Molecular Cloning of a cDNA Encoding Canine Factor IX

By James P. Evans, Herbert H. Watzke, 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 terminal portion of the heavy chains show extensive deficiency states for which cloned genes are available, gene transfer experiments. First, in contrast to many other expression vector into the somatic cells of affected individuals, allowing continuous endogenous production of F.IX. Hemophilia offers at least three advantages as a model for gene transfer experiments. First, in contrast to many other deficiency states for which cloned genes are available, tissue-specific expression (in this case, the liver) is not required. Rather, the sole requirement is that the tissue into which the gene is inserted be capable of secreting the gene product. Second, precise regulation of levels of expression, which is required for substances such as growth hormone or insulin, is not necessary. An increase in F.IX activity levels from <1% to 2% or 3% can produce a profound improvement in the course of the disease, and levels as high as 150% to 200% are not known to cause any ill effects. Third, an animal model exists for hemophilia B; canine hemophilia B is characterized by F.IX activity levels of <1%.2 The canine disease is X-linked and is clinically very similar to the human disease.

To use the canine model as a prototype for gene transfer therapy, one must have access to a full-length cDNA for normal canine F.IX. The characterization of canine F.IX and its comparison with the known sequence for human and bovine F.IX should also provide insight into the structure and function of this critical clotting protein. We describe the isolation and characterization of a cDNA encoding canine F.IX.

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.3 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.4 The RNA was transferred to nitrocellulose (Schleicher and Schuell), prehybridized, and hybridized as described by Thomas.5 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 probe6 with a nick-translation kit.γ

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,6 with several modifications. Ten micrograms poly(A)+ RNA as suspended in 43 μL H2O, and 6.7μg oligo(dT) 12-18 (Pharmacia, Piscataway, NJ) was added. The solution was heated at 95°C for 30 seconds, cooled on ice, and added to a tube containing 10 μg actinomycin D (Pharmacia), 10 μCi 32P α-dCTP (ICN), 90 U RNAse inhibitor (Promega Biotech, Madison, WI), 10μg bovine serum albumin (BSA, Sigma Chemical, St Louis, mol biol grade), 2 mmol/L dithiothreitol (DTT, Sigma), 500 nmol/L dNTPs, 4 mmol/L sodium pyrophosphate, 50 mmol/L KCl, 8 mmol/L MgCl2, and 50 mmol/L Tris pH 8.3 for a final volume of 100 μL. One hundred units of avian myeloblastosis virus reverse transcriptase (Life Sciences) was added, and the reaction was incubated at 42°C for 75 minutes. The mixture was extracted with phenol/chloroform and chloroform, followed by precipitation in ethanol at −20°C overnight. The modifications of the Gubler and Hoffman procedure, including extension of incubation time and
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 37°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

Fig 1. Northern blot analysis of 10 μg total RNA from canine hepatocytes hybridized to human F.IX cDNA. Positions of the major ribosomal RNA species are indicated for size reference.

Fig 2. Restriction map of canine F.IX cDNA showing the sequencing strategy used. Restriction sites used for subcloning and sequencing are indicated. Sites at which oligonucleotides were used as primers in dideoxy sequencing reactions (described in text). Nine overlapping clones are shown. Extent of the clone was determined precisely by sequencing (solid bar); extent of the clone was mapped approximately by restriction analysis (open bar).
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 3'-untranslated region consisting of 224 nucleotides, and a 3'-untranslated region of 1,303 nucleotides, with a polyadenylation signal present ten nucleotides upstream of a long poly(A) tail.

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.

A

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 3'-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

**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.**
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 γ-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 most 3' methionine codon present, 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). F.IX of humans, as noted by Yoshitake et al., possesses 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). F.IX of humans, as noted by Yoshitake et al., possesses less than 50% identity with the coding region.

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 FIX 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.
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