

# Molecular cloning and sequence analysis of the mouse protein C inhibitor gene

Margareta Zechmeister-Machhart, Peter Hufnagl, Pavel Uhrin, Irina Korschineck,  
Bernd R. Binder, Margarethe Geiger \*

Department of Vascular Biology and Thrombosis Research, University of Vienna, Schwarzschanerstrasse 17, A-1090 Vienna, Austria

Received 7 June 1996; revised 29 August 1996; accepted 29 August 1996; Received by D. Schlessinger

## Abstract

The gene encoding mouse protein C inhibitor (*mPCI*) was isolated and its nucleotide sequence determined. Alignment of the genomic sequence with that of a cDNA obtained from mouse testis revealed that the *mPCI* gene (like the human counterpart) is composed of five exons and four introns with highly conserved exon/intron boundaries. It encodes a pre-polypeptide of 405 amino acids, which shows 63% identity with human PCI (hPCI). The putative reactive site is identical to that of hPCI from P5 to P3', suggesting a similar protease specificity. Also the putative heparin binding sites and 'hinge' regions are highly homologous in mouse and hPCI.

**Keywords:** Plasminogen activator inhibitor-3; Protease inhibitor; Serpin; Murine; Blood coagulation; Kallikrein-kinin system

## 1. Introduction

Serpins (*serine protease inhibitors*) are a family of glycoproteins that inactivate serine proteases by forming stable, enzymatically inactive 1:1 enzyme-inhibitor complexes (for review, see Harper and Carrell, 1994). A serpin with a so far undefined biological role is the heparin binding serpin protein C inhibitor (PCI). Human PCI (hPCI) is synthesized in several organs (Laurell et al., 1992; Radtke et al., 1994) and is present in plasma and other body fluids and secretions (Laurell et al., 1992); the highest concentrations are found in seminal plasma (Laurell et al., 1992). In vitro hPCI has been shown to be a non-specific serpin inhibiting a variety of proteases involved in hemostasis (for review, see Geiger, 1988; Suzuki et al., 1989). Furthermore, hPCI also inhibits trypsin (Suzuki et al., 1984), chymotrypsin (Suzuki et al., 1984), tissue kallikrein (Ecke et al., 1992), prostate-specific antigen (Christensson and Lilja, 1994) and the sperm protease acrosin (Hermans et al.,

1994; Zheng et al., 1994). There are several indications that PCI interacts with some of its in vitro target proteases in vivo as well (España and Griffin, 1989; Geiger et al., 1989; Christensson and Lilja, 1994). However, at present the precise biological role of PCI is not known. In this study we cloned and sequenced the gene encoding mouse PCI (*mPCI*). We show that the deduced protein (especially as far as functionally important sites are concerned) is highly homologous to the human counterpart.

## 2. Experimental and discussion

### 2.1. Cloning strategy and restriction mapping of the *mPCI* gene

A mouse liver genomic DNA library was screened by plaque hybridization with a *hPCI* DNA probe. In about 300 000 phages two independent positive clones (*λEMBL150* and *λEMBL135*) were identified. These phage clones were isolated, purified and characterized by restriction digestion with *Bam*HI, *Hind*III, *Bgl*II, *Pst*I, *Kpn*I and *Eco*RI, followed by Southern blot analysis (Southern, 1975). Five *Bam*HI fragments of 4.6, 3.7, 1.6, 1.2 and 0.7 kb were subcloned into pUC19 to yield

\* Corresponding author. Tel. +43 1 40480422; Fax +43 1 4087500; e-mail: Margarethe.Geiger@univie.ac.at

Abbreviations: A, adenosine; cDNA, DNA complementary to RNA; C, cytidine; G, guanosine; h, h, human; m, m, mouse; PCI, *PCI*, protein C inhibitor; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; T, thymidine.

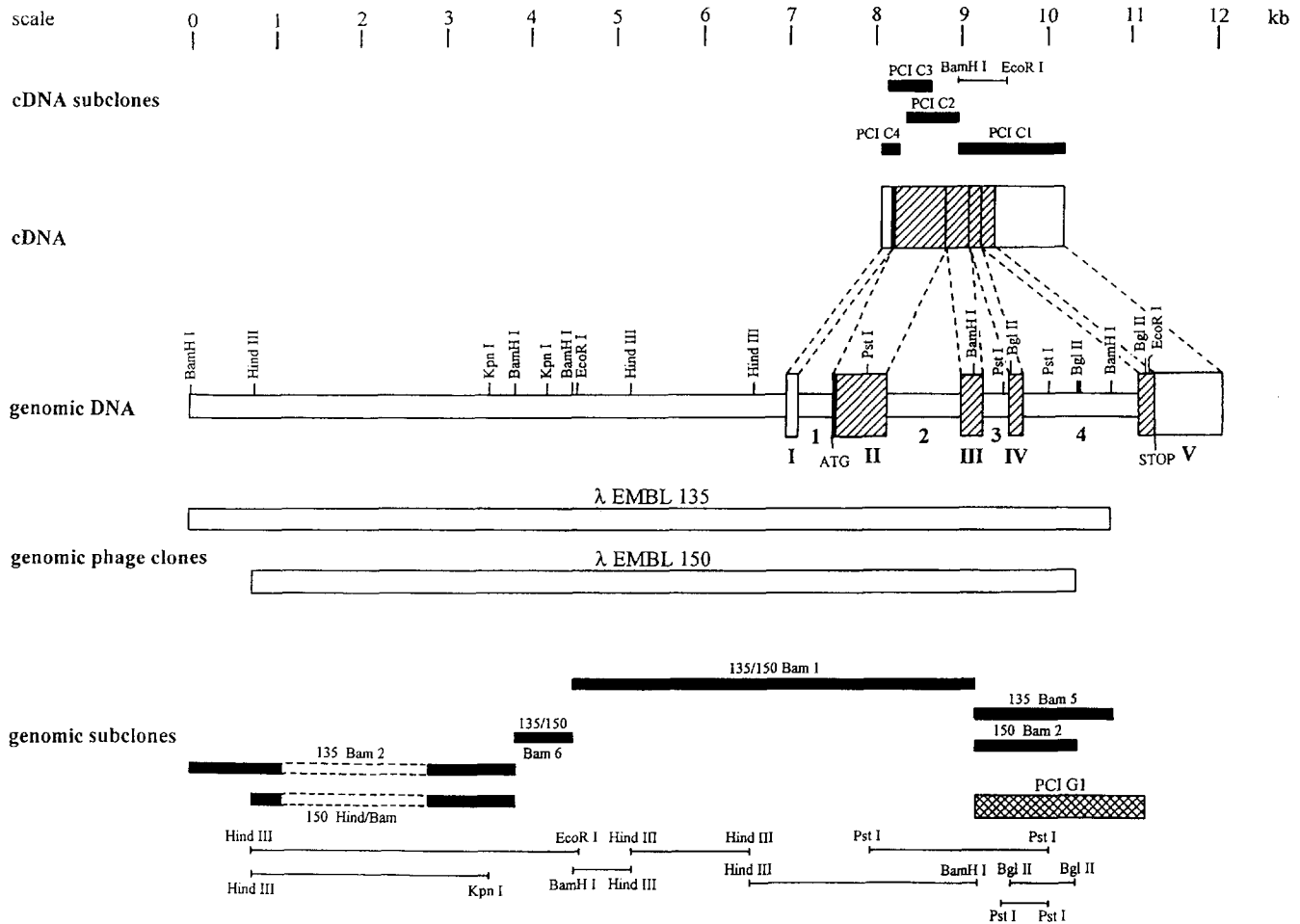


Fig. 1. Physical map of the mPCI gene and subcloning strategy. Mapping of several restriction enzymes for genomic DNA is shown. Connecting lines between the cDNA and genomic DNA maps define corresponding sequences. In the genomic sequence exons are indicated by letters (I–V), introns by numbers (1–4). Untranslated regions are shown as plain boxes, coding regions as hatched boxes, and the signal peptide as a filled box. The ATG initiation codon and the TGA stop codon are indicated. The phage clones ( $\lambda$ EMBL135 and  $\lambda$ EMBL150), the five *Bam*HI subclones (135/150 *Bam*1, 135 *Bam*2, 150 *Bam*2, 135 *Bam*5 and 150/135 *Bam*6), the *Hind*III/*Bam*HI fragment (150 *Hind*/Bam), all generated genomic subclones, and the PCR amplification product (PCI G1) are shown in the lower part of the figure. Broken lines in the subclones 135 *Bam*2 and 150 *Hind*/Bam indicate regions that have not been sequenced. In the upper part of the figure the PCR-amplified cDNA subclones (PCI C1–PCI C4) and the *Bam*HI/*Eco*RI fragment are shown. **Methods:** *Preparation of a human PCI DNA probe:* Using human *PCI* sequence data obtained from EMBL Data Bank (accession numbers M64880–M64884, Meijers and Chung, 1991), this probe was derived from a PCR product from genomic DNA of human umbilical vein endothelial cells. Working under standard conditions (Sambrook et al., 1989) PCR was carried out using a sense primer consisting of bases 7988–8018 and an antisense primer representing the reverse complement of bases 8545–8578. Amplification was performed for 30 cycles as follows: 94°C (1 min) denaturation, 55°C (2 min) annealing, 72°C (3 min) extension with an additional 5-min prolongation of the extension step in the last cycle. *Isolation of genomic clones:* A mouse liver (Balb/C) SP6/T7 genomic library in  $\lambda$ EMBL3 (Clontech, Palo Alto, CA, USA) was screened with the human *PCI* DNA probe, which had been radioactively labelled using a random prime DNA labeling kit (Boehringer-Mannheim, Mannheim, Germany). Approximately 300 000 phages were screened by the plaque hybridization technique (Benton and Davis, 1977; Woo, 1979). Positive clones were plaque purified, and phage DNA was prepared (Blattner et al., 1977). Clones  $\lambda$ EMBL135 and  $\lambda$ EMBL150 were chosen for further analysis. *Cloning of mouse PCI cDNA:* cDNA fragments were amplified from a prepared mouse testis cDNA and from a purchased  $\lambda$ gt11 mouse testis cDNA library (Clontech). Using 7.2  $\mu$ M Oligo-dT primer with an adapter (5'-TGCAGATCTAGAATTCAAGC(dT<sub>22</sub>)-3'), 1  $\mu$ g of total RNA from mouse testis (Chomczynski and Sacchi, 1987) was reverse transcribed into single-stranded cDNA with 75 U AMV reverse transcriptase (Boehringer-Mannheim). *Amplification of the 3'-end of the cDNA by modified RACE method:* PCR was performed using 0.4  $\mu$ M of a nonspecific primer antisense to the adapter and 0.4  $\mu$ M of a gene-specific sense primer located in exon 2 (bases 1072–1089). Using 5  $\mu$ l of the cDNA mixture obtained as described above and 2.5 U Cetus *Taq* DNA polymerase (Perkin-Elmer Cetus, Foster City, CA, USA), PCR was performed for 30 cycles as follows: denaturation, 94°C (30 s); annealing, 52°C (30 s); extension, 72°C (45 s) with an additional 7-min prolongation of the extension step in the last cycle. A second PCR, with a nested gene-specific sense primer in exon III (bases 2164–2183), was used. The PCR reaction generated a 1.2-kb fragment (PCI C1). *Amplification of the 5'-end of the cDNA by a modified RACE method:* Using 10 U T4 RNA ligase (New England Biolabs, Beverly, MA, USA) the anchor (PO<sub>3</sub>-CACGAATT CACTATCGATTCTGGAACCTT CAGAGG-NH<sub>2</sub>) was ligated with the 5'-end of the first-strand cDNA obtained by reverse transcription of testis RNA (Tessier et al., 1986). PCR was performed using 0.2  $\mu$ M of the anchor primer 5'-CTGGTTCGGCCACCTCTG AAGGTTCCAGAATCGATAG-3' and 0.2  $\mu$ M of a gene-specific antisense primer located in exon V (reverse complement of bases 4102–4121). Amplification was performed for 35 cycles using 2U Cetus *Taq* polymerase (Perkin-Elmer Cetus) as follows: 94°C (45 s) denaturation, 57°C

clones 135/150 *Bam*I, 135 *Bam*II, 135 *Bam*III, 150 *Bam*II, and 135/150 *Bam*VI, respectively, as well as a 3.0-kb *Hind*III/*Bam*HI fragment to result in clone 150 *Hind*/Bam. For DNA sequencing, further restriction enzyme fragments of these clones were subcloned into pUC19. These genomic clones and subclones are shown in Fig. 1. The phage clones  $\lambda$ EMBL150 and  $\lambda$ EMBL135 were overlapping from a *Hind*III site (corresponding to the 5'-end of clone 150 *Hind*/Bam) through two nucleotides beyond a *Bgl*II site in intron 4 (610 bp downstream of the exon IV/intron 4 boundary). These are about 9500 overlapping basepairs. Neither clone  $\lambda$ EMBL150 nor clone  $\lambda$ EMBL135 provided sequence information on exon V. In order to obtain information on the missing exon and on the exon/intron boundaries, *mPCI* cDNA was cloned. Total RNA from mouse testis was reverse transcribed and amplified using a modified RACE-PCR technique. On the 3'-end one second-round PCR product (*PCI C1*, 1.2 kb, from base 2164 in exon III to base 5038 in exon V) and on the 5'-end two second-round PCR products (*PCI C2*, 599 bp, from base 719 in exon II to base 2183 in exon III; *PCI C3*, 494 bp, from base 72 in exon I to base 984 in exon II) were amplified. Alignment with an *hPCI* cDNA suggested that clone *PCI C3* did not cover the 5'-end of the *PCI* mRNA. Therefore a further 5'-extended fragment (*PCI C4*, 289 bp, from base 1 in exon I to base 706 in exon II) was amplified from a commercial mouse testis cDNA library using a specific antisense primer in exon II and a sense primer complementary to the adjacent phage sequence. The complete *mPCI* cDNA is covered by these four clones (Fig. 1). The base numbering is according to Fig. 2. In order to obtain the *mPCI* nucleotide sequence not present in the genomic clones  $\lambda$ EMBL 135 and  $\lambda$ EMBL150, and in order to identify the intron 4/exon V boundary, a genomic DNA fragment spanning from exon III to exon V was PCR amplified using information from the cDNA sequence (*PCI C1*). Using a sense primer in exon III and an antisense primer in exon V a 1958-bp fragment was amplified from mouse spleen genomic DNA (clone *PCI G1* in Fig. 1).

## 2.2. Nucleotide sequence analysis and organization of the *mPCI* gene

All clones were sequenced across their entire length, with the exception of clones 150 *Hind*/Bam and 135 *Bam*II, which were partially sequenced. The nucleotide sequence of the *mPCI* gene and the translated amino acid sequence are shown in Fig. 2. The gene (transcription start site to polyadenylation site) is 5038 bp long. Alignment of the genomic sequence with that of the cDNA revealed sequence identity with the exception of base 1130, which was a G in the genomic DNA and an A in the cDNA. *mPCI* consists (as does its human counterpart) of five exons, interrupted by four introns of 417, 866, 312 and 1318 bp in length; the exon/intron boundaries observe the GT-AG rule. The *mPCI* gene encodes a pre-polypeptide of 405 amino acid residues. When the complete amino acid sequence deduced from the cDNA sequence was searched against the SwissProt and PDB databases with the BLAST program (Altschul et al., 1990), it was found to exhibit the highest homology to *hPCI* (63% sequence identity) followed by mouse  $\alpha_1$ -antichymotrypsin-like protein (46% sequence identity) and mouse contrapsin-related protein (42% sequence identity).

Exon I (111 bp) contains the putative transcription initiation site and 5'-untranslated sequence. Exon II (626 bp) contains the translation start codon ATG and encodes the amino-terminal portion of the pre-polypeptide (residues 1–204). Exon III (271 bp) encodes amino acid residues 205–295, exon IV (148 bp) encodes amino acid residues 296–344. Exon V (969 bp) encodes the carboxy-terminal amino acid residues 345–405 and contains the TGA stop codon and 783 bp 3'-untranslated sequence. *mPCI* has one potential *N*-glycosylation site (Asn<sup>247</sup>-Ile<sup>248</sup>-Ser<sup>249</sup>) and one potential *O*-glycosylation site (Ser<sup>372</sup>-Ala<sup>373</sup>-Arg<sup>374</sup>-Pro<sup>375</sup>), as compared to three *N*- and two *O*-glycosylation sites in *hPCI* (Suzuki et al., 1987). The putative reactive site, as determined by alignment with *hPCI*, is encoded by exon V with the P1–P1' residues being Arg<sup>371</sup>-Ser<sup>372</sup>. The amino acid

(45 s) annealing, 72°C (2 min) extension and with an additional 7-min prolongation of the extension step in the last cycle. A second PCR, with a nested gene-specific primer located in exon III (reverse complement of bases 2164–2183) was subsequently performed. The amplification product was 599 bp long and was denominated *PCI C2*. Another product was amplified using the same anchor primer and a gene-specific antisense primer in exon III (reverse complement of bases 2164–2183) for the first PCR. For the second PCR a nested gene-specific antisense primer in exon II (reverse complement of bases 966–983) was used. The PCR reaction generated a 494-bp amplification product (*PCI C3*). *Amplification of the 5'-end of a cDNA from a  $\lambda$ gt11 library:* To obtain further 5'-extended sequence information of *PCI* cDNA, a  $\lambda$ gt11 library prepared from mouse testis mRNA was used (Clontech). PCR was performed using 0.2  $\mu$ M of a gene-specific primer located in exon II (reverse complement of bases 967–984) and 0.2  $\mu$ M of a primer specific for a  $\lambda$ gt11 sequence either located 5' of the insert (5'-ATCGAGCTCGAGATCTAGATTGGTGGCGACGACTCC-3') or located 3' of the insert (5'-TATTATTGTCGACTGCAGACCAACTGGTAATGG-3'). PCR was performed as described above. A second PCR, with a nested gene-specific primer in exon II (reverse complement of bases 689–706) was performed. The PCR reaction generated a 289-bp fragment (*PCI C4*). *PCR amplification of a genomic DNA fragment (*PCI G1*) covering the intron 4/exon V boundary:* Mouse (Balb/C) spleen genomic DNA was isolated (Sambrook et al., 1989); PCR was performed using 1 U rTth *Taq* DNA polymerase (Perkin-Elmer Cetus) and the following sense and antisense primers: the sense primer was located in exon III (bases 2164–2183) and the antisense primer in exon V (reverse complement of bases 4102–4128). The conditions were as follows: denaturation, 95°C (5 min); annealing, 56°C (3 min); and extension, 72°C (3 min) for 30 cycles and with an additional 7-min prolongation of the extension step in the last cycle. The PCR product was denominated *PCI G1*.



sequence of the reactive site of mPCI was found to be identical to that of hPCI from P5 to P3', suggesting a similar protease specificity. The hinge region (Glu<sup>357</sup>-Ar<sup>361</sup> in hPCI (Chai et al., 1993); Glu<sup>355</sup>-Thr<sup>359</sup> in mPCI) is also highly conserved. In hPCI the A<sup>-</sup>-helix and the H-helix have been shown to be involved in heparin binding (Kuhn et al., 1990; Pratt and Church, 1992). When the amino acid sequence of hPCI was aligned with the deduced amino acid sequence of mPCI there were five positively charged amino acid residues in the A<sup>+</sup>-helix of hPCI (between residues 20 and 34) and seven in the respective region of mPCI (between residues 19 and 33). In the H-helix there were seven positively charged residues in hPCI (between residues 283 and 297) and five in mPCI (between residues 281 and 295). Therefore, the putative heparin-binding domains in human and mouse PCI are also highly homologous.

### 3. Conclusions

- (1) The *mPCI* gene has been cloned and sequenced.
- (2) As judged from the presence of specific mRNA, the *mPCI* gene is expressed in testis. It encodes a prepolypeptide of 405 amino acids.
- (3) The deduced amino acid sequence suggests that mPCI is highly homologous to hPCI, especially as far as the putative reactive site, hinge region, and heparin binding sites are concerned.

### Acknowledgement

The authors thank T. Nardelli for his artwork contributions. This work was supported in part by grants from the Austrian Science Foundation (P9478-Med) and from the Anton-Dreher-Gedächtnisschenkung, Vienna University, Medical School.

### References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tools. *J. Mol. Biol.* 215, 403–410.
- Benton, W.D. and Davis, R.W. (1977) Screening  $\lambda$ gt recombinant

- clones by hybridization to single plaques in situ. *Science* 196, 180–182.
- Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.-A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.O. and Smithies, O. (1977) Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. *Science* 196, 161–169.
- Chai, K.X., Chen, L.-M., Chao, J. and Chao, L. (1993) Kallistatin: A novel human serine proteinase inhibitor. Molecular cloning, tissue distribution, and expression in *Escherichia coli*. *J. Biol. Chem.* 268, 24498–24505.
- Chomczynski, P. and Sacchi, N. (1987) Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156–159.
- Christensson, A. and Lilja, H. (1994) Complex formation between protein C inhibitor and prostate-specific antigen in vitro in human semen. *Eur. J. Biochem.* 93, 45–53.
- Ecke, S., Geiger, M., Resch, I., Jerabek, I., Stingl, L., Maier, M. and Binder, B.R. (1992) Inhibition of tissue kallikrein by protein C inhibitor. Evidence for identity of protein C inhibitor with the kallikrein binding protein. *J. Biol. Chem.* 267, 7048–7052.
- España, F. and Griffin J.H. (1989) Determination of functional and antigenic protein C inhibitor and its complexes with activated protein C in plasma by elisa's. *Thromb. Res.* 55, 671–682.
- Geiger, M. (1988) Protein C inhibitor/plasminogen activator inhibitor 3. *Fibrinolysis* 2, 183–188.
- Geiger, M., Huber, K., Wojta, J., Stingl, L., España, F., Griffin, J.H. and Binder, B.R. (1989) Complex formation between urokinase and plasma protein C inhibitor in vitro and in vivo. *Blood* 74, 722–728.
- Harper, P.L. and Carrell, R.W. (1994) The serpins. In: Bloom, A.L., Forbes, C.D., Thomas, D.P. and Tuddenham, E.G.D. (Eds.), *Haemostasis and Thrombosis*, 3rd Edn., Vol. 1, Churchill Livingstone, Edinburgh, pp. 641–653.
- Hermans, J.M., Jones, R. and Stone, S.R. (1994) Rapid inhibition of the sperm protease acrosin by protein C inhibitor. *Biochemistry* 33, 5440–5444.
- Kuhn, L.A., Griffin, J.H., Fisher, C.L., Greengard, J.S., Bouma, B.N., España, F. and Tainer, J.A. (1990) Elucidating the structural chemistry of glycosaminoglycan recognition by protein C inhibitor. *Proc. Natl. Acad. Sci. USA* 87, 8506–8510.
- Laurell, M., Christensson, A., Abrahamsson, P.-A., Stenflo, J. and Lilja, H. (1992) Protein C inhibitor in human body fluids. Seminal plasma is rich in inhibitor antigen deriving from cells throughout the male reproductive tract. *J. Clin. Invest.* 89, 1094–1101.
- Meijers, J.C.M. and Chung, D. W. (1991) Organization of the gene coding for human protein C inhibitor (plasminogen activator inhibitor-3). Assignment of the gene to chromosome 14. *J. Biol. Chem.* 266, 15028–15034.
- Pratt, C.W. and Church, F.C. (1992) Heparin binding to protein C inhibitor. *J. Biol. Chem.* 267, 8789–8794.
- Radtke, K.-P., Fernández, J.A., Greengard, J.S., Tang, W.W., Wilson, C.B., Loskutoff, D.J., Scharrer, I. and Griffin, J.H. (1994) Protein C inhibitor is expressed in tubular cells of human kidney. *J. Clin. Invest.* 94, 2117–2124.

Fig. 2. Nucleotide sequence of the *mPCI* gene. The amino acid sequence is shown below the DNA sequence of the corresponding exons. The numbers at the right margin correspond to the last base in each line, the numbers at the left margin to the first amino acid in each line. The ATG start codon in exon II and the TGA stop codon in Exon V are marked with an asterisk. The numbering of the coding sequence starts with the putative transcription initiation site. The 5'-flanking region is shown with negative numbers. The polyadenylation signal AATAAA (nucleotides 5019–5024) is underlined and the poly(A) addition site (5038) is indicated by a dot. The nucleotide sequence shown in the figure has been deposited with GenBank (accession No. U67878). The sequence from base (–4116) until the 5'-end of clone *I35 Bam2* is not shown in the figure but has been deposited with GenBank (accession No. U67877). **Methods:** All genomic fragments shown in Fig. 1 were subcloned into pUC18, all PCR products into pCR<sup>TM</sup>II, and were sequenced using the dideoxy chain termination method (Sanger et al., 1977). All subclones were sequenced by walking primer strategy (Voss et al., 1991). Sequence analysis was performed using the software package of DNA Star (Madison, WI).

- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Sanger, F., Niklen, S. and Coulsen, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gelelectrophoresis. *J. Mol. Biol.* 98, 503–517.
- Suzuki, K., Deyashiki, Y., Nishioka, J., Kurachi, K., Akira, M., Yamamoto, S. and Hashimoto, S. (1987) Characterization of a cDNA for human protein C inhibitor. A new member of the plasma serine protease inhibitor superfamily. *J. Biol. Chem.* 262, 611–616.
- Suzuki, K., Nishioka, J., Kusumoto, H. and Hashimoto, S. (1984) Mechanism of inhibition of activated protein C by protein C inhibitor. *J. Biochem.* 95, 187–195.
- Suzuki, K., Deshiki, Y., Nishioka, J. and Toma, K. (1989) Protein C inhibitor: structure and function. *Thromb. Haemostasis* 61, 337–342.
- Tessier, C.D., Brousseau, R. and Vernet, T. (1986) Ligation of single-strand oligodeoxyribonucleotides by T4 RNA ligase. *Anal. Biochem.* 158, 171–178.
- Voss, H., Wirkner, U., Jakobi, R., Hewitt, N.A., Schwager, C., Zimmermann, J., Ansorge, W. and Pyerin, W. (1991) Structure of the gene encoding human casein kinase II subunit  $\beta$ . *J. Biol. Chem.* 266, 13706–13711.
- Woo, S.L.C. (1979) A sensitive and rapid method for recombinant phage screening. *Methods Enzymol.* 68, 389–395.
- Zheng, X.L., Geiger, M., Ecke, S., Bielek, E., Donner, P., Eberspächer, U., Schleuning, W.-D. and Binder, B.R. (1994) Inhibition of acrosin by protein C inhibitor and localization of protein C inhibitor to spermatozoa. *Am. J. Physiol.* 267 (Cell Physiol. 36), C466–C472.