J Mol Evol (1998) 47:663–669

JOURNAL OF MOLECULAR EVOLUTION

Identification of Evolutionarily Invariant Sequences in the Protein C Gene Promoter*

C. Arnold Spek,[1,]† Rogier M. Bertina,[1] Pieter H. Reitsma[1,2]

[1] Hemostasis and Thrombosis Research Center, Department of Hematology, University Hospital, Leiden, The Netherlands
[2] Institute of Experimental Internal Medicine, Academic Medical Center, Amsterdam, The Netherlands

Received: 6 August 1998 / Accepted: 23 April 1998

Abstract. Recent studies on human protein C gene expression have revealed the presence of three transcription factor binding sites in close proximity to the transcription start site. Binding sites for the liver-enriched hepatocyte nuclear factors 1 and 3 (HNF-1 and HNF-3, respectively) are located immediately upstream of the transcription start site, whereas just downstream of the start site a presently unidentified transcription factor may bind. To identify other candidate transcription factor binding sites in the protein C promoter, we studied the promoter sequence identity in a number of evolutionarily close and more distant species: *Gorilla gorilla, Pongo pygmaeus, Pan troglodytes, Homo sapiens, Cebus apella, Macaca mulatta, Callithrix jacchus, Papio hamadryas, Macaca fascicularis,* and *Rattus norvegicus.* This analysis showed that a high degree of identity (78%) exists among the different primates. Comparison of the primate consensus sequence with the *Rattus norvegicus* protein C promoter sequence revealed the presence of seven identical regions (I to VII). Two of these regions overlap with established regulatory sequences for HNF-3 and HNF-1 (region VI) and for PCE-1 (region VII), respectively. The functional importance and the transcription factors that may bind to the other five identical regions are now to be determined.

Key words: Protein C — Promoter — Evolution — Transcription factors — Responsive elements

Introduction

Protein C is synthesized in the liver as a vitamin K-dependent zymogen of a serine protease. After activation by the thrombin–thrombomodulin complex, through the release of a dodecapeptide from the amino terminus of the heavy chain (Kisiel, 1979), activated protein C inhibits blood coagulation. In the presence of protein S (Dahlbäck, 1991), calcium ions, and phospholipids, activated protein C inactivates two of the regulatory proteins of the coagulation cascade, factors Va and VIIIa (Walker and Fay, 1992; Esmon, 1992). Furthermore, activated protein C neutralizes plasminogen activator inhibitor-1 (PAI-1) (van Hinsberg et al. 1985), with the stimulation of fibrinolysis as a possible consequence.

The physiological significance of protein C anticoagulant activity is clearly shown in individuals homozygous or compound heterozygous for protein C deficiency. These individuals suffer from massive disseminated intravascular coagulation or neonatal purpura fulminans (Branson et al. 1983; Seligsohn et al. 1984; Marlar et al. 1989). Individuals affected by heterozygous protein C deficiency, although more mildly

* The nucleotide sequence data reported in this paper have been submitted to the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession numbers U77648, U77647, U77650, U77652, U77651, U77649, U77646, U77654, and U77653 for *Gorilla gorilla, Pan troglodytes, Pongo pygmaeus, Cebus apella, Macaca mulatta, Callithrix jacchus, Papio hamadryas, Macaca fascicularis,* and *Rattus norvegicus,* respectively.

† *Present address:* Experimental Internal Medicine, Academic Medical Center, Room G1-113, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Correspondence to: C.A. Spek; *e-mail:* C.A.SPEK@AMC.UVA.NL

Table 1. Primers used for the PCR amplification of the promoter region of the protein C gene in different mammals[a]

Identification	Identical to	Sequence and position
5′pr1	Human	$^{-447}$CATCTTCTCATGATTTTATGTATCAG^{-418}
5′pr2	Human	$^{-396}$CAGCGTC**CCCGGG**CTTGTATGGTGGCACATAAATACATGT^{-357}
5′pr3	Primates	$^{-325}$GAAATATGGAATATTACCT^{-307}
5′pr4	Primates/rat	$^{-302}$TGCTGC/AC/TCA/TTGA/GA/GCAAA/GCTATAATATCT^{-275}
3′pr1	Human	$^{+122}$CTCTTCTCTTCTCCCGGGGGCAGCCCTCCCTCCACACCCCTCATA^{+78}
3′pr2	Human	$^{+51}$GGCGGGTCGTGGAGATACTG^{+32}
3′pr3	Primates	$^{+24}$GCCGTCCTGCCGCCATGAC^{+6}
3′pr4	Rat	$^{+43}$AAGGAGAAACTGCAATTT^{+24}
3′pr5	Primates/rat	$^{+18}$C/GTG/ACCGCCATGT/ACAGNCTG^{-1}

[a] Indicated in boldface in primer 5′pr2 are mutagenic nucleotides, which create a *SmaI* restriction site. All nucleotide numbering is relative to the transcription start site (Plutzky et al. 1986). Primers indicated as identical to primates/rat are chosen after the alignment of the primate sequence with the rat sequence.

affected, are at risk of thrombophlebitis, deep vein thrombosis, and/or pulmonary embolism (Griffin et al. 1981; Broekmans et al. 1983; Allaart et al. 1993).

The human protein C gene, located on chromosome 2q13–q14 (Patracchini et al. 1989), contains nine exons spanning 11 kb of genomic DNA (Foster et al. 1985; Plutzky et al. 1986). Of these nine exons the first and part of the second exon (21 bp) consist of non-protein coding sequences. Genetic analysis of the promoter region of the protein C gene of protein C-deficient individuals revealed the importance of three transcription factor binding sites. From position −33 to −22 a hepatocyte nuclear factor-3 (HNF-3) binding site was identified. Reversely orientated and partly overlapping with this first HNF-3 binding site a second HNF-3 binding site is located from position −26 to −37 (Spek et al. 1995). Finally, it has been shown that from position −22 to −10 a HNF-1 binding site is present (Berg et al. 1994). These three binding sites have been shown to be critically important for protein C gene transcription (Spek et al. 1995; Berg et al. 1994).

Functional characterization of the promoter region of the protein C gene has been carried out by chloramphenicol acetyltransferase reporter assays (Miao et al. 1996). These experiments suggested the presence of two possible liver-specific enhancer regions between −1475 and −175 and a strong silencer element between −175 and −95. Moreover, a unique liver-specific regulatory sequence between +1 and +18 (PCE-1) was identified by DNase I protection assays (Miao et al. 1996).

Functionally important segments of DNA tend to be conserved between species. In the present study we have compared structural features of the protein C promoter region between a number of different species. As a result of this comparison seven evolutionarily identical regions have been identified as potential regulatory regions. Two of the identical regions overlap with the previously reported HNF-3 and HNF-1 binding sites and with the proposed PCE-1 binding site, respectively.

Materials and Methods

Animals

Genomic DNA of different primates, rodents, and carnivores was isolated according to standard protocols. From the primates, DNA of the *Gorilla gorilla, Pongo pygmaeus, Pan troglodytes, Homo sapiens, Cebus apella, Macaca mulatta, Callithrix jacchus, Papio hamadryas,* and *Macaca fascicularis* was kindly provided by Dr. R. Bontrop. Rodent DNA was isolated from blood of *Rattus norvegicus* and *Mus musculus,* whereas carnivore DNA was isolated from *Felis silvestris* and *Canis lupus familiaris.*

Amplification of Genomic DNA

To PCR amplify part of the protein C promoter of the different mammals we used the primers listed in Table 1. Amplifications were performed in a 20-μl reaction mixture containing 10 m*M* Tris–HCl, pH 8.0, 1 m*M* $MgCl_2$, 50 m*M* KCl, 350 ng primers, 100 ng genomic DNA, 250 μ*M* dNTP's, 60 μg/ml BSA, and 0.3 U AmpliTaq DNA polymerase (Perkin Elmer). After an initial incubation at 91°C for 4 min, 32 cycles were carried out at 91°C for 1 min, 58°C for 1 min, and 72°C for 1 min.

Sequencing

PCR-amplified fragments were gel purified and cloned into the pMOS-*Blue* T-vector (Amersham). Subsequently, 2 μg of plasmid DNA was sequenced using the Sequitherm Cycle Sequencing Kit from Epicenter Technologies according to the manufacturer's protocol.

Southern Blotting and Hybridization

The PCR-amplified fragments were electrophoresed on 1.5% agarose gels and transferred to a membrane by alkaline blotting. The Southern blots were prehybridized and hybridized at 60°C in 10% dextran sulfate, 0.5% SDS, 5× Denhards, and 6× SSC. As probe we used a random primed-labeled PCR fragment of the human protein C promoter, ranging from nucleotide −282 to nucleotide −153.

Results

To identify important regulatory regions in the protein C promoter region, we studied the evolutionary identity of

part of this promoter region. Therefore, we attempted to amplify the protein C promoter region in a number of primates, rodents, and carnivores. The primates we analyzed can be subdivided into four groups. First, the humans and great apes (family Hominidae), which is comprised of *Homo sapiens, Gorilla gorilla, Pan troglodytes, and Pongo pygmaeus.* Second, monkeys from the Old World (family Cercopithecinae), i.e., *Macaca mulatta, Papio hamadryas,* and *Macaca fascicularis.* Third, monkeys from the New World (family Cebidae), of which we studied *Cebus apella* and, finally, *Callithrix jacchus* (family Callithrichidae). The rodents we attempted to study were *Rattus norvegicus* and *Mus musculus* (family Muridae), whereas *Felis silvestris* (family Felidae) and *Canis lupus familiaris* (family Canidae) were selected to represent carnivores.

Two upstream (5′pr1 and 5′pr2) and two downstream (3′pr1 and 3′pr2) primers were used to PCR amplify genomic DNA of the different mammals. The primers were deduced from the human protein C sequence. Successful amplification was achieved with all combinations of upstream and downstream primers for the primates only. Irrespective of the combination of primers used, specific fragments could not be amplified from DNA of the other mammals tested (*Rattus norvegicus, Felis silvestris, Canis lupus familiaris,* and *Mus musculus*). Reduction of the annealing temperature during PCR of DNA from the nonprimates from 58 to 56°C resulted in the formation of a number of different products of varying lengths. Southern blotting experiments with the human protein C probe were negative, which makes it unlikely that these are specific protein C promoter fragments.

The derived sequences of the primates and the previous reported region of the human protein C gene (Foster et al. 1985) are shown in Fig. 1. Seventy-six percent of the nucleotides are the same in all species. One of the HNF-3 binding sites is completely identical in all species, whereas the other HNF-3 binding site, the HNF-1 binding site, and the proposed liver-specific enhancer region PCE-1 are not completely identical. However, identity percentages of 83 and 89 for the HNF-3 site and the PCE-1 region, respectively, are higher then the overall identity percentage.

After examining the consensus sequence of the primates we synthesized additional upstream and downstream primers (5′pr3 and 3′pr3, respectively). These primers were chosen on the basis of their maximal identity in all species and on the basis of their location on the far 5′ upstream and 3′ downstream ends. PCR amplification with combinations of primers 5′pr1, 5′pr2, or 5′pr3 and 3′pr1, 3′pr2, or 3′pr3 again did not result in any amplified fragments from the nonprimate DNAs. Decreasing the annealing temperature to 56°C resulted in PCR products of variable length, but again, with Southern blotting none of these products seemed to be specific protein C promoter fragments.

Next we made a 3′ primer (3′pr4) identical to the first 18 base pairs of the reported rat cDNA (Okafuji et al. 1992). Comparison of these 18 base pairs with the human protein C cDNA shows a 78% identity in the first exon, making it likely that these 18 base pairs are located in the first exon of the rat protein C gene. PCR amplification of nonprimate DNA with this primer and primer 5′pr2 resulted in a specific product of approximately 500 base pairs for the rat DNA only. Sequence analysis of this fragment showed an overall identity with the primate consensus of only 52% (Fig. 2). However, seven regions (I to VII) with a significant identity could be distinguished (Table 2). Region VI, which is 97% identical between the primates and *Rattus norvegicus* contains both HNF-3 binding sites and the HNF-1 binding site. These three sites are maximally identical between the species examined. Region VII, ranging from −2 to +15, is 88% identical and contains almost all of the proposed PCE-1 enhancer.

In order to confirm the identity of regions I to VII and thereby the possible functional importance of these regions, we synthesized degenerative primers overlapping with region I (5′pr4) or region VII (3′pr5), respectively. PCR amplification of genomic DNA of cat, dog, and mouse with these primers did not, however, result in specific protein C promoter fragments.

To identify the possible transcription factor(s) binding to the evolutionary identical promoter sequences, we used the MatInspector program (Quandt et al. 1995). This program uses a large library of nucleotide distribution matrices to screen for potential transcription factor binding sites. For identical regions I, III, V, VI, and VII, MatInspector did find a number of potential binding sites with significant matrix similarity scores of over 0.8. However, the presence of these binding sites could not be confirmed by a manual comparison of the relevant sequences with a compilation of vertebrate-encoding transcription factors (Faisst and Meyer, 1992).

Discussion

Recent studies of the protein C gene promoter have identified the liver-enriched transcription factors HNF-1 and HNF-3 as major regulators of protein C gene expression. Directly downstream of these binding sites a DNase I protected region was found of which the functional significance is not yet clear. In this study we examined the evolutionary identity of the protein C promoter region to identify other possible regulatory sequences.

Comparison of the protein C promoter sequence of the different primates revealed that 76% of the nucleotides are the same in all nine species. Eighty-five percent of

```
Human             GTG..TCTTA TAATTAATGG tATtTTAGAT TTGAcGAAAT ATGGAATATT
Chimpanzee        GTGCTTCTTA TAATTAATGG tATCTTAGAT TTGAcGAAAT ATGGAATATT
Gorilla           GTGCTTCTTA TAATTAATGG tATtTTAGAT TTGATGAAAT ATGGAATATT
Java-maquaca      GTGCTTCTTA TAATTAATGG CATCTTAGAg TTGATGAAAT ATGGAATATT
Rhesus monkey     GTGCTTCTTA TAATTAATGG CATCTTAGAg TTGATGAAAT ATGGAATATT
Orangutan         GTGCTTCTTA ....TAATGG CATtTTAGAT TTGATGAAAT ATGGAATATT
Baboon            GTGCTTCTTA TAATTAATGG CATCTTAGAg TTGATGAAAT ATGGAATATT
Marmoset          aTctgTtTTA TAATTAATGG CATCcTAGAT TTGATGAAAT ATGGAATATT
Capuchin          GTctgTCTTA TAATTAATGG CATCcTAGAT TTGATGAAAT ATGGAATATT

Consensus         -T---T-TTA ----TAATGG -AT--TAGA- TTGA-GAAAT ATGGAATATT
                        -351       -341       -331       -321       -311

Human             ACCTGTTGTG CTGATCTTGG GCAAACTATA ATATCTCTGG GcAAAAATGT
Chimpanzee        ACCTGTTGTG CTGATCTTGG GCAAACTATA ATATCTCTGG GTAAAAATGT
Gorilla           ACCTGTTGTG CTGATCTTGG GCAAACTATA ATATCTCTGG GTAAAAATGT
Java-maquaca      ACCTGTTGTG CTGATCTTGG GCAAACTAcA ATATCTCTGG GTAAAAATGT
Rhesus monkey     ACCTGTTGTG CTGATCTTGG GCAAACTATA ATATCTCTGG GcAAAAATGT
Orangutan         ACCTGTTGTG CTGATCTTGG GCAAACTATA ATATCTCTGG GTAAAAATGT
Baboon            ACCTGTgaTG CTGATCTTGG GCAAACTAcA ATATCTCTGG GTAAAAATGT
Marmoset          ACCTGTTGTG CTGATCTTGG aCAAACTAcA ATgTtTCTGG GT.AAAAcGT
Capuchin          ACCTaTT.TG CTGATCTTGG aCAAACTAcA ATgctTCTGG GT.AAAATaT

Consensus         ACCT-T--TG CTGATCTTGG -CAAACTA-A AT---TCTGG G--AAAA--T
                        -301       -291       -281       -271       -261

Human             CCCCATCTGA AAAA.CAGGG ACAACGTTCC TCCCTCAGCC AGCCACTATG
Chimpanzee        CCCCATCTGA AAAA.CAGGG ACAACGTTCC TCCCTCAGCC AGCCACTATG
Gorilla           CCCCATCTGA AAAA.CAGGG ACAACGTTCC TCCCTCAGCC AGCCAgTATG
Java-maquaca      CCCCATCTGA AAAACCAGGG ACAACGTTCC TCCCaCgGCC AGCCACTATG
Rhesus monkey     CCCCATCTGA AAAA.CAGGG ACAACGTTCC TCCCTCAGCC AGCCACTATG
Orangutan         CtCCATCTGA AAAA.CAGGG ACAACGTTCC TCCCTCAGCC AGCCACTATG
Baboon            CCCCATCTG. AgAACCAGGG ACAACGTTCC TCCCTCgGCC AGCCACTATG
Marmoset          CCCCATCTGt AAAACagGGa caAcgtc... TCCCTtAGCC AGCCACT.TG
Capuchin          CCCCATCTGA AAAACagGGa caAcgtccCt cCCCTtAGCC AGCCACT.TG

Consensus         C-CCATCTG- A-AA---GG- --A------- -CCC---GCC AGCCA-T-TG
                        -251       -242       -232       -222       -212

Human             GGGCTAAAAT GAGACCACAT CTGTCAAGGG TTTTGCCCTC ACCTCCCTCC
Chimpanzee        GGGCTAAAAT GAGACCACAT CTGTCAAGGG TTTTGCCCTC ACCTCCCTCC
Gorilla           GGGCTAAAAT GAGACCACAT CTGTCAAGGG TTTgGCCCTC ACCTCCCTCC
Java-maquaca      GGGCTAAAAT GAGACCACAT CTGTCAAGGG TTTTGCCCTC ACCTCCtTCC
Rhesus monkey     GGGCTAAAAT GAGACCACAT CTGTCAAGGG TTTTGCCCTC ACCTCCtTCC
Orangutan         GGGCTAAAAT GAGACCACAT CTGTCAAaGG TTTTGCCCTC ACCTCCCTCC
Baboon            GGGCTAAAAT GAGACCACAT CTGTCAAGGG TTTTGCCCTC ACCTCCtTCC
Marmoset          GGGCTAAAAT GAGACCACAT CTGTCAAGGG TTTTGCCCTC ACCTCCCTCC
Capuchin          GGGCTAAAAT GAGACCACAT CTGTCAAGaG TTTTGCCCTC ACCTCCCTCC

Consensus         GGGCTAAAAT GAGACCACAT CTGTCAA--G TTT-GCCCTC ACCTCC-TCC
                        -202       -192       -182       -172       -162

Human             CTGCTGGACG GCATCCTTGG TGGGCAGAGG TGGGCTTCGG GCAGAaCAAG
Chimpanzee        CcGCTGGACG GCATCCTTGG TGGGCAGAGG TGGGCTTCaG GCAGACCAAG
Gorilla           CTGCTGGACG GCATCCTTGG TGGGCAGAGG TGGGCcTCGG GCAGACCAAG
Java-maquaca      CTGCTGGACG GCATCCTTGG TGGGCAGAGG TGGGCTTCGG GtAGACCAAG
Rhesus monkey     CTGCTGGACG GCATCCTTGG TGGGCAGAGG TGGGCTTCGG GtAGACCAAG
Orangutan         CTGCTGGACG GCATCCTTGG TGGGCAGAGG TGGGCTTCca GCAGACCAAG
Baboon            CTGCTGGACG GCATCCTTGG TGGGCAGAGG TGGGCTTCGG GtAGACCAAG
Marmoset          CcaCTGGACa GCATCCgTGG TGGGCAGAGG TGGGCTTCaG GCAtcCCAAG
Capuchin          CTGCTGGAtG GCATCtgTGG TGGGCAGAGG TGGGCTTCaG GCAacCCAAG

Consensus         C-GCTGGA-- GCATC--TGG TGGGCAGAGG TGGGC-TC-- G-A---CAAG
                        -152       -142       -132       -122       -112

Human             CCGTGCTGAG CTAGGACCAG GAGTGCTAGT GCCAC...TG TTTcTCTATG
Chimpanzee        CCGTGCTGAG CTAGGACCAG GAGTGCTAGT GCCAC...TG TTTGTCTATG
Gorilla           CCGTGCTGAG CTAGGACCAG GAGTGCTAGT GCCAC...TG TTTGTCTATG
Java-maquaca      CCGTGCTGAG CTAGGACCAG GAGTGCTAGT GCCAC...TG TTTGTCTgcG
Rhesus monkey     CCGTGCTGAG CTAGGACCAG GAGTGCTAGT GCCAC...TG TTTGTCTgcG
Orangutan         CCGTGCTGAG CTAGGACCAG GAGTGCTAGT GCCAC...TG TTTGTCTATG
Baboon            CtGTGCTGAG CTAGGACCAG GAGTGCTAGT GCCAC...TG TTTGTCTgca
Marmoset          aCGTGCTGAG CTAGGACCAG GAGTGCTAcT GCCACTTTTG TTTGTCTATG
Capuchin          CCGTGCTGAG CTAGGACCAG GAGTGCTATt GCCACTTTTG TTTGTCTATG

Consensus         --GTGCTGAG CTAGGACCAG GAGTGCTA-T GCCAC---TG TTT-TCT---
                        -102        -92        -82        -75        -65

Human             GAGAGGGAGG CCTCAGTGCT GAGGGCC[AAG CAAATATTTG TGGTTATGGA]
Chimpanzee        GAGAGGGAGG CCTCAGTGCT GAGGGCC[AAG CAAATATTTG TGGTTATGGA]
Gorilla           GAGAGGGAGG CCTCAGTGCT GAGGGCC[AAG CAAATATTTG TGGTTATGGA]
Java-maquaca      GAGAGGGAGG CCTCAGTGCT GAGGGCC[AAG CAAATATTTG TGGTTATGGA]
Rhesus monkey     GAGAGGGAGG CCTCAGTGCT GAGGGCC[AAG CAAATATTTG TGGTTATGGA]
Orangutan         GAGAGGGAGG CCTCAGTGCT GAGGGCC[AAG CAAATATTTG TGGTTATGGA]
Baboon            GAGAGGGAGG CCTCAGTGCT GAGGGCC[AAG CAAATATTTG TGGTTATGGA]
Marmoset          GAGAGGGAGG CCTCAGcGCT GAGGGCC[AgG CAAATATTTG TGGTTATGcA]
Capuchin          GAGAGGGAGG gCTCAGcGCT GAGGGCC[gAG CAAATATTTG TGGTTgTGGA]

Consensus         GAGAGGGAGG -CTCAG-GCT GAGGGCC[--G CAAATATTTG TGGTT-TG-A]
                         -55        -45        -35        -25        -15

Human             [TTAAC]TCGAA CTCC[AGGCTG TCATGGCGGC AG]GACGGC.G AACTTG
Chimpanzee        [TTAAC]TCGAA CTCC[AGGCTG TCATGGCGGC AG]GACGGC.G AACTTG
Gorilla           [TTAAC]TCGAA CTCC[AGGCTG TCATGGCGGC AG]GACGGC.G AACTTG
Java-maquaca      [cTAAC]TCGAA CTCC[AGGCTG TCATGGCGGC AG]GACGGC.G AACTTG
Rhesus monkey     [TTAAC]TCGAA CTCC[AGGCTG TCATGGCGGC AG]GACGGC.G AACTTG
Orangutan         [TTAAC]TCGAA CTCC[AGtCTG TCATGGCGGC AG]GACGGC.G AACTTG
Baboon            [TTAAC]TCaAA CT.C[AGGCTG TCATGGCGGC AG]GACGGC.G AACTTG
Marmoset          [TTAAC]TCcAc CTgC[AGcCTG TCATGGCGGC AG]GACGGCaG AACTTG
Capuchin          [TTAAC]TCGAA CTCC[A.cCTG TCATGGCGGC AG]GACGGC.G AAtTTG

Consensus         [-TAAC]TC-A- CT-C[A--CTG TCATGGCGGC AG]GACGGC-G AA-TTG
                         -5         +6         +16        +25
```

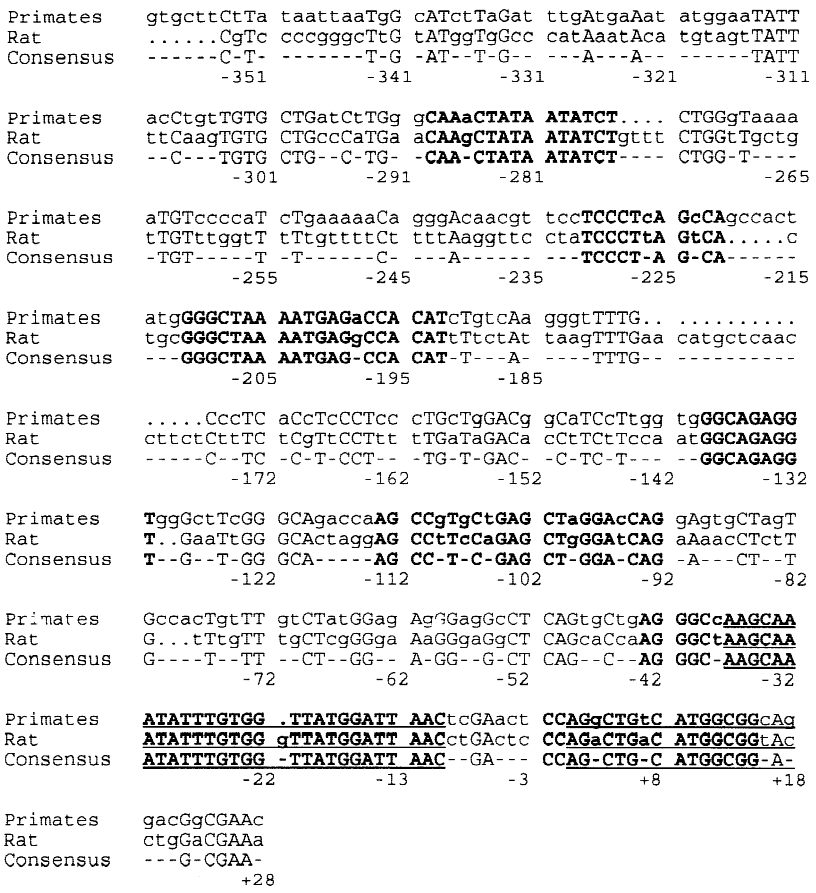

Fig. 2. Comparison of the protein C promoter region of the rat with the consensus sequence of the primate protein C promoter region. Identical nucleotides are indicated in *uppercase*, whereas nonidentical nucleotides are indicated in *lowercase.* The identical regions I to VIII are indicated in *boldface.* The known transcription factor binding sites (HNF-3 and HNF-1) and the PCE-1 region are *underlined.* The nucleotide numbering is relative to the transcription start site of the human protein C gene (Plutzky et al. 1986). The PileUp program (GCG Sequence Analysis Software Package Version 8.1) was used to align the sequences. The nucleotide sequence data of the rat will appear in the GenBank, EMBL, and DDBJ nucleotide sequence databases with the accession number U77653.

the nucleotides are the same in eight of the nine primates, whereas 95% of the nucleotides of the protein C promoter are identical in seven primates. The high degree of sequence similarity makes it difficult to point out important regulatory sequences. However, essential sequences for human protein C gene transcription, which have been reported previously, are identical above average (>83%). This indicates the functional relevance of evolutionary identical sequences.

In Fig. 3 a phylogenetic tree based on the protein C promoter region is shown. Despite the relatively small non-protein coding DNA fragment on which it is based, the tree completely matches the common classification of primates into four families, i.e., Hominidae (*Homo sapiens*, *Gorilla gorilla*, *Pan troglodytes*, and *Pongo pygmaeus*), Cercopithecinae (*Macaca mulatta*, *Papio ha-*

Table 2. Identical regions of the protein C promoter region between primates and rat

Region	Localization[a]	Identity	
		Number (*n*)	Percentage (%)
I	−289 to −275	14/15	93
II	−231 to −221	9/11	82
III	−211 to −192	19/20	95
IV	−139 to −131	9/9	100
V	−113 to −92	17/22	77
VI	−43 to −10	33/34	97
VII	−2 to +15	15/17	88

[a] Numbering refers to the human protein C sequence [all nucleotide numbering is relative to the transcription start site (Plutzky et al. 1986)].

Fig. 1. Comparison of the protein C gene promoter region in a number of primates. Identical nucleotides are indicated in *uppercase,* whereas nonidentical nucleotides are indicated in *lower case.* Represented by consensus are nucleotides that are invariant between all species. Indicated in *boldface* are the PCE-1 region (+1 to +18), the HNF-3 binding sites (−33 to −22 and −26 to −37), and the HNF-1 binding site (−22 to −10). The nucleotide numbering is relative to the transcription start site of the human protein C gene (Plutzky et al. 1986). The PileUp program (GCG Sequence Analysis Software Package Version 8.1) was used to align the sequences. The nucleotide sequence data will appear in the GenBank, EMBL, and DDBJ nucleotide sequence databases with the following accession numbers: U77648 (*Gorilla gorilla*), U77647 (*Pan troglodytes*), U77650 (*Pongo pygmaeus*), U77652 (*Cebus apella*), U77651 (*Macaca mulatta*), U77649 (*Callithrix jacchus*), U77646 (*Papio hamadryas*), and U77654 (*Macaca fascicularis*). The human protein C sequence is derived from GenBank accession number U47685, nucleotides 1117–1505.

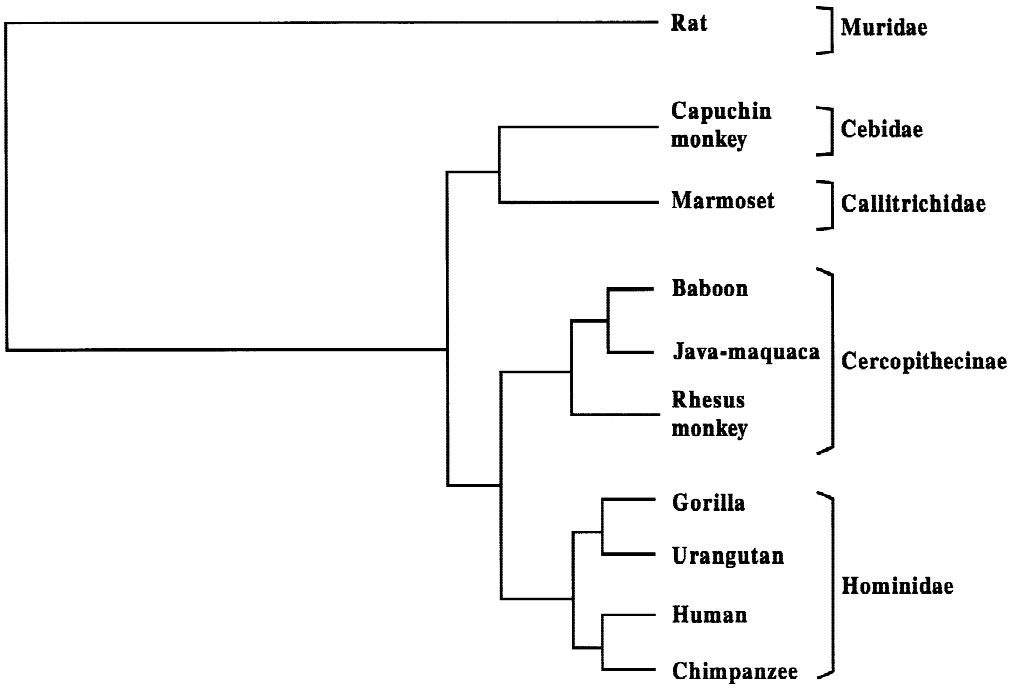

Fig. 3. Phylogenetic tree based on the protein C gene promoter region. The Clustal V program (*Higgins et al. 1991*) was used to create the phylogenetic tree.

madryas, and *Macaca fascicularis*), Cebidae (*Cebus apella*), and Callitrichidae (*Callithrix jacchus*).

Analysis of the rat protein C gene clearly illustrates a loss of similarity in the promoter region when compared with its primate homologue. However, more detailed inspection of the aligned sequences reveals seven identical, in both sequence and position, areas among primate and rat genes. One of these regions, located from position −44 to −10, overlaps with both the HNF-3 and the HNF-1 binding sites and is completely identical. Region VII (−2 to +15) contains most of the DNase protected PCE-1 region (+1 to +18). The fact that DNase protected regions tend to be larger than the corresponding transcription factor binding site may explain the discrepancy between the DNase protection assay and the evolutionary identity.

Recently, Maio et al. (1996) proposed a silencer element between −193 and −75. Regions IV and V are located in this area. However, comparison of these regions with a number of transcription factor consensus sequences does not give candidate transcription factors binding to these regions. Therefore, the involvement of these regions in downregulating protein C gene expression is uncertain. Whether these, and the other identical regions, are involved in transcriptional regulation of the protein C gene should be determined in functional assays.

The impossibility of PCR amplifying rat protein C promoter fragments on the basis of human sequence information is understandable considering the low overall sequence identity between primates and rat. Outside the seven identical regions, only 35% identity exists between the species, indicating that selection of PCR primers based solely on the human protein C sequence will be unsuccessful. As shown for the rat, the success rate of the amplifications can be increased with the use of one human-specific primer and one species-specific primer. This also explains why we were unable to amplify the promoter region of cat, dog, and mouse protein C to which no species-specific sequence information is available. The inability to PCR amplify the promoter region in cat, dog, and rat with primers overlapping regions I and VII does not imply that these two regions are not alike in these three species. It is well conceivable that the degenerative primers are not suited to amplify DNA from these species. This suggestion is strengthened by the fact that human DNA is only very poorly amplified by these primers (data not shown).

In conclusion, we have identified seven evolutionary identical regions in the protein C promoter region. The functional relevance of the majority of these regions in protein C gene expression remains to be determined.

Acknowledgment. This work was supported by Grant 92.004 from the Trombosestichting Nederland.

References

Allaart CF, Poort SR, Rosendaal FR, Reitsma PH, Bertina RM, Briët E (1993) Increased risk of venous thrombosis in carriers of hereditary protein C deficiency defect. Lancet 341:134–138

Berg L-P, Scopes DA, Alhaq A, Kakkar VV, Cooper DN (1994) Disruption of a binding site for hepatocyte nuclear factor 1 in the protein C gene promoter is associated with hereditary thrombophilia. Hum Mol Genet 3:2147–2152

Branson HE, Marble R, Katz J, Griffin JH (1983) Inherited protein C deficiency and coumarin-responsive chronic relapsing purpura fulminans in a newborn infant. Lancet 1165–1168

Broekmans AW, Veltkamp JJ, Bertina RM (1983) Congenital protein C deficiency and venous thromboembolism—A study of three dutch families. N Engl J Med 309:340–344

Dahlbäck B (1991) Protein S and C4b-binding protein: Components involved in the regulation of the protein C anticoagulant system. Thromb Haemost 66:49–61

Esmon CT (1992) The protein C anticoagulant pathway. Arterioscler Thromb 12:135–145

Faisst S, Meyer S (1992) Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res 20:3–26

Foster DC, Yoshitake S, Davie EW (1985) The nucleotide sequence of the gene for human protein C. Proc Natl Acad Sci USA 82:4673–4677

Griffin JH, Evatt B, Zimmerman TS, Kleiss AJ, Wideman C (1981) Deficiency of protein C in congenital thrombotic disease. J Clin Invest 68:1370–1373

Higgins DG, Bleasby AJ, Fuchs R (1991) CLUSTAL V: Improved software for multiple sequence alignment. CABIOS 8:189–191

Kisiel W (1979) Human plasma protein C: Isolation, characterization, and mechanism of activation by α-thrombin. J Clin Invest 64:761–769

Marlar RA, Montgomery RR, Broekmans AW (1989) Diagnosis and treatment of homozygous protein C deficiency. J Pedriatr 114:528–534

Miao CH, Ho W, Greenberg DL, Davie EW (1996) Transcriptional regulation of the gene coding for human protein C. J Biol Chem 271:9587–9594

Okafuji T, Maekawa K, Nawa K, Marumoto Y (1992) The cDNA cloning and mRNA expression of rattus norvegicus protein C. Biochim Biophys Acta Gene Struct Express 1131:329–332

Patracchini P, Aiello V, Palazzi P, Calzolari E, Bernardi F (1989) Sublocalization of the human protein C gene on chromosome 2q13-q14. Hum Genet 81:191–192

Plutzky J, Hoskins J, Long GL, Crabtree GR (1986) Evolution and organization of the human protein C gene. Proc Natl Acad Sci USA 83:546–550

Quandt K, Frech K, Karas H, Wingender E, Werner T (1995) MatInd and MatInspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res 23:4878–4884

Seligsohn U, Berger A, Abend M, Rubin L, Attias D, Zivilin A, Rapaport SI (1984) Homozygous protein C deficiency manifested by massive venous thrombosis in the newborn. N Engl J Med 310: 559–562

Spek CA, Greengard JS, Griffin JH, Bertina RM, Reitsma PH (1995) Two mutations in the promoter region of the human protein C gene both cause type I protein C deficiency by disruption of two HNF-3 binding sites. J Biol Chem 270:24216–24221

van Hinsberg VWM, Bertina RM, van Wijngaarden A, van Tilburg NH, Emeis JJ, Haverkate F (1985) Activated protein C decreases plasminogen activator inhibitor activity in endothelial cell conditioned media. Blood 65:1914–1920

Walker FJ, Fay PJ (1992) Regulation of blood coagulation by the protein C system. FASEB J 6:2561–2567