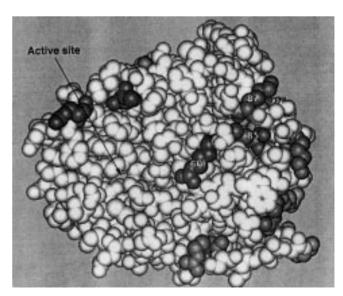


Fig. 4 Space-filling image of the model of crotalase generated using INSIGHT II. Basic amino acids are delineated as dark areas. Arrow points to active site cleft. Same orientation as Fig. 3

al considerations point in the same direction. First, crotalase is not inactivated by the thrombin inhibitor hirudin. During 30 min of incubation (Experimental Procedures section, Hirudin Inhibition) hirudin-crotalase mixtures showed no change from control (hirudinfree) fibrinogen clotting times. In contrast, the thrombin-hirudin mixtures exhibited immediate loss of fibrinogen-clotting activity. Since it has now been shown directly that hirudin inhibits thrombin by binding to its FRE (31, 32), the failure of hirudin to affect the clotting activity of crotalase suggests a different fibrinogen recognition site. And hirudin-insensitivity is shared with other TLEs (3). A second functional consideration is that crotalase does not promote platelet aggregation (27). Triggering of platelet aggregation by thrombin entails cleavage of platelet thrombin receptor, mediated by binding of the receptor to the FRE of the enzyme (33). Thus, absence of a thrombinlike FRE in crotalase would constitute a sufficient explanation for the inability of crotalase to bring about platelet aggregation. Finally, Binnie and Lord (34) have shown that, while a fibringen peptide complementary to the FRE of thrombin competitively inhibits FPA release and clotting by thrombin, this peptide had no effect on these activities of crotalase's close homolog batroxobin.

The evidence for lack of a thrombin-like FRE in the crotalase molecule prompted a search for an alternative cationic fibrinogen recognition site. Fig. 4, generated using the coordinates of Massova et al. (4), shows the basic residues that present at the surface of the crotalase molecule. The right-hand lower part of the model shows the region that corresponds to the FRE of thrombin, as has already been discussed. The basic residues at the upper left can also be excluded as possible components of an alternative FRE as they are not conserved among the TLEs. The remaining residues, Arg 60F, Lys 85, Lys 87, and Arg 107 line a groove that extends from the active site cleft toward the northeast. Three of these residues are absolutely conserved among the TLEs and one is largely conserved. In the thrombin molecule, three of these positions are also occupied by basic residues but a crotalase-like groove is absent, as it is occupied by thrombin's 60A-60I loop (4). Thus, the cationic groove on the surface of the crotalase molecule, favorably oriented to its active site cleft, is a plausible candidate for an alternative fibrinogen recognition exosite.

It still must be considered, however, that fibrinogen recognition by crotalase may depend on totally different forces, e.g. hydrophobic interactions at yet a different exosite. Also, it is conceivable that interactions near the active site of crotalase may be sufficiently constrained to account for the observed degree of fibrinogen specificity.

Acknowledgments

This work was supported in part by NIH Grants HL-13598, HL-22875, HL-31267 (HP), HL-42412 (AH) and HL-57527 (BE). We thank Dr. Klara Osapay for generating images of the three-dimensional model of crotalase.

References

- 1. Markland FS, Damus PS. Purification and properties of a thrombin-like enzyme from the venom of *Crotalus adamanteus* (eastern diamondback rattlesnake). J Biol Chem 1971; 246: 6460-73.
- Markland FS, Pirkle H. Thrombin-like enzyme from the venom of *Crotalus adamanteus* (eastern diamondback rattlesnake). Thromb Res 1977; 10: 487-94.
- 3. Pirkle H. Thrombin-like enzymes from snake venoms: An updated inventory. Thromb Haemost 1998; 79: 675-83.

- Massova I, Pirkle H, Edwards BFP, Mobashery S. Insights into the threedimensional structure of crotalase: Implications for biological activity and substrate specificity. Bioorg Med Chem Lett 1997; 7: 3139-44.
- Bajwa WW, Markland FS Jr. A new method for purification of the thrombin-like enzyme from the venom of the eastern diamondback rattlesnake. Thromb Res 1979; 16: 11-23.
- 6. Pirkle H, Theodor I, Miyada D, Simmons G. Thrombin-like enzyme from the venom of *Bitis gabonica*. Purification, properties, and coagulant actions. J Biol Chem 1986; 261: 8830-5.
- 7. Hirs CHW. Reduction and S-carboxymethylation of proteins. Methods Enzymol 1967; 11: 199-203.
- 8. Habeeb AFSA. Reaction of protein sulfhydryl groups with Ellman's reagent. Methods Enzymol 1972; 25: 457-64.
- 9. Habeeb AFSA, Cassidy HG, Singer SJ. Molecular structural effects produced in proteins by reaction with succinic anhydride. Biochim Biophys Acta 1958; 29: 587-93.
- 10. Fontana A, Dalzoppo D, Grandi C, Zambonin M. Cleavage at tryptophan with *o*-iodosobenzoic acid. Methods Enzymol 1983; 91: 311-7.
- Huber R, Bode W. Structural basis of the activation and action of trypsin. Acc Chem Res 1978; 11: 114-22.
- Itoh N, Tanaka N, Mihashi S, Yamashina I. Molecular cloning and sequence analysis of cDNA for batroxobin, a thrombin-like snake venom enzyme. J Biol Chem 1978; 262: 3132-5.
- Nikai T, Ohara A, Komori Y, Fox JW, Sugihara H. Primary structure of a coagulant enzyme, bilineobin, from *Agkistrodon bilineatus* venom. Arch Biochem Biophys 1995; 318: 89-96.
- Pirkle H, Theodor I. Thrombin-like enzymes. In: Snake Venom Enzymes. Bailey GS (ed). Alaken, Inc., Fort Collins, Colorado 1998; 39-69.
- Bode W, Turk D, Karshikov AJ. The refined 1.9-Å X-ray structure of D-Phe-Pro-Arg chloromethylketone-inhibited α-thrombin: Structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. Protein Sci 1992; 1: 426-71.
- Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. Ann Rev Biochem 1985; 54: 631-64.
- Tanaka N, Nakada H, Itoh N, Mizuno Y, Takanishi M, Kawasaki T, Tate S-I, Inagaki F, Yamashina I. Novel structure of the N-acetylgalactosamine containing N-glycosidic carbohydrate chain of batroxobin, a thrombin-like snake venom enzyme. J Biochem 1992; 112: 68-74.
- Lochnit G, Geyer R. Carbohydrate structure analysis of batroxobin, a thrombin-like serine protease from *Bothrops moojeni* venom. Eur J Biochem 1995; 228: 805-16.
- Pfeiffer G, Dabrowski U, Dabrowski J, Stirm S, Strube KH, Geyer R. Carbohydrate structure of a thrombin-like serine protease from *Agkistrodon rhodostoma*. Structure elucidation of oligosaccharides by methylation analysis, liquid secondary-ion mass spectrometry and proton magnetic resonance. Eur J Biochem 1992; 205: 961-78.
- 20. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 1995; 4: 2411-23.
- Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 1989; 182: 319-26.
- 22. Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J. The refined 1,9-Å crystal structure of human α -thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. EMBO J 1989; 8: 3467-73.
- 23. Le Bonniec BF, Guinto ER, MacGillivray RTA, Stone SR, Esmon CT. The role of thrombin's Tyr-Pro-Pro-Trp motif in the interaction with fibrinogen, thrombomodulin, protein C, antithrombin III, and the Kunitz Inhibitors. J Biol Chem 1993; 268: 19055-61.
- 24. Le Bonniec BF, Guinto ER, Esmon CT. Interaction of thrombin des-ETW with antithrombin III, the Kunitz inhibitors, thrombomodulin and protein C. Structural link between the autolysis loop and the Tyr-Pro-Pro-Trp insertion of thrombin. J Biol Chem 1992; 267: 19341-8.
- 25. Dang QD, Sabetta M, Di Cera L. Selective loss of fibrinogen clotting in a loop-less thrombin. J Biol Chem 1997; 272: 19649-51.

- 26. Martin PD, Robertson W, Turk D, Huber R, Bode W, Edwards BFP. The structure of residues 7-16 of the Aα-chain of human fibrinogen bound to bovine thrombin at 2.3-Å resolution. J Biol Chem 1992; 267: 7911-20.
- Damus PS, Markland FS Jr, Davidson TM, Shanley JD. A purified procoagulant enzyme from the venom of the eastern diamondback rattlesnake (*Crotalus adamanteus*): In vivo and in vitro studies. J Lab Clin Med 1972; 79: 906-23.
- Fenton JW II, Bing DH. Thrombin active-site regions. Sem Thromb Hemost 1986; 12: 200-8.
- 29. Stubbs MT, Bode W. A player of many parts: The spotlight falls on thrombin's structure. Thromb Res 1993; 69: 1-58.
- 30. Chang JY. Deciphering the structural elements of hirudin C-terminal peptide that bind to the fibrinogen recognition site of α -thrombin. Biochemistry 1991; 30: 6656-61.
- 31. Rydel TJ, Tulinsky A., Bode W, Huber R. The structure of a complex of recombinant hirudin and human α -thrombin. J Mol Biol 1991; 221: 583-601.
- 32. Vitali J, Martin PD, Malkowski MG, Robertson WD, Lazar JB, Winant RC, Johnson PH, Edwards BFP. The structure of a complex of bovine α-thrombin and recombinant hirudin at 2.8-Å resulution. J Biol Chem 1992; 267: 17670-8.
- Coughlin SR, Vu T-KH, Hung DT, Wheaton VI. Characterization of a functional thrombin receptor. Issues and opportunities. J Clin Invest 1992; 89: 351-5.
- 34. Binnie CG, Lord ST. A synthetic analog of fibrinogen a27-50 is an inhibitor of thrombin. Thromb Haemost 1991; 65: 165-8.
- Hartley BS, Kauffman D. Corrections to the amino acid sequence of bovine chymotrypsinogen A. Biochem J 1966; 101: 229-31.
- Swift GH, Dagorn J-C, Ashley PL, Cummings SW, MacDonald RJ. Rat pancreatic kallikrein mRNA: Nucleotide sequence and amino acid sequence of the encoded preproenzyme. Proc Natl Acad Sci USA 1982; 79: 7263-7.

- Walsh KA, Neurath H. Trypsinogen and chymotrypsinogen as homologous proteins. Proc Natl Acad Sci USA 1964; 52: 884-9.
- 38. Magnusson S, Peterson TE, Sottrup-Jensen L, Claeys H. Complete primary structure of prothrombin: Isolation, structure and reactivity of ten carboxylated glutamic acid residues and regulation of prothrombin activation by thrombin. In: Proteases and Biological Control. Reich E, Rifkin DB, Shaw E (eds). Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 1975; pp 123-49.
- Pirkle H, Theodor I, Henschen A. Crotalase, a fibrinogen-clotting venom enzyme: Primary structure and evidence for lack of a fibrinogen recognition exosite homologous to that of thrombin. Haemostas 1996; 26 (Suppl 3): 452 (Abstract).
- 40. Burkhart W, Smith GFH, Su J-L, Parikh I, LeVine H III. Amino acid sequence determination of Ancrod, the thrombin-like α-fibrinogenase from the venom of *Akistrodon rhodostoma*. FEBS Lett 1992; 297: 297-301.
- Magalhães A, Da Fonseca BCB, Diniz DR, Gilroy J, Richardson M. The complete amino acid sequence of a thrombin-like enzyme/gyroxin analogue from venom of the bushmaster snake (*Lachesis muta muta*). FEBS Lett 1993; 329: 116-20.
- 42. Nishida S, Fujimura Y, Miura S, Ozaki Y, Usami Y, Suzuki M, Titani K, Yoshida E, Sugimoto M, Yoshioka A, Fukui H. Purification and characterization of bothrombin, a fibrinogen-clotting serine protease from the venom of *Bothrops jararaca*. Biochemistry 1994; 33: 1843-9.
- 43. Shieh T-C, Kawabata S-I, Kihara H, Ohno M, Iwanaga S. Amino acid sequence of a coagulant enzyme, flavoxobin, from *Trimeresurus flavoviridis* venom. J Biochem 1988; 103: 596-605.
- Hahn B-S, Yang K-Y, Park E-M, Chang I-M, Kim Y-S. Purification and molecular cloning of calobin, a thrombin-like enzyme from *Agkistrodon caliginosus* (Korean viper). J Biochem 1996; 119: 835-43.

Received May 26, 1998

Accepted after revision September 22, 1998

http://www.schattauer.com