Cloning of subunits of convulxin, a collagen-like platelet-aggregating protein from *Crotalus durissus terrificus* venom

Mireille LEDUC and Cassian BON¹

Unité des Venins, Institut Pasteur, 25 Rue du Dr. Roux, 75724 Paris Cedex 15, France

Convulxin (CVX) is a potent platelet-aggregating glycoprotein from the venom of the snake *Crotalus durissus terrificus*. It consists of two subunits, α and β , joined by disulphide bridges in a hexameric structure. A cDNA library from venom gland was constructed in the vector pT3T7. The cloned cDNAs encoding the two chains of CVX were sequenced. Both are preceded by an identical 23-amino acid peptide signal sequence and encode sequences of 135 amino acids for the α chain and 125 amino acids for the β chain. These polypeptides include a carbohydrate-recognition domain (CRD) in which some of the specific amino

acids required for binding Ca^{2+} and galactose or mannose are absent. The presence of such a domain means that CVX can be included in the family of C-type lectins along with other snake venom proteins, although it is not a true lectin. Assuming that the localization of intracatenary disulphide bridges of each CVX chain is similar to that of the CRD and that an intercatenary bridge between the α and β chains is similar to that of the C-type lectin botrocetin, we postulate the existence of an additional intercatenary bridge, which explains the tridimeric structure $(\alpha\beta)_3$ of CVX.

INTRODUCTION

Convulxin (CVX), a very potent platelet-aggregating protein from *Crotalus durissus terrificus* venom, is a 72 kDa heterodimeric $\alpha_3\beta_3$ complex made of two subunits, α (13.9 kDa) and β (12.6 kDa), linked by disulphide bridges [1–3]. It is a non-enzymic protein which activates blood platelets [4,5] through a Ca^{2+} -dependent mechanism [6] and fibrinogen, ADP- and cyclooxygenase-independent mechanisms [5]. Recently, the binding of ¹²⁵I-labelled CVX to blood platelets was analysed under equilibrium conditions showing that the protein binds to platelet membranes, in a Ca^{2+} -independent manner at a small number of sites (1000 binding sites per cell) with very high affinity (K_d 30 pM) [7,8]. CVX does not bind to other types of cell such as neutrophils and erythrocytes [7]. Binding is not inhibited by α -thrombin, fibrinogen, laminin-binding inhibitor, RGDS peptide, ADP, PAF-acether, 5-hydroxytryptamine or adrenaline [7]. All these results indicate that CVX binds to a specific platelet receptor, recently proposed to be the glycoprotein GPVI, which is also one of the collagen receptors [8,9].

We report here the cloning and sequences of two cDNAs encoding the two chains of CVX and their deduced polypeptide sequences. The polypeptide sequences were analysed and compared with those of other C-type lectins. An arrangement of the CVX disulphide bridges is proposed to explain the hexameric structure of this protein.

MATERIALS AND METHODS

Isolation of CVX α and β chains and determination of partial amino acid sequences of CVX subunits

Purified CVX was first submitted to SDS/PAGE under reducing conditions and then electrotransferred to an Immobilon membrane [7]. The protein bands corresponding to α and β chains

were cut out and digested with trypsin. The tryptic peptide fragments were separated by HPLC on a 218T52 Vydak (Hesperia) column (2.1 × 250 mm) with acetonitrile/water/trifluoroacetic acid [10]. The various peptides were sequenced by means of Edman degradation with an Applied Biosystems 470-A gas-phase sequencer. Phenylthiohydantoin amino acid derivatives were identified by on-line reverse-phase HPLC on an RP18 column, with an Applied Biosystems 120-A analyser.

RNA purification

Total RNA was prepared from *C. durissus terrificus* glands by the method of Chomczynski and Sacchi [11] using a total RNA separator kit (Clontech).

cDNA library

Polyadenylated RNAs were prepared from total RNA with oligotex-dT columns (Qiagen). From 4 μg of these mRNAs a directional cDNA library was constructed using the Superscript plasmid system (Life Technologies) with minor modifications. We used an *EcoRI* adaptor (Pharmacia) instead of the *SalI* proposed. Size-selected cDNAs were ligated in cloning vector pT3T7 *NotI/EcoRI*/BAP (Pharmacia). The resulting plasmids were used to transform highly competent *Escherichia coli* cells (Epicurian XL-blue-MRF'; Stratagene). The transformed bacteria were fractionated into 19 pools, which were screened by PCR, with one primer annealing on the vector and the other on the insert sequence. PCR was performed with *Taq* polymerase (Amersham) and 10 pmol of each nucleotide in a final volume of 50 μl . When necessary, PCR products were cloned with the pGEM T-system (Promega). Double-strand sequencing was performed with the T7 sequencing kit (Pharmacia).

Sequence alignment was performed with the Clustal V program, and the search for homologous sequences in databanks was carried out using the Fasta system.

Abbreviations used: CVX, convulxin; CRD, carbohydrate-recognition domain; IX/X-BP, IX/X-binding protein; LmsL, lectin from *Lachesis muta stenophrys*; RSL, rattlesnake lectin.

¹ To whom correspondence should be addressed (e-mail: cbon@pasteur.fr).

Nucleotide primer sequences

Nucleotide primer sequences were as follows:

CVX α :

A1: 5'-GATTGGTATTACTATGATCAG-3'

RCA2: 5'-TTCATTGAAGATCCGGTAGCAATGCTG-3'

CVX β :

B13: 5'-TGCACACAACAGCACACAGGCAGC-3'

RCB15: 5'-TGGTTATCAAATGTCCTGGATATGAG-3'

Primer in the signal peptide sequence:

AB5P: 5'-GAAGGAAGACCATGGGGCGATTTCATCTTCG-3'

Primers in the vector pT3T7:

T3: 5'-AATTAACCCTCACTAAAGGG-3'

T7: 5'-CATTATGCTGAGTGATATCCCCG-3'

RESULTS AND DISCUSSION

Cloning of CVX

From the N-terminal polypeptide sequences of CVX α and CVX β we designed degenerated PCR primers as well as the corresponding antisense primers. These primers combined with primers in the vector were tested on the cDNA library constructed in the vector pBR322 and previously used to clone cDNAs encoding other proteins from this venom [12]. The sequencing of the PCR products revealed 5' end non-coding regions, a signal peptide and incomplete coding sequences of proteins for both CVX subunits. Nevertheless, the sequences of these clones allowed the design of specific PCR primers for CVX α (A1 and RCA2) and CVX β (B13 and RCB15), which were used to screen a new cDNA library. mRNAs from a venom gland of *C. durissus terrificus* were prepared and used to construct a new cDNA library in the vector pT3T7 between the *EcoRI* and *NotI* sites. Several inserts of interest were amplified from the library by PCR using various pairs of specific primers in CVX α or CVX β , and specific ones in the vector (T3 or T7). The amplified material obtained after each PCR was sequenced, allowing us to determine the complete cDNA sequences of the CVX chains (Figure 1). These sequences were consistent with the partial amino acid sequences determined by Edman degradation from purified CVX. The two cDNAs encoding CVX α and CVX β were then obtained from the pT3T7 library pools by three consecutive PCRs performed as described in Figure 2. The final PCR products were cloned in the pGEM-T system and sequenced again to confirm that they correctly encoded CVX α and CVX β proteins.

To test the possibility that the two chains of CVX are encoded by the same RNA transcript, a PCR amplification was performed on total RNAs isolated from the venom gland with a pair of specific primers, A1 in CVX α and RCB15 in CVX β or with another pair of primers, RCA2 in CVX α and B13 in CVX β . No PCR products were obtained in either case, showing that the two CVX subunits were encoded by two different transcripts. Control experiments performed using two specific primers in either CVX α (A1 and RCA2) or CVX β (B13 and RCB15) gave positive results.

Reamplification of the cDNA library prepared in the pT3T7 vector led to the loss of the clones containing cDNAs coding for CVX α and CVX β . This was mentioned by Alfandari and Darribère [13], and the method they proposed to recover the clones was also unsuccessful in our hands.

A

```

1      GGTGCAGGCAATGTCTGGTTCAGCAGGGGATCCTTGGGAGGCAAGGACAAAGGAATCT
61     AGCAGGGGGATCTCTGGGAAGAGAGCCCTTTGGGGGCCCTCTTCTCTCTGACGGGAA
           M G R F I F V S F G L L V L F L
121    GGAAGGAAGACCAATGGGGCGATTTCATCTCGTGAGCTTCGGCTTGGTGGTCTGTCTCTC
           S L S G T G A G L H C P S D W Y Y Y D Q
181    TCCCTGAGTGGTACTGGAGCTGGTTTACATTGCCCTCTGATTGTATTACTATGATCAG
[14]   H C Y R I F N E E M N W E D A E W F C T
241    CATTGTACCGGATCTTCAATGAAGAGATGAACCTGGGAAGATGCAGAGTGGTCTGCACG
[34]   K Q A K G A H L V S I K S A K E A D F V
301    AAGCAGGCGAAGGGCGCGCATCTCGTCTCTATCAAAAGCCCAAGAGCAGACTTTGTG
[54]   A W M V T Q N I E E S F S H V S I G L R
361    GCCTGGATGGTCACTCAGAACATAGAGGAATCCTTTCCCATGCTCGATTGGAGCTGAGG
[74]   V Q N K E K Q C S T K W S D G S S V S Y
421    GTTCAAACAAAGAAAGCAATGCAGCAGGAAGTGGAGCGATGGCTCCAGCGTCAGTTAT
[94]   D N L L D L Y I T K C S L L K K E T G F
481    GACAACTTGTGGATCTATATATTACAAAGTGTAGTCTGCTGAAAAAGAGACAGGGTTT
[114]  R K W F V A S C I G K I P F V C K F P P
541    CGTAAGTGGTTCGTGCTAGCTGTATAGGAAGATTCTTTCGTGCAAGTCCCGCCGA
[134]  Q C .
601    CAGTGTAAAGTACCAGCTTTGTGAAGTCTGGAGAAGCAAGAGCCCCCCCCCTCCCAACC
661    CCCGCCCTGCCCAATCTCTGCTCTGCACCTGTCTTCATGGGTGCTCTCTGTGGCTG
721    GATCTGGTCTCTGCTCTGATGGGCCAGAGGTCCATAAATTCTGCCTAGCATGAAA
781    AAAAAAAAAAAAAAAAAAAAA

```

B

```

1      AAGCCTGGAGTGTGCTCTGAGCAGACTTGTCTACCTGTGGAGGCCGAGGAACGGTTCTCTC
           M G R F I F V S F G L L V
61     TGCAGGGAAGGAAGGAAGACCAATGGGGCGATTTCATCTCGTGAGCTTCGGCTGCTGGTC
[-10]  V F L S L S G S E A G F C C P S H W S S
121    GTTCTCTCTCCCTGAGTGGAGTGAAGCTGGTTTCTGTGTGCTCCCTCCATTGGTCTTCC
[11]   Y D R Y C Y K V F K O E M T W A D A E K
181    TATGATCGGTATTGCTACAAGTCTTCAAACAGAGATGACCTGGGCCGATGCAGAGAAA
[31]   F C T Q Q H T G S H L V S F H S T E E V
241    TTCTGCACACACAGCACAGGCGAGCACTGTGGTCTCTTCACAGCACTGAAGAAGTA
[51]   D F V V K M T H Q S L K S T F F W I G A
301    GATTTGTGGTCAAGATGACCCACCAAGTTTGAAGTCCACTTTTCTGGATTGGAGCG
[71]   N N I W N K C N W O W S D G T K P E Y K
361    AACAAATATCTGGAATAAATGCAACTGGCAGTGGAGCGATGGCAACCAAGCCTGAGTACAAA
[91]   E W H E E F E C L I S R T F D N O W L S
421    GAATGCCATGAAGAATTGAATGTCTCATATCCAGGACATTTGTATAACCAAGTGGTTAAGT
[111]  A P C S D T Y S F V C K F E A .
481    GCACCCCTGCACTGATACTTACTCTTTCTGCTGCAAGTTCGAGGCAATGCTGAAGATCCG
541    GCTGTGTAAGTCTGGAGAAGCAAGAACGCCCCACCCCTCCCAACCCCACTGTGCCG
601    AATCTCTGCTGCCCCCTTCACTCAACGGATGCTCTGTGGCTGATGCTGTTCTGCT
661    GCTCCTGATGGGCCAGAGGTCCATAAATTCTGCCTAGCAGGAAAAAAGAAAAA
721    AAAAAAAAAAAAAAAAAAAAA

```

Figure 1 Nucleotide sequences of cDNAs encoding CVX α (A) and CVX β (B) and the deduced polypeptide sequences

Numbering of the nucleotides is in the 5' to 3' direction. Numbering of the amino acids, indicated in brackets, starts at the N-terminal amino acid of the mature CVX α and β chains. Polypeptide sequences determined by Edman sequencing are underlined.

Comparison of the cDNA sequences encoding CVX α and β chains and that of the deduced polypeptides

Comparison of the nucleotide sequences of the cDNAs encoding CVX subunits shows that they are very similar, with an overall identity of 74%, which appears to be higher in the non-coding sequences, 83.4% and 91.2% for the 5' and 3' ends respectively, than in the coding region (63.8% only). A similar finding was

Figure 2 Constructs of CVX α and CVX β by PCR

Constructs were obtained from the relevant library pool by a series of three consecutive PCRs: the first between primer AB5P in the peptide signal sequence and an antisense primer RCA2 (CVX α) or RCB15 (CVX β); the second between a specific primer A1 (CVX α) or B13 (CVX β) in the sequence and the antisense primer T3 in the vector; the third on an equal mixture of the two previous PCR products as template between primer AB5P and primer T3 for CVX α and CVX β .

first reported for cDNAs encoding the two subunits of crotoxin, the major protein of the same venom. This was, however, much more marked, with 98 and 90 % identity for the non-coding 5' and 3' end sequences compared with 50 % in the coding region [12]. The low percentage of identity in the coding sequences compared with the high percentage in the non-coding sequences remains largely unexplained. It may result from recent gene divergence, as suggested for the crotoxin subunits [12].

The encoded protein sequences correspond to polypeptides of 158 and 148 amino acids for CVX α and CVX β respectively (Figure 1). A comparison of the N-terminal sequence of these polypeptides with that determined by Edman degradation of the isolated protein shows that the sequences deduced from cDNA include a signal peptide of 23 amino acids, which is removed in the mature protein. No difference was found between the N-terminal sequence of mature protein and the internal polypeptide sequences for each subunit obtained by Edman degradation. Indeed the N-terminal sequences of the CVX subunits determined by Polgar et al. [9] are in agreement with the polypeptide sequences deduced from the cDNAs identified in this study. The sequences of the mature subunits give a molecular mass of 14850 Da for CVX α and 13750 Da for CVX β . These values are in rough agreement with the apparent molecular masses determined by Marlas et al. [2] of 13900 and 12600 Da respectively. In our nomenclature, CVX α and CVX β were defined according to a homologous family of proteins in which the α subunits show homology to CVX α and β subunits to CVX β . Furthermore, no consensus sequence for the N-glycosylation site was found in the sequence of either CVX α or CVX β . Since CVX has been reported to be a glycoprotein [2], the glycan moiety is probably associated through O-linked peptides.

C-type lectin structure of CVX and covalent association of subunits by disulphide bridges

The amino acid sequences of the CVX subunits possess most of

the characteristics defined by Drickamer [14], Spiess [15] and Hirabayashi [16] for the carbohydrate-recognition domain (CRD) of C-type lectins. Accordingly, CVX α and CVX β could be included in the family of C-type lectins, along with other snake venom proteins such as factor IX/X-binding protein (IX/X-BP) [17], alboaggregin-B [18], echicetin [19,20], botrocetin [21], bothrojaracin [10] and CHH-B [22]. A few differences were, however, apparent in CVX α , in which a serine residue replaces the invariant Trp-69 in the CRD and a valine residue replaces the conserved Pro-87. Moreover, as generally found in C-lectins [23,24], the CRD of CVX α and CVX β does not possess the appropriate amino acids for binding of Ca^{2+} and galactose or mannose. In agreement with this finding, it has been shown that CVX does not possess agglutinating activity [7]. CVX binding to its platelet receptor is not modified by the presence or absence of Ca^{2+} [7].

CVX α and CVX β each contain eight cysteine residues. According to the structural model of the CRD proposed by Spiess [15], six of these cysteines are involved in the intracatenary disulphide bridges (Cys⁴–Cys¹⁵, Cys³²–Cys¹²⁹ and Cys¹⁰⁴–Cys¹²¹ for CVX α and Cys⁴–Cys¹⁵, Cys³²–Cys¹²¹ and Cys⁹⁸–Cys¹¹³ for CVX β). Such a model was proved to be correct in the case of the snake venom proteins botrocetin and rattle snake lectin (RSL), in which the disulphide bridges have been chemically located [21,23]. In these two proteins, besides these six intracatenary disulphide bridges, a disulphide bridge between the α and β chains has been shown to associate cysteine residues around position 80 [21,23]. Such an intercatenary bridge may exist in CVX between Cys-80 of the α chain and Cys-77 of the β chain. When compared with RSL and botrocetin, CVX possesses one additional cysteine residue in the α chain (Cys-135) and one in the β chain (Cys-3). So it may be that an additional intercatenary disulphide bridge between these two cysteine residues stabilizes the hexameric structure of CVX reported by Marlas et al. [3] in a tridimeric ($\alpha\beta$)₃ structure. GPVI cross-linking to the CVX tridimeric structure could be compared with GPVI cross-linking

CVX a	GLHCPDSDWYYDQHCYRIFNEEMNWEDAWEFCTKQAKGAHLVSIKS-AKE
Botro a	--DCPSGWSSYEGNCYKFFQKMNWADAERFCSEQAAGHLSVSIKYSKE
CHH-Ba	DLECPDSDWSSYDRYCYKPFQKQEMTWADAERFCSEQAAGHLLSVET-ALE
BjA	--DCPSDWSSHEGHCHYKFFQKMNWADAERFCSEQAAGHLSVFSQS-DGE
IX/Xbp A	--DCLSGWSSYEGHCHYKAFKQKTWEDAERVCTEQAKGAHLVSIKS-SGE
Echic a	DQDCLSGWSSYEGHCHYQLFR-LKTWDEAEKYC-NQWDGGHLSVSIKS-NAK
ALBa	--DCPSDWSSFKQYCYQIIKQLKTWEDAERFCMDQVKGHLVSIKS-SYR-E
CVX b	GFCCPSHWSSYDRYCYKVFQKQEMTWADAERFCTQQHTGSHLSVFSHS-TEE
Botro b	D--CPPDWSSYEGHCHYRFFKEWMHWDDAEFCTEQQTGAHLVFSQS-KEE
CHH-Bb	D--CPSDWSSYEGHCHYRVFQKQEMTWADAERFCTQQHTGSHLSVFSRS-SEE
BjB	D--CPPDWSSYEGHCHYRVFEQKMNWADAERFCSQQQTGSHLSVFSQS-SEE
IX/X-bp B	D--CPSDWSSYEGHCHYKPFSEPKNWADAENFCTQQAAGHLSVFSQS-SEE
Echic b	N--CLPDWV-YEG--YKVFKERMNWADAERFCMKQVKDGHLSVFSRN-SKE
ALBb	D--CPSDWSSYDLYCYRVFQEKKNWERAERFCTQQHTDSHLSVFSDS-SEE
RSL	N-NCPLDWLPMNGLCYKIFNQLKTWEDAEMFCRKYKPGCHLASFHRYGES
LmsL	N-NCPLDWLPMNGLCYKIFDEQKAWEDAEMFCRKYKPGCHLASFHRYGES
	* * * * *
CVX a	ADFVAVMVTQNIIEESFSH--VSIQLRVQNKQKQSTKWSGSSSVSY----
Botro a	KDFVGLDVTQNIQSSDLY--AWIGLRVENKEKQCSSEWSGSSSVSY----
CHH-Ba	ASFVDNVLNKEYLTRY--IWIGLRVQNKQKQPCSSIIYSENLDVDF----
BjA	TDFVVLNLTQNIQSTDL--AWIGLRVQNKQKQCSKWSGSSSVSY----
IX/Xbp A	ADFVAQLVTQNMKRLDFY--IWIGLRVQCKVQKCNSEWSGSSSVSY----
Echic a	AEFVAQLISRKLPSAIEDRVWIGLRDRSKREQCGLHWTDNSFVHY----
ALBa	AVFVAQLSENKTKTKYD--VWIGLSVVKQKQCSSEWSGSSSVSY----
CVX b	VDFVVKMTHQSLKS--TFFWIGANNI--NKCWNQWSDGDKPEY-----
Botro b	ADFVRSLTSEMLKG--DVVWIGLSDVW--NKRFEWTDGMEFDY-----
CHH-Bb	VDFLVS----ILKF--DLFWMGWRDIW--NERRLQWSDGDKVNY-----
BjB	ADFVVSILTSPILRD--SFVWTLGSDVW--KECSFEWSGSDLSYKD----
IX/X-bp B	ADFVVKLAFQTFGH--SIFWMGLSNVW--NQCWNQWNSAAMLRY-----
Echic b	VDFMISLAFPMMLKM--ELVWIGLSDYV--RDCYWEWSGDAQLDY-----
ALBb	VDFVASKTFPVLKH--DLVWIGLGSVW--NACKLQWSDGTELKY-----
RSL	LE-IAEYISDYHKGQ-ENVWIGLRD--KKKDFSWEWTDRCSTDYLTWDKN
LmsL	LE-IAEYISDYHKGQ-AEVWIGLRD--KKKDFSWEWTDRCSTDYLTWDKN
	*
CVX a	DNLLDLYITKCSLLK--KETGFRKWFVASCIGKIPFVCKFPFQC-
Botro a	ENVVERTVKKCFAL--KDLGFLWINLYCAQKNPFVCKSPPP--
CHH-Ba	E-----CFMVS--DRTRLREWFKVDCEQHSFICKFTRPRR
BjA	ENVVGRVTKKCFAL--KEQEFFVWINIYCGQNPVCKSPPP--
IX/Xbp A	ENWIEAESKTCGLE--KETDFRKWVNIYCGQNPVFCFA-----
Echic a	EHVVP--TKCFVLE--RQTEFRKWIAVNCEFKFPFVCKAKIP-R
ALBa	ENLVKPLSKKCFVLK--KGTEFRKWFNVACEQKHLFMCKFLRP--
CVX b	KE--WHEEFECLIS---R-TFDNQWLSAPCSDTYSFVCKFEA---
Botro b	DDYLLIAEYECVAS---KPTN-NKWWIIPCTRFKNFVCEFA---
CHH-Bb	KA--WSAEPECIVC---RATD-NQWLSTSCSKTHNVCKF-----
BjB	NYQFVFSEYECVAS---K-TKNNKWRIIPCTKLEYFVCEFA---
IX/X-bp B	KA--WAEESYCVYF---KSTN-NKWSRACRMMAQFVCEFA---
Echic b	KA--WDNERHCFAA---KTTD-NQWMRRKCSGEFYFVCKC-P---
ALBb	NA--WSAESECITS---KSTD-NEWLTRSCSRTYFVCKFQA---
RSL	QPDHYQNKFEVCVELV--SLTGYRLWNDQVCESKDAFLCQCKF---
LmsL	QPDHYEGKEFCVELV--SLTGYRLWNDQVCESKNAFLCQCKF---
	* * *

Figure 3 Polypeptide sequence alignment between different snake venom C-type lectins

The two CVX chains are aligned with the two chains of botrocetin from *Bothrops jararaca* [21] (Botro a and b), the two chains of IX/X-BP from *Trimeresurus flavoviridis* [17] (IX/Xbp A and B), the two chains of bothrojaracin from *B. jararaca* [10] (BjA and BjB), the two chains of alboagreggin from *Trimeresurus albolabris* [22] (ALBa and ALBb), the two chains of CHH-B from *Crotalus horridus horridus* [22] (CHH-Ba and CHH-Bb), the two chains of echicetin [19,20] (Echic a and b), and the single chain of RSL from *Crotalus atrox* [23] and the single chain of LmsL from *Lachesis muta stenophrys* [24]. Gaps have been inserted to maximize homology. Asterisks indicate similarity between residues.

Figure 4 Neighbour-joining tree with C-type lectins and C-lectins from snake venom and human lithostatin [27]

The numbers at the node represent the bootstrap values, which were 1000 replicates according to the Must Package [28]. EchicetinA, chain of echicetin; CHHBa, chain a of protein CHH-B; ALBa, chain a of alboaggregin; CVX a, chain α of CVX; IX/Xbp A, chain a of IX/XBP; BjA, chain a of bothrojaracin; Botro a, chain a of botrocetin; EchicetinB, chain B of echicetin; IX/Xbp B, chain b of IX/XBP; BjB, chain b of bothrojaracin; Botro b, chain b of botrocetin; CVX b, chain β of CVX; ALBb, chain b of alboaggregin; CHHBb, chain b of CHH-B protein.

with $F(ab')_2$ fragments of anti-GPVI IgG, which has been shown to be similar to collagen stimulation [25].

Sequence similarities between CVX and other C-type lectins from snake venom

The cDNA sequences of CVX subunits were compared with those of two other dimeric snake venom C-type lectins, bothrojaracin [10] and IX/X-BP [26]. All are homologous with a high degree of identity. The sequence similarities are greater in untranslated regions (86–94%) than in coding sequences (64–77%).

The polypeptide sequences of the CVX subunits were compared with those of other C-type lectins from snake venom, bothrojaracin [10], botrocetin [21], IX/X-BP [17], echicetin [19,20], alboaggregin-B [18], CHH-B [22], and two snake venom C-lectins, RSL [23] and LmsL [24], which are highly similar in their primary sequences but possess very different physiological activity, as they interact specifically with quite different proteins (Figure 3). An evolutionary tree [27] constructed from these sequences (Figure 4) clearly demonstrates the duplication of an ancestral gene pre-dating the diversification of the *Crotalidae*. It should also be noticed that the tree has very long branches, and only bootstrap values over 60% have been detected.

From alignment of the C-type lectin sequences, nothing could

be detected in the polypeptide sequence involved in the interaction between the C-type lectin and its receptor. However, a knowledge of this sequence is essential to understand biological activity. To acquire such information, the crystal structure is very helpful, and the 2.5 Å crystal structure of IX/X-BP was recently reported [28]. The crystal structure of CVX will allow us to suggest which polypeptide sequences of the tridimeric structure are likely to interact with the glycoprotein GPVI on the platelets. With such sequences, directed mutagenesis can be used to produce mutants which will help us to understand the interaction between CVX and its receptor.

We are very grateful to Dr. A. Wisner (Unité des Venins, Institut Pasteur, Paris) and to Dr. X. Cousin (INRA-ENSA, Montpellier) for fruitful discussions, to Dr. L. Morin (Laboratoire de Biologie cellulaire 4, Université Paris-Sud, Orsay) for his expertise in the elaboration of the phylogenetic tree and to Dr. J. d'Alayer (Laboratoire de Microséquençage des Protéines, Institut Pasteur, Paris) for performing the amino acid sequences. We are thankful to Professor B. Lejeune and to Dr. M.-F. Jubier (IBP, Bâtiment 630, Université Paris-Sud, Orsay) for fruitful discussions. This work was supported by funds from the European Community (contract no. C11-CT94-0073).

REFERENCES

- Marlas, G., Joseph, D. and Huet, C. (1983) *Biochimie* **65**, 405–416
- Marlas, G., Joseph, D. and Huet, C. (1983) *Biochimie* **65**, 619–628
- Marlas, G. (1985) *Biochimie* **67**, 1231–1239
- Vargaftig, B. B., Prado-Franceschi, J., Chignard, M., Lefort, J. and Marlas, G. (1980) *Eur. J. Pharmacol.* **68**, 451–454
- Vargaftig, B. B., Joseph, D., Wal, F., Marlas, G., Chignard, M. and Chevance, G. L. (1983) *Eur. J. Pharmacol.* **92**, 57–68
- Sano-Martins, I. S. and Daimon, T. (1992) *Toxicon* **30**, 141–150
- Franciscetti, I. M. B., Saliou, B., Leduc, M., Carlini, C. R., Hatmi, M., Randon, J., Faill, A. and Bon, C. (1997) *Toxicon* **35**, 1217–1228
- Jandrot-Perrus, M., Lagrue, A.-H., Okuma, M. and Bon, C. (1997) *J. Biol. Chem.* **272**, 27035–27041
- Polgar, J., Clemetson, J. M., Kehrel, B. E., Wiedermann, M., Magnenat, E. M., Wells, T. N. C. and Clemetson, K. J. (1997) *J. Biol. Chem.* **272**, 13576–13583
- Arocas, V., Wisner, A., Zingali, R. B., Guillin, M.-C., Jandrot-Perrus, M. and Bon, C. (1997) *Eur. J. Biochem.* **248**, 550–557
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Bouchier, C., Boulain, J.-C., Bon, C. and Menez, A. (1991) *Biochim. Biophys. Acta* **1040**, 217–224
- Alfandari, D. and Darribère, T. (1994) *PCR Methods and Applications* **4**, 46–49
- Drickamer, K. (1988) *J. Biol. Chem.* **263**, 9557–9560
- Spiehs, M. (1990) *Biochemistry* **29**, 10009–10018
- Hirabayashi, J. (1994) *Lectins: Biology, Biochemistry and Clinical Biochemistry* **10**, 205–219
- Atoda, H., Hyuga, M. and Morita, T. (1991) *J. Biol. Chem.* **266**, 14903–14911
- Usami, Y., Susuki, M., Yoshida, E., Sakurai, Y., Hirano, K., Kawasaki, T., Fusimura, Y. and Titani, K. (1996) *Biochem. Biophys. Res. Commun.* **219**, 727–733
- Peng, M., Holt, J. C. and Niewarowski, S. (1994) *Biochem. Biophys. Res. Commun.* **205**, 68–72
- Polgar, J., Magnenat, E. M., Peitsch, M. C., Wells, T. N. C., Saqi, M. S. A. and Clemetson, K. J. (1997) *Biochem. J.* **323**, 533–537
- Usami, Y., Fujimura, Y., Susuki, M., Oseki, Y., Nishio, H., Fukui, H. and Titani, K. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 928–932
- Andrews, R. K., Kroll, M. H., Ward, C. M., Rose, J. W., Scarborough, R. M., Smith, A. I., Lopez, J. A. and Berndt, M. C. (1996) *Biochemistry* **35**, 12629–12639
- Hirabayashi, J., Kusunoki, T. and Kasai, K. (1991) *J. Biol. Chem.* **266**, 2320–2326
- Aragon-Ortiz, F., Mentele, R. and Auerswald, E. A. (1996) *Toxicon* **34**, 763–769
- Tsuji, M., Ezumi, Y., Arai, M. and Takayama, H. (1997) *J. Biol. Chem.* **272**, 23528–23531
- Matsuzaki, R., Yoshiara, E., Yamada, M., Shima, K., Atoda, H. and Morita, T. (1996) *Biochem. Biophys. Res. Commun.* **220**, 382–387
- Philippe, H. (1996) *Nucleic Acids Res.* **21**, 5264–5274
- Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H. and Morita, T. (1997) *Nature Struct. Biol.* **4**, 438–441