Molecular Cloning of a cDNA Encoding Canine Factor IX

By James P. Evans, Herbert H. Watze, Jerry L. Ware, Darrel W. Stafford, and Katherine A. High

Factor IX (F.IX) is a vitamin K-dependent plasma protein, a deficiency of which results in hemophilia B. A canine model of hemophilia B exists; attempts to use this model for gene transfer experiments or characterization of the hemophilic defect require elucidation of normal canine F.IX structure. We report the isolation and characterization of the coding region for canine F.IX cDNA. Canine F.IX possesses 86% identity at the amino-acid level with human F.IX. The leader peptide, Gla domain, EGF domains, and the carboxy-terminal portion of the heavy chains show extensive sequence conservation between the canine and human. All Glu residues undergoing γ-carboxylation in humans are conserved in canines. The complete coding sequence for canine F.IX has been determined, and the derived translation product has been analyzed. A similar approach should allow identification of the causative mutation in canine hemophilia B. Furthermore, this clone may prove a valuable resource in gene transfer experiments for this disease. © 1989 by Grune & Stratton, Inc.

MATERIALS AND METHODS

Canines. A canine liver was obtained from the Division of Laboratory Animal Medicine of the University of North Carolina. The animal, which was killed in another experiment, was from a normal colony and had no known defects of coagulation.

Isolation of hepatic RNA; Northern analysis. Total RNA was isolated from frozen (−70°C) dog liver using the guanidinium thiocyanate method of Chirgwin et al.2 RNA was passed over an oligo dT-cellulose column, and the poly(A)+ fraction recovered. Poly(A)+ RNA was denatured by heating at 65°C for ten minutes in 50% (vol/vol) formamide and was fractionated by electrophoresis through a 1% agarose formaldehyde gel. The RNA was transferred to nitrocellulose (Schleicher and Schuell), prehybridized, and hybridized as described by Thomas.3 Blots were exposed to x-ray film (Kodak XAR-5) at −70°C with a Cronex (Dupont, Wilmington, DE) intensifying screen. 32P dCTP was incorporated into a human F.IX cDNA probe4 with a nick-translation kit.

Construction of canine hepatic cDNA library. Poly (A)+ RNA was used as a template for synthesis of double-stranded cDNA with oligo(dT) as a primer. First-strand synthesis was performed essentially as outlined by Gubler and Hoffman,4 with several modifications. Ten micrograms poly(A)+ RNA as suspended in 43 μL H2O, and 6.7 μg oligo(dT) 12-18 (Pharmacia) was incorporated into a human F.IX cDNA probe with a nick-translation kit.

From the Departments of Medicine and Pathology, The Center for Thrombosis and Hemostasis, UNC School of Medicine; and the Department of Biology, University of North Carolina at Chapel Hill.

Submitted July 11, 1988; accepted March 1, 1989.

Supported by a grant from the North Carolina Biotechnology Center. J.P.E. is the recipient of a Judith Graham Pool award from the National Hemophilia Foundation. K.A.H. is the recipient of Clinical Investigator Award No. K08-HL01922 from the National Heart, Lung and Blood Institute. H.H.W. is supported by the Max Kade Foundation, Incorporated.

Address reprint requests to Katherine A. High, MD, Department of Medicine, Division of Hematology, CB 7035, 416 Burnett-Womack Building, University of North Carolina, Chapel Hill, NC 27599.


The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

© 1989 by Grune & Stratton, Inc.
0006-4971/89/7401-0002$3.00/0

lowering of the ratio of reverse transcriptase to RNA resulted in synthesis of longer reverse transcripts. Synthesized DNA was measured by trichloroacetic acid (TCA) precipitable incorporation, yielding a value of 0.3 μg, or 3% of the starting material.

For second-strand synthesis, the first-strand material was dissolved in 40 μL 10× second-strand buffer (200 mmol/L Tris pH 7.4, 50 mmol/L MgCl₂, 100 mmol/L (NH₄)₂SO₄, 1 mol/L KCl). The following reagents were added: 20 μL 1 mg/mL BSA, 3.2 μL 5 mmol/L dNTPs, 319 μL H₂O, 3 μCi ³²P dCTP, 4.3 U RNase H (Pharmacia), and 92 U Escherichia coli Pol I (New England Biolabs, Beverly, MA) for a final volume of 400 μL. The reaction was incubated at 15°C for 60 minutes, at room temperature for 60 minutes, and at 70°C for ten minutes. The mixture was made 0.1 mmol/L in dNTPs by adding 4.8 μL 5 mmol/L dNTPs; 20 U T4 DNA polymerase was then added. The reaction was performed at 37°C for ten minutes and extracted with phenol:chloroform isoamyl alcohol. cDNA was precipitated in ethanol at −20°C overnight. Incorporation was again measured by TCA precipitation, yielding 0.3 μg synthesized second-strand material.

The cDNA was methylated by dissolving the double-stranded material in 100 μL 100 mmol/L Tris pH 8, 10 mmol/L EDTA, S-Adenosyl methionine was added to a final concentration of 80 μmol/L, and 80 U EcoRI methylase (New England Biolabs) was added. The reaction was incubated for one hour at 37°C, extracted, and precipitated as above.

Ligation of cDNA to 4 μg EcoRI linkers (pCCGAATTCCG, Pharmacia) was performed in 50 mmol/L Tris pH 7.5, 10 mmol/L MgCl₂, 5 mmol/L DTT, 1 mmol/L ATP; 1,200 U T4 DNA ligase was added, and the reaction was incubated overnight at 15°C. The resultant reaction was digested with 20 U EcoRI, and cDNA was separated from free linkers on an 8-mL Sepharose CL 4B column. Two percent of the resultant cDNA was ligated to dephosphorylated λ gt 11 arms (Promega Biotech) and packaged (Gigapack Gold, Stratagene, La Jolla, CA).

Screening of recombinant phage; DNA sequencing. The recombinant library was screened by standard techniques with human F.IX cDNA as a probe; 300,000 phage were screened, yielding 18 F.IX cDNA recombinant phage, nine of which were plaque-purified. All nine inserts were subcloned into M13, and the longest of these was sequenced in its entirety in both orientations by the Sanger dideoxy method. To complete sequencing of the second strand in regions not easily accessed by restriction sites, four 17-base oligonucleotides were synthesized to use as primers in M13 sequencing. The sequences for the oligonucleotides were derived from the (already sequenced) opposite strand. The locations at which synthetic oligonucleotides were used as sequencing primers are shown in Fig 1.

The 5’ and 3’ termini of the remaining eight clones were mapped and are shown in Fig 2. Sequence data were analyzed with the Beckman Microgenie programs.

RESULTS

Characterization of canine F.IX mRNA. To determine the size of the F.IX transcript in canine liver, we performed
Northern analysis. The results (Fig 2) show that canine F.IX is indeed synthesized in liver and that the transcript is ~3 kb, the same size as human F.IX mRNA.

Characterization of cDNA clone for canine F.IX. A restriction map of the reverse transcripts and the strategy used for sequence determination are shown in Fig 1. The nucleotide sequence and derived amino acid sequence obtained from the longest clone are shown in Fig 3.

The longest cDNA isolated was 2,845 nucleotides, agreeing well with the size of the F.IX mRNA transcript detected by Northern analysis (Fig 2). A single long open-reading frame of 1,356 nucleotides exists, predicting a protein of 452 amino acids. The processed secreted protein consists of 413 amino acids with a mol wt of 46,609. The open reading frame is flanked by a 5'-untranslated region consisting of 224 nucleotides, and a 3'-untranslated region of 1,303 nucleotides. A polyadenylation signal is present ten nucleotides upstream of a long poly(A) tail.

Structure of canine F.IX. Comparison of the nucleotide and amino-acid sequences of canine, human, and bovine F.IX allows identification of conserved regions of the molecule. Interspecies identity varies widely among the domains.

Fig 3. Nucleic acid sequence of canine F.IX cDNA. Encoded amino acids are shown below the DNA sequence. The first methionine and the polyadenylation signal are underlined. The leader peptide, Gla domain, EGF-like domains, and activation peptide are shown.
of the molecule. As shown in Fig 4, the amino-acid sequence of the Gla domain is extensively conserved among the three species; 95% of the residues (43 of 45) are identical. All glutamic acid residues which undergo \( \gamma \)-carboxylation are conserved. The EGF-like domains, whose function is unknown and whose boundaries are only approximate, also appear to be highly conserved, with 83% of the residues (70 of 84) identical in the three species. Not unexpectedly, identity among the amino-acid residues of the activation peptide is very low. Although the cleavage sites, Arg-Ala and Arg-Val, are identical among the three, only 12 of the intervening 33 residues are the same. The canine activation peptide is shorter than human or bovine by three residues. Within the activation peptide, the canine sequence predicts a threonine at residue 149. This corresponds to the polymorphic site (threonine \( v \) alanine) at residue 148 in humans.\(^6\)

Overall conservation of the heavy chain is high, with 83% of residues conserved among the three species. Conservation at the carboxy-terminal portion of the heavy chain is remarkable, with 57 of the last 59 residues identical in all three species. Throughout the coding region, all of the cysteine residues are conserved.

An extensive 5'-untranslated region was present on the longest of the reverse transcripts isolated, which possesses 83% identity to the human. This region is clearly untranslated, as two stop codons interrupt the reading frame which
codes for the canine F.IX protein. Only one possible methionine translation initiation codon exists and is shown in Fig 3.

**DISCUSSION**

The site at which translation is initiated in human F.IX is uncertain. This is due to the presence of three possible methionine codons located 46, 41, and 39 amino acids upstream of the amino terminal tyrosine of the mature, secreted protein. Which of these three possible methionines is the actual start site of translation in the human has not been determined. Only one of these potential initiation sites is conserved in the canine, suggesting that this position (the methionine at position 39) may serve as the start site for translation in humans as well. This conjecture is further supported by the presence of a Kozak consensus sequence surrounding this methionine in humans (6 of 9 nucleotides match) as well as canines (6 of 9 nucleotides match). Such a consensus sequence is not found surrounding the other two potential methionine codons in humans, as noted by Yoshitake et al.

Assuming that translation in humans is indeed initiated from the methionine at position 39, comparison of the predicted pre-pro leader sequences of canines and humans shows a high degree of identity; 91% of nucleotides and 85% of amino acids are the same.

Both human and canine F.IX possess extensive 3'-untranslated regions. Identity between these regions is considerably less than in the coding regions. There is 72% identity through the first 800 nucleotides of the 3'-untranslated region, after which identity disappears altogether (through the last 500 nucleotides). Identity between these regions is considerably less than in the coding regions. There is 72% identity through the first 800 nucleotides of the 3'-untranslated region, after which identity disappears altogether (through the last 500 nucleotides). In human F.IX, a region less than in the coding regions. There is 72% identity through the first 800 nucleotides of the 3'-untranslated region, after which identity disappears altogether (through the last 500 nucleotides).

The canine FIX cDNA provides considerable information regarding conserved regions of the molecule. In addition, since the hemophilic canines produce a FIX transcript, determination of the mutation responsible for their disease should be possible by comparison of a hemophilic F.IX cDNA and this normal canine cDNA. Finally, the isolated clone should prove a valuable resource in trials of gene transfer therapy.

**ACKNOWLEDGMENT**

We acknowledge Sherry Stanfield-Oakley for technical assistance and Dianne Partin for manuscript preparation. We also thank Dr David Joseph for valuable discussions and suggestions.

---

**Fig 4.** Canine, human, and bovine F.IX amino-acid sequences; major domains of the molecules are indicated. *Positions at which an amino acid is not conserved between species. The sequence of the leader peptide of bovine factor IX has not been determined.
REFERENCES

5. Thomas PS: Hybridization of denatured RNA and a small DNA fragments transferred to nitrocellulose. Proc Natl Acad Sci USA 77:5201, 1980
Molecular cloning of a cDNA encoding canine factor IX
JP Evans, HH Watzke, JL Ware, DW Stafford and KA High