Hemophilia B (Factor IX<sub>Seattle</sub>) Due to a Single Nucleotide Deletion in the Gene for Factor IX

Barbara G. Schach, Shinji Yoshitake, and Earl W. Davie
Department of Biochemistry, University of Washington, Seattle, Washington 98195

Abstract

To understand the molecular basis for hemophilia B in patients with little or no circulating Factor IX antigen, a patient who had <0.2% circulating Factor IX antigen (Factor IX<sub>Seattle</sub>) was selected for analysis of his Factor IX gene. Genomic DNA fragments from the abnormal gene were cloned into bacteriophage λ vectors and recombinant phage were identified using radiolabeled genomic probes obtained from the normal Factor IX gene. The exons and flanking regions of the abnormal gene were sequenced by the dideoxy chain-termination method and this sequence was compared with that of the normal gene. Only one significant difference was observed, the deletion of a single adenine nucleotide in exon V. This resulted in a frameshift that converted an aspartic acid at position 85 in the protein to a valine and the formation of a stop signal at position 86. These data indicate that the gene for Factor IX<sub>Seattle</sub> codes for an 85 residue polypeptide that terminates after the first epidermal growth factor domain. Thus, the putative Factor IX<sub>Seattle</sub> polypeptide lacks the second epidermal growth factor domain, the activation peptide, and the catalytic domain present in the normal protein. This provides an explanation for the coagulation disorder in this patient and represents the first report of a single nucleotide deletion and frameshift resulting in hemophilia B.

Introduction

Hemophilia B (Christmas disease) is a bleeding disorder that results from a deficiency of Factor IX coagulant activity. Factor IX is one of several vitamin K-dependent proteins that are converted to serine proteases during coagulation process (1). Patients with hemophilia B can be grouped into three general categories based upon the amount of immunologically reactive Factor IX (cross-reactive material [CRM]) circulating in their blood (2). Individuals with normal levels of Factor IX antigen, but reduced Factor IX activity are designated CRM<sup>+</sup>, whereas those with decreased levels of antigen and a corresponding decrease in activity are designated CRM<sup>-</sup>. Individuals with little or no Factor IX antigen circulating in their blood are designated CRM<sup>-</sup>-.

Recently, the molecular basis for hemophilia B has been determined for several CRM<sup>+</sup> patients. As anticipated, an amino acid substitution in a critical region of Factor IX, such as the activation site (3) in one CRM<sup>+</sup> pedigree, is sufficient for decreased Factor IX activity. Three unrelated pedigrees have amino acid substitutions in the leader sequence (4–6). Two of the reports suggest a role for the leader sequence in directing the γ-carboxylation of the amino-terminal glutamic acid residues (4, 5). An amino acid substitution near the first epidermal growth factor domain also appears to be responsible for diminished Factor IX activity in another CRM<sup>+</sup> patient (7). It is not readily evident, however, how this amino acid substitution disrupts the Factor IX activity.

The molecular basis for hemophilia B has also been determined for some CRM<sup>-</sup> patients. Defects thus far reported range from single nucleotide substitutions in splice junctions (8, 9) to deletions of large segments of the gene (10–12). An association between large deletions of the gene and production of antibodies directed against Factor IX has been noted by some investigators (10, 11), but not by others (12).

Cloning of the complementary DNA (cDNA) (13–15) and gene (16–18) for human Factor IX has enabled investigators to analyze the gene from patients with hemophilia B. This approach has complemented classic methods of protein analysis in CRM<sup>-</sup> patients and is the only approach currently available to study CRM<sup>-</sup> patients.

To better understand the molecular basis for hemophilia B in patients with little or no circulating Factor IX antigen, the gene coding for Factor IX was selected for study from a patient with virtually no detectable Factor IX protein. It was anticipated that a small genetic defect similar to those previously reported for patients with thalassemia (19) might account for the absence of circulating Factor IX. Therefore, fragments of the gene coding for the deficient protein, designated Factor IX<sub>Seattle</sub>, were cloned and the coding regions, splice junctions, proposed promoter region, and the 3' processing region were sequenced. This led to the identification of a single adenine nucleotide deletion at nucleotide number 17,699 in exon V, which accounts for the molecular basis of hemophilia B in this patient.

Methods

Materials. The DNA restriction endonuclease EcoRI, T4 DNA ligase, and λgtWES-EcoRI arms were purchased from Bethesda Research Laboratories, Gaithersburg, MD. M13mp18 and M13mp19 sequencing vectors and sequencing kits were purchased from New England Biolabs, Beverly, MA. [γ<sup>32</sup>P]dATP and λ DNA in vitro packaging kits were purchased from Amersham Corp., Arlington Heights, IL. Proto-
clone GT System αgt11 DNA and λEML3 DNA were products of Promega Biotec, Madison, WI. [32P]-labeled deoxynucleotides were purchased from New England Nuclear, Boston, MA, and nitrocellu-
lose, from Schleicher & Schuell, Inc., Keene, NH. X-ray film (XRP1 and XAR) was obtained from Eastman Kodak, Co., Rochester, NY, and synthetic DNA oligonucleotide primers specific for Factor IX were prepared by Dr. Foon Lee at the Howard Hughes Medical Institute Synthesis Facility at the University of Washington, Seattle, WA. DNA sequence analysis was performed by using the program Genepro (Riv-
erside Scientific, Seattle, WA).

DNA preparation and analysis. Genomic DNA was prepared from the peripheral lymphocytes obtained from a patient with little or no circulating Factor IX (Factor IXsanto) and from two normal controls, according to established methods (20). 10 μg of DNA from the patient and each normal control was digested with the restriction endonucle-
ase EcoRI. The restriction fragments were separated by size on a 0.7% agarose gel in 90 mM Tris buffer, pH 8.3, containing 90 mM borate acid and 2 mM EDTA and transferred to nitrocellulose according to the method of Southern (21). The blot was probed with a nick-translated cDNA coding for Factor IX, labeled according to the method of Rigby et al. (22). Prehybridization was carried out in 50 mM Tris buffer, pH 7.0, containing 0.3 M NaCl and 100 μg/ml salmon testes DNA (TS buffer) and 50% deionized formamide at 42°C for 5 h. Hybridization was performed in a fresh solution of TS buffer with 50% deionized formamide and 1 x 106 cpm/ml radiolabeled Factor IX cDNA. After ~24 h, the blot was washed at a final stringency of 0.1x standard saline citrate (SSC),2 and 0.5% sodium dodecyl sulfate (SDS) at room temperature for 15 min. Excess wash solution was blotted from the nitrocellulose, and the blot was then enclosed in Saran Wrap. The blot was exposed to XRP1 x-ray film with an intensifying screen at ~70°C for 4 d.

Construction of genomic libraries. Approximately 100 μg of DNA from the patient was digested to completion with the restriction endonuclease EcoRI. The restriction fragments were separated according to size on a 0.5% agarose gel in 90 mM Tris buffer, pH 8.3, containing 90 mM borate acid and 2 mM EDTA. Gel slices containing fragments in the range of 10-15, 6.7-9.4, 5.3-7, 3.9-5, and 1.4-2 kilobase (kb) were isolated from the gel and treated, as previously described by Yoshitake et al. (16).

A separate library was constructed for each size-fractionated frag-
ment. The mixture containing the 12.5 kb fragment coding for exon I was cloned into bacteriophage λEMBL3 (23), as described by Yoshita-
ke et al. (16). The mixture containing the 4.9 kb fragment coding for exons II, III, and IV and the 4.8-kb fragment coding for exon V was cloned into bacteriophage αgt11 (24). Likewise, the mixture containing the 1.8-kb fragment coding for the 3' end of exon VIII was cloned into αgt11. The mixtures containing the 6.8-kb fragment coding for exon VI and the 5.5-kb fragment coding for exons VII and VIII were cloned into λgtWES (25). A typical ligation reaction was performed overnight at 15°C in 5 μl of ligase buffer (50 mM Tris buffer, pH 7.4, 10 mM MgCl2, 10 mM dithiothreitol, and 1 mM ATP, 50 μg/ml bovine serum albumin [BSA]) with 0.5-1 Weiss units of T4 DNA ligase. The reaction mixture also contained a total of ~0.5 μg of vector plus genomic DNA. The molar ratios of genomic to vector DNA varied between 1:1 for the 6.8-kb insert to 6:1 for the 1.8-kb fragment. Reaction condi-
tions were optimized for each insert size and vector combination using theoretical considerations (26) and empirical adjustments. Ligation were packaged in vitro according to the manufacturer's instructions.

Library screening. The λEML3 library was grown in Escherichia coli strain Q359, and the αgt11 and λgtWES libraries, in strain Y1090. Nick-translated genomic probes prepared from the normal Factor IX gene were used to screen the libraries. The plaque hybridization method of Benton and Davis (27) was used with the following modifications. Prehybridization was performed in 2X SSC, 2X Denhardt's (0.04% each of polyvinylpolyglycolide 3600, ficoll 600, and BSA), 0.5% SDS, 1 mM EDTA, pH 7.5, at 60°C until the nitrocellulose filters were thoroughly wet (from 15 min to 2 h). Hybridization was carried out in fresh solution containing 2 x 105 cpm/ml probe for ~16-24 h. The filters were rinsed in 2X SSC, 0.5% SDS, pH 7.5, at room temperature, followed by a wash in fresh solution at 60°C for 10 min. The filters were dried, covered with Saran Wrap, and exposed to Kodak XAR film in the presence of an intensifying screen at ~70°C for 5-20 h. Positive clones were plaque-purified and then amplified by the liquid culture lysis method of Silhavy et al. (28).

DNA sequencing. Restriction fragments containing exons with some flanking intron sequences were subcloned into M13 phage vec-
tors (29, 30). Sequencing was carried out by the dideoxy chain-termina-
mation method of Sanger et al. (31), as modified for [35S]dATP (32). Oligonucleotide primers specific for Factor IX were designed and used in sequencing several of the exons.

Results and Discussion

A Southern blot of an EcoRI restriction digest of genomic DNA from a patient with severe Factor IX deficiency (Factor IXsanto) was probed with a radiolabeled cDNA for Factor IX. Four bands of ~12.5, 6.8, 5.5, and 4.8 kb were identified by Southern blotting and were indistinguishable from bands present in adjacent lanes from the normal controls. The origin of these bands is illustrated by the EcoRI restriction map shown in Fig. 1. Note that the 4.8-kb fragment coding for exon V cannot be distinguished from the 4.9-kb band coding for exons II, III, and IV by Southern blot analysis of a 0.7% agarose gel. These preliminary experiments indicated that essentially all of the coding regions for the normal Factor IX gene were also present in the gene from the patient with Factor IX deficiency.

A number of λ phage libraries were then constructed from the genomic DNA from the patient, as described under Methods. The number of phage screened for all library constructions is shown in Table I. Between 4 x 105 and 2.5 x 106 phage were screened for each Factor IX genomic fragment. From one to six positive clones were identified for each fragment and plaque purified, except for the library containing exon VII and the majority of exon VIII. A total of 4,162 bases of Factor IXsanto 2 gene were sequenced employing the genomic inserts in the phage. This included 857 bases of DNA sequence from the exons and several hundred bases of flanking sequence for each exon.

When comparing the DNA sequence of the gene for Factor IXsanto 2 with the normal Factor IX gene (16), five differences were noted. The only significant difference, however, involved the deletion of a single adenine nucleotide at position 17,699 in the gene for Factor IXsanto 2 (Fig. 2). This deletion corresponds to the first base in exon V where the adenine nucleotide is part of the triplet GAT coding for aspartic acid at position 85.

![Figure 1. EcoRI restriction enzyme map of the gene for human Fact-

or IX. Exons are represented by the open boxes (not to scale) and introns by the horizontal line. The exon numbers are indicated by the roman numerals. EcoRI restriction sites are indicated by the vertical lines. Bars below the gene indicate the restriction fragments that contain exons. Their approximate sizes are shown in kilobases.](image-url)
in the normal protein. In the gene for Factor IXSeattle 2, the deletion results in a frameshift such that a GTT codes for a valine at position 85 and the next triplet becomes TAA, which is a stop codon (Fig. 3). The deletion was verified by sequencing two independently isolated clones for exon V from a non-amplified library. Thus, it is highly unlikely that the deletion was the result of a cloning artifact in the same position in these two independent clones. Also, the deletion was demonstrated by sequencing both DNA strands in the region of the defect. The deletion of this nucleotide cannot be demonstrated by DNA restriction endonuclease digestion since it does not create or alter any known restriction site.

This single-base deletion is the first reported frameshift mutation responsible for Factor IX deficiency in a patient with hemophilia B. Similar frameshift mutations have previously been reported for β globin deficiencies in cases of β thalassemia (33–36). Probable frameshift mutations with less damaging consequences have also been reported for other proteins, such as albumin (37).

The mechanism responsible for the nucleotide deletion in this case is not known. The deleted adenine nucleotide occurs in the first base of the seven-base tandem repeat of ATGTAAC. There is no similarity of this sequence pattern with patterns previously associated with deletions in the DNA slippage hypothesis (38, 39) or other hypotheses (40, 41). Instead the deletion may be explained by a mechanism like dislocation mutagenesis, which is not dependent on a specific DNA sequence (42, 43). As more frameshift mutations are analyzed, it is anticipated that the mechanisms responsible for small deletions will be better understood.

The second difference between the normal and abnormal genes identified a potential FokI restriction fragment length polymorphism at position 6,843 in which the G in the normal gene has been replaced with an A in the gene for Factor IXSeattle 2. This eliminates the FokI restriction site that begins at nucleotide number 6,842. Individuals possessing this FokI site would have two fragments of ~270 and 1,690 base pairs after digestion with FokI, whereas individuals without the site would have a single fragment of 1,960 base pairs. Since FokI cuts the gene for Factor IX frequently and only 21 bases of the cDNA would hybridize in the area of interest, genomic probes specific for this region would be better suited than cDNA probes when screening for this polymorphism. Alternatively, synthetic oligonucleotide probes designed to match the polymorphic site could be used for screening. The frequency of this polymorphism in the population is not currently known. Like other restriction fragment length polymorphisms, its usefulness may be limited because of linkage disequilibrium with other polymorphic sites already identified. This difference occurs within an intron about 1 kb downstream from the third exon and has no effect on the coding region of the gene.

The last three differences between the normal and abnormal genes were found to be errors in the sequence in the 5' flanking region of the normal gene, which was published earlier from our laboratory (16). These included the replacements of a T with an A at −2,176, −2,104, and −2,105. These errors occur over 2 kb upstream from the promoter region and have no effect on the amino acid sequence of the protein.

A comparison of the tentative structure of the normal circulating Factor IX protein with the structure predicted by the gene for Factor IXSeattle 2 is shown in Fig. 4. The Factor IXSeattle 2 protein contains the γ-carboxylglutamic acid domain, the first epidermal growth factor domain, and terminates with a valine
Figure 4. Comparison of the tentative protein structure for normal Factor IX with the structure predicted for Factor IX\textsubscript{Sente 2}. Normal Factor IX is on the left, and Factor IX\textsubscript{Sente 2} is on the right. The arrow in the normal Factor IX indicates the location of the amino acid substitution and termination point for Factor IX\textsubscript{Sente 2}. 
at position 85. Thus, the Factor IX<sub>Seattle</sub> protein is lacking the second epidermal growth factor domain, the activation peptide, and the catalytic domain.

It has not been established that this truncated protein circulates in the plasma of the Factor IX<sub>Seattle</sub> patient. The very low level of protein (<0.2%) detected in the plasma of the patient is at the lower limit of the sensitivity of the radioimmunoassay used to detect the Factor IX antigen (44, 45). Also, this patient is on home care therapy and the possibility of residual Factor IX from previous treatments cannot be excluded. Possible explanations for the virtually undetectable Factor IX<sub>Seattle</sub> antigen levels include an unstable messenger RNA (mRNA) that is rapidly turned over without being translated, or an unstable protein that is not secreted or that is rapidly degraded inside the cell or in circulation. Also, rapid clearance of the truncated protein via renal filtration is a likely possibility. A processed but unstable mRNA, secondary to a frameshift mutation, has been identified in the reticulocytes of a β thalassemia patient (33).

Several potential polyadenylation signals are located downstream of the newly generated stop codon in the gene for Factor IX<sub>Seattle</sub> at positions 18,585, 18,652, 18,717, and beyond. However, assessment of the Factor IX mRNA level in the Factor IX<sub>Seattle</sub> patient is not feasible since it would require a liver biopsy specimen. Analysis of the patient's urine for Factor IX antigen, however, is more feasible. Bray and Thompson (46) have already identified a hemophilia B pedigree in which renal clearance of a truncated Factor IX protein, Factor IX<sub>Seattle</sub>, resulted in very low levels of circulating Factor IX antigen.

Regardless of the mechanism behind the markedly reduced Factor IX antigen, the Factor IX<sub>Seattle</sub> patient has a severe deficiency of Factor IX activity. The deficiency would persist even if the truncated protein specified by the gene were circulating at normal levels because the gene for Factor IX<sub>Seattle</sub> codes for a nonfunctional protein. This protein lacks the catalytic domain and other domains that are probably essential for binding Factor IX to co-factors or substrates. The possibility that other point mutations may occur elsewhere in the gene cannot be ruled out since the whole gene was not sequenced. However, it is highly probable that the primary cause of hemophilia B in this patient is due to the deletion of a single nucleotide in exon V.

Acknowledgments

The authors thank Dr. Dominic Chung, Dr. Donald Foster, Dr. Kotoku Kurachi, Dr. Steven Leytus, and Dr. Arthur Thompson for helpful discussions. Also, we would like to express our gratitude to Mr. Stanley Yates for his assistance in this investigation.

This work was supported in part by research grant HL 16919 from the National Institutes of Health, and by a Judith Graham Postgraduate Research Fellowship from the National Hemophilia Foundation (Dr. Schach).

References


