ANALYSIS OF THE MOLECULAR DEFECTS CAUSING HAEMOPHILIA B IN SIX PATIENTS

by

Beatrice M. Tam

B. Sc (Hon.), University of British Columbia, 1990.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Masters of Science

in

THE FACULTY OF GRADUATE STUDIES

Department of Biochemistry

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Biochemistry

The University of British Columbia
Vancouver, Canada

Date July 21, 1992

ABSTRACT

The factor IX genes from six haemophilia B patients were analyzed in order to determine the molecular defect responsible for causing the disease in each of the cases.

Blood samples were obtained from the six patients and genomic DNA was extracted from the white blood cells. Regions of the factor IX gene were amplified from the genomic DNA via the polymerase chain reaction (PCR) for use in either single stranded conformational polymorphism (SSCP) analysis or for subcloning. The exons, including the intron/exon splice junctions, were selectively amplified. DNA representing 150 base pairs 3' of the first exon (containing the putative promoter region) was amplified together with exon 1. SSCP analysis was performed on the amplified exons to screen rapidly for the presence of base pair mutations as compared to wild type FIX sequence. The exact nature and locations of the mutations were then determined by DNA sequencing of the subcloned exons. The remaining exons of each FIX gene were also subcloned and sequenced and no other sequence discrepancies were found. A single base pair alteration was found in each FIX gene and was therefore assumed to cause the defect in factor IX and thus cause haemophilia B. Factor IX antigen levels in the patients' plasmas were determined using sandwich ELISA assays with polyclonal anti-factor IX antibodies. The coagulant activity of the mutant factor IX polypeptides were determined using the standard APTT (activated partial thromboplastin time) assays.

A C to A change at nt. position 17,700 was found in FIX Edmonton 1. The predicted amino acid sequence at residue 95 changed from cysteine to a stop codon. Likewise, in FIX Edmonton 2, a C to T change at nt. position 31,133 introduced a stop codon in place of arginine at residue 338. Both haemophilias are severe with less than 1% activity and antigen. Disruption of protein structure probably caused these two truncated polypeptides to be degraded within the hepatocytes. A G to C mutation at nt. position 17,756 in FIX Leamington resulted in the conversion of glycine$_{114}$ to alanine. Activity and antigen levels of FIX Leamington are both below 1%. Glycine$_{114}$ is situated in the second

EGF-like domain of factor IX. From sequence homology and models based on human EGF and the first EGF-like domains of FIX and FX, glycine$_{114}$ is conserved and likely occupies the third position in a type II β turn. Alanine at this position is thought to disrupt the turn and possibly affect disulfide bonding between cysteine$_{111}$ and cysteine$_{124}$. In FIX Creston, arginine$_{180}$ is mutated to proline due to a G to C alteration at nt. position 20,519. Although normal levels of FIX were present, only ~2% activity was exhibited. The mutation disrupted the arginine$_{180}$-valine$_{181}$ cleavage site required by FXIa and FVIIa to activate FIX. In FIX Edmonton 3, a G to A transition at nt. position 30,150 converted alanine$_{233}$ to threonine and resulted in a mutant enzyme with 15% activity. Computer models of the catalytic domain (based on crystal structures of the pancreatic serine proteases) show alanine$_{233}$ to be removed from the active site and substrate binding pocket. The threonine mutation may therefore disrupt FVIIIa binding rather than interfere with catalysis. Alternatively, according to computer models the leucine$_{379}$ to phenylalanine mutation (A to C at nt. position 31,258) is situated near the active site region in FIX Brantford. Phenylalanine, having a larger and more constrained side chain, likely disrupts the surrounding tightly packed residues thus affecting substrate binding and/or catalysis. FIX Brantford exhibited 5% activty.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Adenosine
Ag	Antigen
amp	Ampicillin
ATP	Adenosine triphosphate
bp	Base pair(s)
C	Cytidine
CRM	Cross reacting material
DMF	Dimethyl formamide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ELISA	Enzyme linked immunosorbent assay
G	Guanosine
Gla	γ-carboxyglutamic acid
Hya	β-hydroxyaspartate
IPTG	Isopropyl-β-D-Thiogalactopyranoside
kb	Kilobase pairs
LB	Luria broth
mRNA	Messenger ribonucleic acid
nt.	Nucleotide
OD	Optical density
PCR	Polymerase chain reaction

RNase	Ribonuclease
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
SSCP	Single stranded conformational polymorphism
T	Thymidine
TEMED	N, N, N', N'-tetramethylethylenediamine
U	Units
V	Volts
Xgal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

ACKNOWLEDGEMENTS

To all the members of the MacGillivray lab, those who were and those who are, I wish to extend my sincerest thanks for all the help, encouragement and advice given over the years. Honourable mentions to Billy Chow who couldn't put together a proper English sentence to save his life but was a damn fine scientist anyways, David Banfield whose jokes provided many of the lighter moments in the lab and whose advice (just write the thing, nobody will read it anyways) should have been more closely followed, Willie Pewarchuk the computer guru extraordinaire and Helene Cote my fellow woman scientist. The wonderful computer graphics were produced in the Brayer lab and photographed by Terry Lo; thanks a bunch. Thanks also to my friends who kept me sufficiently diverted from work to protect my sanity.

This thesis would not be complete without expressing my gratitude to Bruce Ritchie and Ross MacGillivray. Bruce, thank you for not only watching over my academic and scientific well-being but also my personal well-being. Ross, thank you for opening the door and giving me the opportunity. Your knowledge and appreciation of science is astounding but it is your overwhelming generosity and patience that I will remember most.

INTRODUCTION

I. BLOOD COAGULATION

A. General Overview

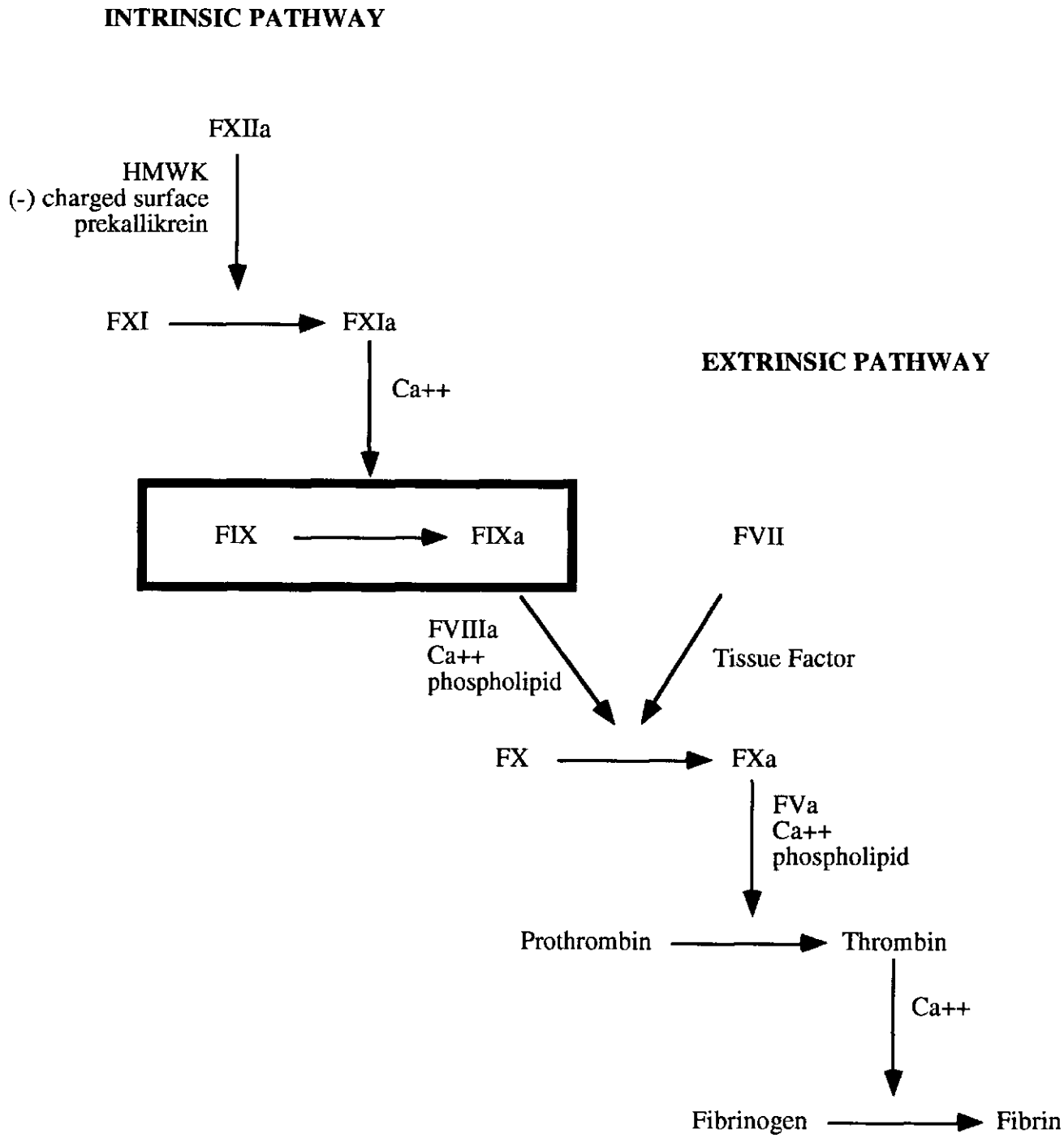
Over the course of evolution, vertebrates have developed closed circulatory systems for the transport of blood, thereby necessitating a system for repairing breaches to the system resulting from blood vessel injury. What has evolved is a complex series of enzymatic reactions leading to the formation of a blood clot. A clot is made up of activated platelets within a fibrin meshwork and acts essentially as a physical barrier which prevents blood loss. Although not yet fully understood, many aspects of the coagulation process have been elucidated, including the modes of action of most of the proteins involved. Together these proteins constitute a tightly regulated system of haemostasis commonly referred to as the blood clotting cascade (Jackson and Nemerson, 1980). Simplified, the cascade may be described as the orderly interaction of plasma serine proteases in which inactive zymogen precursors are sequentially activated by limited proteolysis (Davie et al., 1979). The result is amplification of the coagulation response at each step of the cascade. Traditionally the cascade has been divided into two pathways, extrinsic and intrinsic, depending on the event initiating coagulation (MacFarlane, 1964; Davie and Ratnoff, 1964). However, both pathways eventually converge on the activation of factor X. The physiological contribution of each pathway is not precisely known but both must interact in a coordinated fashion for proper haemostasis (Zur and Nemerson, 1981).

B. Blood Clotting Cascade

Although factor XII can initiate the intrinsic pathway (Ratnoff and Rosenblum, 1958), recent evidence suggests that the physiological role of FXII may not be in blood

Figure 1. Blood clotting cascade

The major reactions involved in the intrinsic and extrinsic pathways of the mammalian blood coagulation cascade are shown. Both pathways converge upon the activation of FX to FXa resulting in the formation of an insoluble fibrin clot. Factor IX plays an integral role in the intrinsic pathway as shown by the occurence of haemophilia B in the absence or deficiency of FIXa activity.


INTRINSIC PATHWAY
FXIIa
HMWK
(-) charged surface
prekallikrein
FXI
FXIa
Ca++
EXTRINSIC PATHWAY
FIX
FIXa
FVII
FVIIIa
Ca++
phospholipid
Tissue Factor
FX
FXa
FVa
Ca++
phospholipid
Prothrombin
Thrombin
Ca++
Fibrinogen
Fibrin

coagulation but perhaps rather in inflammatory response. In vitro, FXII activation can be initiated by concomitant binding with high molecular weight kininogen (HMWK) to a negatively charged surface. This complex is responsible for cleavage of FXI and prekallikrein to produce FXIa and kallikrein; FXIa and kallikrein reciprocate by further activating FXII. FXIa also catalyzes the activation of FIX to FIXa in the presence of calcium. FIXa activates FX by cleavage of a single peptide bond in the presence of FVIII, calcium and phospholipid (Fujikawa et al, 1975) (Figure 1).

The extrinsic pathway is initiated in the presence of tissue factor, a transmembrane lipoprotein expressed on endothelial cells when tissue is damaged (Colman et al., 1987). Tissue factor binds FVIIa and together they activate FX. The identity of the enzyme responsible for FVII activation in vivo is not known but FIXa (Masys et al., 1982), FXa (Radcliffe and Nemerson, 1976), thrombin (Broze and Majerus, 1980), FXIIa (Kisiel et al., 1977) and even FVIIa (Pedersen et al., 1989) itself are activators of FVII in vitro. In a reaction analogous to the reaction catalyzed by FIXa, FXa binds to its cofactor, FVa, on a membrane surface in the presence of calcium and phospholipid. This catalytic unit is known as the prothrombinase complex. Prothrombin is cleaved twice by FXa and thrombin is released into solution. In the final step, thrombin cleaves fibrinogen to fibrin. Fibrinogen is a dimer consisting of two identical halves; each half is made up of an $A\alpha$ chain, a $B\beta$ chain and a γ chain. Thrombin catalyzed cleavages release fibrinopeptides A and B leaving a fibrin monomer. These monomers polymerize spontaneously and are crosslinked through their γ chains by a transglutaminase, FXIIIa, to create a strong fibrin network (Doolittle, 1984).

It is important to note that the above scheme represents only the major reactions; many of the proteins involved are self regulatory or regulate coagulation at another point in the cascade.

C. Localization of Coagulation to Biological Membranes

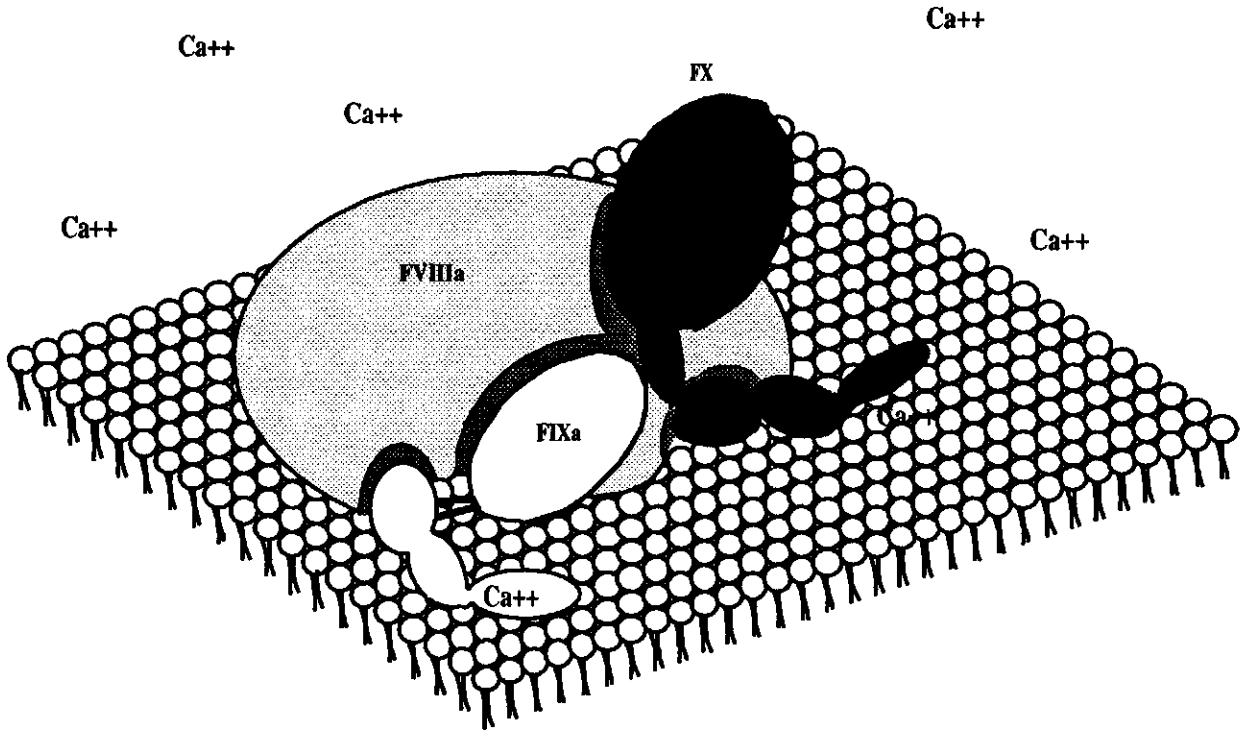
In solution, the blood clotting proteases exhibit little appreciable activity on a physiologically relevant time scale. The formation of coagulation complexes on biological membranes, however, greatly enhances the kinetics of coagulation (Mann et al., 1990). It has been postulated that anchoring of protein cofactors to a membrane surface allows binding and positioning of the coagulation protease and its substrate in the optimal orientation for catalysis (Figure 2). Alternatively, binding of the cofactor may cause conformational changes which favour interactions with either the protease or the substrate or both.

Under normal conditions, platelets do not adhere to vascular endothelial cells but when vessel walls are damaged, platelets attach to the damaged endothelium or subendothelium via adhesive proteins and aggregate to form platelet plugs (Weksler, 1987). Exposure of the aggregated platelets to subendothelial collagen activates the platelets leading to conformational changes which expose binding sites for the clotting factors on the platelet surface. This system of binding the clotting proteases and their cofactors to damaged endothelium localizes coagulation to sites of injury.

D. Regulation of the Cascade and Fibrinolysis

Once the clotting process has been initiated, mechanisms are required which will control and eventually halt the coagulation reaction. Regulation is maintained by proteins which either inactivate or inhibit the coagulation proteases. Protein S is a membrane binding protein which complexes with another plasma component, protein C. Bound protein C is activated by a thrombin-thrombomodulin (an integral membrane protein) complex on endothelial cells. Activated protein C abolishes both FVa and FVIIIa activity through limited proteolysis (Esmon, 1987). In this way, thrombin is self regulatory as it shuts off its own mechanism of activation. Another class of regulatory proteins are the serpins (serine protease inhibitors) such as the C1 inhibitor and α_1 antitrypsin which inhibit

Figure 2. The FIXa/FVIIIa/FX coagulation complex



In the presence of calcium, FIXa, FVIIIa and FX form a complex on biological membranes. Such a complex serves to enhance the kinetics of FX activation as well as localizing coagulation to the site of injury. The cofactor, FVIIIa, binds FIXa and FX in the optimal orientation for catalysis to occur.

FXIIa and FXIa respectively. The major serpin involved in regulating the clotting cascade is antithrombin III (ATIII) (Collen, 1981). ATIII counteracts the effects of thrombin by forming a 1:1 stoichiometric complex with thrombin, an association which is accelerated in the presence of heparin (an intestinal mucosal polysaccharide) (Jesty, 1978). Heparin also causes multiple rounds of complex formation of ATIII with factors IXa and Xa.

As well as down-regulating the coagulation response, the body must also destroy the fibrin clot (ie fibrinolysis) as wounds heal. The plasma protein, plasminogen, is converted to its active form, plasmin, by two endothelial proteins, urokinase and tissue plasminogen activator. In blood, plasmin is the principal enzyme responsible for the degradation of fibrin polymers (Colman, 1987). Together, the regulatory proteins and the fibrinolytic proteins prevent the over deposition of fibrin, a condition known as thrombosis.

II. DOMAIN ORGANIZATION OF THE BLOOD CLOTTING PROTEINS

A. Evolutionary Origin

The family of coagulation serine proteases has evolved from the simple digestive proteases such as trypsin and chymotrypsin. Both classes of enzymes function via the degradation of peptide bonds. However, the digestive proteins work on a wide range of polypeptides whereas the clotting proteases exert their activity on a limited number of substrates. Apart from the catalytic domain, the nascent trypsinogen and chymotrypsinogen polypeptides include only a signal peptide and a short activation peptide. These three structural elements remain common to all the serine proteases but during evolution, other large non-catalytic domains have been aquired between the signal peptide and the activation peptide (Jackson and Nemerson, 1980). Addition of these N terminal segments creates more complex levels of regulation and allow for greater substrate specificity. It is likely that the different domains existed previously as autonomous protein

units which, at various times in evolution, fused to the serine proteases (ie exon insertion) (Patthy, 1985). Many of the coagulation serine proteases share common structural and gene organizations. Once a domain was acquired by one of the clotting factors, exon shuffling could have occured by mispairing and double cross-over or by gene conversion with homologous sequences on other clotting proteins. Some of the proteases contain multiple copies of a single type of domain and this is thought to be due to tandem duplications (Doolittle, 1979). The family of blood clotting proteases represents a striking example of gene duplication of an ancestral gene with subsequent divergence of the sequence and function of the protein (Hewett-Emmett et al., 1981).

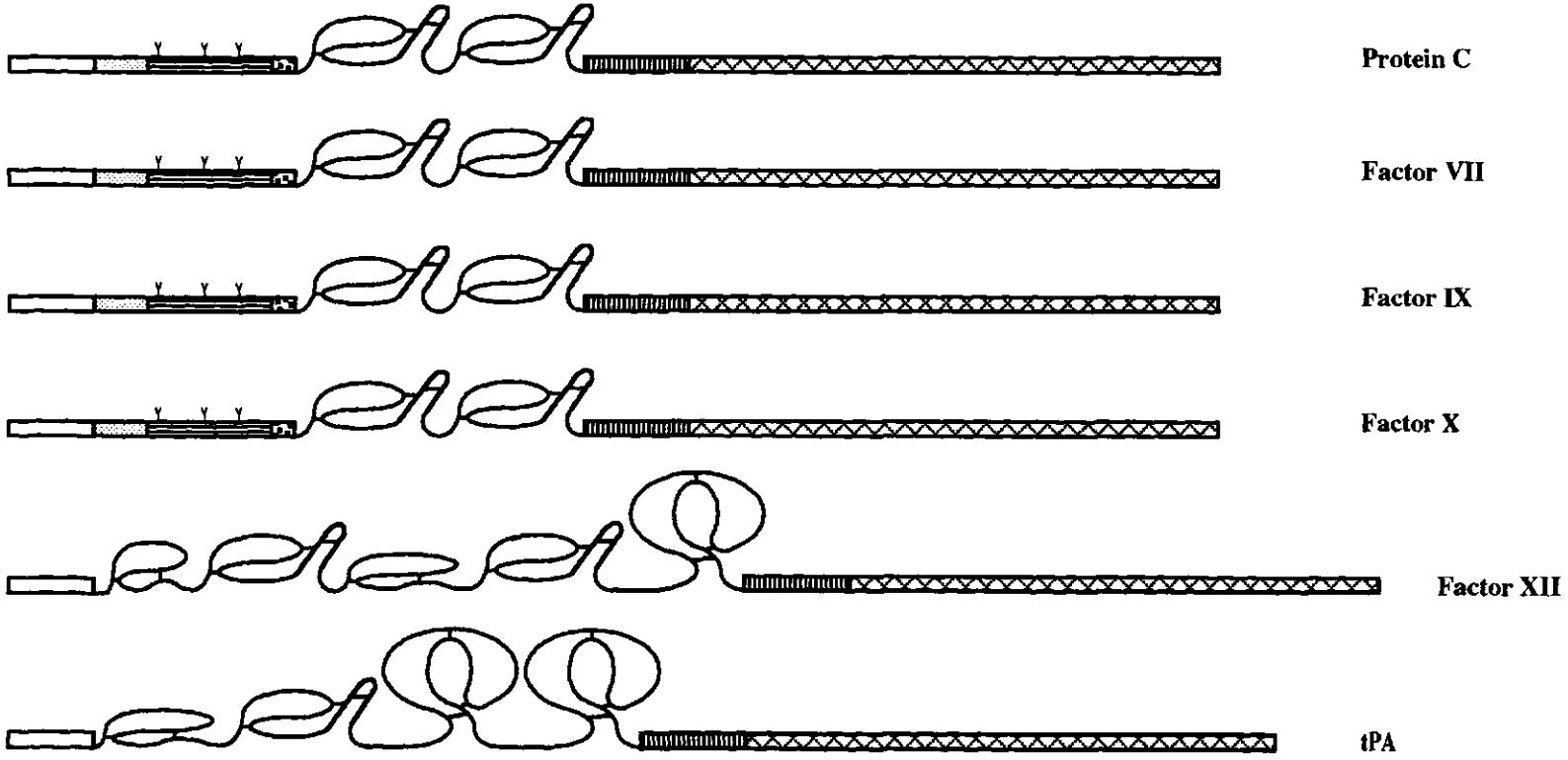
B. Common Domain Structures

Besides the signal peptide, activation peptide and catalytic domain, there are basically four other main types of domains shared by the blood clotting proteases: vitamin K-dependent calcium binding domains, epidermal growth factor-like domains, kringle domains and fibronectin type domains (Figure 3). The vitamin K-dependent calcium binding domains are situated at the N-termini of prothrombin, FIX, FVII, FX and protein C. This region is also known as the Gla domain due to the γ-carboxylation of the glutamic acid residues within the first forty two amino acids. Modification of these residues is a post-translational process catalyzed by a vitamin-K dependent carboxylase. The Gla residues create calcium binding sites which, upon binding calcium, cause a conformational change to occur within the proteins (Suttie, 1985). It is postulated that the conformational change exposes membrane binding sites on the proteins.

In Factors VII, IX, X and protein C, the Gla domain is followed by two epidermal growth factor (EGF)-like domains. FXII, urokinase and t-PA also contain EGF-like regions. As the name suggests, these domains share homology with one or more modules

Figure 3. Common structural domains found in haemostatic proteases

Similar structural domains are found in many of the coagulant and fibrinolytic proteins. These domains are identified in the key and described in the text. Cartoon representations of the protein structure are given for prothrombin (Degen et al., 1983), protein C (Foster and Davie, 1984; Beckman et al., 1985; Long et al., 1984), FVII (Hagen et al., 1986), FIX (Kurachi and Davie, 1982; Choo et al., 1982), FX (Leytus et al., 1984; Fung et al., 1985), FXII (Cool et al., 1985) and tissue-type plasminogen activator (Ny et al., 1984). From Furie and Furie, 1988.


Protein C
Factor VII
Factor IX
Factor X
Factor XII
tPA
Signal peptide
Gla domain
Kringle
EGF-like domain
Propeptide
Fibronectin type 1 domain
Fibronectin type 2 domain
Activation peptide
Catalytic domain
Aromatic stack

in the epidermal growth factor precursor (Doolittle et al., 1984). In the clotting factors, however, these domains do not exhibit any mitogenic properties as does EGF. The most striking characteristic of EGF and EGF-like domains is the conserved spacing of six cysteine residues which together form three disulfide bonds. Although functions have not yet been assigned for the EGF-like domains in each of the proteins, they are thought to be involved in interactions of the proteases with their cofactors or with cell surfaces.

Kringle modules are approximately 100 residues long and contain three conserved disulfide bonds with the first and sixth conserved cysteines forming the boundaries of the domain (Magnussen et al., 1975). The sulphur atoms of two of the disulfide bonds form a cluster in the centre of the kringle, giving this domain its characteristic structure (Park and Tulinsky, 1986). A single kringle is found in urokinase and FXII while multiple copies exist in prothrombin, tPA and plasminogen. The functions of kringle domains are still not defined but research is being directed towards their possible role in macromolecular assembly.

There are two types of domains each with homology to fibronectin (Peterson et al., 1983). Type I domains (also known as finger domains) are approximately 50 residues long containing two internal disulfide bonds. FXII contains a single finger domain (Cool et al., 1985). Type II domains are approximately 60 residues long and also contain two disulfides. In both types of domains, the disulfides are arranged so that the first and third and second and fourth cysteines form bonds. Type II domains are found in FXII and tPA (Cool et al, 1985; Pennica et al.).

III. FACTOR IX

A. Gene Structure

The gene for FIX has been localized to the distal end of the long arm of the X chromosome, specifically at band Xq27.1 (Camerino et al., 1984; Boyd et al., 1984). It is closely linked to the locus for the fragile X mental retardation syndrome (Xq27.3) which itself is flanked on the distal end by the FVIII gene locus (Purrello et al., 1985). In total, the FIX gene encompasses approximately 34 kb and contains 8 exons interrupted by 7 introns (Figure 4). The mRNA includes only 2.8 kb with about half of this being 3' untranslated sequence. The human FIX gene has been cloned (Choo et al, 1982; Anson et al., 1984)) and the entire nucleotide sequence obtained (Yoshitake et al., 1985). Each exon corresponds to a distinct structural domain of the protein.

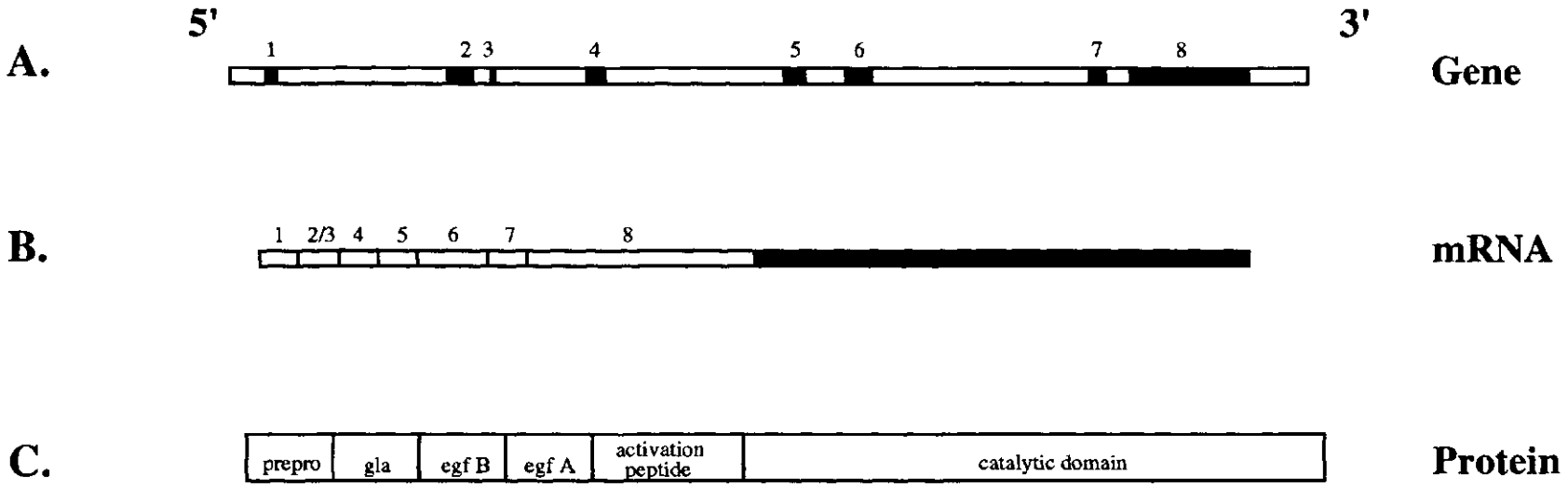
The FIX minimal promoter is thought to lie in the region -98 to +21 (Crossley and Brownlee, 1990). Through DNAse I footprinting assays, it has been shown that the sequence -99 to -77 is a potential nuclear factor-1 liver (NF1-L) binding site and that the sequence +1 to +18 is a potential CCAAT/enhancer binding protein (C/EBP) binding site. In the human FIX gene, the mRNA contains three in frame AUG codons in the 5' transcribed region which could potentially be the translation start codon. The third of these most closely matches the optimal consensus sequence for initiation codons as defined by Kozak. Also, only the third is conserved amongst the human, macaque, dog, rat and mouse FIX genes (Pang et al, 1990). The transcription start site is situated 29 nucleotides 5' of the initial AUG codon (Anson et al., 1984).

B. Protein Structure and Function

Human FIX is a vitamin K-dependent glycoprotein of MW ~57,000 synthesized in the liver. The zymogen precursor is 461 amino acids long and undergoes extensive post-

Figure 4. Structure of the human factor IX gene, mRNA and protein

A. The FIX gene is represented schematically in the 5' to 3' direction. Solid dark bars represent the eight exons which are numbered above. Altogether, the gene spans ~34 kb. B. The mRNA is ~2.8 kb, approximately half of which is 3' non-translated sequence (denoted by the dark bar). Again the exons are numbered. C. The FIX protein is 461 amino acids in length. The 'prepro' consists of the signal peptide and the propeptide, both of which are cleaved to give the mature zymogen. The 'Gla' domain contains modified γ-carboxy glutamic acid residues. The 'EGF-like' domains are homologous to human epidermal growth factor. The 'activation peptide' contains the two cleavage sites required for FIX activation. The catalytic domain contains the catalytic triad responsible for the proteolytic function of FIXa.


A.
5'
3'
1
2 3
4
5
6
7
8
Gene
B.
1
2/3
4
5
6
7
8
mRNA
C.
prepro
gla
egf B
egf A
activation peptide
catalytic domain
Protein

translational modification before being secreted. The mature protein is 415 amino acids long and exists as two chains held together by (an undetermined number of) disulfide bonds. Its circulating concentration in plasma is approximately 5μg/ml (Furie and Furie, 1990).

1. Signal and propeptide

Exon 1 of the FIX gene codes for the hydrophobic signal peptide which is required for translocation of FIX into the rough endoplasmic reticulum where this 26 amino acid peptide is cleaved. Exon 2 codes for the propeptide. There is marked sequence homology in this region amongst all of the vitamin K-dependent clotting proteins. The propeptide contains the γ-carboxylation recognition site (γ-CRS) which is involved in directing the post-translational modification of the 12 glutamic acid residues in the adjacent domain (DiScipio and Davie, 1979). Jorgensen et al. (1987) have shown that the absence of this 18 amino acid portion in FIX molecules completely abolishes γ-carboxylation. As well, alteration of residues -16 or -10 resulted in impairment of this function (Jorgensen et al., 1987). The propeptide is normally cleaved prior to secretion of FIX.

2. Gla domain

The first domain (exons 3 and 4) in the mature circulating FIX molecule is the Gla domain spanning residues 1 through 47. This region is associated with a class of low affinity calcium binding sites as well as interaction with phospholipids. In the presence of calcium, the Gla domain undergoes a large conformational change (Borowski et al, 1986; Liebman et al., 1985). Two or three of the Gla residues bind a calcium ion and form a noncovalent intramolecular bridge between regions of the polypeptide backbone thus stabilizing the tertiary structure (Tai et al., 1984). It is thought that this process exposes a site which is involved in the binding of FIX to a putative endothelial cell receptor.

At the C terminus of the Gla region is a short 8 amino acid segment known as the aromatic stack or connecting peptide. In FIX, it obeys the conserved sequence: Phe-Trp-

x-x-Tyr. Blake et al. (1987) suggest that interaction between the aromatic stacks in FIX and FX is required for correct alignment of the enzyme with its substrate.

3. Growth factor-like domains

The following two exons encode two EGF-like domains. The N terminal unit is comprised of residues 47-85 and shares close homology with the e subunit of the EGF precursor. The C terminal domain extends from residue 86 to 128 and is more closely related to the j subunit. The 2 EGF-like domains have probably not arisen through tandem duplication of a single domain. It is hypothesized that the C terminal EGF-like domain was initially inserted into the ancestral protease followed by insertion of the N terminal domain (Patthy, 1985).

The first EGF-like domain is associated with a high affinity calcium binding site and thus with binding to an endothelial cell receptor. Previously it was believed that β-hydroxyaspartate (Hya) at residue 64 (Drakenberg et al, 1983) was responsible for the calcium binding site (Rees et al., 1988). It has since been determined that lack of post-translational hydroxylation at this particular residue in recombinant FIX does not affect calcium binding assays and Hya_{64} must therefore have another role. Furthermore, hydroxylation is incomplete in humans with only ~0.3mol Hya/mol FIX (Fernlund and Stenflo, 1983). Hase et al. (1990) and Nishimure et al. (1989) have identified a unique carbohydrate side chain linked to Ser_{53} (Xylα1-3Glu-O-Ser) which may be involved in interactions with FVIIIa. A portion of the light chain of FVIIIa contains ~20% sequence identity with the slime mold lectin discoidin I (Kane et al, 1988). Astermark and Stenflo (1991) postulate that this region of FVIIIa may function as a lectin with affinity for the carbohydrate chain on FIX.

The second EGF-like domain does not contain any Hya residues and has no calcium binding properties. The function of this domain remains largely unknown.

4. Activation peptide

The activation peptide of FIX is 35 amino acids long (residues 146-180) and is cleaved to yield active FIXa. It contains two glycosylated asparagines at positions 157 and 167 and has no coagulant activity. In both the extrinsic and intrinsic pathway, the Arg_{145}-Ala_{146} bond is cleaved prior to the Arg_{180}-Val_{181} bond. However, limited proteolysis by FXIa occurs at the two cleavage sites almost simultaneously whereas in activation by FVIIa, the second bond is cleaved considerably more slowly (Enfield and Thompson, 1984). Upon activation, the catalytic domain undergoes a large conformational change. Inclusion of the activation peptide in the zymogen probably acts to maintain the zymogen in a structure unable to form an enzyme-substrate complex or for that complex to reach its transition state.

5. Catalytic domain

The protease portion of FIX comprises a little more than half of the mature molecule (residues 181-415) and shows 44% sequence identity with trypsin. Crystallographic data of trypsin and chymotrypsin has revealed that in addition to sharing sequence identity, these proteins also share homology in the three dimensional structure of the peptide backbone. Other proteins belonging to the serine protease family showing modest sequence homology also share very similar secondary structures. Based on these facts, Furie et al. (1982) created computer models of the catalytic domains of FIX, FX and prothrombin based on the 3D structures of trypsin and chymotrypsin. Comparisons of the surface features of FIX with trypsin revealed regions of similarity (constant regions) and dissimilarity (variable regions). Because these proteins are mechanistically similar, the constant regions probably contain amino acids essential to FIXa enzyme function. Conversely, the variable regions are more likely to contain amino acids involved in binding and recognition of the substrate.

Like the other serine proteases, FIX hydrolyzes peptide bonds via the charge relay system described by Blow. Specifically, the catalytic domain of FIXa is responsible for cleavage of the Arg_{51}-Ile_{52} peptide bond in FX to produce active FXa. The conserved catalytic triad are situated at $histidine_{221}$, aspartic $acid_{269}$ and $serine_{365}$. Other than its catalytic function, the protease domain is also thought to play a role in FVIIIa interaction.

IV. HAEMOPHILIA B

A. Overview of haemophilia B

Haemophilia B (also known as Christmas disease) is an X-linked bleeding disorder occuring in approximately 1 in every 30,000 males. Haemophilia B results from a deficiency of active FIX in the coagulation cascade and is characterized by impairment of the ability of blood to clot. The severity of cases covers a wide spectrum ranging from <1% coagulant (severe) to 1-5% activity (moderate) to 5-15% activity (mild), indicating the heterogeneity of the disease. Due to the complexity and number of functions which the FIX enzyme must fulfill, such heterogeneity is not surprising as disruption of any one of these functions is likely to affect FIXa activity. In approximately one third of haemophiliacs there is a complete absence of FIX antigen while the other two thirds produce reduced or normal amounts of defective protein. The former is referred to as CRM^- (cross reacting material negative) and the latter as CRM^+ (cross reacting material positive).

FIX activity is measured by the activated partial thromboplastin time (APTT) which is a standard haematological measure of the activity of the extrinsic pathway.

B. Haemophilia B Leyden

A rare subclass of haemophiliacs, first diagnosed in a Dutch family and designated as Haemophilia B Leyden, possess FIX antigen levels which increase with age. At birth,

affected males have little or no detectable FIX activity and antigen. Around the onset of puberty however, FIX activity increases at a rate of approximately 4-5% per year (Veltkamp et al., 1970). By adulthood, FIX levels in the plasma may range from 40-80% of normal, resulting in the near or total disappearance of clinical symptoms. Mutations which cause Haemophilia B Leyden occur in the putative promoter region and are thus thought to disrupt FIX transcription. A possible explanation for the rise in FIX levels is the existence of a nearby androgen-responsive element which, at puberty, may contribute to the strength of the promoter (Crossley and Brownlee, 1990).

C. Haemophilia B_m

Another interesting subclass of CRM^+ haemophiliacs are known as Haemophilia B_m, named after the Murphy family in whom it was originally discovered (Hougie and Twomey, 1967). In cases involving Haemophilia B_m, normal prothrombin times are exhibited when using human brain thromboplastin as the source of tissue factor. However, when ox brain thromboplastin was used, the prothrombin times were markedly prolonged. This is unusual because FIX is not expected to affect the prothrombin time but in the presence of ox brain tissue factor, the mutant FIX interferes with the FVII-dependent activation of FX. It is postulated that this heterogeneous group of aberrant FIX molecules somehow disrupt the activation of FVII or FX or, more likely, the binding of FVII to the bovine tissue factor.

V. DATA BASE OF KNOWN MUTATIONS

A. Deletions

Deletions that have been found to cause haemophilia B vary in size from a single base to 250 kb. Patients with gross deletions within the FIX gene exhibit, as a rule, severe haemophilia B and have essentially no FIX antigen in their plasma. At least 15 complete

gene deletions have been found, including FIX Pisa, FIX Boston I, FIX HB5 and FIX Rheine. In most cases the breakpoints have not been identified so the exact size of the entire deletions are not known. Partial gene deletions have also been identified including FIX Strasbourg, FIX Chicago 1, FIX Seattle 1 and FIX Hannover. FIX Chicago 1 is especially interesting in that the FIX gene contains two separate deletions. Many shorter deletions of less than 20 nucleotides have also been detected. Deletions often cause frameshifts in the coding region and usually cause severe CRM^- haemophilia (FIX UK2, FIX UK11, FIX Bonn 2, FIX Malmo 1). Often alteration of the frame results in a stop codon and truncation of the aberrant polypeptide. In FIX UK 10, a 3 base pair deletion of an arginine codon at residue 37 leaves the coding sequence in frame resulting in 12% FIX antigen but less than 1% activity. In FIX Bottrop, a 3 base pair in frame deletion of a glycine codon at residue 184 results in severe CRM^+ haemophilia. A 4 base pair deletion starting at residue 17,667 in FIX HB6 disrupts the consensus sequence for the acceptor splice site resulting in mild haemophilia with 20% activity.

B. Insertions

A large insertion of over 6 kb of unrelated DNA is inserted in the intron between exons IV and V in FIX El Salvador causing severe CRM^- haemophilia. There are also many other CRM^- haemophilias caused by single base pair insertions which result in frameshifts and introduce premature stop codons.

C. Point mutations

The FIX Leyden phenotype is so far known to arise from T to A nucleotide changes at position -20, G to A nucleotide changes at position -6, a G to C nucleotide change also at position -6, A to G nucleotide changes at position +13 or most recently a T to C transition at position +8. Although all of the above mutations give rise to the Leyden phenotype, the -20 and -6 mutations are thought to differ mechanistically from the +8 and +13 mutation.

By far, most of the mutations which are known to give rise to haemophilia B are single point mutations in the coding regions of the FIX gene. Nonsense mutations resulting from point mutations often involve conversion of an arginine codon to a stop codon (FIX Malmo3,4 and 7, FIX Leiria, FIX HB29, FIX Bonn 1). Other stop mutations involve conversion of glutamine (FIX HB17, FIX HB5 Japan, FIX), tryptophan (unnamed, Wang et al., 1990), or glutamic acid (FIX Oxford b3).

A common missense mutation arises from conversion of Arg_{-4} to Trp (FIX Malmo 6), Leu (FIX Bendorf), or Glu (FIX Oxford 3). This residue is thought to be part of the γ-CRS directing carboxylation of the Glu residues in the Gla domain. In FIX San Dimas (Arg_{-4} to Glu) and FIX Cambridge (Arg_{-1} to Ser) the FIX molecule circulates with the propeptide as well as exhibiting abnormal carboxylation. In the Gla domain, mutation of Glu residues to Ala and Asp at positions 7 and 33 (FIX Oxford b3 and FIX HB9) results in moderate haemophilia (~5% activity and 4% activity) while mutation of a Glu residue to Lys and Val at position 33 (FIX Seattle 3, and FIX Chongqing) result in severe haemophilia. These mutations preclude the possibility for γ-carboxylation at these residues.

In the first EGF-like region, FIX Alabama (Asp_{47} to Gly) the mutated Asp residue is thought to be involved in calcium binding. Loss of this residue results in mild haemophilia B possibly due to impairment of the ability to bind to FVIII or the endothelial cell membrane. FIX Hollywood is also thought to cause a loss in either of these two functions. A β-turn involving residues Asn-Pro_{55}-Cys-Leu is strongly predicted in normal FIX. In FIX Hollywood, Pro_{55} is mutated to Ala resulting in a predicted four-fold decrease in the β turn probability. This is likely to disrupt the structure of this domain and thus the calcium binding ability. Although the Hya_{64} residue has been shown not to be involved in the high affinity calcium binding site in the first EGF-like domain, mutations at this position (FIX UK6 and FIX Oxford d1) result in mild to moderate haemophilia. Hya_{64} must have another important function in the coagulant role of FIX.

Mutations disrupting the Arg_{145}-Ala_{146} bond in the activation peptide abolish cleavage by FXIa and produce the intermediate molecule, FIXaα. In cases where Arg_{145} is mutated to cysteine (FIX Albuquerque, FIX Cardiff 1) the result is severe CRM^+ haemophilia. In FIX Chapel Hill and FIX Chicago 2, Arg_{145} is mutated to histidine resulting in mild CRM^+ haemophilia. Loss of either residues involved in the second cleavage site, Arg_{180}-Val_{181}, gives rise to severe CRM^+ haemophilia B_m (FIX Hilo, FIX Milano). In Cardiff 2, a conservative Val_{182} to Leu mutation results in a B_m variant with 15% activity.

B_m variants also are caused by mutations at Pro_{368} (FIX Bergamo), Ala_{390} (Lake Elsinore) and Gly_{396} (Angers) in the catalytic domain. These last two mutations are thought to disrupt an extended substrate recognition site. Conversion of Ala_{233} to Thr leads to reduced levels (8-22%) of fully functional FIX (FIX HB1, FIX Malmo 28-31). Conversion of Thr_{296} to Met results in moderate to mild haemophilia (FIX HB 19, FIX UK 32). In FIX London 2, Arg_{333} is mutated to Glu resulting in severe CRM^+ haemophilia. Although not well conserved in most other serine proteases, this residue is absolutely conserved in human and bovine FIX, FX and prothrombin and is thought to reside on the surface of these proteins. Tsang et al. postulate that the positive charge normally contributed by arginine plays an important role in a dispersed substrate or cofactor binding site. Similarly, the substitution of Thr for Ile_{397} in FIX Vancouver and FIX Long Beach is thought to interfere with substrate binding. Based on computer models of the FIX catalytic domain, introduction of threonine was predicted to provide the possibility of hydrogen bonding between the side chain hydroxyl group of Thr_{397} and the carbonyl oxygen of Trp_{385}. In support of this hypothesis, K_m and k_{cat} values of FIX Vancouver were found to be 3.4 times higher and 8 times lower than normal FIXa, respectively.

VI. METHODS USED IN STUDYING HAEMOPHILIA B

A. Cloning

Prior to recent advances in technology which now allow rapid amplification of known sequences, studies on naturally occuring mutations causing haemophilia B were a laborious process. Previous studies required making a genomic library from DNA obtained from the haemophilia patient. The genomic DNA was partially digested with an appropriate enzyme before fragments of intermediate length were isolated and ligated into the arms of a vector such as λ EMBL3. The reconstituted EMBL3 would subsequently be packaged into phage and plated onto a lawn of *E. coli*. Screening of the library with a labelled FIX cDNA probe would then be necessary to identify clones which contained FIX sequence. Most likely several clones would have to be found in order to obtain all eight coding exons. After mapping the clones to the FIX genes by restriction enzyme analysis, selected fragments containing the exons were isolated and subcloned into vectors for DNA sequencing. Altogether the above protocol could take many months to accomplish.

B. Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a powerful technique developed in the last several years (Saiki et al., 1985) which allows rapid amplification of known nucleic acid sequences. PCR is mechanistically very similar to the actual *in vivo* process of DNA replication. Initially, two strands of template DNA are denatured by high temperature after which the temperature is dropped to allow annealing of two extension primers. Each of these short synthetic oligonucleotides are complementary to sequences on one or the other strand of DNA and flank the region to be amplified. In the presence of dNTPs extension occurs in the 5' to 3' direction using a thermostable DNA polymerase, commonly from *Thermus aquaticus* (Taq polymerase). One cycle of PCR represents the sequential

performance of these three steps (ie. denaturation, annealing and extension). Repeated cycling results in the exponential amplification of the target sequence.

Since the entire gene sequence is known for FIX, it is relatively simple to design primers for the amplification of the exonic sequences for direct DNA sequence analysis or subcloning into a plasmid for sequencing (Figure 5). This new process negates the requirements for making a genomic library and requires only days once DNA has been collected.

C. Single Stranded Conformational Polymorphism Analysis

In 1989, Orita et al. (1989) reported a rapid and sensitive method for the detection of single base pair mutations called single stranded conformational polymorphism (SSCP) analysis (Figure 6). SSCP analysis is based upon the principle that single stranded DNA acquires a unique secondary conformation that is both sequence specific and size specific when electrophoresed on a non-denaturing polyacrylamide gel. Mobility of DNA in such a gel is dependent on conformation, therefore sequence changes would become evident as a shift in mobility.

To detect point mutations in FIX, the exons can be amplified using PCR in the presence of radiolabelled dNTPs and electrophoresed on a non-denaturing polyacrylamide gel. After drying and exposure to film, the autoradiogram should indicate point mutations through shifts in mobility of the DNA fragments as compared to normal samples. Although this is not a substitute for sequencing, it is a rapid screening technique and is particularly useful in pedigree analysis as it is informative in all cases.

Figure 5. PCR strategy for amplification of factor IX sequences

The FIX gene is represented schematically in the 5' to 3' direction with the approximate positions of the exons indicated by vertical bars. Sets of synthetic oligonucleotide primers (20 nucleotides in length) were designed to flank the exonic sequences. Exon 1 was amplified along with ~150 base pairs of 5' sequence. Exons 2 and 3 were amplified together. Exons 4, 5, 6, 7 and 8 were amplified separately. The size of each PCR amplified fragment is indicated below the exon.

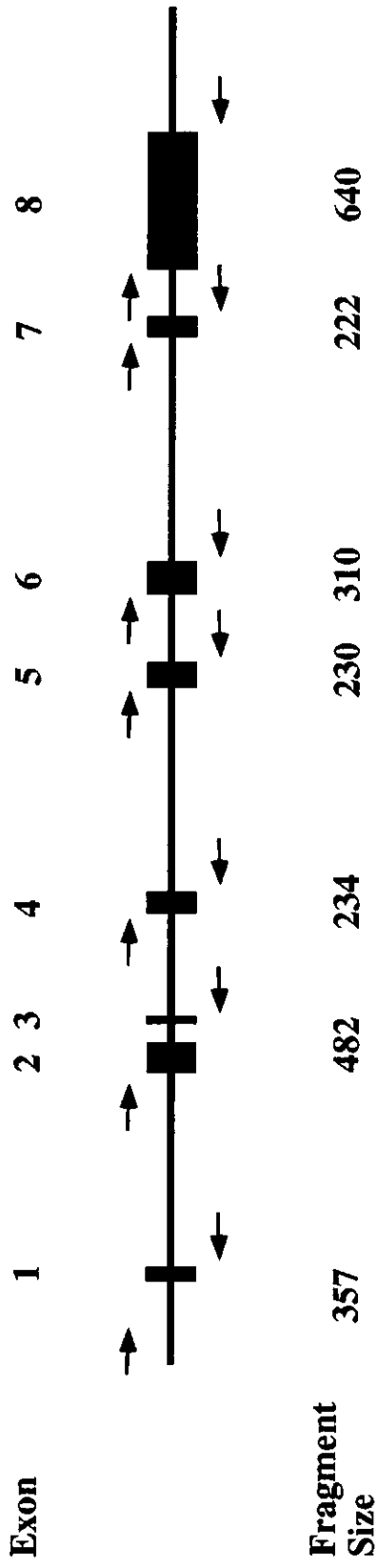
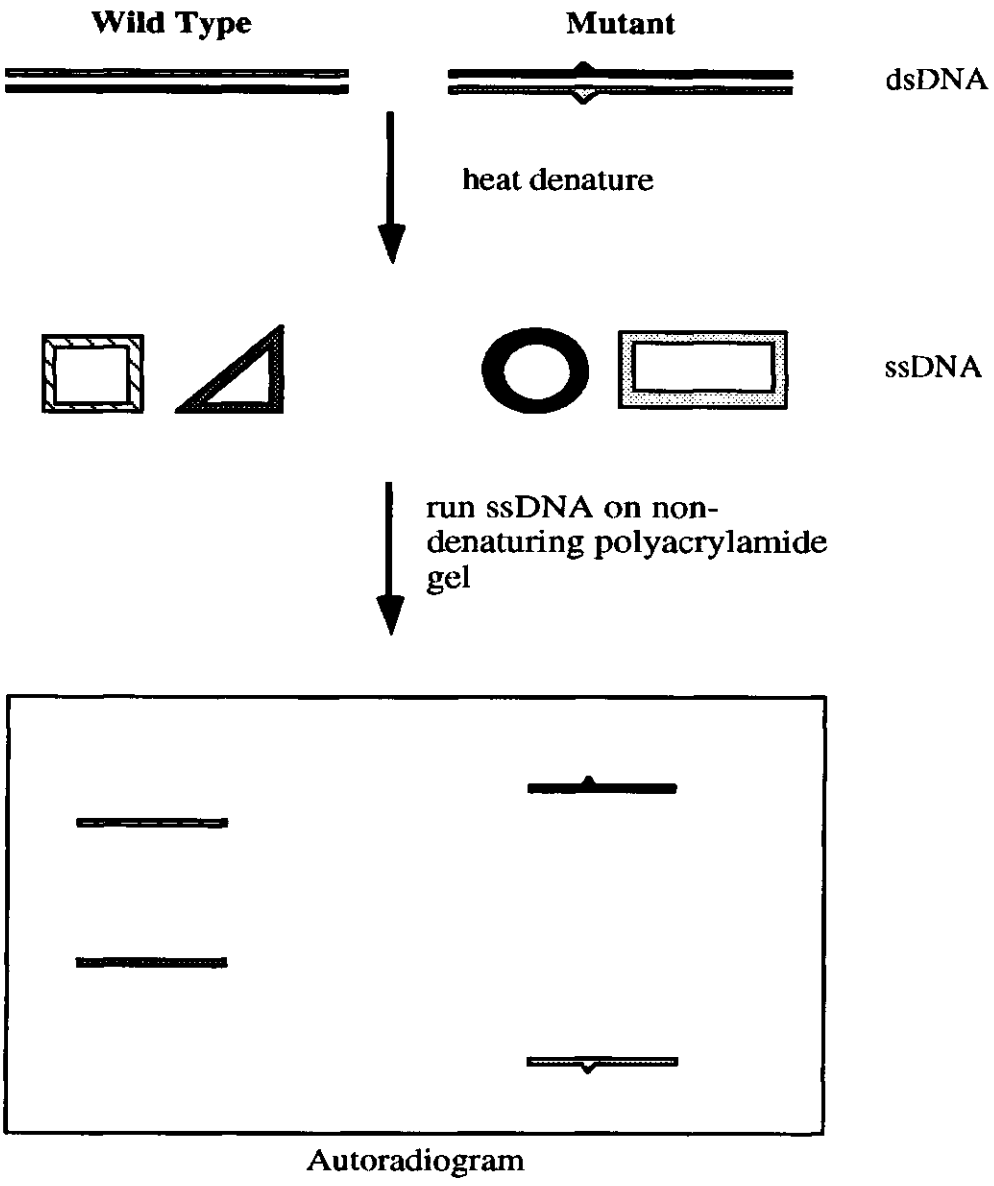

Exon
1
2 3
4
5
6
7
8
Fragment Size
357
482
234
230
310
222
640

Figure 6. Single stranded conformational polymorphism analysis (SSCP)



VII. OBJECTIVES OF THIS STUDY

The main objective of the current study was to screen a panel of FIX genes from western Canadian haemophilia B patients in order to identify the molecular defects causing the bleeding disorder. By using molecular biological techniques such as PCR, SSCP analysis and DNA sequencing, it was our goal to determine the exact nature and location of the mutations. Activity and antigen levels of the aberrant FIX molecules were also determined. In conjunction with this data, an attempt was made to correlate the mutations with their possible effects on the various functions that FIX fulfills in order to obtain a clearer picture of the structure-function relationship. This study leads the way to recombinant expression of mutant FIX molecules for further characterization studies.

MATERIALS AND METHODS

I. REAGENTS

Yeast extract, bacto-tryptone and bacto-agar were purchased from Difco Laboratories. Agarose, acrylamide, bisacrylamide, ammonium persulfate, TEMED and urea were purchased from Bio-Rad Laboratories. Ammonium sulfate and phenol were obtained from British Drug Houses Ltd. Deoxy and dideoxy-ribonucleotides were obtained from Pharmacia. β-mercaptoethanol, DTT, RNase A, DNase I and lysozyme were purchased from Sigma Chemical Co. α-[thio-^{35}S]-dATP was purchased from New England Nuclear. All other reagents were of reagent grade or higher and were obtained from Bio-Rad, BDH, Pharmacia or Sigma.

T_4 polymerase and Klenow-large fragment, Hinc II, Xba I, Pst I and 1 kb DNA ladder were obtained from Bethseda Research Laboratories. T_7 polymerase was purchased from Pharmacia. Recombinant Taq polymerase (Amplitaq) was obtained from Perkin Elmer Cetus.

Kodak X-Omat and Kodak XAR film was used for autoradiography.

II. STRAINS, VECTORS, AND MEDIA

A. Vectors

Factor IX exons, as amplified by PCR, were ligated into the polycloning site of pUC19. pUC contains the pBR322 ampicillin resistance gene, the pBR322 origin of replication and part of the lac Z gene of *E. coli.* These small plasmids are 2.7 kb long and grow to a relatively high copy number.

B. Bacterial strains

E.coli strain DH5α was used for transformation . The genotype of these cells is: F^-, endA1, hsdR17(r_k^-, m_k^+), supE44, thi-1, λ-, recA1, gyrA96, Δ(argF-Laczya)U169, ϕ80LacZΔM15. The ϕ80LacZΔM15 marker allows α complementation of the β-galactosidase gene from pUC. Blue/white colour selection was used for screening colonies on bacterial plates.

C. Media

LB plates (to 950 ml water add 10g bacto-tryptone, 5g bacto-yeast extract, 10g NaCl, adjust to pH 7.4, add 15g agar, and water to 1 litre, autoclave to sterilize) were used to support growth of the transformed cells. The plates also contained 50 mg/ml (filter sterilized) of ampicillin, 0.1M IPTG, and 2% Xgal in DMF which were added after autoclaving.

III. OLIGONUCLEOTIDES

Synthetic oligonucleotides were designed for use in DNA amplification (polymerase chain reaction) and in DNA sequence analysis (Table 1). Forward (F) and reverse primers (P) were placed approximately 20 base pairs 5' and 3' of the intron/exon splice junction. In the case of exon 1 however, the forward primer was placed ~150 base pairs 5' to the transcription start site to include the putative promoter. Exons 2 and 3 were amplified together as both exons are relatively small and are situated very close together in the gene. Internal primers (I) were designed for exons 2/3 and 8 to facilitate sequencing.

The oligonucleotides were synthesized on the Applied Biosystems 391 DNA Synthesizer and purified by reverse phase chromatography on Sep-Pak C18 cartridges as described by Atkinson and Smith (1984). Purified oligonucleotides were resuspended in dH_20 and stored at -20°C.

Table 1. Oligonucleotide primers used for the polymerase chain reaction and DNA sequence analysis

Oligonucleotide	Exon	Nucleotide position	Forward/Reverse/ Internal	Sequence
FIX-1	1	-147 to -166	Forward	AAATCAGCCACAGTGGCAGA
FIX-2	1	152 to 171	Reverse	TTTCTATATCTAAAAGGCAA
FIX-3	2/3	6272 to 6291	Forward	TTCATGATGTTTTCTTTTTT
FIX-4	2/3	6734 to 6753	Reverse	TAATTCTCATATGTTTCATA
FIX-19	2/3	6466 to 6485	Internal	GTTTTTGAAAACACTGAAAG
FIX-20	2/3	6521 to 6540	Internal	ATTTTCTATTCTATGCTCTG
FIX-5	4	10344 to 10353	Forward	AGGACCGGGCATTCTAAGCA
FIX-6	4	10548 to 10567	Reverse	CAGTTTCAACTTGTTTCAGA
FIX-7	5	17618 to 17637	Forward	TTAGAAATGCATGTTAAATG
FIX-8	5	17828 to 17847	Reverse	TGAAGTTTCAGATACAGATT
FIX-9	6	20308 to 20327	Forward	GTTCCATTTGCCAATGAGAA
FIX-10	6	20598 to 20617	Reverse	CTGTGTCTTGCCAGCTGAGC
FIX-11	7	29988 to 30007	Forward	TTTCTAGATCAAATGTATTA
FIX-12	7	30190 to 30209	Reverse	CAATCATATTAAAGAGCTAG
FIX-13	8	30780 to 30799	Forward	GAAATACTGTTTGTGACTTA
FIX-14	8	31400 to 31419	Reverse	CCCTGTTAATTTTCAATTCC
FIX-15	8	30948 to 30967	Internal	ACGAACCCTTAGTGCTAAAC
FIX-16	8	31010 to 31029	Internal	CCAAATTTGAGGAAGATGTT
FIX-17	8	31163 to 31182	Internal	TGTTCTGTGCTGGCTTCCAT
FIX-18	8	31226 to 31245	Internal	CCTTCCACTTCAGTAACATG

IV. ISOLATION OF DNA

A. Isolation of genomic DNA from peripheral blood cells

Blood (20 ml) was collected in EDTA (5mM final concentration) and gently mixed. The plasma fraction was removed by diluting the sample with isotonic saline (0.85% NaCl) to 50 ml in a Falcon tube and centrifuging at 500 x g for 15 minutes. The supernatant was removed without disturbing the buffy layer (ie. white blood cell layer). Red blood cells were lysed by filling the Falcon tube to 45 ml with NH_4Cl:Tris solution (0.14 M NH_4Cl, 0.017 M Tris) and incubating at 37^{o}C with occasional gentle mixing for 10 minutes. The mixture was again spun at 500 x g for 15 minutes and the supernatant removed without disturbing the white cell pellet. The pellet was washed with NH_4Cl:Tris solution and then twice more with isotonic saline until the supernatant was fairly clear. To lyse the white cells, the pellet was resuspended in 5 ml of 0.1 M Tris pH 8.0, 0.04 M EDTA after which an equal volume of the same solution containing 0.2% SDS was quickly injected with an 18 gauge needle. The mixture was digested with 1/100 volume proteinase K (10 mg/ml stock solution) and incubated overnight at 50^{o}C. The protein was extracted with water saturated phenol three times until there was no longer a significant white precipitate at the interface. This was followed once with a phenol/chloroform extraction and thrice more with chloroform. A 1/10 volume of 4M ammonium acetate and 2 volumes of ethanol was added to precipitate the DNA. The tube was gently rocked until the DNA had condensed. The DNA was removed with a glass wand and transferred to a fresh microfuge tube. The precipitate was washed with 70% ethanol, dried under a vacuum and resuspended in 1.0 ml of TE. The DNA was quantified by digesting a 10 µl aliquot with DNAse I in low salt buffer for 15 minutes. The absorbance was read at 280nm and 260nm. Finally, undigested and digested aliquots of DNA were electrophoresed on a 1% agarose gel to check for degradation.

B. Isolation of plasmid DNA

Plasmid DNA was isolated using a modified boiling lysis procedure (Gatermann et al., 1988). LB broth (1.5 ml) was inoculated with a single colony and grown at 37°C overnight. The cells were centrifuged at 12,000 x g for 1 minute and the resulting pellet was resuspended in 100 µl of STET buffer (8% sucrose, 5% Triton X-100, 50 mM EDTA, 0.5 mg/ml lysozyme). To lyse the cells, the tubes were placed in a boiling waterbath for 2 minutes. Cell debris and genomic DNA were collected by centrifuging at 12,000 x g for 15 minutes at 4°C and removed using a sterile toothpick. Isopropanol (100 µl) was added to the supernantant and vigorously vortexed. After centrifuging at 12,000 x g for 15 minutes at 4°C, the precipitate was washed with 70% ethanol. The pellet was allowed to air dry before resuspending in 40µl of distilled water.

V. GEL ELECTROPHORESIS

A. Agarose gel electrophorsesis

DNA fragments (ie genomic DNA, PCR products, plasmids and plasmid inserts) were visualized by electrophoresis on agarose gels. The running buffer used was TAE (1.6 M Tris, 0.8 M sodium acetate, 40 mM EDTA, pH to 7.2). Ethidium bromide was added to both the gel and the running buffer at a concentration of 1 µg/ml. Gels were electrophoresed at 100 V until the bromophenol blue was approximately 2 cm from the end of the gel. The DNA fragments were visualized under ultraviolet light.

B. Denaturing polyacrylamide gels

DNA fragments from sequencing reactions were electrophoresed on 6% polyacrylamide gels at 50 watts constant power for two hours. Gels were prepared by adding 8.3 M urea and the appropriate volume of 38:2 acrylamide:bisacrylamide stock solution to the running buffer, 1xTBE (50 mM Tris base, 50 mM boric acid, 1 mM

EDTA). Polymerization was initiated by addition of 0.066% (w/v) ammonium persulphate and 0.024% (w/v) TEMED. After electrophoresis, gels were dried onto 3MM paper and exposed to x-ray film.

C. Non-denaturing polyacrylamide gels

Radiolabelled DNA fragments from PCR reactions were electrophoresed on 6% non-denaturing polyacrylamide gels for single strand conformational polymorphism (SSCP) analysis. The gels were prepared similarly to denaturing polyacrylamide gels but addition of the 8.3 M urea was omitted. Also 10% (v/v) glycerol was added to the gels prior to initiation of polymerization. Non-denaturing gels were run at only 30 watts constant power for 6-12 hours. Fans were used to cool the glass plates during electrophoresis.

VI. POLYMERASE CHAIN REACTION

A. For use in subcloning

Exons of the FIX gene were amplified using the polymerase chain reaction (PCR). The reactions were performed as described by Saiki et al. with a few minor alterations. The following reagents were added sequentially into a 0.5 ml eppendorf tube: 72.9 µl dH_20, 10.0 µl 10x reaction buffer, 1.6 µl dNTPs (1.25 mM each dNTP), 5.0 µl forward primer (20 µM), 5.0 µl reverse primer (20 µM), 5.0 µl genomic DNA template (100 ng/µl) and 0.5 Taq polymerase (5 Units/µl). Total reaction volume was 100 µl. The contents were gently mixed and overlaid with light mineral oil to prevent evaporation. The tubes were incubated at 94°C for 5 minutes prior to 40 cycles in a Perkin Elmer Cetus thermocycler; each cycle consisted of denaturing at 94°C for 15 seconds, annealing at 48-53°C for 30 seconds and extending at 72°C for 60 seconds.

Annealing temperatures were dependent on the base composition of the oligonucleotides and were estimated using the following equation: T_A=[4°C(GC)+2°C(AT)]-5°C. The optimal annealing temperature for each set of primers was determined emperically using the estimates obtained from the equation. Two different 10x reaction buffers were used. One was recommended by Perkin Elmer Cetus (500 mM KCl, 100 mM Tris-Cl pH 8.3, 15 mM $MgCl_2$, and 0.1% (w/v) gelatin). Another buffer (67 mM Tris, 1.0 mM $MgSO_4$, 16.6 mM NH_4SO_4, 9.8 mM β-mercaptoethanol) used in later experiments, was found to give higher yields of amplified DNA .

B. Asymmetric amplification for direct sequencing

The production of single stranded DNA using PCR required adding unequal amounts of the two primers to the reaction. The concentration of the limiting primer was 8X less than that of the non-limiting primer. All other conditions remained as in part A.

C. For use in SSCP analysis

Fragments amplified for use in SSCP analysis were radiolabelled using α[thio^{35}S]dATP. Label (0.5 μL) was added to the basic reaction which was scaled down to 20 ml. Twenty-five cycles were performed instead of 40. All other conditions remained the same.

VII. SUBCLONING DNA

A. Purification of DNA fragments to be subcloned

In PCR reactions where there appeared to be little non-specific amplification (as judged from electrophoresis on an agarose gel), the reaction mixture was extracted with an equal volume of SEVAG (24:1 chloroform:isoamylalcohol) to removed the mineral oil.

The DNA was precipitated by adding one volume of 4 M ammonium acetate and two volumes of isopropanol to the aqueous phase and incubating on ice for 10 minutes. The precipitate was collected by centrifugation at 4°C for 10 minutes after which it was washed with 70% EtOH and dried under vacuum. The DNA pellet was resuspended in 20 µl dH_20.

In PCR reactions where there appeared to be considerable amount of background, DNA fragments were further purified. After isopropanol precipitation, the DNA was electrophoresed on a 1% agarose gel. The fragment of interest was excised form the agarose gel using a sterile scalpel and the DNA was recovered using the GeneClean (Bio101) protocol as described by the manufacturer. The DNA was eluted in a final volume of 20 µl dH_20.

B. Blunt ending PCR fragments

A portion of DNA fragments produced by PCR are known to possess 3' overhangs usually of a single base. The Klenow fragment of Pol I was used to make the PCR fragments blunt-ended. Klenow (1µl of 5 units/µl), 3 µl 10x polymerase buffer (70 mM Tris pH 7.4, 500 mM NaCl, 70 mM $MgCl_2$) and 3 µl 10 mM DTT was added to the 20 µl DNA and incubated at 37°C for 15 minutes. dNTP (3 µl of 0.5 mM each) was then added and incubated for a further 15 minutes. The enzyme was heat inactivated at 68°C for 10 minutes. A phenol/chloroform extraction was performed by adding equal volumes of phenol and chloroform. The solution was vortexed and spun at 4°C for 1 minute. The top aqueous layer was transferred to a new 1.5 ml eppendorf and ethanol precipitated. After drying down the pellet, the DNA was resuspended in 10 µl of TE.

C. Cutting and dephosphorylating the vector

The vector (pUC19) was dephosphorylated after being cut with a restriction enzyme in order to increase the efficiency of blunt end ligations. Only vectors incorporating an

insert would religate and thus be transformed. pUC19 (10 µg) was cut with the blunt cutter, Hinc II (20 units) in Pharmacia One-Phor-All buffer. The vector was digested for 1/2 hour at 37°C. Calf intestinal alkaline phosphatase (1 µl of 6.7 units/µl) was added and the digestion continued for another 1/2 hour. The phosphatase was removed by a phenol/chloroform extraction. After ethanol precipitation, the linearized, dephosphorylated pUC19 was resuspended in 100 µl of TE to a final concentration of 100 ng DNA/µl.

D. Ligation of DNA fragments into pUC19

Ligation reactions contained 3 µl of linearized, dephosphorylated pUC19 (300 ng), 8 µl of insert DNA, 4 µl 5x ligase buffer, 1µl T_4 ligase (5 units)and 4 µl distilled water in a 0.5 ml eppendorf tube. The plasmid to insert ratio was approximately 4 to 1 which is considered optimal for blunt end ligations. A negative control was performed in which the DNA insert was substituted with distilled water. A positive control was also included in which pUC19 (linearized but not dephosphorylated) was religated to itself. The tubes were incubated overnight in a 15°C waterbath and then frozen at -20°C until required for transformation.

VIII. TRANSFORMATION OF RECOMBINANT PUC19 INTO E.COLI

Competent bacterial cells were prepared by growing a 40 ml liquid culture of DH5α to an OD_{600} of ~0.6. The cells were centrifuged at 5,000 x g for 10 minutes at 4°C and the pellet was resuspended in 20 ml of ice cold 50 mM $CaCl_2$. After incubation on ice for 30 minutes, the cells were centrifuged again and the pellet resuspended in 4 ml ice cold 50 mM $CaCl_2$, 20% glycerol. The cells were aliquoted, frozen in a dry ice/EtOH bath and stored at -70°C.

A 2 µl aliquot of the ligation mixture was combined with 50 µl of competent DH5α in a 1.5 ml eppendorf tube. The cells were incubated on ice for 30 minutes and then heat shocked at 37°C for 20 seconds and finally placed on ice for 2 minutes. LB broth (0.95

ml) was added to the tube and the cells were grown at 37°C with constant aeration for 30 minutes. The cells were then centrifuged for 1 minute, 900 µl of the supernatant was removed and the cells were resuspended in the remaining 100 µl of broth. Transformed cells were selected by plating on LB/amp/IPTG/Xgal agar plates and incubating at 37°C overnight. Isolated white colonies were chosen for subsequent DNA isolation. To confirm the competency of the cells, 10 ng of uncut pUC19 was also used to transform *E. coli*.

IX. DNA SEQUENCE ANALYSIS

A. Sequencing from plasmid DNA

The chain termination method (Sanger et al., 1977) was used in DNA sequence analysis. All solutions used are as described in the Sequenase protocol (3rd ed.). Double stranded DNA template was prepared for sequencing by boiling 8 µl of plasmid DNA with 1 µl sequencing primer (20 pmole), 2 µl of 2M NaOH and 11 µl of distilled water for 2 minutes. This was followed by an ethanol precipitation and the pellet was resuspended in 8 µl TE in a 0.5 ml eppendorf tube. To the DNA solution, 2 µl of 5x sequencing buffer was added and the tubes were incubated at 37°C for 15 minutes to anneal the primers. To each sequencing reaction, 1 µl 100 mM DTT, 1 µl [α-^{35}S]dATP (3000Ci/mmol), 2 µl labelling mix (1:5 diluted in dH_2O)and 2 µl T_7 polymerase (1:8 diluted in TE) was added. The labelling reactions continued for 2 minutes after which 3.5 µl of the reaction mixture was added to 2.5 µl of each termination mix (ie G,A,T,C) in a microtitre plate. The entire plate was incubated at 37°C for 10 minutes. Stop buffer dye (4 µl) was mixed into each well to stop the reactions. After boiling for 4 minutes, 2 µl of each mixture was loaded onto a 6% polyacrylamide sequencing gel (preheated for 30 minutes).

B. Direct sequencing from PCR DNA

After asymmetric amplification, PCR products were extracted with chloroform and precipitated with isopropanol. The DNA was resuspended in 9 µl of water, 2 µl of which was analyzed by electrophoresis on a 1.0% agarose gel. The remaining 7 µl was used directly for sequencing. The single stranded sequencing protocol differed slightly from above in that the DNA was incubated with 1 µl of primer (20 pmole) and 2 µl of 5x sequencing primer at 37°C for 15 minutes. The rest of the protocol remained basically the same, except that the labelling step was omitted and only 3.0 µl of each reaction was added to the termination mixes. Because the yield of DNA from PCR was relatively low (compared to the yield from plasmid preparations), exposure of the film required anywhere from 1 to 4 days.

X. ENZYME LINKED IMMUNO-SORBANT ASSAY

A sandwich ELISA was performed on plasma from the haemophilia B patients to determine FIX antigen levels. All reagents required were supplied in the Asserachrom IX:Ag kit and the protocol used is described in detail in the kit. Essentially, rabbit anti-FIX $F(ab')_2$ fragments were coated onto NUNC Type 1 microwell plates overnight. The FIX deficient plasma was then bound to the primary antibody via various antigenic determinants by incubating the antigen and antibody together for 2 hours. A secondary antibody, anti-FIX antibody coupled to peroxidase, forms a sandwich by binding to the remaining epitopes of FIX. Both antibodies used in the ELISA were polyclonal. The substrate, ortho-phenylenediamine (OPD), in the presence of hydrogen peroxide was incubated with the samples for 3 minutes. Colour development was terminated with 3M sulphuric acid and subsequently measured on a microtitre plate reader. A control curve was constructed using normal plasma obtained from Asserachrom (LOT NUMBER 9C073, FIX level: 90%). A 1:50 stock dilution (by definition 100%) was made and the dilution was adapted

so as to bring it to a level of 100%. From this stock, 0, 5, 10, 25, 50 and 100% plasma solutions were made. The test plasma was diluted to 1:10 and 1:5.

RESULTS

I. ACTIVITY AND ANTIGEN STATUS

Based on ELISA results of plasma samples from each of the patients, FIX antigen levels were determined. In addition, FIX coagulant activities were assessed using standard activated partial thromboplastin time (APTT) assays (results summarized in Table 2). FIX Edmonton 1, FIX Edmonton 2 and FIX Leamington all resulted in severe CRM^- haemophilia B. In each case, less than 1% coagulant activity was detected (as compared to that found in normal plasma). As well, there were no detectable amounts of FIX antigen. FIX Creston also exhibited significantly reduced FIX activity (~2% activity) but had normal levels of FIX antigen and was therefore classified as moderately severe CRM^+ haemophilia B. FIX Brantford and FIX Edmonton 3 were both cases of mild haemophilia exhibiting ~5% and ~15% coagulant activity respectively. Plasma samples from these two cases were not available and therefore their antigen status is unknown.

II. PCR AMPLIFICATION OF THE FIX GENE

All eight exons and the putative promoter region of the FIX gene from each of the six haemophilia patients were successfully amplified using the polymerase chain reaction. The amplified fragments were of the predicted lengths indicating that there were neither any gross deletions nor insertions within the genes.

The efficiency of PCR amplification varied for each of the fragments owing to the nature of the flanking oligonucleotides. In all cases sufficient quantities of DNA were produced for subsequent subcloning. Asymmetric PCR amplification for production of single stranded DNA was significantly less efficient than symmetric amplification. Again, however, it was found to be sufficient for use in direct sequencing of the PCR product.

Table 2. Summary of results

Patient	Nt. (a)	A.A. (a)	% Activity	% Antigen (b)	Comments
$FIX_{Leamington}$	G to C 17,756	Gly to Ala 114	<1	<1	-highly conserved glycine thought to be critical in type II b turn
$FIX_{Brantford}$	A to C 31,258	Leu to Phe 379	~5	N.D.	-introduction of bulky side-chain proximal to active site
$FIX_{Edmonton1}$	C to A 17,700	Cys to stop 95	<1	<1	-truncated protein lacking the catalytic domain
$FIX_{Edmonton2}$	C to T 31,1333	Arg to stop 338	~1	<1	-truncated protein lacking Asp of the catalytic triad
$FIX_{Edmonton3}$	G to A 30,150	Ala to Thr 233	~15	N.D.	-introduction of hydrophilic side chain into hydrophobic pocket in catalytic domain
$FIX_{Creston}$	G to C 20,519	Arg to Pro 180	<2	~100	-abolishes FXI_a cleavage site in activation peptide

(a) Numbering according to Yoshitake et al., 1985
(b) N.D. = not determined

III. DNA SEQUENCE ANALYSIS

The putative promoter and coding regions of the FIX gene, including the regions of the intron/exon splice junctions, from each of the haemophiliacs were sequenced. The DNA sequence obtained was compared to that of the wild type FIX gene (Yoshitake et al, 1985) (results summarized in Table 2).

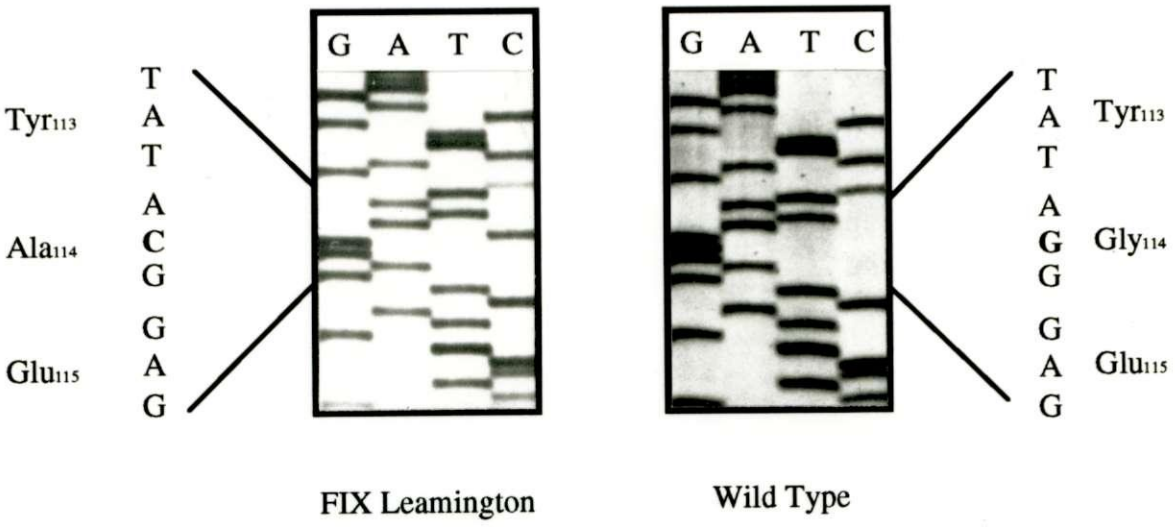
A C to A transversion was detected at nucleotide 17,700 in FIX Edmonton 1 which is predicted to convert the cysteine$_{95}$ codon to a stop codon. In FIX Leamington, a G to C transversion found at nucleotide 17,756 is predicted to alter the wild type coding sequence from glycine$_{114}$ to alanine. In FIX Creston, a G to C transversion found at nucleotide 20,519 is predicted to mutate arginine$_{180}$ to proline. A G to A transition was detected in FIX Edmonton 3 at nucleotide 30,150. The predicted coding sequence is changed from alanine to threonine at residue 233. In FIX Edmonton 2, a C to T transition was found at nucleotide 31,133 predicting the introduction of a premature stop codon in place of arginine$_{338}$. In FIX Brantford, an A to C transversion was found at nucleotide 31,258 predicting the mutation of leucine$_{379}$ to phenylalanine. Other than the single base pair mutation identified in each of the FIX genes, no other sequence discrepancies were present in the remainder of the gene analyzed. This was taken to indicate that the mutation found was probably the causative factor for haemophilia in each of the cases.

Autoradiograms showing the mutations found in this study are depicted in Figures 7 and 8.

Figure 7. A base pair mutation in exon 5 of FIX Leamington

Exon 5 of the FIX gene from FIX Leamington was analyzed using the chain termination method of DNA sequencing. A G was observed at nucleotide 17,756 (numbering according to Yoshitake et al., 1985) on the sense strand of DNA from FIX Leamington as compared to a C observed in normal FIX sequence (panel A). This base pair alteration results in the substitution of alanine for glycine$_{114}$. The corresponding mutation is also shown on the antisense strand of DNA (panel B).

A


G A T C
G A T C
Tyr113
Ala114
Glu115
T A T A C G G A G
T A T A G G G A G
Tyr113
Gly114
Glu115
FIX Leamington
Wild Type

B

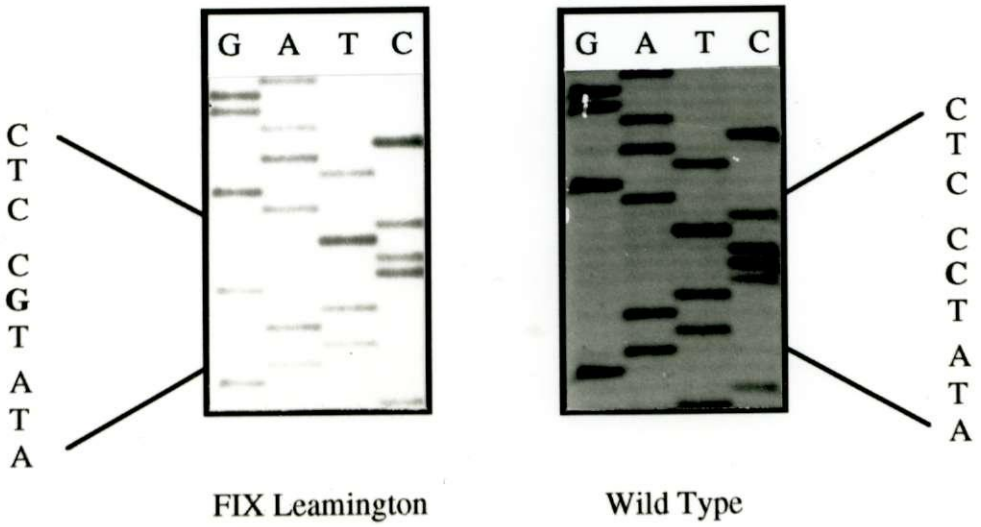
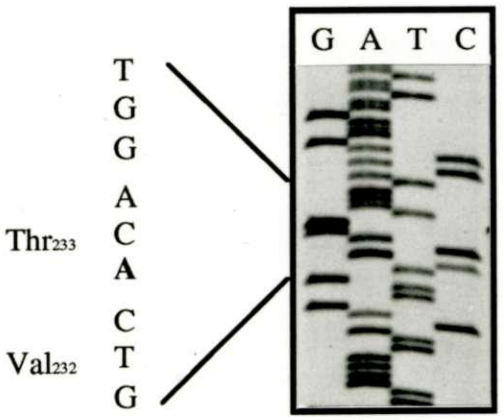
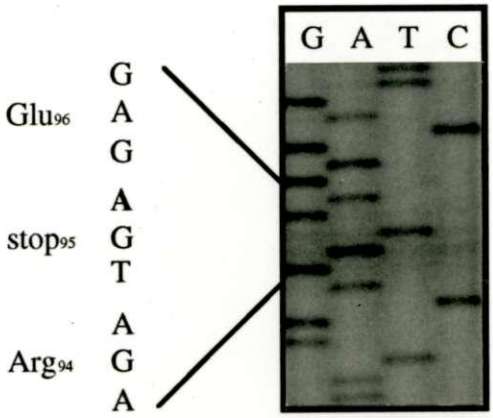

G A T C
G A T C
C T C C G T A T A
C T C C C T A T A
FIX Leamington
Wild Type

Figure 8. Base pair alterations in FIX Edmonton 1, FIX Edmonton 2, FIX Edmonton 3, FIXCreston and FIX Brantford

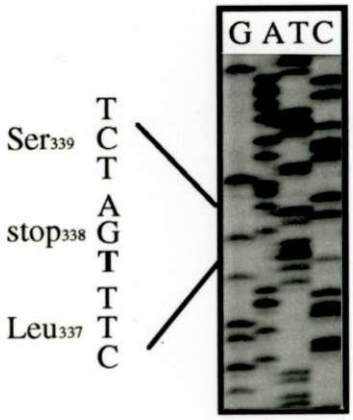
Autoradiograms showing the sequence alterations in the FIX genes of FIX Edmonton 1, FIX Edmonton 2, FIX Edmonton 3, FIX Creston and FIX Brantford. DNA sequencing was performed on both sense and antisense strands of DNA but only the sense strand is shown. The nucleotide base shown in bold is the one which is mutated from wild type.



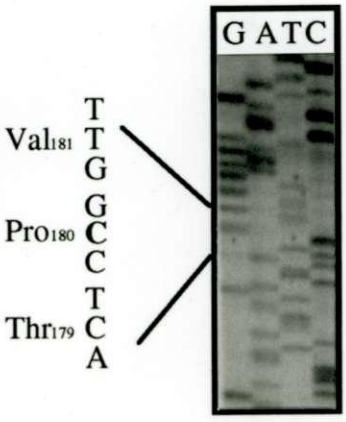
FIX Edmonton 3



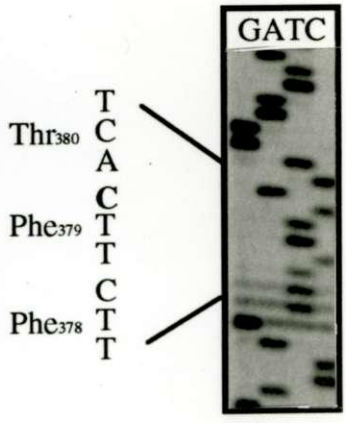
FIX Edmonton 1



FIX Edmonton 2



FIX Creston



FIX Brantford

IV. SSCP ANALYSIS

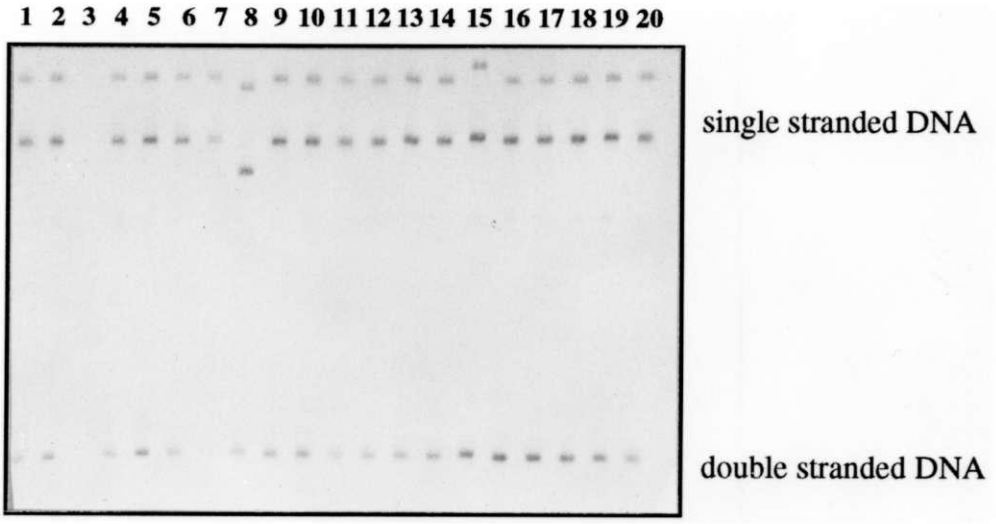
A. Detection of mutations in the propositus

All of the mutations described in this study were detected using SSCP analysis. Figure 9 shows a picture of the SSCP autoradiogram for exon 5 of the FIX gene. Lanes showing mobility shifts in one or both single stranded DNA bands indicate mutations in those DNA samples. All other lanes indicate normal wild type exon 5 sequence.

B. Carrier Detection and Pedigree Analysis

Southern blotting analysis using restriction fragment length polymorphisms (RFLPs) was one method used for carrier detection and pedigree analysis. SSCP analysis was also used for this purpose. This latter technique is always informative as the detection is dependent directly on the presence of the mutation itself as opposed to linkage to a polymorphic site elsewhere. Figure 10 (Panel A) shows autoradiograms of both SSCP and Southern blotting analysis of a haemophiliac family. From the banding pattern in SSCP anlaysis, FIX-47, FIX-50, FIX-54 and FIX-56 (lanes 1,4,5, and7 respectively) all carry the mutant allele for haemophilia. FIX-48 and FIX-53 (lanes 2 and 8 respectively) carry the unaffected allele. FIX-55 (lane 6) carries both wild type FIX alleles while FIX49 (lane 3) is a carrier for haemophilia. FIX-53 represents a normal or control sample as he is an unaffected male related to the family only by marriage. The same information is given in the Southern analysis in a slightly different form (Figure 10, Panel B). The mutant allele is represented by the 6.3 kb band while the normal allele is represented by the 8.0 kb band. The propositus, FIX-50, is FIX Leamington in our study.

Figure 9. SSCP anlaysis of exon 5 from 20 factor IX genes

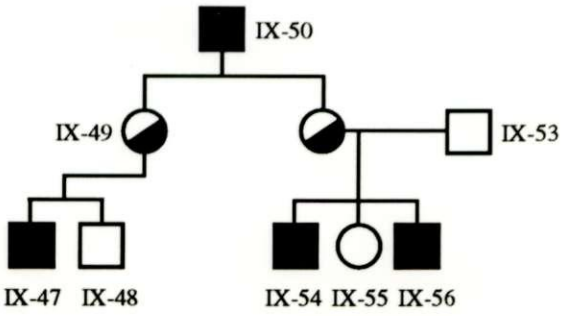


Exon 5 from the FIX gene of 20 different haemophilia B patients was amplified in the presence of α-[thio-^{35}S]-dATP using the polymerase chain reaction. Aliquots of the DNA were electrophoresed on a 6% non-denaturing polyacrylamide gel containing 10% glycerol (lanes 1-20). Mobility shifts in the bands representing single stranded DNA in lane 8 (FIX Leamington) and lane 15 (FIX Edmonton 1) indicate the presence of mutations in both these exons. All other lanes represent wild type sequence.

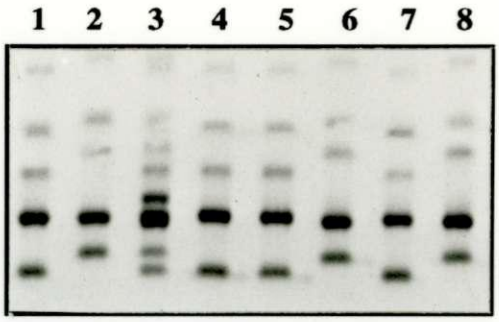
Figure 10. Pedigree analysis of a haemophilia B family

A. The family tree of FIX Leamington (FIX-50) is shown. Squares denote males and circles denote females. Black squares/circles represent haemophiliacs while white squares/circles represent non-haemophiliacs. Half filled in circles represent carriers. B. SSCP analysis was performed on relatives of FIX Leamington. The presence of the first, third, fifth and ninth bands (from top to bottom) indicated unaffected family members, as in lanes 2, 6 and 8. The presence of the second, fourth, sixth and tenth bands indicated affected males, as in lanes 1, 4, 5 and 7. Lane 3 contained all bands indicating that she was a carrier of the FIX Leamington mutation. The eighth band represents double stranded DNA and is present in all lanes. C. Southern blotting analysis was also performed on DNA from the same family. The same conclusions as above were arrived at. Presence of the 8.0 kb band indicated affected males. Presence of the 6.3 kb band indicated unaffected persons. Presence of both bands indicated a female carrier.

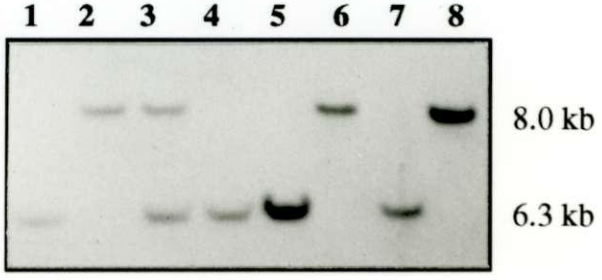
A


IX-50
IX-49
IX-53
IX-47
IX-48
IX-54
IX-55
IX-56

B


1 2 3 4 5 6 7 8

C


1 2 3 4 5 6 7 8
8.0 kb
6.3 kb

DISCUSSION

I. STRUCTURE/FUNCTION ANALYSIS OF FIX MUTANTS

To date, attempts to obtain crystallographic data on the blood clotting factors have been largely unsuccessful. Therefore, little is known about the three dimensional structures of these proteins and their various domains. However, as some of these domains share homology with proteins of known structure, models have been constructed which attempt to approximate their actual three dimensional conformations. For example, the catalytic regions of the coagulation serine proteases share a high degree of amino acid sequence identity with the pancreatic serine proteases. The crystal structures of trypsin, α-chymotrypsin and elastase have all been determined and when compared, the α-carbon backbones of these three proteins were discovered to possess a high degree of similarity. Furie and Furie (1982) used these structures to construct computer models of the protease regions of FIX, FX and thrombin. A similar approach was taken by Geddes et al.(1987) and Cool et al. (1987) to model the structures of the catalytic regions of factors IXa and XIIa, respectively. Also, Cooke et al. (1987) solved the NMR solution structure of human EGF and this was initially used to model the first EGF-like domain of FIX. Since then, Huang et al. (1991) have solved the NMR solution structure of the first EGF-like domain of FIX. Selander et al (1990) have done the same for the first EGF-like domain of FX. From such models, a prediction can be made as to what kind of effect a certain mutation will have on the structure and thus on the function of the protein. However, these predictions can only be substantiated by performing functional assays.

A. FIX mutations causing CRM$^-$ haemophilia B

Cross reacting material negative (CRM$^-$) haemophilia refers to the absence of any appreciable amount of FIX antigen in the plasma, as detected by immunological means.

Mutations resulting in a lack of protein synthesis generally disrupt the transcription process and have been found in the promoter region. Alternatively, protein may be produced but not secreted from the hepatocytes. Mutations interfering with protein secretion have been localized to the signal peptide, the region which is responsible for directing the export of proteins from the cell. Proteins may also fail to be secreted if they are improperly folded. Cells possess the ability to recognize aberrant protein structures and rapidly hydrolyze them. This is an important defence mechanism since the buildup of abnormal polypeptides is potentially harmful to the cell. Mutations which disrupt protein structure have been found throughout the FIX gene. Although CRM^- mutations aid in the study of FIX folding and structure, they are not useful in identifing those residues critical in the functional aspects of FIX.

B. FIX mutations causing CRM^+ haemophilia B

Cross reacting material positive (CRM^+) haemophilia B refers to the subgroup in which FIX antigen is present in the plasma but is not fully functional. In order to participate in the coagulation cascade, FIX must fulfill many roles: activation by FXIa and FVIIa, calcium mediated binding to phospholipid membranes, binding FVIII, binding FX, and catalyzing the conversion of FX to FXa. Impairment of any of these functions will effectively result in a decrease in clotting activity. Therefore CRM^+ mutations have been located throughout the FIX gene. They may also affect protein conformation and stability such that the half life in plasma may be reduced but not to the extent that the FIX is completely degraded within the cell.

II. ANALYSIS OF FIX MUTATIONS FOUND IN THIS STUDY

A. Nonsense mutations

1. Factor IX Edmonton 1 and Factor IX Edmonton 2

FIX Edmonton 1 and FIX Edmonton 2 involve mutations which introduce premature stop codons at amino acid residues 95 and 338 respectively, and thus produce truncated polypeptides. Both cases represent severe CRM^- haemophilia with less than 1% coagulant activity and FIX antigen.

In FIX Edmonton 1, the polypeptide is predicted to terminate near the beginning of the second EGF-like domain. The wild type cysteine residue would normally be involved in disulfide bridging with $cysteine_{110}$ which is not present in the mutant FIX. Also, $cysteine_{88}$ can no longer form a disulfide bond with $cysteine_{99}$ in the mutant FIX. The result is a ten residue stretch with no ordered secondary structure and which probably acts as a signal for degradation. The truncated product also lacks most of the second EGF-like domain as well as the entire activation peptide and catalytic domain. Therefore, even if present in the plasma, the FIX Edmonton 1 polypeptide would have no activity.

The mutation present in FIX Edmonton 2 has been documented previously in other families by Ludwig et al. (1989), Driscoll et al. (1989), Green et al. (1990), and Freedenberg et al. (1989). The stop codon is introduced fairly late in the protein, in the catalytic domain. Absence of the terminal 77 residues probably exposes hydrophobic residues normally found in the interior of the protein. Exposure of the hydrophobic core to solvent likely results in gross denaturation of this domain and thus the inability to form its native structure. As with FIX Edmonton 1, this is somewhat irrelevant as the protein is terminated prior to Ser_{365} of the catalytic triad which is absolutely required for enzymic activity. Thus, even if the truncated protein were able to maintain its proper structure, it would still be unable to catalyze the activation of FX.

B. Missense mutations

1. Factor IX Leamington

Unlike the two examples given above, severe CRM^- haemophilia in FIX Leamington is caused by a missense mutation which alters a single amino acid residue; $glycine_{114}$ was mutated to alanine. The integrity of the remainder of the protein is, however, maintained. $Glycine_{114}$ is found in the C terminal EGF-like domain. As mentioned before, the most striking feature about EGF and EGF-like domains is the presence of six cysteine residues which form three disulfide bridges. Typically, the first and third, second and fourth and the fifth and sixth cysteines in the sequence disulfide bond with each other thus creating three closed loops. As these domains are not very large (~40 residues long), the disulfide bonding causes their structures to be highly constrained. From solution NMR data of hEGF and the first EGF-like domain of FIX and FX, other significant structural features which have been identified are a major antiparallel β sheet and a minor antiparallel β sheet. At least three β-type turns have also been identified. One turn connects the two strands which make up the major sheet. A second turn, designated a type II β turn, directly follows the fifth cysteine. A third turn allowing formation of the third disulfide bridge and the minor β sheet was also found (Figure 11).

Alignment of the sequences from hEGF, the first EGF-like domains of FIX and FX, and the second EGF-like domains of human protein C, FVII, FIX and FX shows $glycine_{114}$ of FIX to be absolutely conserved (Figure 12). The analogous residue in hEGF and FX forms the third residue in a type II β turn. β turns are common structural features, consisting of four consecutive residues, that interconnect or interupt various secondary structural elements. They are mainly concentrated near the surfaces of proteins and are thus

Figure 11. Structures of hEGF and the EGF-like Domains of FIX and FX

Cartoon representations are given of the structures of human epidermal growth factor (panel A), the first EGF like domain of FX (panel B) and the second EGF-like domain of FIX (panel C). Amino acid sequence alignment of these three peptides shows that the residue indicated by the darker circle is glycine in all three cases. Solution NMR structures of hEGF and the first EGF-like domain of FX indicates that this glycine residue is involved in a type II β turn. Disulfide bonds are indicated by dark lines. β is β–hydroxyaspartate.

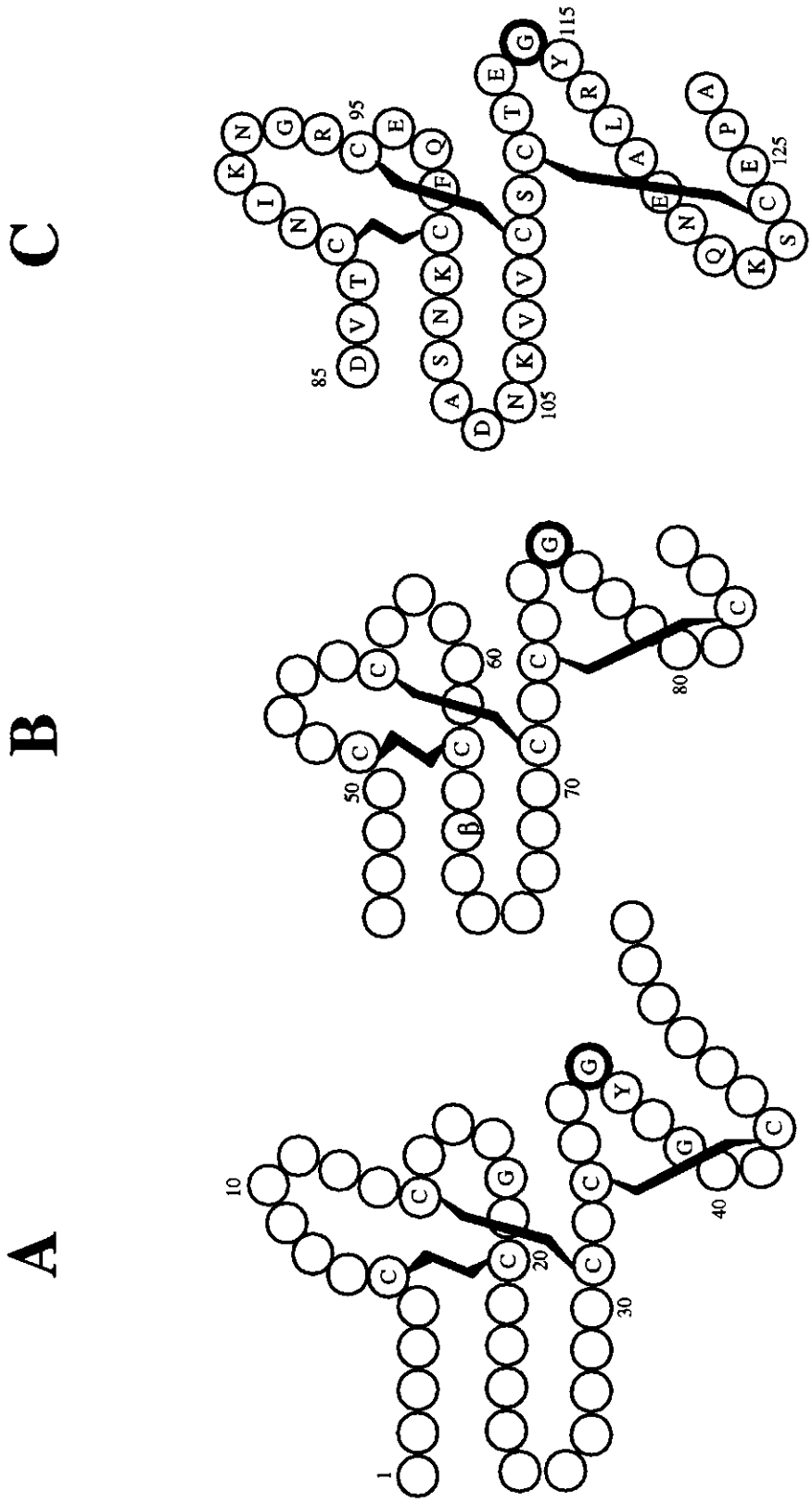
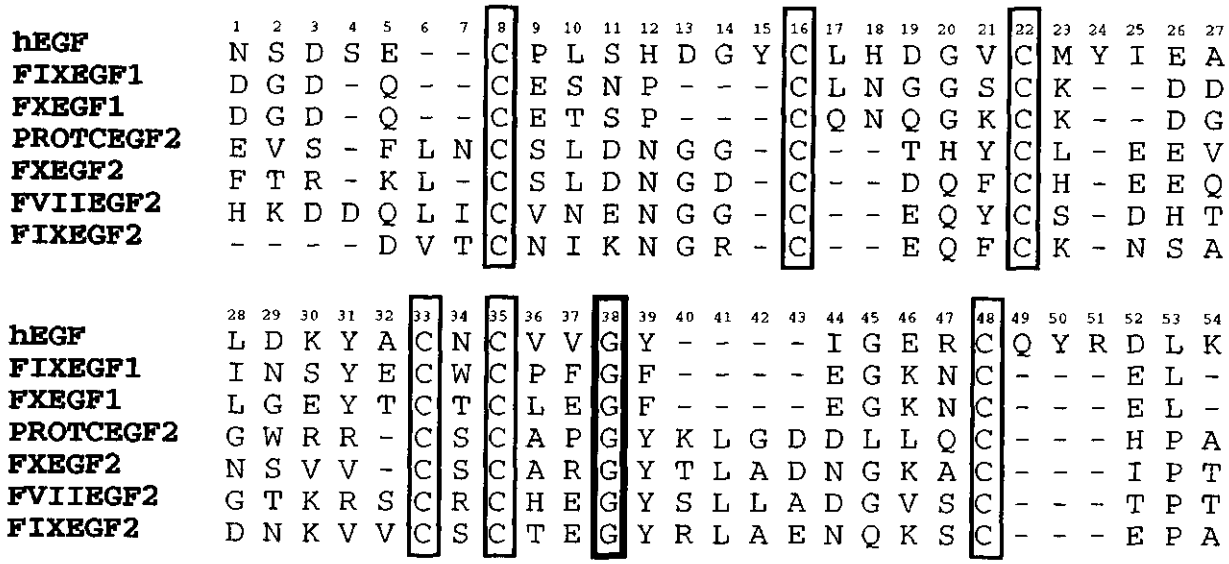

A
B
C
1
10
20
30
40
50
60
70
80
85
95
105
115
125

Figure 12. Amino acid sequence alignment of hEGF and EGF-like peptides



The amino acid sequences of hEGF (Gregory, 1975), the first EGF-like domains of FIX (Yoshitake et al., 1985), and FX (Leytus et al., 1986; Fung et al., 1988) and the second EGF-like domains of protein C (Foster and Davie, 1984; Foster et al., 1985), FX(Leytus et al., 1986), FVII (O'Hara et al., 1987) and FIX (Yoshitake et al., 1985) are shown. The numbering system used here is based on the sequence for hEGF. The six conserved cysteines are shown in thin boxes. The glycine mutated in FIX Leaminton is conserved amongst all 7 peptides and is shown in the thick box. Deletions are indicated by dashes.

comprised predominantly of hydrophilic residues. Wilmot et al. (1988) compared 735 β turns from 59 different proteins and classified them into 7 turn types according to their φψ angles. Type II turns are relatively flat with their central peptide lying in a plane perpendicular to the plane of the turn. Wilmot et al. found no significant sequence preference at the first position of type II turns. Proline was most favoured to occupy the second position while amino acids with branched C^{β} side chains (Thr, Val and Ile) were unfavourable due to steric hindrance. Glycine was overwhelmingly preferred (73%) at position three of the turn as it most readily adopts the optimal φψ angles. Amino acid residues with larger side chains were found to be sterically unfavourable. Of the 101 type II β turns analyzed, only one contained an alanine at position three as is found in FIX Leamington. Alanine introduces a slightly bulkier side chain (ie. methyl group in place of a hydrogen). Also, alanine is a more hydrophobic residue which, as mentioned already, is generally unfavourable. One or both of these factors may disrupt the turn such that the turn cannot be achieved or that the turn is not as tight as it normally would be. The constraints on the structure of this domain imposed by the disulfide bridges might be expected to minimize the significance of the such effects. However, it is possible that the disruption in structure caused by the mutation may also interfere with disulfide bridge formation.

In protein synthesis, the formation of secondary structures is followed by folding of the polypeptide into its tertiary structure through hydrophobic interactions. Through this process, cysteine residues are brought into close proximity for disulfide bond formation. In order for proper bonding of two cysteine residues, they must be in the proper distance from and geometrical relationship to each other (Watanabe et al., 1991). Substitution of alanine for glycine may change one or both of these parameters possibly resulting in the loss of a disulfide bridge between cysteine_{111} and cysteine_{124}. This would have a much more drastic effect than simply disruption of the β turn and may in fact be the signal for degradation. It is interesting to note, however, that the analogous residue in the first EGF-

like domain of wild type FVII is an alanine. Why alanine is compatible in the FVII polypeptide but not in FIX is not clear.

2. Factor IX Creston

Normal levels of FIX are present in FIX Creston but almost no enzymatic activity is associated with the protein. Mutation of the wild type $arginine_{180}$ to proline results in loss of the second cleavage site required for activation of FIX by FXIa and FVIIa. Similar families have been reported in which $arginine_{180}$ has been mutated to tryptophan (Suehiro et al, 1989; Bertina et al, 1989; Ludwig et al) or to glutamine (Huang et al, 1989; Vidaud, 1990; Bertina et al, 1989; Ludwig et al). In these cases, normal levels of FIX were also recorded with <1% associated coagulant activity. As well, they have been found to be examples of the Haemophilia B_m variant so it is likely that FIX Creston is also.

Residue 180 likely plays a non-critical role in the structure of FIX as shown by the lack of degradation when mutated to tryptophan, glutamine and proline.

3. Factor IX Edmonton 3

In FIX Edmonton 3, the $alanine_{233}$ to threonine mutation causes very mild haemophilia with ~15% activity. Although antigen levels have not been obtained, at least 15% of normal FIX levels, and possibly even more, must be present in the plasma. $Alanine_{233}$ resides in the catalytic domain of FIX. In order to determine the potential effects of the mutation on FIX activity, a model of the FIX serine protease domain has been constructed based upon the known X-ray crytallographic structures of α-chymotrypsin, trypsin and elastase. Compared to one another, the pancreatic serine proteases share ~41% amino acid sequence identity. Even more striking, when their backbone chains are superimposed, ~85% of the α-carbon atoms are topologically equivalent (James et al., 1978). As FIX shares ~36% amino acid sequence similarity with these pancreatic enzymes

(Figure 13), it is reasonable to assume that a model based on them would also be a valid approximation for the structure of FIX.

Furie et al. (1982) created computer generated models for FXa, FIXa and thrombin. Comparing the amino acid sequences, they found that generally the cores of the proteins were highly conserved while regions located on the surface of the proteins exhibited greater variation. The blood clotting factors are much more highly specific for their protein substrates as compared to the pancreatic peptidases. Since all serine proteases work via the same catalytic mechanism, it is these variable surface areas that are thought to be responsible for substrate specificity via protein-protein interactions. Furie et al. identified seven constant regions making up the interior of the protein and six variable regions found mainly on the surface of the protein.

Alanine$_{233}$(68) is found in one of these variable regions (the numbering is consistent with that of the sequence for FIX with the corresponding α-chymotrypsin numbering in brackets). From computer models it can be seen that the wild type residue normally sits just beneath the surface of the protein and at quite a distance from the active site and substrate binding pocket. Ala$_{233}$(68) is situated in a highly hydrophobic pocket consisting of the side chains of Val$_{196}$(31), Ile$_{233}$(46), Val$_{231}$(65), Leu$_{275}$(108), Leu$_{279}$(112) and Val$_{285}$(118). FIX Edmonton 3, which contains threonine$_{233}$(68) in place of alanine, introduces a hydrophilic side chain into this hydrophobic environment which is likely to be disruptive to the local polypeptide structure (Figure 14). However, because this residue lies just beneath the surface of the molecule, the backbone chain could possibly shift so that the side chain could interact with the surrounding solvent. Also, judging from the computer model, there are no obvious favourable hydrogen bonding partners with the hydroxyl group of threonine. Spatially, the substitution of threonine for alanine poses no steric problems. Because the antigen levels of FIX Edmonton 3 are not known, it is not possible to say whether the mutation causes instability of the protein. Alternatively, this region of the catalytic domain may be involved in FVIII binding as it is relatively far away

Figure 13. Amino acid sequence alignment of the catalytic domains of trypsin, chymotrypsin, elastase and factor IX

The amino acid sequences of bovine trypsin (Marquart et al., 1983), bovine chymotrypsin (Cohen et al., 1981), porcine elastase (Sawyer et al., 1978), and human FIX (Yoshitake et al. 1985) are shown. The numbering system used is based on that of chymotrypsinogen A (Hartley et al., 1966) with insertions in the sequences of related enzymes denoted by letters (eg. 36A, 36B). Residues conserved amongst all 4 peptides are boxed. Deletions are indicated by dashes. In FIX Edmonton 3, Ala_{233}(68) is mutated to Thr. In FIX Brantford, Leu_{379}(209) is mutated to Phe.

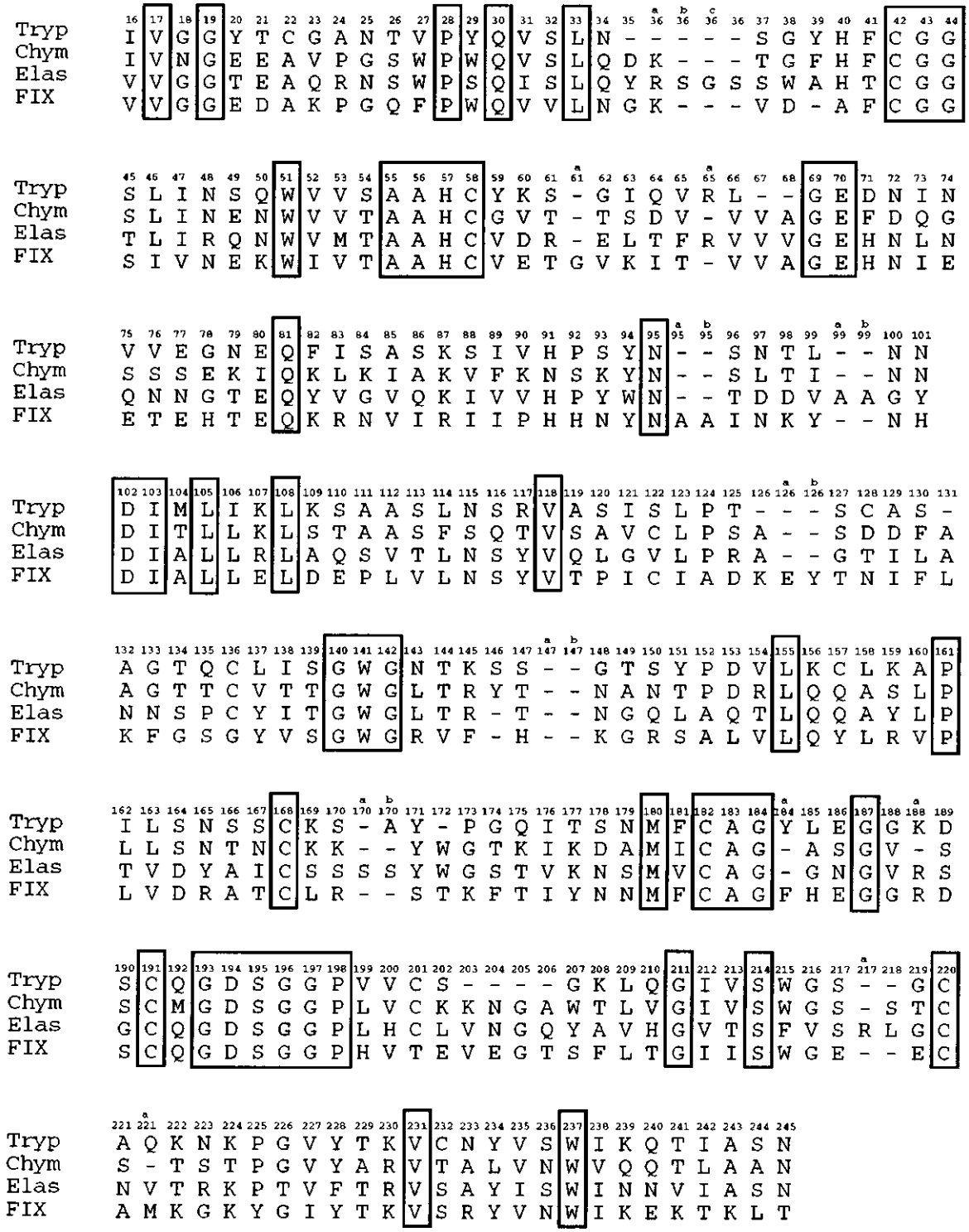
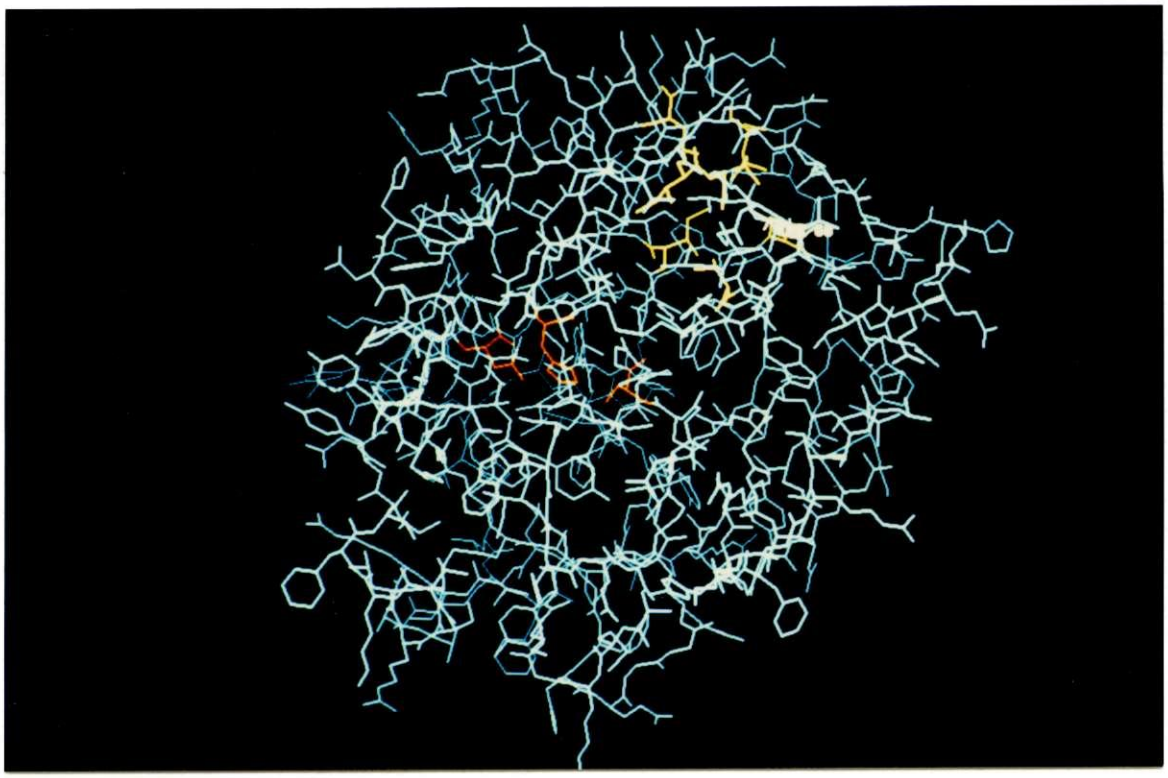

Tryp I V G G Y T C G A N T V P Y Q V S L N - - - - - S G Y H F C G G
Chym I V N G E E A V P G S W P W Q V S L Q D K - - - T G F H F C G G
Elas V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H T C G G
FIX V V G G E D A K P G Q F P W Q V V L N G K - - - V D - A F C G G
Tryp S L I N S Q W V V S A A H C Y K S - G I Q V R L - - G E D N I N
Chym S L I N E N W V V T A A H C G V T - T S D V - V V A G E F D Q G
Elas T L I R Q N W V M T A A H C V D R - E L T F R V V V G E H N L N
FIX S I V N E K W I V T A A H C V E T G V K I T - V V A G E H N I E
Tryp V V E G N E Q F I S A S K S I V H P S Y N - - S N T L - - N N
Chym S S S E K I Q K L K I A K V F K N S K Y N - - S L T I - - N N
Elas Q N N G T E Q Y V G V Q K I V V H P Y W N - - T D D V A A G Y
FIX E T E H T E Q K R N V I R I I P H H N Y N A A I N K Y - - N H
Tryp D I M L I K L K S A A S L N S R V A S I S L P T - - - S C A S -
Chym D I T L L K L S T A A S F S Q T V S A V C L P S A - - S D D F A
Elas D I A L L R L A Q S V T L N S Y V Q L G V L P R A - - G T I L A
FIX D I A L L E L D E P L V L N S Y V T P I C I A D K E Y T N I F L
Tryp A G T Q C L I S G W G N T K S S - - G T S Y P D V L K C L K A P
Chym A G T T C V T T G W G L T R Y T - - N A N T P D R L Q Q A S L P
Elas N N S P C Y I T G W G L T R - T - - N G Q L A Q T L Q Q A Y L P
FIX K F G S G Y V S G W G R V F - H - - K G R S A L V L Q Y L R V P
Tryp I L S N S S C K S - A Y - P G Q I T S N M F C A G Y L E G G K D
Chym L L S N T N C K K - - Y W G T K I K D A M I C A G - A S G V - S
Elas T V D Y A I C S S S S Y W G S T V K N S M V C A G - G N G V R S
FIX L V D R A T C L R - - S T K F T I Y N N M F C A G F H E G G R D
Tryp S C Q G D S G G P V V C S - - - - G K L Q G I V S W G S - - G C
Chym S C M G D S G G P L V C K K N G A W T L V G I V S W G S - S T C
Elas G C Q G D S G G P L H C L V N G Q Y A V H G V T S F V S R L G C
FIX S C Q G D S G G P H V T E V E G T S F L T G I I S W G E - - E C
Tryp A Q K N K P G V Y T K V C N Y V S W I K Q T I A S N
Chym S - T S T P G V Y A R V T A L V N W V Q Q T L A A N
Elas N V T R K P T V F T R V S A Y I S W I N N V I A S N
FIX A M K G K Y G I Y T K V S R Y V N W I K E K T K L T

Figure 14. Computer modelling of the mutation causing FIX Edmonton 3

Photographs were taken of the computer screen showing a model of the FIXa catalytic region based upon crystallographic data from the digestive pancreatic proteases. Panel A shows the entire catalytic domain. The catalytic residues are indicated by red lines. The mutant $Thr_{233}(68)$ is indicated by white lines. The hydrophobic pocket in which $Thr_{233}(68)$ is situated is indicated by yellow lines. The rest of the molecule is indicated by green lines. Panel B shows a close up of $Thr_{233}(68)$ and the surrounding hydrophobic residues. Intramolecular distances to the two closest residues are indicated by dotted lines and are measured in Angstroms.




ILE 48
LEU 100
LEU 112
VAL 110
2.39
C
2.64
3.04
GLN 81
VAL 31

from the active site. Impairment of FVIII binding would lower the apparent coagulant activity of FIX.

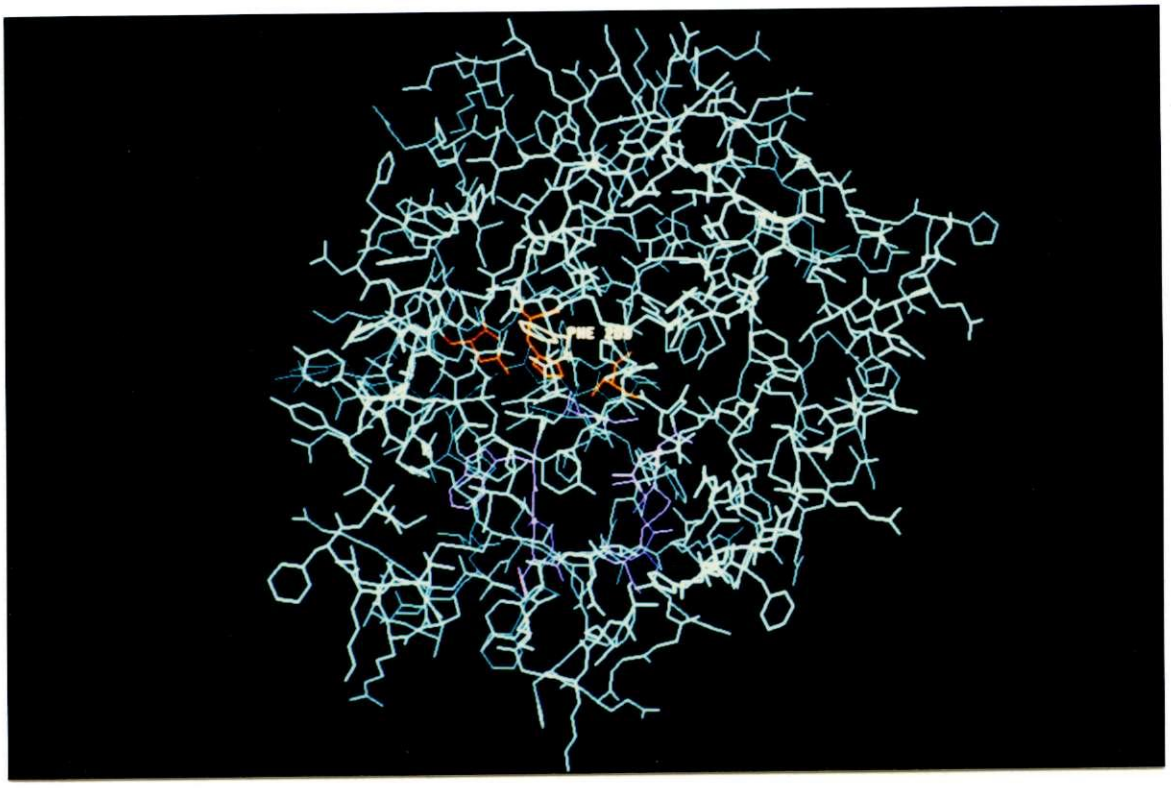
4. Factor IX Brantford

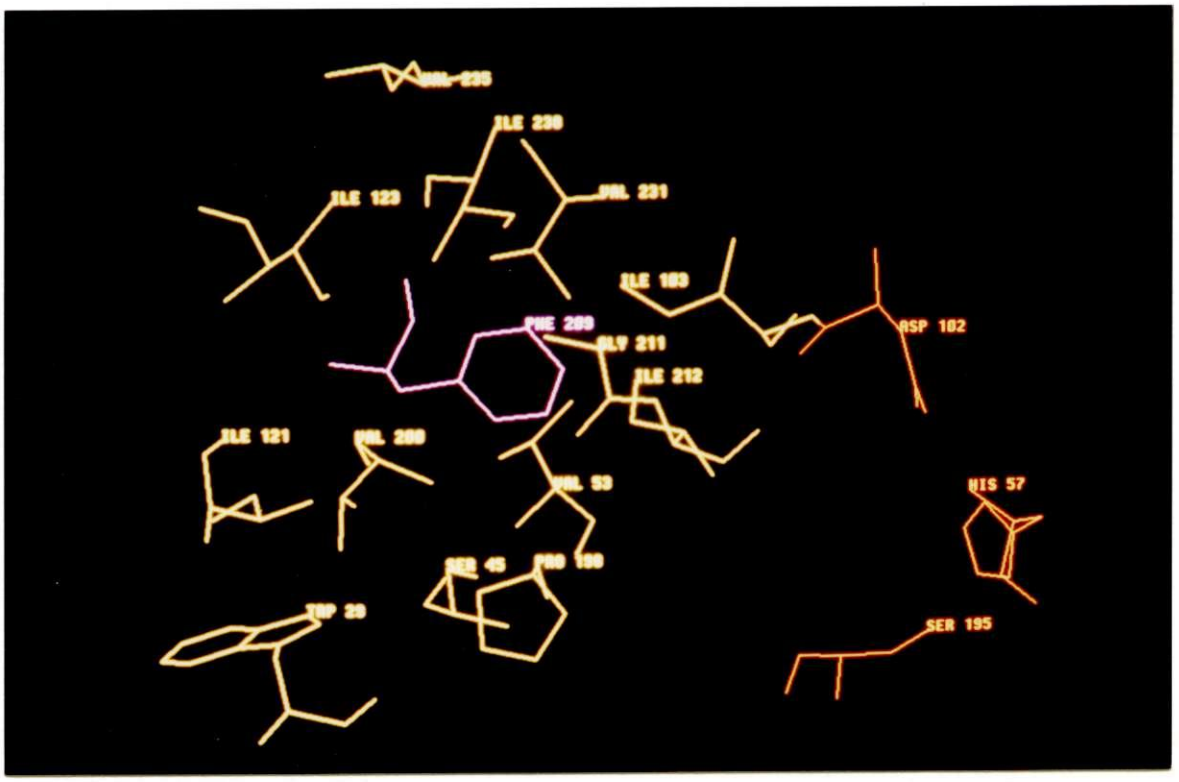
A leucine$_{379}$(209) to phenylalanine mutation in the catalytic domain of FIX Brantford results in mild haemophilia with 5% activity. Again, the FIX levels have not been determined but at least a corresponding 5% antigen must be present. The wild type residue is found in the interior of the protein fairly proximal to the active site and substrate binding pocket (~10A). Steitz et al. have identified several amino acid residues in α-chymotrypsin which are involved with substrate binding including Ser$_{189}$, Ser$_{190}$, Cys$_{191}$, Met$_{192}$, Val$_{213}$, Trp$_{215}$, Gly$_{216}$, Ser$_{217}$, Cys$_{220}$, and Gly$_{226}$.

Although leucine and phenylalanine are both neutral hydrophobic residues, the mutation is nevertheless a significant one structurally. Whereas leucine has a flexible side chain, phenylalanine must maintain a fairly restricted conformation; the χ_1 torsional angles between the C_α and C_β is more restricted than leucine. As well, the aromatic ring must maintain a rigid planar conformation. From computer modelling, it is apparent that the substitution of phenylalanine for leucine creates serious steric conflicts (Figure 15). The hydrophobic cores of proteins are usually tightly packed, therefore shifting of the surrounding residues to accomodate phenylalanine may indirectly affect residues a distance away. In this fashion, the position of those residues involved in catalysis (ie catalytic triad and substrate binding) may be altered. As in the case of FIX Edmonton 3, it is not known whether this mutation decreases coagulant activity by destabilizing the protein structure leading to degradation or by decreasing the optimum orientation of residues for catalysis.

Figure 15. Computer modelling of the mutation causing FIX Brantford

Photographs were taken of the computer screen showing a model of the FIXa catalytic region based upon crystallographic data from the digestive pancreatic proteases. Panel A shows the entire catalytic domain. The catalytic residues are indicated by red lines. The mutant $Phe_{379}(209)$ is indicated by white lines. Some of the residues which possibly make up the substrate binding pocket are indicated by purple lines. The rest of the molecule is shown in green. Panel B shows a close-up of the region surrounding $Phe_{379}(209)$ and its proximity to the catalytic triad.




ILE 230
ILE 123
VAL 231
ILE 189
PHE 209
ASP 102
GLY 211
ILE 212
ILE 121
VAL 209
VAL 93
HIS 57
SER 46
PRO 190
TRP 29
SER 195

III. SSCP ANALYSIS

Under the non-denaturing conditions used, at least two different conformational species of DNA were produced per sample, one corresponding to the sense strand and the other corresponding to the non-sense strand of DNA. These species of DNA were visualized as discrete bands after exposure of the gels to X-ray film. In all autoradiograms a third band corresponding to non-denatured double stranded DNA was observed. In many of the autoradiograms, even more than three bands were observed. Presumably, these additional bands represented minor or alternative conformations adopted by the single stranded DNA. The single base pair mutations described in this study were detected as mobility shifts in the bands as compared to those of a wild type sample.

Results obtained from SSCP analysis were unpredictable with regard to the number of bands produced for a given fragment. As well, the rates of mobility of the single stranded DNA were highly variable. No correlations were found between type of mutation (ie. transition or transversion) and band shifts. Electrophoresis of a given sample of DNA multiple times under near identical conditions often resulted in both the number of bands and their mobility pattern being altered. SSCP analysis was found to be very sensitive to temperature. Even slight variations of 2-3 degrees in room temperature caused significant changes in the number and/or mobility of bands. Another factor which was inconsistent between electrophoretic runs was the glycerol concentration of each gel. Because glycerol is an extremely viscous liquid, accurate measurement was difficult and consequently the amount of glycerol varied slightly from gel to gel. Although lack of reproducibility of the results was a problem, SSCP analysis was still considered to be reliable in that mutations showed up consistently if not always in precisely the same manner. Also, as this technique was neither used quantitatively to determine the position of the mutation within the

fragment nor qualitatively to identify the type of mutation, it was considered an effective and rapid method of screening for mutations.

D. FUTURE STUDIES

Further studies can be performed to prove or disprove some of the hypotheses given in the discussion. In the absence of crystallographic data on the blood clotting factors, functional assays can be performed. The mutant FIX proteins must first be obtained in sufficient amounts and purity. If present in adequate levels, FIX can be purified from plasma obtained from the patient. Alternatively, recombinant FIX can be produced in mammalian tissue culture using tissue culture cells such as baby hamster kidney (BHK) cells or chinese hamster ovary (CHO) cells. Both of these cell lines have been shown to produce biologically active FIX (Anson et al., 1985; De la Salle et al., 1985; Busby et al., 1985; Kaufman et al., 1986). Our laboratory has used the pNUT vector in BHK cells to produce recombinant prothrombin (LeBonniec et al., 1991) and recombinant transferrin (Funk et al., 1990). The FIX cDNA could be ligated into the Sma I site of the expression vector pNUT. Expression, using this vector, is driven by the metallothionine I promoter which responds to heavy metals such as zinc which would be present in the tissue culture media. The DHFR cDNA is present on the vector for selection purposes. After transfection of the vector into the cells, methotrexate is added to the media; only those BHK cells which have incorporated the vector will be able to survive. The desired mutations could then be introduced into the FIX cDNA using PCR site-directed mutagenesis. A serum free medium (ultraser G) could be used to simplify purification.

For expression of FIX involving CRM^- mutations, the protocol described above would not be a useful approach since the enzymes are most likely degraded and may not even have a chance to accumulate inside the cell. Moreover, more detailed examination of the truncated polypeptides in FIX Edmonton 1 and FIX Edmonton 2 would do little to

further what is known about the FIX structure/function relationship. In FIX Leamington however, it might be interesting to confirm the extent that alanine disrupts the type II β turn and whether disulfide bonding is actually interfered with. In this case, an expression system would have to be devised where hydrolysis of the protein does not occur, possibly by inclusion of a compound which inhibits protein breakdown. Ideally, the second EGF-like domain containing the mutation would be expressed on its own and following purification, the isolated domain could be analyzed using solution NMR.

Production of recombinant FIX would be a more appropriate method for the CRM^+ mutants. The effect of altering the FXIa cleavage site in the activation peptide of FIX has been well documented and characterized so that further studies on FIX Creston are unneccessary. Since FIX Edmonton 3 and FIX Brantford have at least 15% and 5% FIX levels in plasma respectively, they may be produced in sufficient amounts in tissue culture. Because both of these mutations occur in the catalytic domain, it would be appropriate to do kinetic studies on them. FIX Brantford is thought to directly affect the catalytic mechanism (ie. FX binding or FX activation) whereas FIX Edmonton 3 may affect activity via alteration of FVIIIa binding properties. These two scenarios may be distinguished by assaying their catalytic activities (using chromogenic substrates) and comparing them to that of wild type FIXa.

REFERENCES

Anson, D.S., Choo, K.H., Rees, D.J.G., Giannelli, F., Gould, K., Huddleston, J.A., Brownlee, G.G. (1984) *EMBO J.***3**, 1053-1060

Anson, D.S., Austen, D.E.G., Brownlee, G.G. (1985) *Nature* **315**, 683

Astermark, J., Stenflo, J. (1991) *J. Biol. Chem.* **266**, 2438-2443

Blake, C.C., Harlos, K., Holland, S.K. (1987) *Cold Spring Harbor Symposia on Quantitative Biology* **52**, 925-931

Beckmann, R.J., Schmid, R.J., Santerre, R.F., Plutzky, J., Crabtree, G.R., Long, G.L. (1985) *Nucl. Acids. Res.* **13**, 5233-5247

Borowski, M., Furie, B.C., Bauminger, S., Furie, B. (1986) *J. Biol. Chem.* **261**, 14969-14975

Boyd, Y., Buckle, V.J., Munro, E.A., Choo, K.H., Migeon, B.R., Craig, I.W. (1984) *Ann. Hum. Genet.* **48**, 145-152

Busby, S., Kumar, A., Joseph, M., Halfpap, L., Insley, M., Berkner, K., Kurachi, K., Woodbury, R. (1985) *Nature* **316**, 271

Camerino, G., Grzeschik, K.H., Jaye, M., De la Salle, H., Tolstoshev, P., Lecocq, J-P., Heilig, R., Mandel, J.L. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 498-502

Choo, K.H., Gould, K.G., Rees, D.J.G., Brownlee, G.G. (1982) *Nature* **229**, 178-180

Cohen, G.H., Silverton, E.W., Davies, D.R. (1981) *J. Mol. Biol.* **148**, 449-479

Collen, D. (1981) In *Haemostasis and Thrombosis*, Bloom, A.L. and Thomas, D.P., Eds., 225-235 (Churchill Livingstone, Edinburgh)

Cool, D.E., Edgell, C.J.S., Louie, G.V., Zoller, M.J., Brayer, G.D., MacGillivray, R.T.A. (1985) *J. Biol. Chem.* **260**, 13666-13676

Crossley, M., Brownlee, G.G. (1990) *Nature* **345**, 444-446

Crossley, P.M., Winship, P.R., Black, A., Rizza, C., Brownlee, G.G. (1989) *Lancet* **1**, 60

Davie, E.W., Fujikawa, K., Kurachi, K., Kisiel, W. (1979) *Adv. Enzymol.* **48**, 277-318

Davie, E.W., Ratnoff, O.D. (1964) *Science* **145**, 1310-1312

Degen, S.J., MacGillivray, R.T.A., Davie, E.W. (1983) *Biochemistry* **22**, 2087-2097

De la Salle, H., Altenberger, W., Elkaim, R., Dott, K., Dieterle, A., Drillien, R., Cazenave P-P., Tolstoshev, P., Lecocq, J-P. (1985) *Nature* **316**, 268

DiScipio, R.G., Davie, E.W. (1979) *Biochemistry* **18**, 899-904

Doolittle, R.F. (1979) In The *Proteins*, Neurath, H. and Hill, R.L., Eds., 1-118 (New York: Academic Press)

Doolittle, R.F. (1984) *Ann. Rev. Bioc.* **53**, 195-229

Doolittle, R.F., Feng, D.F., Johnson, M.S. (1984) *Nature* **307**, 558-560

Drakenberg, T., Fernlund, P., Roepstorff, P. Stenflo, J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1802

Enfield, D.L., Thompson, A.R. (1984) *Blood* **64**, 821

Esmon, C.T. (1987) *Science* **235**, 1348-1352

Fernlund, P., Stenflo, J. (1983) *J. Biol. Chem.* **258**, 12509-12512

Foster, D., Davie, E.W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4766-4770

Funk, W.D., MacGillivray, R.T.A., Mason, A.B., Brown, S.A., Woodworth, R.C. (1990) *Biochemistry* **29**, 1654-1660

Fung, M.R., Hay, C.W., MacGillivray, R.T.A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3591-3595

Furie, B. and Furie, B.C. (1990) *Seminars in Hematology* **27**, 270-285

Furie, B., Bing, D.H., Feldman, R.J., Robison, D.J., Burnier, J.P., Furie, B.C. (1982) *J. Biol. Chem.* **257**, 3875-3882

Furie, B., Furie, B.C. (1988) *Cell* **53**, 505-518

Green, P.M., Bentley, D.R., Mibashan, R.S., Nilsson, I.M., Giannelli, F. (1989) *EMBO J.* **8**, 1067-1072

Hagen, F.S., Gray, C.L., O'Hara, P., Grant, F.J., Saari, G.C., Woodbury, R.G., Hart, C.E., Insley, M., Kisiel, W., Kurachi, K., Davie, E.W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2412-2416

Hase, S., Nishimura, H., Kawabata, S., Iwanagas, S., Ikenaka, T. (1990) *J.Biol. Chem.* **265**, 1858-1861

Hewett-Emmett, D., Czelusniak, J., Gookman, M. (1981) *Ann. N.Y. Acad. Sci.* **370**, 511-527

Hougie, C., Twomey, J.J. (1967) *Lancet* **1**, 698

Jackson, C.M., Nemerson, Y. (1980) *Annu. Rev. Biochem.* **49**, 765-811

Jesty, J. (1978) *Arch. Biochem. Biiophys.* **185**, 165-173

Jorgensen, M.J., Cantor, A.B., Furie, B.C., Brown, C.L., Shoemaker, C.B., Furie, B. (1987) *Cell* **48**, 185-191

Kane, W.H., Davie, E.W. (1988) *Blood* **71**, 539-555

Kaufman, R.J., Wasley, L.C., Furie, B.C., Furie, B., Shoemaker, C.B. (1986) *J. Biol. Chem.* **261**, 9622

Kisiel, W. Fujikawa, K., Davie, E.W. (1977) *Biochemistry* **16**, 4189-4194

Koeberl, K.K., Bottema, C.D.K., Buerstedde, J.M., Sommer, S.S. (1989) *Am. J. Hum. Genetics* **45**, 448-457

Kurachi, K., Davie, E.W. (1982) *Proc.Natl. Acad. Sci. USA* **79**, 6461-6464

Le Bonniec, B.F., MacGillivray, R.T.A., Esmon, C.T. (1991) *J. Biol. Chem.* **266**, 13796-13803

Leytus, S.P., Chung, DW., Kisiel, W., Kurachi, K., Davie, E.W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3699-3702

Leytus, S.P., Foster, d.C., Kurachi, K., Davie, E.W. (1986) *Biochemistry* **25**, 5098-5102

Liebman, H.A., Eklund, D.M., Furie, B.C., Furie, B. (1985) *Throm. Haemost.* **54**, 226

Long, G.L., Belagaje, R.M., MacGillivray, R.T.A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5653-5656

MacFarlane, R.G. (1964) *Nature* **202**, 498-499

Magnusson, S., Sottrup-Jensen, L., Blaeys, H. (1975) In *Proteases and Biological Control*, Reich, E., Rifkin, D.B., and Shaw, E., Eds., 123-149 (Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.)

Maniatis, T., Fritsch, E.F., Sambrook, J. (1989) In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor

Mann, K.G., Nesheim, M.E., Church, W.R., Haley, P., Krishnaswamy, P. (1990) *Blood* **76**, 1-16

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., Huber, R. (1983) *Acta. Crystallograph. Sect. B. Struct. Sci.* **39**, 480

Marslarz, F., Siniscalco, M. (1985) *EMBO J.* **4**, 725-729

Masys, D.R., Bajaj, S.P., Rapaport, S.I. (1982) *Blood* **60**, 1143-1150

Nishimura, H., Kawabata, S., Kisiel, W., Hase, S., Ikenaka, T., Takao, T., Shimonishi, Y., Iwanaga, S. (1989) *J.Biol. Chem.* **264**, 20320-20325

Ny, T., Elgh, F., Lund, B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5355-5359

Orita, M., Suzuki, Y., Sekiua, T., Hayashi, K. (1989) *Genomics* **5**, 874-879

Pang, C.P., Crossley, M., Kent, G., Brownlee, G.G. (1990) *Nuc. Acids Res.* **18**, 6731

Park, C.H., Tulinsky, A. (1986) *Biochemistry* **25**, 2977-3982

Patthy, L. (1985) *Cell* **41**, 657-663

Pennica, D., Holmes, W.E., Kohr, W.J., Harkins, R.N., Vehar, G.A., Ward, C.A., Bennett, W.F., Yelverton, E., Seeburg, P.H., Heyneker, H.L., Goeddel, D.V., Collen, D. (1983) *Nature* **301**, 214-221

Peterson, T.E., Thogersen, H.C., Shorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L., Magnusson, S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 137-141

Purrello, M., Athadeff, B., Esposito, D., Szabo, P., Rocchi, M., Truett, M., Radcliffe, R., Nemerson, Y. (1976) *J. Biol. Chem.* **251**, 4797-4802

Rees, D.J.G., Jones, I.M., Hanford, P.A., Walter, S.J., Esnouf, M.P., Smith, K.J., Brownlee, G.G. (1988) *EMBO J.* **7**, 2053-2061

Reitsma, P.H., Bertina, R.M., Ploos van Amstel, J.K., Riemens, A., Briet, E. (1988) *Blood* **72**, 1074-1076

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N. (1985) *Science* **230**, 1350-1354

Sanger, F., Nicklen, S., Coulson, R.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467

Sawyer, L., Shotton, D.M., Campbell, J.W., Wendell, P.L., Muirhead, H., Watson, H.C., Diamond, R., Ladner, R.C. (1978) *J. Mol. Biol.* **118**, 137-208

Suttie, J.W. (1985) *Ann. Rev. Bioc.* **54**, 459-477

Tai, M.M., Furie B.C., Furie, B. (1984) *J. Biol. Chem.* **259**, 4102-4168

Tsang, T.C., Bentley, D.R., Mibashan, R.S., Giannelli, F. (1988) *EMBO J.* **7**, 3009-3015

Tschesche, H., Mair, G., Godec, G., Fiedler, F., Ehret, W., Hirchauer, C., Lemon, M., Fritz, H., Schmidt-Kastner, G., Kutzback, K. (1979) In *Advances in Experimental Medicine and Biology* ,Fujii, S., Moriya, H., and Suzuki, T., Eds., 245-260 (Plenum Publishing Corp., New York)

Veltkamp, J.J., Meilof, J., Remmelts, H.G., van der Vlark, D., Loeliger, E.A. (1970) *Scandinavian Journal of Haemotology* **7**, 82-90

Yoshitake, S., Schach, B.G., Foster, D.C., Davie, E.W., Kurachie, K. (1985) *Biochemistry* **24**, 3736-3750

Zur, M., Nemerson, Y. (1981) In *Haemostasis and Thrombosis*, Bloom, A.L. and Thomas, D.P., Eds., 124-139 (Churchill Livingstone, Edinburgh)