Molecular Marine Biology and Biotechnology (1995) 4(1), 90–103

Molecular cloning and sequence analysis of Factor C cDNA from the Singapore horseshoe crab, *Carcinoscorpius rotundicauda*

Jeak L. Ding* and M. Anthony A. Navas III†
Marine Biotechnology Laboratory, and BioScience Centre, Department of Zoology, National University of Singapore, Kent Ridge, Singapore 0511

Bow Ho
Department of Microbiology, National University of Singapore, Kent Ridge, Singapore 0511

Abstract

Two forms of Factor C cDNAs: CrFC21 (3448 bp) and CrFC26 (4182 bp) have been cloned into λgt22. CrFC26 includes 568 nucleotides of 5′ untranslated region (5′ UTR) containing seven ATGs before the real initiation site, an open reading frame (ORF) of 3249 nucleotides, a stop codon, and 365 nucleotides of 3′ untranslated sequence. There are four polyadenylation signals and six potential glycosylation sites. The ORF codes for a signal peptide of 24 amino acids and a Factor C zymogen of 1059 residues. The CrFC21 lacks most of the 5′ UTR, and has some base changes in its ORF. The predicted secondary mRNA structures of the 5′ end of CrFC26 showed numerous stem-and-loop structures, thus obscuring its real start codon. In contrast, CrFC21 has a well-exposed AUG start site, and expresses Factor C in transcription-translation reactions in vitro. There is a typical serine protease catalytic triad of Asp-His-Ser, which is structurally like prothrombin, but catalytically more similar to trypsin. Although an overall homology of 97.7% was observed in comparison with the *Tachypleus tridentatus* Factor C (TtFC) cDNA, there were notable differences in the restriction sites and subtle base substitutions in the CrFC cDNA. The high degree of homology between Factor C from *T. tridentatus* and *C. rotundicauda* substantiates, at the molecular level, the proximity of these two species in the course of evolution. This finding contravenes the apparent disparities with respect to their morphology, ecological habitat, and taxonomical classification.

*Correspondence should be sent to this author.
†Present address: Department of Biochemistry, Baylor College, Houston, Texas, USA.

Introduction

There are four extant species of horseshoe crabs: *Limulus polyphemus, Tachypleus tridentatus, Tachypleus gigas,* and *Carcinoscorpius rotundicauda.* Within the predominant type of blood cell, amoebocytes, are enzymes belonging to the family of serine proteases that sequentially undergo limited proteolysis (Iwanaga et al., 1985) upon endotoxin activation to finally convert coagulogen into an insoluble gel, coagulin (Solum, 1973). This cascade of enzymes and coagulogen constitutes the commercially available amoebocyte lysate used for detecting endotoxin. The latter is extremely ubiquitous and indomitable, contaminating parenteral preparations, water, food, and pharmaceutical products. The *Limulus* amoebocyte lysate (LAL) is a rapid and sensitive in vitro assay for detecting femtogram levels of endotoxin (Ho, 1983).

The elucidation of the molecular events that are responsible for the subsequent gelation of the amoebocyte lysate was attributed to the isolation and characterization of the serine proteases from either *L. polyphemus* or *T. tridentatus*. These include Factor C (Nakamura et al., 1986) and proclotting enzyme (Nakamura et al., 1985). However, to date, there is a lack of information on similar enzymes from their mud-dwelling Southeast Asian counterpart, *C. rotundicauda,* except our recent reports on purification and characterization of its Factor C enzyme (Navas et al., 1990; Ding et al., 1993a; Ho et al., 1993).

Factor C is the endotoxin-sensitive intracellular serine protease zymogen that initiates the coagulation cascade system. It exists as single- and double-chain forms in *C. rotundicuada* (Ding et al., 1993a). While the catalytic site of the molecule is found in the light chain, the endotoxin-binding site is located at the NH_2-terminal heavy chain region (Nakamura et al., 1988). Muta et al. (1991) have cloned the *T. tridentatus* Factor C, albeit in two partial overlapping fragments: pFC 41 and pFC 53, henceforth re-

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Seikagaku Corporation Exhibit 2008 - Page 1 of 14
Charles River Laboratories, Inc. v. Seikagaku Corporation
PGR2025-00023

ferred to as TtFC41 and TtFC53, respectively. These two fragments gave a composite sequence of 3474 bp. We report, here, directional cloning of the full-length *C. rotundicauda* Factor C (CrFC) cDNA in λgt22 vector, its characterization by restriction mapping and DNA sequencing, homology comparison studies with TtFC cDNAs and other serine proteases in the EMBL gene bank, the computational analysis of the 5′ region of CrFC cDNAs and their corresponding predicted mRNA secondary structures, and the evolutionary significance of CrFC. The commercial significance of recombinant Factor C is also implied.

Results and Discussion

Northern analysis of C. rotundicauda *amoebocyte RNA*

Northern hybridization of *C. rotundicauda* RNA using TtFC probe under low stringency of 42°C showed three bands corresponding to 15.8, 5.9, and 4.0 kNt (Figure 1a). However, subsequent hybridization under high stringency with homologous EE fragment of CrFC26 revealed only the 4.0 kNt RNA band (Figure 1b). The other forms of CrFC mRNAs could represent transcripts resulting from alternative splicing as indicated by the presence of alternative polyadenylation signals (see Fig. 5). Since total RNA was used in the earlier blot, the 15.8-kNt band could have been the hnRNA transcript.

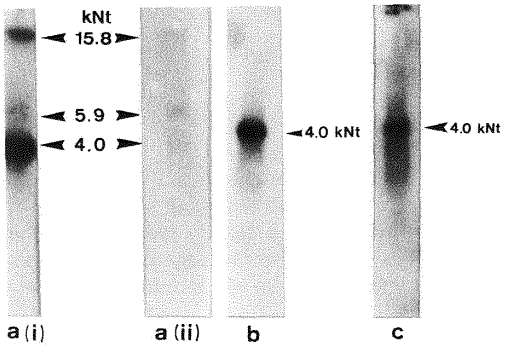


Figure 1. Northern blot analysis using TtFC53 as probe on 10 μg total RNA in lane a (i) and 1 μg of mRNA in lane a (ii); pCrFC26 (*Eco*RI–*Eco*RI fragment) as probe on 10 μg of total RNA in lane b; and the 368-bp *Eco*RI–*Nde*I (5′ end) of pCrFC26 as probe on 10 μg of total RNA in lane c. A single band of approximately 4 kNt in both blots (b) and (c) indicates that the cDNA isolated is full-length and the entire 5′ end of crFC26 unequivocally belongs to this species of Factor C.

The Factor C cDNA λ recombinant clones

From 250,000 recombinant clones, only 0.03% of the cDNA library was positive for Factor C cDNA, indicating its status as a rare copy gene. Following secondary and tertiary screenings, 16 λ recombinant clones, henceforth referred to as CrFC cDNAs, were further characterized. Their orientations were determined, and initial restriction maps were analyzed for *Eco*RI and *Not*I sites (Figure 2).

Double digestion of the CrFC clones with *Eco*RI and *Not*I followed by Southern blot hybridization with the *T. tridentatus* Factor C (TtFC53) cDNA probe showed positive signals for all the CrFC clones (data not shown). When the same membrane was deprobed and reprobed with the 3′-end EN fragment of CrFC26, or the whole of TtFC41, two of the CrFC clones, CrFC45 and CrFC63, did not

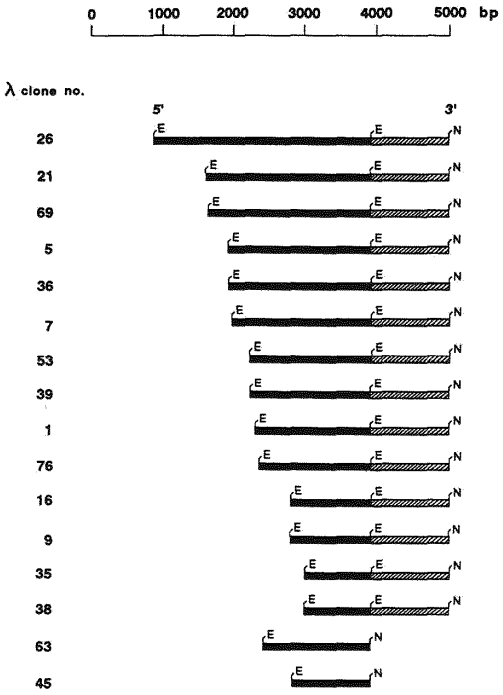


Figure 2. The initial restriction map of the various Factor C cDNAs isolated and purified from the *C. rotundicauda* amoebocyte cDNA library based on Southern analysis with TtFC53 and the 1100-bp *Eco*RI–*Not*I fragment. The cDNAs are arranged in decreasing order of size and are labeled with the restriction sites *Eco*RI (E) and *Not*I (N). Fragments homologus to TtFC53 are drawn with solid boxes while those homologous to the E–N fragment are drawn with hatched boxes.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

cross-hybridize with either probe, suggesting that these two clones are not similar to the other CrFC cDNA clones. They do not have an internal *Eco*RI site, but are homologous to only the 5′-end region (EE fragment) of other CrFC clones and the TtFC53.

Comparison of the restriction maps of CrFC and TtFC cDNAs

The complete restriction maps of the two longest clones of *C. rotundicauda* Factor C cDNAs, CrFC21 and CrFC26, are shown in Figure 3, in comparison with the maps of the two overlapping clones of *T. tridentatus* Factor C, TtFC53 and TtFC41 (Muta et al., 1991). CrFC26 has a longer 5′ terminus compared with the TtFC53. Based on the restriction maps of TtFC53 and TtFC41 clones, the cDNAs from both species have the same locations for restriction sites *Sty*I, *Sal*I, and *Sac*I. However, notable differences were also observed between the two genes: CrFC has two *Bam*HI sites compared with only one in TtFC53. On the other hand, the former only has two *Hind*III sites compared with three in the latter.

The TtFC41, on the other hand, encompasses the 3′ region of the TtFC. It extends from the *Sst*I/*Sac*I site up to the 3′ end, and it was found to contain the internal *Eco*RI site relative to that of CrFC. The other common site between TtFC41 and CrFC is the *Pst*I site proximal to the 3′ end, but the second *Pst*I site near the 5′ end of TtFC41 was absent in the CrFC cDNAs. The *Bam*HI and *Hind*III sites in TtFC41 were absent in the CrFC26.

The DNA sequence and derived amino acid sequence of CrFC cDNA

The DNA sequences of CrFC21 and CrFC26 were determined on both strands by the Sanger dideoxy sequencing method. Figure 4 shows the sequencing

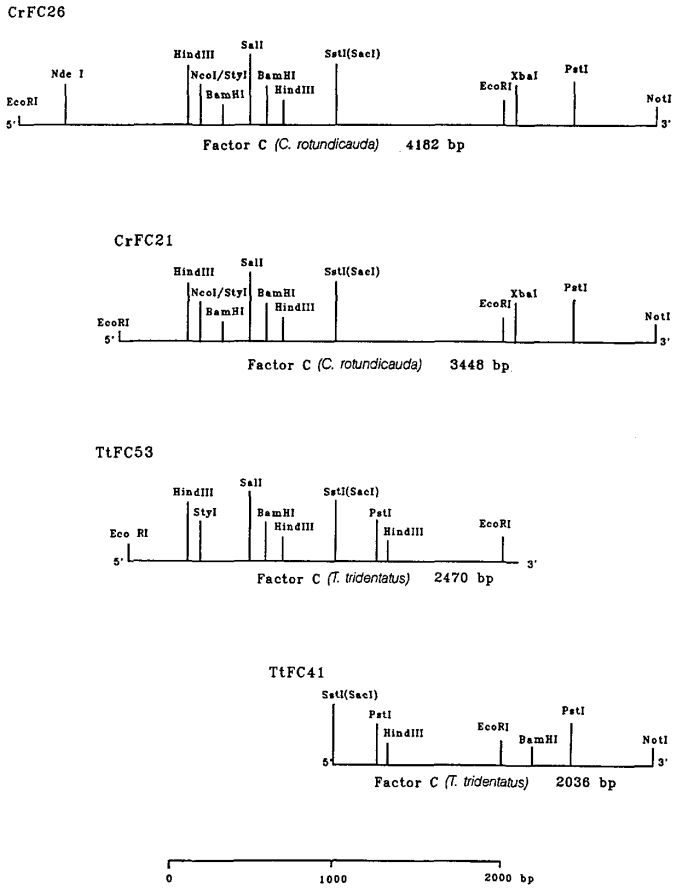


Figure 3. A comparison between the complete restriction maps of CrFC26, CrFC21 *(C. rotundicauda)*, and TtFC53 and TtFC41 (*T. tridentatus*, adapted from Muta et al., 1991).

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

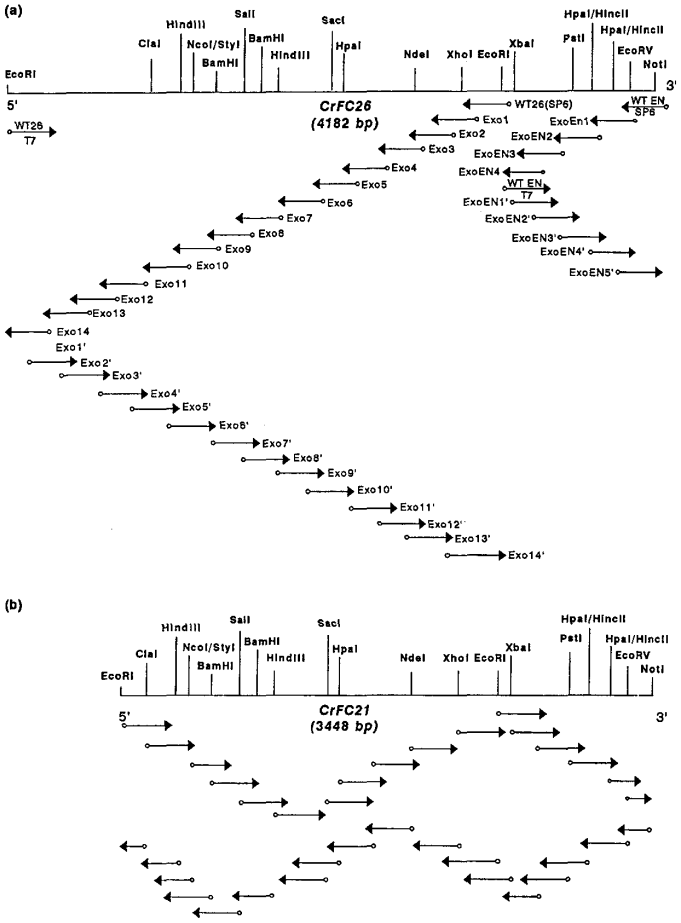


Figure 4. Sequencing strategies of (a) CrFC26 and (b) CrFC21. Deletion subclones were prepared from pEE and pEN in both directions. The arrows indicate the direction and the extent of the sequences obtained. T7 promoter primer (o⟶); SP6 promoter primer (⟵o).

strategies. The CrFC26 was found to be 4182 bp long (Figure 5). This is consistent with the Northern analysis of the amoebocyte total RNA with homologous EE insert, which showed only one hybridizing band at approximately 4 kNt (Figure 1b). This complete cDNA includes 568 nucleotides of 5′ untranslated sequence containing seven ATGs before the real initiation site, an ORF of 3249 nucleotides coding for a protein of 1083 amino acids, a stop codon, and 365 nucleotides of 3′ untranslated sequence. At the 3′ end of the cDNA, the canonical hexanucleotide sequence AATAAA is present 19 nucleotides upstream from the polyadenylation site (at nucleotide position 4142). There are three other AATAAA sequences found at nucleotide positions 183, 239, and 2474, respectively. The ORF codes for a signal peptide of 24 amino acids and a Factor C zymogen of 1059 residues, which, prior to *N*-glycosylation, has a calculated molecular weight of 120,244 daltons. This is close to the estimated molecular weight of single- and double-chain Factor C enzymes purified from the amoebocytes (Ding et al., 1993a). The CrFC21 cDNA of 3448 nucleotides in length, its structural features, and its derived amino acid sequence are shown in Figure 5. With the exception of the absence of the majority of 5′ UTR and some subtle base changes, the structural features within the ORF of CrFC21 bear close resemblance to those of CrFC26.

The unique proteolytic site due to endotoxin-activation of the Factor C enzyme is found between Phe-Ile: F^{762}-I^{763} for CrFC21, and F^{826}-I^{827} for CrFC26, thus indicating the integrity of the cloned Factor C cDNA, and its potentials in endotoxin detection.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

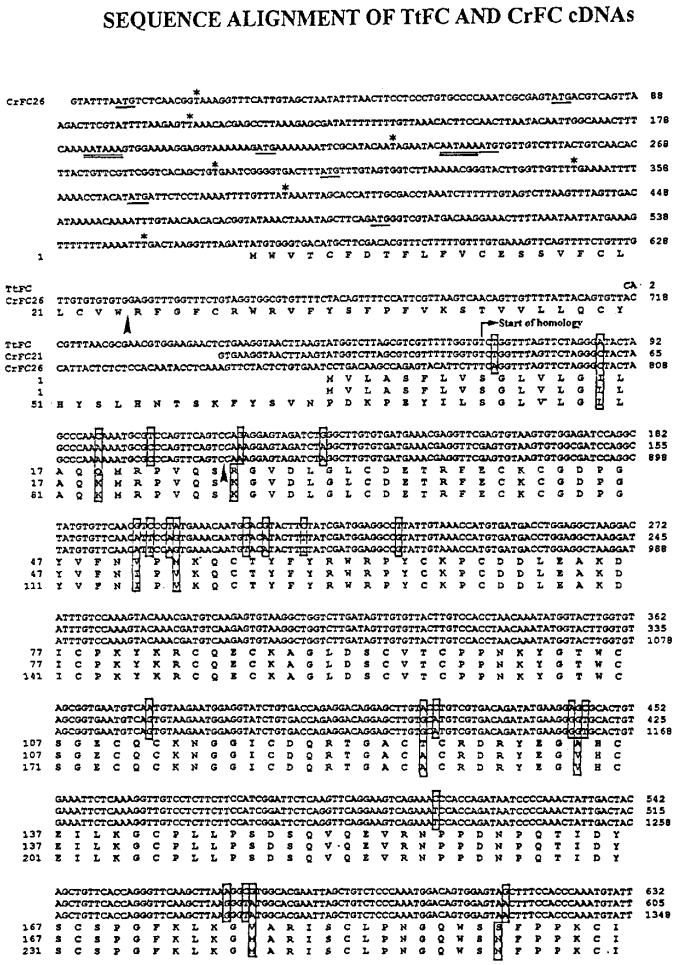


Figure 5. Alignment of the complete DNA sequence and deduced amino acid sequence of *C. rotundicauda* Factor C (CrFC26 and CrFC21) with the Factor C of *T. tridentatus* (TtFC, upper sequence is adapted from Muta et al. 1991). The numbering of the amino acid residues is found at the left while that of the DNA (in bp) is found at the right. The site of truncation of the signal peptide is indicated by a small arrowhead (▲). The "start of homology" region is indicated by a directional arrow (↱). The differences in both the DNA and amino acid sequences are boxed (□). The cleavage site of the Factor C protein into heavy- and light-chain intermediates is indicated by a hollow thick arrow (⬡) between residues R and S, while proteolysis of the light chain into A and B due to endotoxin-activation is indicated by a solid thick arrow (⬟) between residues F and I. The potential glycosylation sites are marked with closed diamonds (◆). The amino acid residues constituting the catalytic triad by analogy with trypsin are indicated by white stars (✪). The four AATAAA sites are marked with double underlines (═). The seven false start ATG sites are indicated with single underlines. These sites are terminated shortly by in-frame stop codons indicated by black stars (★) located several bases downstream.

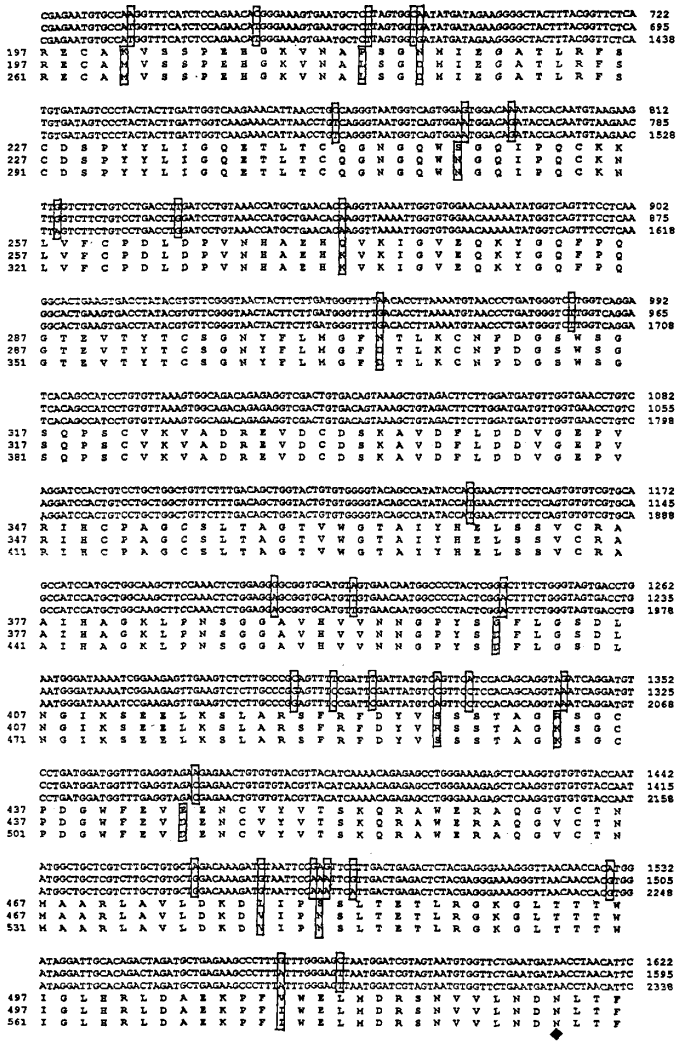
Homology between the DNA sequences of CrFC and TtFC

In comparing CrFC26 and TtFC53 (Figure 5), the *C. rotundicauda* cDNA was found to be longer by 716 nucleotides at the 5′ end, and it also has an extra 64 amino acids upstream from the starting *met*. The start of homology of both the DNA and amino acid sequences between the two species was found at nucleotide position 785 for CrFC26 and at position 69 for TtFC. The homologous region extends through to the poly(A) tail of both genes, giving an overall homology of 97.7%. Most of the dissimilarities, however, were found in the 3′ UTR, where a lower percentage homology of 86% was observed. On the other hand, similar to TtFC, there are six potential *N*-glycosylation sites within the CrFC sequence Asn-Xaa-Ser/Thr.

A majority of the differences between the two sequences found within the ORF were due to base substitutions (Figure 5). In fact, these substitutions explain the perceived differences in the restriction sites previously obtained by restriction mapping (see Figure 3). For instance, the absence of the *Bam*HI (GGATCC) site in TtFC, which in CrFC26 is located 1549 bp from the 5′ end, was due to the substitution of the first G with a T in TtFC. The converse is shown with the *Bam*HI site at 2430 of TtFC in which the second C is substituted with a T in CrFC. Similarly, the *Pst*I site (CTGCAG) of TtFC

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Figure 5. *Continued*



(at 1683 bp from its 5′ end) was not found in CrFC because the first C was substituted with a T. Again, a C-to-T base substitution (at 1700 and 2443 bp positions for CrFC21 and CrFC26, respectively) was responsible for the loss of the third *Hind*III (AAGCTT) in CrFC, which in TtFC was located 1724 bp from the 5′ end. A change from A (at 646 of TtFC) to T (at 1362 of CrFC26) gave rise to either an *Sty*I or *Nco*I site in CrFC26, whereas TtFC is recognized by only *Sty*I. Base substitution was also shown in the *Xba*I (TCTAGA) site at 3101 bp of CrFC, where the first A is substituted with a C in TtFC.

The 5′ noncoding region of CrFC and its computational analysis

Figure 6 shows the secondary structure predicted for the 5′ ends of the mRNA of CrFC21 and CrFC26. Numerous hairpin stems and loops are seen for the CrFC26, obscuring its real start codon, whereas CrFC21 showed a well-exposed AUG start site for ribosome binding.

Exceptionally long 5′ untranslated sequences occur on unusually interesting mRNAs such as epidermal growth factor (EGF) (Scott et al., 1983), oncogenes (Watt et al., 1983), and heat shock proteins

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

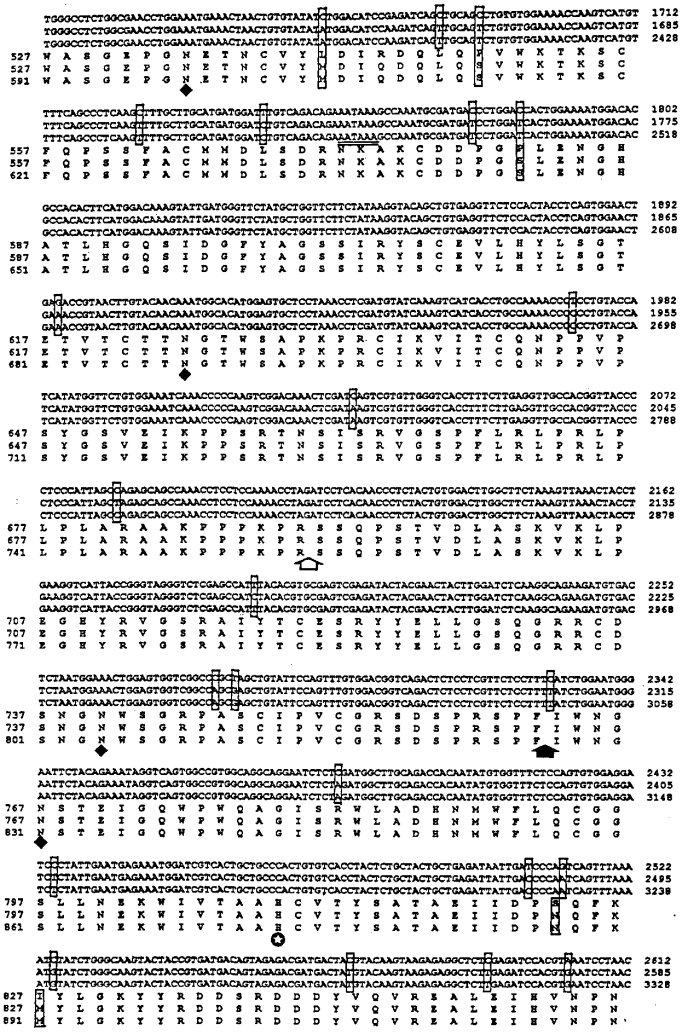


Figure 5. *Continued*

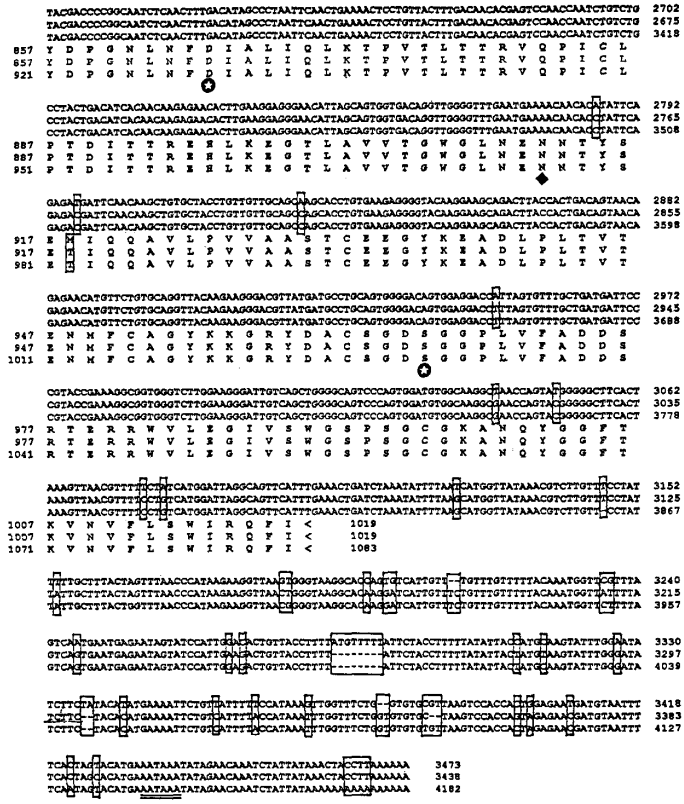
(Ingolia and Craig, 1981). This invites speculation that long structures of the 5′ noncoding region participate in the regulated expression of these genes, including that of Factor C. However, a long 5′ noncoding sequence in some mRNAs may have a deleterious effect (Kozak, 1983), especially if it contains secondary structures that may deflect ribosomes from the authentic initiation site that lies farther downstream. This was true for the Semliki Forest virus (Lehtovaara et al., 1982) and the recombinant fish antifreeze protein III gene (Li et al., 1991). This could also apply to the CrFC26 since there are seven false ATG sites (boxed in Figure 5) upstream from the real ATG site at nucleotide position 569. All these false start sites were followed shortly by in-frame stop sites downstream. This configuration agrees with the ribosome scanning-reinitiation model of translation in eukaryotic genes (Kozak, 1986). This is probably one reason why there exists another form of Factor C mRNA having a much shorter 5′ noncoding sequence as in CrFC21, where it may be translated more efficiently.

Variant forms of the CrFC gene

The Factor C cDNA sequence indicates the existence of at least two types of Factor C mRNA in *C. rotundicauda*. The 5′-end terminal sequences of CrFC26 and CrFC21 differed significantly. Compari-

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Figure 5. *Continued*

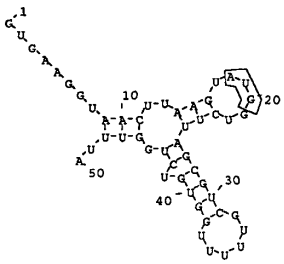


son of the sequences of *C. rotundicauda* Factor C clones CrFC21 and CrFC26 with that of *T. tridentatus* (TtFC53) shows that CrFC21 and TtFC53 share precise homology at their 5′ ends. Yet, the most notable difference in the DNA sequences of Factor C from the two species of horseshoe crabs is the existence of an extra 716 nucleotides at the 5′ end of the CrFC26 (Figure 7). Thus, CrFC21 was not merely a truncated species of CrFC26; rather, it was derived from a totally distinct species of mRNA that is exactly identical to the one that gave rise to the *T. tridentatus* Factor C cDNA. To demonstrate unequivocally the authenticity of the 5′ noncoding region of CrFC26, a further Northern blot analysis was performed, but using only the first 368 bp of CrFC26 (fragment flanked by *Eco*RI and *Nde*I) as probe. The Northern blot result in Figure 1c clearly demonstrates that the entire 5′ end of CrFC26 truly belongs to this particular species of Factor C mRNA. Thus, the existence of the two types of *C. rotundicauda* Factor C cDNA could be attributable to differential splicing of the initial primary transcript around its 5′ terminal.

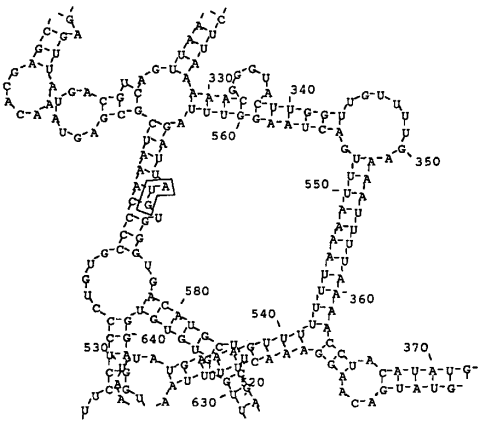
Structural domains and evolutionary significance of C. rotundicauda *Factor C*

The secondary folding structures of several blood clotting factors in the human coagulation system are usually identified by common domains derived from homologous regions in the DNA sequence (Furie and Furie, 1988). Typically, all these factors have a signal peptide at its amino-terminus, and a catalytic domain at the carboxy-terminus, which for serine proteases like *C. rotundicauda* Factor C, also contain the catalytic triad, His^{809}-Asp^{865}-Ser^{966} for CrFC21 and His^{873}-Asp^{929}-Ser^{1030} for CrFC26 (Figure 5). A hydropathy profile based on the formula of Kyte and Doolittle (1982) of the first 59 amino acids of CrFC26 showed the first 24 residues to be the putative signal peptide rich in hydrophobic amino acids (Figure 8). CrFC26 has five short consensus repeats (residues 206–259, 263–318,

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)



(a) CrFC 21



(b) CrFC 26

Figure 6. The computational predictions of the secondary structures of mRNAs of (a) CrFC21 and (b) CrFC26. The leader sequence of CrFC26 shows numerous hairpin stems and loops. The authentic AUG start codon is boxed. The start codon in CrFC21 is exposed in a loop, and appears more accessible.

324–385, 640–698, and 758–812), one lectin-domain (500–632), one EGF-like domain (167–200), a Cys-rich region (90–165), and one Pro-rich region (732–754). The serine protease domain is found at positions 827–1083. No propeptide sequence, kringle domains (Park and Tulinsky, 1986), amino acids stacks, or finger domains were found.

Scanning the EMBL gene bank in a homology search for the serine protease catalytic domain of CrFC revealed that it is structurally closest to *T. tridentatus* Factor C, with 97.7% homology in 777 overlapping nucleotides (Table 1).

From the overall comparison with serine proteases from other species, the CrFC is structurally closest to prothrombin and thrombin. This finding agrees well with the preference of Factor C for hydrolysis of synthetic substrates of thrombin. Surprisingly, a human liver hepsin (Leytus et al., 1986) exhibited a slightly higher level of homology with CrFC than did other blood coagulation factors. Although the function of hepsin is still unknown, this serine protease, which lacks a typical signal peptide but instead has a transmembrane domain, was nev-

Table 1. Percentage homology comparison of the catalytic domain of *Carcinoscorpius rotundicauda* Factor C with other serine proteases based on entries in the EMBL data bank.

Gene	Homology (%)	Overlap (bp)
Tachypleus tridentatus Factor C	97.7	777
Onchorhynchus mykiss thrombin	54.4	709
Rattus norvegicus thrombin	52.2	624
Rat prothrombin	52.2	624
Acipencer transmontanus thrombin	52.8	623
Oryctolagus cuniculus thrombin	57.3	595
Mouse prothrombin	52.7	509
Human blood coagulation Factor VII	55.7	420
Guinea pig Factor IX	57.4	296
Human liver hepsin	59.0	278

```
                                                                                                  Start of homology
                                                                                                  ------------------->
CrFC21    1                                  GTGAAGGTAACTTAAGTATGGTCTTAGCGTCGTTTTTGGTG | TCTGGTTTAGTT
TtFC53    1  CACGTTTAACGCGAACGTGGAAGAACTCTGAAGGTAACTTAAGTATGGTCTTAGCGTCGTTTTTGGTG | TCTGGTTTAGTT
CrFC26  717  ACCATTACTCTCTCCACAATACCTCAAAGTTCTACTCTGTGAATCCTGACAAGCCAGAGTACATTCTT | TCAGGTTTAGTT
```

Figure 7. Comparison of the amino-terminal nucleotide sequences of different Factor C cDNAs isolated from the horseshoe crabs. The clones TtFC53 and TtFC41 is from *T. tridentatus* (Muta et al., 1991) while clones CrFC21 and CrFC26 are from *C. rotundicauda* (this work). CrFC21 and TtFC53 have identical 5′-end sequence, while CrFC26 has an extra 716 nucleotides at its 5′ end.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

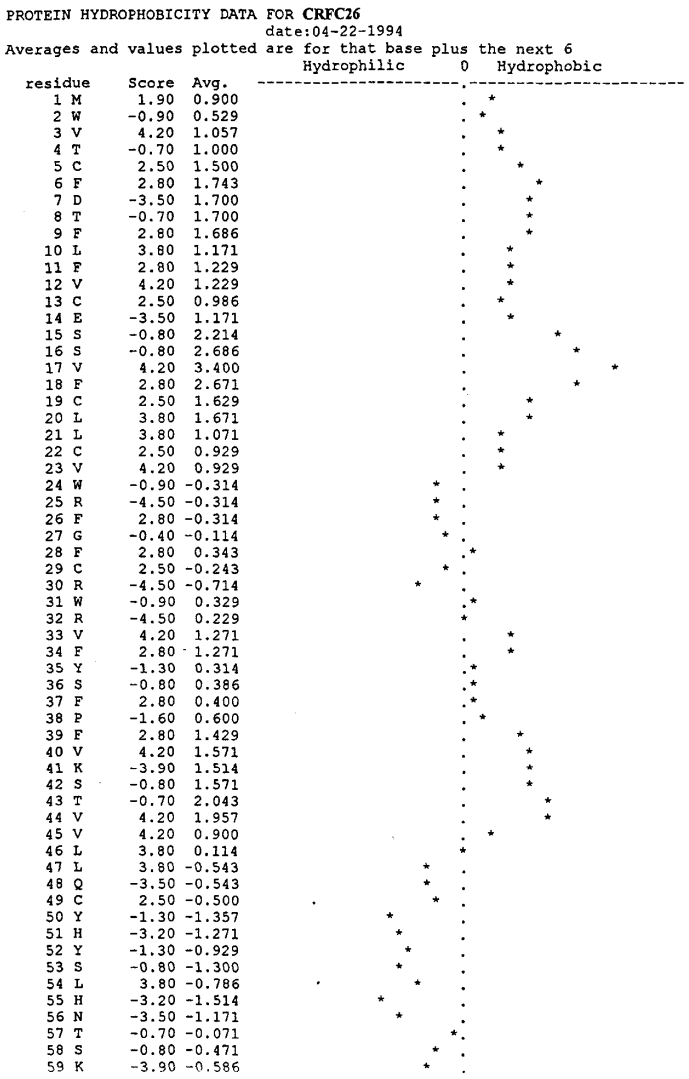


Figure 8. Hydropathy analysis of the first 59 deduced amino acid sequence of *C. rotundicauda* Factor C. The hydrophobic peak spanning residues 1–24 represents the putative signal sequence.

ertheless shown to have characteristics typical of trypsin. Based on the primary structure of CrFC, it was also found to the catalytically similar to trypsin-type serine protease, although structurally closer to thrombin. This may be explained by the Asp^{1024} (as opposed to Ser) in CrFC26, analogous to Asp^{189} in trypsin (Hartley, 1970), which strongly suggests that CrFC possesses a substrate specificity similar to that of trypsin.

The high degree of homology of the DNA sequences of CrFC and TtFC substantiates, at the molecular level, the proximity of these two species of horseshoe crab in the course of evolution. Calculations based on mutation distances among the amino acid sequences of coagulogen purified from the four extant species of horseshoe crab have indicated that *C. rotundicauda* could have diverged more recently *from T. tridentatus* than did *L. polyphemus* and *T. gigas* (Srimal et al., 1985). This is despite the disparity of their morphology and ecological environment. *C. rotundicauda* is approximately less than half the size of the other three species, and it is the only one that lives in brackish waters whereas the rest inhabit the open seas. In addition to other recent

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

findings on the molecular structure and enzymatic components of the coagulation enzymes, our present study on Factor C favors the widespread conservation of the system across the different species of horseshoe crabs.

Expression potentials of recombinant Factor C

In view of the harsh estuarine-brackish habitat of the *C. rotundicauda,* where extremely high populations of Gram-negative bacteria up to 10^{12} CFU/g of sediment is expected to challenge its survival, this species of horseshoe crab has a potent antibacterial defense mechanism (Yeo et al., 1993; Ding et al., 1993b; Ho et al., 1994). Indeed its amoebocyte lysate has been found to be more sensitive to endotoxin than the *Limulus* or *Tachypleus* amoebocyte lysates (Kim et al., 1988). Being the first enzyme in the coagulation cascade reaction that is activated by endotoxin, recombinant Factor C from *C. rotundicauda* would be of great significance to the pharmaceutical industry, in which pyrogen-detection constitutes an inherent part of quality assurance. The commercial implications of Factor C could be realized by its synthesis from expression vectors to produce recombinant Factor C to be used to formulate a rapid and sensitive detection kit for Gram-negative bacterial endotoxin using synthetic substrates. Subcloning the CrFC cDNAs in expression vectors and in vitro expression studies with transcription-coupled-translation (TnT) reactions in which the CrFC cDNA inserts were driven by T7 promoters have revealed higher expression of Factor C from CrFC21 constructs (our unpublished data, patent pending). In comparison with native CrFC26, the higher expression capability of CrFC21 is probably due to the absence of multiple false-start ATGs in the 5′ region, the more easily accessible AUG, and the presence of a consensus Kozak sequence in the CrFC21.

Experimental Procedures

Procurement of C. rotundicauda *amoebocytes*

The horseshoe crabs were partially bled by cardiac puncture and subsequently returned alive to their habitat. Blood was collected into equal volumes of pyrogen-free NaCl (3%) containing 0.125% *N*-ethylmaleimide (Calbiochem) as anticoagulant. Amoebocytes were pelleted by centrifugation at 150 × *g* for 30 minutes at room temperature. The cells were washed with 3% NaCl and quick-frozen in liquid nitrogen before being used for RNA preparation. All procedures were performed under pyrogen-free conditions at 0°–4°C unless otherwise stated.

Poly (A)⁻ RNA preparation

Total RNA was purified according to Chirgwin et al. (1979). Amoebocytes of 2.5 g wet weight were resuspended in 18 ml of 4 M guanidinium isothiocynate (Fluka) containing 20 mM sodium acetate, pH 5.2, 0.1 mM DTT, and 0.5% sarkosyl. Cell lysis was enhanced by passing the suspension several times through an 18-gauge needle and four to five times through a 23-gauge needle until a homogenous lysate was obtained. The lysate was then layered onto a cushion of 4 ml of 5.7 M CsCl solution in a Beckman quickseal tube. Ultracentrifugation was effected at 400,000 × *g* for 16 hours at 18°C. The total cellular RNA pellet was air-dried and resuspended in 1 ml of TES (10 mM Tris-HCl, pH 7.4, containing 5 mM EDTA and 1% SDS) and extracted twice with chloroform:butanol (4:1). The RNA was precipitated with ethanol, resuspended in DEPC-treated water, digested with RNase-free DNase I (BRL), and extracted with phenol-chloroform. Poly(A)$^+$ RNA was purified by chromatography through Oligo-dT Tris Acryl (IBF, France).

cDNA synthesis and cloning

The mRNAs purified from amoebocytes were used to synthesize cDNAs following a modification of the method of Gubler and Hoffman (1983). *Not*I primer adaptors were used to prime the first-strand cDNA synthesis to ensure orientation-specific ligation to vector, λgt22 (Promega, USA). A replacement reaction was employed for the second-strand synthesis while *Eco*RI adaptors were used to produce sticky-ended double-stranded cDNAs. After digestion with *Not*I, the double-stranded cDNAs flanked by *Eco*RI–*Not*I restriction sites were ligated to *Eco*RI–*Not*I-digested λgt22 vector arms. The recombinant DNA was packaged and transduced into *Escherichia coli,* LE392, and Y1090. The cloning strategy is outlined in Figure 9.

Screening of λgt22 cDNA library

The λgt22 cDNA library was screened using, as probe, *T. tridentatus* Factor C cDNA (TtFC53), which is a partial fragment of the Factor C cDNA (Muta et al., 1991; a kind gift from Prof. S. Iwanaga). As a rare copy gene, *C. rotundicauda* Factor C cDNA represented only 0.03% of the amoebocyte cDNA library. A total of 40 presumptive clones were isolated, of which 16 were further studied. The *C. rotundicauda* λ Factor C recombinants (CrFC) were

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

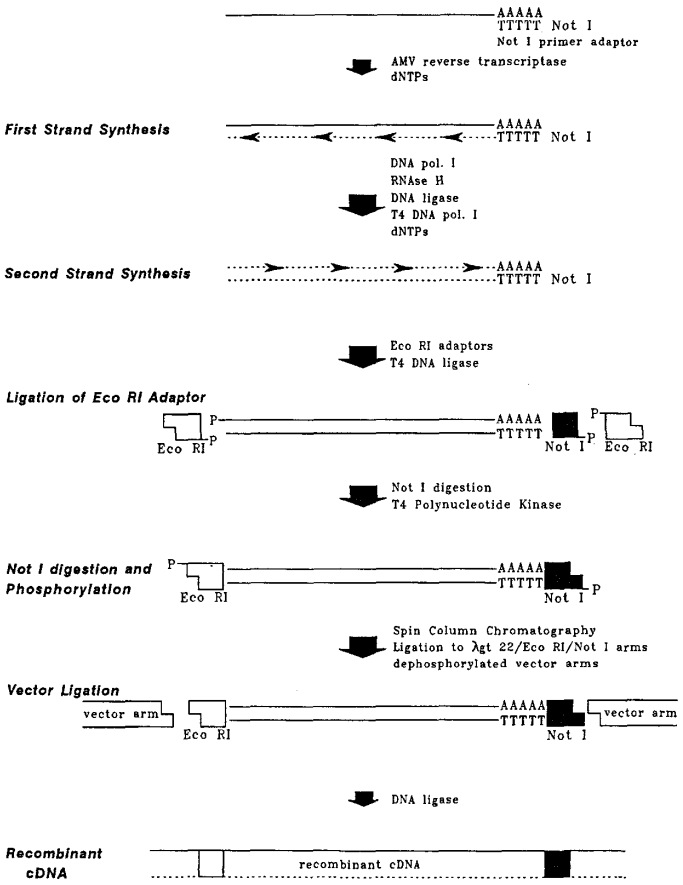


Figure 9. The overall strategy utilized for the synthesis and cloning of the amoebocyte cDNA. *Not*I primer adaptors were used to prime the first-strand synthesis and allow orientation-specific or directional ligation into the vector, bacteriophage λgt22. A replacement reaction was employed for the second-strand synthesis reaction using RNase H, DNA polymerase I, and DNA ligase while *Eco*RI adaptors were used to produce a sticky-ended double-stranded cDNA. Afer digestion with *Not*I, the unique restriction ends were phosphorylated with T_4 kinase to allow directional ligation to the dephosphorylated *Eco*RI–*Not*I-digested vector arms. After packaging in vitro, the phage particles were transduced into bacterial host cells (*E. coli,* LE392, and Y1090) for subsequent propagation or cloning.

purified, and the size of their inserts were determined after digestion with *Eco*RI and *Not*I restriction enzymes.

Northern and Southern blot analysis

RNA samples were denatured with glyoxal and Me_2SO (McMaster and Carmichael, 1977), separated on 1% agarose gel, and transferred to nylon membrane (PALL Biodyne A, 0.22 μm) (Reed and Mann, 1985). Likewise, DNA samples were fractionated on 1% agarose gels, depurinated, and transferred as above under alkaline conditions. ^{32}P-labeled DNA probes (either heterologous TtFC53 or, later, homologous CrFC cDNA) with specific activities of 10^9 cpm/μg of DNA were routinely used for hybridization under standard conditions (Maniatis et al., 1982). With the heterologous TtFC53 probe, hybridization was carried out at 42°C in the presence of 50% formamide. Posthybridization washing was effected with 1× SSC/0.1% SDS at room temperature followed by 0.1× SSC/0.1% SDS at 42°C until low background radioactivity was attained. With homologous CrFC cDNA probes, higher-stringency conditions of hybridization and washings were employed at 65°C.

Subcloning of CrFC cDNA fragments in pGEM 11Zf(+)

To facilitate further manipulations of the CrFC cDNA, the *Eco*RI-*Eco*RI (EE) and *Eco*RI-*Not*I (EN) fragments were subcloned into plasmid pGEM 11Zf(+) and transformed into *E. coli* JM109. The orientation of the EE fragments subcloned into the *Eco*RI site of pGEM 11Zf(+) was determined before restriction mapping using the unique *Xho*I site at the 3′ end of EE fragment and pGEM 11Zf(+). The

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

subclones containing EE and EN fragments are henceforth named pEE and pEN, respectively. The full-length CrFC21 and CrFC26 were also inserted into pGEM 11Zf(+).

DNA sequencing of CrFC21 and CrFC26

The DNA sequences of CrFC21 and CrFC26 were determined on both strands by the dideoxy method of Sanger et al. (1977). Figure 2 shows the sequencing strategies for CrFC26 and CrFC26. To countercheck the sequences obtained, the 5′ ends of some other CrFC subclones; CrFC1, CrFC16, CrFC35, and CrFC69 (see Figure 2) were also sequenced using T7 promoter as the primer binding site.

Acknowledgments

This work was supported by grant RP: 88/357 (B.H. and J.L.D.). M.A.A.N. III was in receipt of an National University of Singapore Research Scholarship. We thank Mr. C. Chai for technical assistance and Mr. H.K. Yip for photography.

References

Chirgwin, J.M., Przbyla, A.E., McDonald, R.J., and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294–5299.

Ding, J.L., Navas, M.A.A., and Ho, B. (1993a). Two forms of Factor C from the amoebocytes of *Carcinoscorpius rotundicauda:* purification and characterisation. *Biochim Biophys Acta* 1202:149–156.

Ding, J.L., Sababathy, T.K., and Ho, B. (1993b). Morphological changes of *Carcinoscorpius rotundicauda* amoebocyte and *Escherichia coli* during their interaction. *Cytobios* 75:21–32.

Furie, B., and Furie, B.C. (1988). The molecular basis of blood coagulation. *Cell* 53:505–518.

Gubler, U., and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* 25:263–269.

Hartley, B.S. (1970). Homologies in serine proteases. *Philos Trans R Soc Lond Biol* 257:77–87.

Ho, B. (1983). An improved *Limulus* gelation assay. *Microbios Lett* 24:81–84.

Ho, B., Kim, J.C., and Ding, J.L. (1993). Electrophoretic analysis of endotoxoin-activated gelation reaction of *Carcinoscorpius rotundicauda* amoebocyte lysate. *Biochem Mol Biol Int* 29:687–694.

Ho, B., Yeo, D.S.A., and Ding, J.L. (1994). A tetrodotoxin neutralizing system in the haemolymph of the horseshoe crab, *Carcinoscorpius rotundicauda. Toxicon* 32:755–762.

Ingolia, T.D., and Craig, E.A. (1981). Primary sequence of the 5′ flanking regions of the *Drosphila* heat shock genes in chromosome subdivision 67B. *Nucleic Acids Res* 9:1627–1642.

Iwanaga, S., Morita, T., Miyata, T., and Nakamura, T. (1985). Haemolymph coagulation system in *Limulus*. In : *Microbiology,* Leive, L., Bonventre, P.F., Morello, J.A., Schlesinger, S., Silver, S.D., and Wu, H.C. (eds.). Washington, D.C.: American Society of Microbiology, pp. 29–32.

Kim, J.C., Ding, J.L., and Ho, B. (1988). Endotoxin activation of clotting proteins from *Carcinoscorpius rotundicauda* ameobocyte lysate. In: *Advances in Biochemistry & Biotechnology in Asia and Oceania,* Sipat, A., Ampon, K., Perumal, R., Aziz, S., and Thambyrajah, V. (eds.). Malaysian Biochemical Society (FAOB), Kuala Lumpur, Malaysia: pp. F1.

Kozak, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiol Rev* 47:1–45.

Kozak, M. (1986). Bifunctional messenger RNAs in eucaryotes. *Cell* 47:481–483.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying hydropathic character of a protein. *J Mol Biol* 157:105–132.

Lehtovaara, P., Soderlund, H., Keranen, S., Petterson, R., and Kaarianen, LL. (1982). Extreme ends of the genome are conserved and rearranged in the defective interfering RNAs of Semliki Forest virus. *J Mol Biol* 156:731–748.

Leytus, S.P., Foster, D.C., Kurachi, K., and Davie, E.W. (1986). Gene for human Factor X: a blood coagulation factor whose gene organization is essentially identical with that of Factor IX and protein C. *Biochemistry* 25:5098–5102.

Li, X., Trinh, K.Y., and Hew, C.S. (1991). Expression and characterization of an active and thermally more stable recombinant antifreeze polypeptide from ocean pout, *Macrozoarces americanus,* in *Escherichia coli:* improved expression by the modification of the secondary structure of the mRNA. *Protein Eng* 4:995–1002.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor, Laboratory Press.

McMaster, G.K., and Carmichael, G.G. (1977). Analysis of single and double-stranded nucleic acids on polyacrylamide and agarose gel by using glyoxal and acridine orange. *Proc Natl Acad Sci USA* 74:4835–4838.

Muta, T., Miyata, T., Misumi, Y., Tokunaga, F., Nakamura, T., Toh, Y., Ikehara, Y., and Iwanaga, S. (1991). *Limulus* Factor C: an endotoxin sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like and lectin-like domains. *J Biol Chem* 266:6554–6561.

Nakamura, T., Morita, T., and Iwanaga, S. (1985). Intracellular proclotting enzyme in *Limulus (Tachypleus tridentatus)* hemocytes: its purification and properties. *J Biochem* 97:1561–1574.

Nakamura, T., Morita, T., and Iwanaga, S. (1986). Lipopolysaccharide sensitive serine-protease zymogen (Factor C) found in *Limulus* hemocytes: isolation and characterization. *Eur J Biochem* 154:511–521.

Nakamura, T., Tokunaga, F., Morita, T., Iwanaga, S., Kusumoto, S., Shiba, T., Kobayashi, T., and Inoue, K. (1988). Intracellular serine-protease zymogen, Factor C, from horseshoe crab hemocytes: its activation by synthetic lipid A analogues and acidic phospholipids. *Eur J Biochem* 176:89–94.

Navas, M.A.A., Ding, J.L., and Ho, B. (1990). Inactivation of

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)

Factor C by dimethyl sulfoxide inhibits coagulation of the *Carcinoscorpius* amoebocyte lysate. *Biochem Int* 21:805–813.

Park, C.H., and Tulinsky, A. (1986). Three-dimensional structure of the kringle sequence: structure of prothrombin-fragment I. *Biochemistry* 25:3977–3982.

Reed, R.C., and Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res* 13:7207–7221.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.

Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R, Fong, N., Selby, M., Rutter, W.J., and Bell, G.I. (1983). Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins. *Science* 221:236–240.

Solum, N.O. (1973). The coagulogen of *Limulus polyphemus* hemocytes, a comparison of the clotted and the non-clotted forms of the molecule. *Thromb Res* 2:55–70.

Srimal, S., Miyata, T., Kawabata, S., Miyata, T., and Iwanaga, S. (1985). The complete amino acid sequence of coagulogen isolated from Southeast Asian horseshoe crab, *Carcinoscorpius rotundicauda. J Biochem* 98:305–318.

Watt, R., Stanton, L.W., Marcu, K.B., Gallo, R.C., Croce, C.M., and Rovera, G. (1983). Nucleotide sequence of cloned cDNA of human c-myc oncogene. *Nature* 303:725–728.

Yeo, D.S.A., Ding, J.L., and Ho, B. (1993). An antimicrobial factor from the plasma of the horseshoe crab, *Carcinoscorpius rotundicauda. Microbios* 73:45–58.

Kopie von subito e.V., geliefert für df-mp Dörries Frank-Molnia & Pohlman Patentanwälte Rechtsanwälte PartG mbB (COM04X01498)