CHARACTERIZATION OF THE HUMAN FACTOR XII (HAGEMAN FACTOR)

CDNA AND THE GENE

by

DEBORAH E. COOL

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Biochemistry

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October, 1987

© Deborah E. Cool, 1987
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
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date **Oct. 14, 1987**
ABSTRACT

A human liver cDNA library was screened by colony hybridization with two mixtures of synthetic oligodeoxyribonucleotides as probes. These oligonucleotides encoded regions of β-factor XIIa as predicted from the amino acid sequence. Four positive clones were isolated that contained DNA coding for most of factor XII mRNA. A second human liver cDNA library was screened by colony hybridization with 32P-labeled cDNA clones obtained from the first screen and two identical clones were isolated.

DNA sequence analysis of these overlapping clones showed that they contained DNA coding for the signal peptide sequence, the complete amino acid sequence of plasma factor XII, a TGA stop codon, a 3' untranslated region of 150 nucleotides, and a poly A+ tail. The cDNA sequence predicts that plasma factor XII consists of 596 amino acid residues. Within the predicted amino acid sequence of factor XII, were identified three peptide bonds that are cleaved by kallikrein during the formation of β-factor XIIa.

Comparison of the structure of factor XII with other proteins revealed extensive sequence identity with regions of tissue-type plasminogen activator (the epidermal growth factor-like region and the kringle region) and fibronectin (type I and type II homologies). As the type II region of fibronectin contains a collagen-binding site, the homologous region in factor XII may be responsible for the binding of factor XII to collagen. The carboxyl-terminal region of factor XII shares considerable amino acid sequence homology with other serine proteases including trypsin and many clotting factors.

A human genomic phage library was screened by using a human factor XII cDNA as a
hybridization probe. Two overlapping phage clones were isolated which contain the entire human factor XII gene. DNA sequence and restriction enzyme analysis of the clones indicate that the gene is approximately 12 kbp in size and is comprised of 13 introns and 14 exons. Exons 3 through 14 are contained in a genomic region of only 4.2 kbp with introns ranging in size from 80 to 554 bp.

The multiple regions found in the coding sequence of FXII that are homologous to putative domains in fibronectin and tissue-type plasminogen activator are contained on separate exons in the factor XII gene. The intron/exon gene organization is similar to the serine protease gene family of plasminogen activators and not to the clotting factor family.

Analysis of the 5' flanking region of the gene shows that it does not contain the typical TATA and CAAT sequences found in other genes. This is consistent with the finding that transcription of the gene is initiated at multiple start sites.
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosinetriphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base Pair(s)</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>Calcium ions</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotidetriphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>ddNTP</td>
<td>Dideoxynucleotidetriphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic Acid</td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium Bromide</td>
</tr>
<tr>
<td>G</td>
<td>Guanosine</td>
</tr>
<tr>
<td>Gla</td>
<td>γ-Carboxyglutamic Acid</td>
</tr>
<tr>
<td>GuHCl</td>
<td>Guanidine Hydrochloride</td>
</tr>
<tr>
<td>hnRNA</td>
<td>Heterogenous Nuclear RNA</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-beta-D-Thiogalactopyranoside</td>
</tr>
<tr>
<td>Kbp</td>
<td>Kilobase Pair(s)</td>
</tr>
<tr>
<td>Krpm</td>
<td>Thousand Revolutions Per Minute</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Broth</td>
</tr>
<tr>
<td>mA</td>
<td>Milliamps</td>
</tr>
<tr>
<td>min</td>
<td>Minute(s)</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>N</td>
<td>Any Nucleotide (G, A, T or C)</td>
</tr>
<tr>
<td>OD</td>
<td>Optical Density</td>
</tr>
<tr>
<td>pfu</td>
<td>Plaque Forming Unit</td>
</tr>
<tr>
<td>PA</td>
<td>Plasminogen Activator(s)</td>
</tr>
<tr>
<td>R</td>
<td>Purine (A or G)</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-Tetramethylethylenediamine</td>
</tr>
<tr>
<td>tPA</td>
<td>Tissue-Type Plasminogen Activator</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transer RNA</td>
</tr>
<tr>
<td>U</td>
<td>Uridine</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra Violet</td>
</tr>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>T</td>
<td>Thimidine</td>
</tr>
<tr>
<td>W</td>
<td>Watts</td>
</tr>
<tr>
<td>X-Gal</td>
<td>5-Bromo-4-Chloro-3-Indoyl-β-D-Galactopyranoside</td>
</tr>
<tr>
<td>Y</td>
<td>Pyrimidine (T or C)</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

Abstract ................................................................................................................................. ii 

List of Abbreviations ................................................................................................................ iv

Table of Contents ....................................................................................................................... vi

List of Tables ............................................................................................................................ ix 

List of Figures ........................................................................................................................... x

Acknowledgements ................................................................................................................... xii

I. Introduction ........................................................................................................................... 1
   A. THE CONTACT SYSTEM OF PLASMA ............................................................................. 1
      1. The Biochemistry of the Surface Activation Components ........................................... 2
      a. The Catalytic Mechanism of Serine Proteases .......................................................... 5
      b. Factor XII .................................................................................................................. 6
      c. Prekallikrein and Factor XI ...................................................................................... 12
         d. High Molecular Weight Kininogen ......................................................................... 14
      2. A Model of the Surface Activation Mechanism ............................................................ 15
   B. PHYSIOLOGICAL FUNCTIONS OF THE CONTACT SYSTEM ....................................... 18
      1. Blood Coagulation ....................................................................................................... 18
      2. Fibrinolysis ................................................................................................................ 21
         a. Plasminogen and Its Activators ............................................................................. 22
      3. The Inflammatory Response ..................................................................................... 24
      4. Conversion of Prorenin to Renin ............................................................................. 26
   C. THE STRUCTURE OF EUKARYOTIC STRUCTURAL GENES ........................................... 26
      1. Promoters .................................................................................................................. 26
      2. Introns and Exons ...................................................................................................... 28
      3. Transcription and Processing ................................................................................... 28
   D. MECHANISMS OF GENE EVOLUTION .......................................................................... 29
      1. Gene Duplication ........................................................................................................ 29
      2. Gene Fusions ............................................................................................................. 30
      3. Exon Shuffling .......................................................................................................... 30
   E. EVOLUTION OF AMINO ACID AND DNA SEQUENCE ................................................ 30
      1. Molecular Clock ....................................................................................................... 30
      2. Intron Sliding ............................................................................................................ 31
   F. PROTEIN DOMAINS OF THE COAGULATION AND PLASMINOGEN ACTIVATOR GENES .................................................................................................................. 32
      1. Trypsin-Like Family of Serine Proteases .................................................................. 32
   G. EVOLUTION OF THE SERINE PROTEASES ................................................................. 36
   H. OBJECTIVES ................................................................................................................... 37

II. Materials and Methods ......................................................................................................... 39
   A. MATERIALS .................................................................................................................. 39
   B. STRAINS, VECTORS AND MEDIA ................................................................................. 40
      1. Bacterial Strains ....................................................................................................... 40
      2. Vectors ....................................................................................................................... 41
### III. Results

#### A. CHARACTERIZATION OF THE HUMAN FXII cDNA

1. Isolation of Factor XII cDNA Clones ................................. 69
2. Sequence Analysis of Human Factor XII cDNA Clones .......... 76
3. Predicted Amino Acid Sequence of Human Factor XII ...... 79
4. 5' End of the Factor XII cDNA ............................................. 80
5. A 2.40 kb mRNA From Liver Codes For Human Factor XII ... 83

#### B. THE HUMAN FACTOR XII GENE

1. Isolation and Characterization of Factor XII Genomic Clones .......... 86
2. Southern Blot Analysis of the Human FXII Gene .................. 91
3. Localization of Intron/Exon Junctions ............................... 94
4. Nucleotide Sequence of the Human Factor XII Gene ............ 97
5. Transcription Initiation Site of the Factor XII Gene .......... 101

### IV. Discussion

#### A. CHARACTERIZATION OF THE HUMAN FACTOR XII cDNA .......................... 109
LIST OF TABLES

I. DNA Sequencing Mixes .......................................................... 64

II. Nucleotide Sequence of Intron/Exon Junctions in the FXII Gene .................. 95

III. Frequencies of Nucleotides at Intron/Exon Junctions ............................... 96

IV. Size and Position of Exons and Introns in the FXII Gene .............................. 97

V. Per-cent Homologies Between the Amino Acid Sequence in the \( \beta \)-factor XIIa and Pancreatic Serine Proteases Catalytic Domains. .............................. 123

VI. Summary of TATA-Like Elements Found in Liver Specific and Coagulation Genes . 130
**LIST OF FIGURES**

1. The Blood Coagulation Cascade ................................................................. 3
2. Schematic Diagram of β-factor XIIa ............................................................... 8
3. Schematic Diagram of the Surface-Dependent Contact Activation Mechanism .......... 16
4. Amino Acid Sequence Homologies in Coagulation Factor Zymogens ..................... 33
5. Predicted Oligonucleotide Sequences for β-factor XIIa .................................... 70
6. Double Oligonucleotide Screen For Factor XII cDNA ..................................... 72
7. Restriction Enzyme Map and Sequencing Strategy for Factor XII cDNA ............... 74
8. Nucleotide Sequence of Factor XII cDNA ..................................................... 77
9. Nucleotide Sequence of a cDNA for the FXII Signal Peptide ............................ 82
10. Size Determination of the Factor XII mRNA Using rRNA Markers ...................... 84
11. Southern Blot Analysis of FXII Genomic Clones .......................................... 87
12. Intron/Exon Organization of the FXII Gene ................................................ 90
13. Genomic DNA Southern Blot Analysis of the FXII Gene .................................. 92
14. Partial DNA Sequence of the FXII Gene ..................................................... 98
15. Nuclease S1 Mapping of the FXII Gene ....................................................... 102
16. Primer Extension Analysis of the FXII Gene ............................................... 104
17. Northern Blot Analysis of the 5' End of the FXII Gene .................................. 107
18. Proposed Model for Factor XII ....................................................................... 112
19. Amino Acid Sequence Homologies of FXII With Other Proteins ....................... 115
20. Amino Acid Sequence Alignment of the Catalytic Domain of FXII With Other Serine Proteases ...................................................................................... 121
21. Schematic Diagram of FXII and Positions of Introns ....................................... 134
22. N-Terminal Exon Organization in the Plasminogen Activators and FXII ............ 138
23. Comparison of Exon Organization of the Serine Protease Domain in Some Serine Proteases ...................................................................................... 143
ACKNOWLEDGEMENTS

I would like to thank the members of my supervisory committee, Drs. Jim Richards and Peter Candido for their kind support and advice throughout this project. I would also like to thank the members of the MacGillivray lab, those past and present, for their good natured humour, for tolerating Molly and for the much appreciated technical advice and for their tremendous support during my work in the lab. I would especially like to acknowledge my great supervisor, Dr. Ross MacGillivray (or "Ross") who has taught me always to try to achieve a high standard of excellence in science and to take pride and pleasure in the work. Thank you Ross, I am grateful for all the assistance you have given me. I would like to thank Dr. M.J. Smith who was my Masters of Science Degree supervisor and who became a good friend and encouraged me throughout my Ph.D. degree. Finally, I would especially like to thank my husband, for his unending encouragement during my graduate degrees and for all his help in preparation of my manuscripts and this thesis.
I. INTRODUCTION

A. THE CONTACT SYSTEM OF PLASMA

Organisms can respond to invasion by foreign substances, to trauma and to internal disorders by the activation of four plasma systems: they are the complement systems, the contact (Hageman) activated clotting system, the extrinsic clotting system and the fibrinolytic system (Ratnoff and Saito, 1979; Cochrane and Griffin, 1982; Pesce and Dosekun, 1983; Sundsmo and Fair, 1983; Colman, 1984). Maintenance of a homeostatic environment \textit{in vivo} may be a result of the synergism that exists between these plasma systems (Cochrane, 1980; Bennett and Ogston, 1981). At least the coagulation and fibrinolytic systems can be activated both \textit{in vivo} and \textit{in vitro} by a mechanism described as "surface-mediated" or "contact phase" activation which requires a negatively charged surface as well as the appropriate plasma proteins (reviewed in Ratnoff and Saito, 1979). \textit{In vitro} studies have also demonstrated that the contact activation mechanism plays a role in the inflammatory response, in the production of kinins and in the conversion of prorenin to renin (Bennett and Ogston, 1981; Cochrane and Griffin, 1982). The proteins which are primarily responsible for the surface activation reactions are factor XII (Hageman factor), prekallikrein (Fletcher factor), high molecular weight (HMW) kininogen (Fitzerald factor) and factor XI (PTA, plasma thromboplastin antecedent) (Davie \textit{et al.}, 1979). Factor XII and HMW kininogen have surface binding properties and therefore are essential in the surface activation mechanism. The following review will describe the physical and biochemical properties of the components involved in contact activation; this will include mainly the role of factor XII in surface activation of the plasma systems listed above and the subsequent relevance of factor XII to the maintenance of homeostasis in living
organisms.

1. The Biochemistry of the Surface Activation Components

The characteristic enzymatic reaction in the plasma systems described above is the action of serine proteases. Similar to the pancreatic serine proteases, trypsin (Marquart et al., 1983), chymotrypsin (Cohen et al., 1981) and elastase (Sawyer et al., 1978), the proteases involved are synthesized in precursor forms and are activated by limited proteolysis (Davie et al., 1979; Jackson and Nemerson, 1980). The active enzymes have a trypsin-like specificity with a preference for arginyl peptide bonds and contain the conserved activation and catalytic site residues described for the pancreatic enzymes (Jackson and Nemerson, 1980). Unlike the digestive enzymes, the proteases required in the complex plasma systems have long amino terminal polypeptide domains that appear to be important in functions such as metal-binding, surface binding or fibrin binding and in their regulation (Jackson and Nemerson, 1980). The zymogens in the contact system are activated by limited proteolysis (discussed below) resulting in a cascade of enzymatic reactions in which inactive zymogens are converted to active serine proteases that act upon each other in a controlled and step-wise manner (Jackson and Nemerson, 1980). The physiological significance of the cascade mechanism is the amplification of reactions by a few initiator molecules resulting in a massive amount of end-product such as fibrin from the blood coagulation cascade (Fig. 1) (Davie et al., 1979; Jackson and Nemerson, 1980) or plasmin from the fibrinolytic cascade (Collen, 1979).
Figure 1: The Blood Coagulation Cascade

Outline of the mammalian blood coagulation cascade with the intrinsic pathway (left) and extrinsic pathway (right) converging at the activation of factor X to factor Xa, and ending with the formation of the insoluble fibrin clot. Bars represent the polypeptide chains (proportional to polypeptide chain length) with molecular weights indicated below. Intra molecular disulfide bridges are indicated by lines between the two chains (from Neurath, 1984).
a. The Catalytic Mechanism of Serine Proteases

The catalytic mechanism of peptide bond cleavage is the same for all serine proteases and has been intensively studied. The classical model of serine protease catalysis is the "charge-transfer relay" mechanism (see Creighton, 1984) involving the catalytic triad of amino acid residues which consists of His$_{57}$, Asp$_{102}$ and the active site Ser$_{195}$ in chymotrypsin. These amino acids are conserved in all serine proteases and are essential residues in the hydrolysis of amide and ester bonds. In the charge-transfer relay mechanism the polypeptide binds to the S1 binding site pocket, and the oxygen of the carbonyl group in the amide linkage undergoes nucleophilic attack by the hydroxyl group of Ser$_{195}$, which has been polarized by His$_{57}$; a tetrahedral intermediate is formed and is further stabilized by the transfer of a hydrogen ion to Asp$_{102}$. The hydrogen ion from Ser$_{195}$ is then transferred to the leaving group, and an acyl-enzyme intermediate is formed. This intermediate undergoes nucleophilic attack by an incoming water molecule, and the tetrahedral intermediate is formed again resulting in the transfer of the polypeptide chain to water and restoration of the serine hydroxyl group in the protease active site.

Aspects of the charge-transfer relay theory have been disputed, however, based on high resolution X-ray diffraction studies on trypsin-substrate complexes and on neutron diffraction studies which can readily position hydrogen atoms (Creighton, 1984). One fundamental change is that interactions between the enzyme and substrate take place which make the active site more energetically suited to the tetrahedral intermediate conformation than to the substrate, products or the acyl-enzyme intermediate conformations. The carbonyl group in the scissile bond becomes distorted when entering the active site and begins assuming the tetrahedral conformation. This is stabilized by three
hydrogen bonds. The first hydrogen bond is between the backbone carbonyl group at residue 214 and the backbone -NH- group of the P₁ residue (the amino terminal residue to the scissile bond). The other two hydrogen bonds are between the carbonyl oxygen of P₁ and the two backbone -NH- groups of residues 213 and 215 of the enzyme, which form the oxyanion binding site. Polarization between the negative charge on the oxygen atom and the positive charge on the carbon atom of the carbonyl group is large enough that the carbon atom will be susceptible to nucleophilic attack by the oxygen in the hydroxyl group of Ser₁₉₅. When this occurs the hydrogen atom of the serine residue is transferred to His₅₇, and the acyl-enzyme intermediate is formed. An important feature of this reaction is that the Asp₁₂₂ residue is not needed for the extra positive charge assumed transiently by His₅₇, but the residue appears to be essential in orienting the substrate in the active site through hydrogen bonding and in maintaining the His₅₇ conformation during catalysis so that the hydrogen atom of the imidazole ring is not on the N° (which accepts the Ser₁₉₅ hydrogen atom) but on the N°₁ atom of the ring (see Creighton, 1984).

b. Factor XII

Factor XII, or Hageman factor, was so named after the individual in which the deficiency was first discovered (Ratnoff and Colopy, 1955). Although factor XII circulates in plasma at very low concentrations (approximately 30 µg/ml), it has been highly purified from human, bovine, rabbit and guinea pig plasmas (Cochrane and Wuepper, 1971; Movat and Ozge-Anway, 1974; Revak et al., 1974; Griffin and Cochrane, 1976; Fujikawa et al., 1977a,b; Claeys and Collen, 1977; Fujikawa and Davie, 1981; Yamamoto and Cochrane, 1981). The complete amino acid sequence of 596 amino acid residues and the identification of glycosylation sites of human factor XII have only recently been determined for the human
protein (Fujikawa and McMullen, 1983; McMullen and Fujikawa, 1985) because of the difficulty in isolating large quantities of the protein from plasma for amino acid sequence analysis. Prior to the commencement of the work described in this thesis, the only known amino acid sequence for bovine factor XII was at the amino-terminus of the zymogen and around the active site serine residue in the serine protease domain. Soon after this project was initiated, however, the entire amino acid sequence of the serine protease domain of the human protein, $\beta$-factor XIIa (Fig. 2) was reported (Fujikawa and McMullen, 1983) which aided in the identification of cloned cDNAs of human factor XII.

Human factor XII is a glycoprotein containing 16.8% carbohydrate including 4.2% hexose, 4.7% hexosamine and 7.9% N-acetylneuraminic acid. The protein is synthesized in the liver and secreted into blood as a single polypeptide chain with N-glycosylation on Asn and O-glycosylation on Thr and Ser residues (Fujikawa and McMullen, 1983; McMullen and Fujikawa, 1985). Glycosylation is intimately involved in protein secretion (Blobel et al., 1979). Carbohydrate attachment occurs during protein synthesis (N-glycosylation) and in the Golgi apparatus (O-glycosylation). Proteins which are to be glycosylated are synthesized on the rough endoplasmic reticulum and contain an amino-terminal extension called a "signal sequence" which is 16 to 30 residues in length and has a hydrophobic core (Blobel et al., 1979). During protein synthesis, the signal peptide sequence is bound by a "signal recognition protein", and the complex is subsequently directed to the endoplasmic reticulum membrane. The nascent strand is transported through the membrane and is cleaved on the lumen side by a resident signal peptidase. The specificity of the enzyme recognition site has been described (Watson, 1984). The protein is then shuttled to the Golgi apparatus by an unknown mechanism where it is glycosylated. Secreted proteins are transported from the Golgi apparatus to the plasma membrane in secretory vesicles and
Figure 2: Schematic Diagram of β-factor XIIa

The open bars represent the human factor XII molecule with "Site 1" and "Site 2" arrows indicating the kallikrein cleavage sites of the zymogen generating the three polypeptide chains of $M_r$ 40,000 (amino terminal fragment), 12,000 and 30,000 (carboxy terminal fragment with serine protease activity) (from Fujikawa and McMullen, 1983). The amino acid sequence shows the amino terminal nanopeptide (L-chain of β-factor XIIa) generated by kallikrein cleavage at unknown sites in either the $M_r$ 40,000 or the 12,000 fragments; partial amino acid sequence of the serine protease domain is given in the H-chain of β-factor XIIa.
L-Chain of α-Factor XIIa

H-Chain of α-Factor XIIa

AsnGlyProLeuSerCysGlyGlnArg

ValValGlyGly

ValValGlyGlyCysGlyAspSerGlyGlyThrValSer

IleProProTrp

40,000

12,000

ValValGlyGly

30,000

Site 2

Site 1

or
enter plasma upon membrane fusion. The form of factor XII found in plasma is a single-chain polypeptide (M_r 80,000) with an isoleucine residue at the amino terminus. In the prefactor XII, this Ile residue is presumably preceded by the bond cleaved by signal peptidase.

The zymogen form of factor XII is activated in plasma by partial proteolysis by kallikrein generating an active serine protease (Davie et al., 1979; Jackson and Nemerson, 1980). There are multiple sites in factor XII that are susceptible to cleavage by a number of proteases including kallikrein, plasmin and trypsin (Margolis, 1958; Kaplan and Austen, 1971; Meier et al., 1977; Dunn and Kaplan, 1982). Limited proteolysis by kallikrein activates surface-bound factor XII and produces two active enzyme forms, α-factor XIIa and β-factor XIIa. A single cleavage of the zymogen generates α-factor XIIa while two more subsequent cleavages by kallikrein yield β-factor XIIa (Revak et al., 1974; Dunn et al., 1982; Fujikawa and McMullen, 1983). Both active enzymes consist of two polypeptide chains held together by a disulfide bond (Revak et al., 1974, 1977). Alpha-factor XIIa is comprised of two polypeptide chains of M_r 52,000 and 28,000 while β-factor XIIa consists of two polypeptide chains of M_r 2,000 and 28,000. Amino-terminal sequence analysis has shown that the M_r 52,000 fragment in α-factor XIIa is derived from the amino-terminal region of factor XII (Fujikawa et al., 1980b) and contains the surface-binding site (Revak and Cochrane, 1976; Revak et al., 1977). However, the M_r 2,000 fragment of β-factor XIIa does not bind to surfaces. The M_r 28,000 fragments of both α- and β-factor XIIa are identical and contain the catalytic domain derived from the carboxy-terminal region of factor XII (Revak et al., 1974, 1977). The catalytic domain of α- and β-factor XIIa shares extensive amino acid sequence homology with other serine proteases, including many blood clotting factors (see Jackson and Nemerson, 1980; Fujikawa and McMullen, 1983). The
sequence homology is highest at the activation regions and near the active sites of these proteases.

Factor XII and α-factor XIIa readily bind to anionic surfaces such as silicates, dextran sulfate and sulfatides \textit{in vitro} (Ratnoff and Rosenblum, 1958; Fujikawa \textit{et al.}, 1977, 1980a). \textit{In vivo}, these binding properties may be responsible for the activation of factor XII when it comes in contact with collagen or platelet membranes (Wilner \textit{et al.}, 1968; Harpel, 1972). Surface-bound factor XII has enzyme activity towards its protein substrates, prekallikrein and factor XI (McMillin \textit{et al.}, 1974; Saito, 1977; Heimark \textit{et al.}, 1980) (see below). Surface-bound α-factor XIIa is further cleaved by kallikrein resulting in β-factor XIIa which is free to enter plasma since it no longer has the fragment of $M_r$ 52,000 which contains the surface binding property of the molecule.

Various \textit{in vivo} and \textit{in vitro} substrates for activated factor XII have been described in the literature and include prekallikrein, factor XI and factor VII (reviewed in Cochrane and Griffin, 1982). The serine protease recognizes the sequence Thr-Arg-Ile (in prekallikrein) and Pro-Arg-Ile (in FXI) and cleaves the polypeptide bond carboxy-terminal to the Arg residues.

The serine proteases including the digestion enzymes show little amino acid sequence identity (less than 40%) except in a few regions which are important in the catalytic reaction. However this amount of homology may transfer to an 85% tertiary structural homology (James \textit{et al.}, 1978). This observation has allowed the construction of computer generated models of the serine protease domains of a number of clotting factors including thrombin, FIXa, FXa (Furie \textit{et al.}, 1982) and FXII (Cool \textit{et al.}, 1985). The means of
substrate specificity of the enzymes remains unclear (Furie et al., 1982) but probably results from the surface topology which would be unique for each protein. This kind of specificity is essential in the ordered amplification of the coagulation and fibrinolytic cascades (see Davie et al., 1979; Jackson and Nemerson, 1980).

Activated factor XII is regulated in plasma by inhibitors. The major plasma inhibitor for \( \beta \)-factor XIIa is the C1-inhibitor (de Agostini et al., 1984) which is the major proteolytic inhibitor of the proteolytic enzymes derived from the first component of complement (Ziccardi, 1981). The C1-inhibitor is also the primary inactivator of plasma kallikrein (Schapira et al., 1982; Van der Graaf et al., 1983). Kinetic studies (de Agostini et al., 1984) with human plasma and purified C1-inhibitor, \( \alpha_2 \)-antiplasmin and antithrombin III showed that 74\% of \( \beta \)-factor XIIa was inhibited by C1-inhibitor and the remaining activated protein inhibited equally by \( \alpha_2 \)-antiplasmin antithrombin III. A complex of 145,000 daltons was detected by SDS-polyacrylamide gel electrophoresis which represents a 1:1 stoichiometric relationship between \( \beta \)-factor XIIa and C1-inhibitor (de Agostini et al., 1984).

c. Prekallikrein and Factor XI

Prekallikrein and factor XI circulate in plasma as zymogen forms of serine proteases and are involved in the surface-dependent initiation of the coagulation cascade (Davie et al., 1979; Jackson and Nemerson, 1980). Prekallikrein is also involved in the contact activation of fibrinolysis (Bouma et al., 1980; Mandle and Kaplan, 1977; Colman, 1969), kinin generation (Kerbiriou and Griffin, 1979; Nakayasu and Nagasawa, 1979; Mori and Nagasawa, 1981) and the inflammatory response (Kaplan, 1978; Collen, 1979; Cochrane, 1980; Bouma and Griffin, 1981; Colman, 1984).
Prekallikrein and factor XI circulate in plasma as inactive zymogens in complexes with HMW kininogen (Mandel et al., 1976; Thompson et al., 1977). The proteinases, purified from rabbit, bovine and human plasmas (reviewed in Cochrane and Griffin, 1982), are also glycoproteins synthesized by the liver and secreted (Davie et al., 1979; Jackson and Nemerson, 1980). Prekallikrein enters plasma as a single chain polypeptide with a molecular weight of 88,000 (Mandle and Kaplan, 1977; Bouma et al., 1980; Heimark and Davie, 1981) whereas factor XI (Bouma and Griffin, 1977; Kurachi and Davie, 1977, 1981) is the only serine protease secreted as a dimer with a molecular weight of 55,000 to 63,000 for the monomer. The proteins are activated by a single proteolytic cleavage by factor XIIa (Revak et al., 1974; Fujikawa et al., 1980c; Bock et al., 1981; Tans and Griffin, 1982; Griffin and Cochrane, 1976; Bouman and Griffin, 1977; Kurachi et al., 1980; Ohkubo et al., 1982) producing two chain molecules with amino-terminal heavy chains containing the substrate-HMW kininogen-binding sites and carboxy-terminal light chains which bear the serine protease domains (Mandle and Kaplan, 1977; van der Graaf et al., 1983b).

Recently the complete amino acid sequence of both human prekallikrein (Chung et al., 1986) and factor XI (Fujikawa et al., 1986) have been determined by partial amino acid sequence of the purified proteins and by prediction of amino acid sequences from isolated cDNA clones. The two proteins were found to share 86% amino acid sequence identity. Alignment of half-cysteine residues shows that these two proteins share a structural domain of 91 amino acid residues which is repeated four times in both proteins. This repeat is present in the heavy chains of the prekallikrein and factor XI molecules and does not exhibit sequence homology with any other known protein. It is possible that this repeat is important in the HMW kininogen interaction in the molecule.
d. High Molecular Weight Kininogen

High molecular weight kininogen is a nonenzymatic cofactor which greatly enhances the surface activation of the various plasma systems described below (Griffin and Cochrane, 1976; Meier et al., 1977; Revak et al., 1977; Griffin, 1978; van der Graaf et al., 1982). The protein is synthesized in the liver and is secreted into plasma as a glycoprotein (M<sub>r</sub> 110,000) containing 13% carbohydrate (Jackson and Nemerson, 1980; Burgess and Esmouf, 1985). Human HMW kininogen is cleaved twice by kallikrein giving a two chain molecule held together by a disulfide bond and releasing a small peptide (approximately 1200 daltons) called bradykinin (Saito, 1977; Jackson and Nemerson, 1980). This cleavage reaction occurs only when the HMW kininogen/ prekallikrein complex has bound to a negatively charged surface and prekallikrein is activated by factor XIIa. The sites of interaction between prekallikrein (or factor XI) and HMW kininogen are on the light chain in HMW kininogen (the carboxy-terminal region) and on the heavy chain of prekallikrein (the amino-terminal half of the molecule) (or FXI) (Bock et al., 1985). The interaction is noncovalent, and a 1:1 complex is formed between these proteins. The light chain of HMW kininogen has a histidine rich region that is adjacent and carboxy-terminal to the bradykinin domain in the molecule and is probably responsible for binding to surfaces since deletion of this region removes coagulation enhancement ability of the molecule (Cochrane and Griffin, 1979; Kerbiriou et al., 1980; Silverberg et al., 1980).

An important product of the kallikrein-HMW kininogen reaction is the nonapeptide, bradykinin. The physiological functions of kinins have been studied intensely; their roles in the organism are varied and include participation in the inflammatory response, in renal function and in the regulation of the cardiovascular system (Bennett and Ogston, 1981; Diz,
2. A Model of the Surface Activation Mechanism

The enzymatic mechanism for surface activation (Cochrane and Griffin, 1982) is still speculative but a schematic diagram (Fig. 3) from Griffin and Cochrane (1976) illustrating coagulation initiation can be used to describe the essential steps (Bouma and Griffin, 1981). The complexes, prekallikrein/HMW kininogen and factor XI/HMW kininogen, and factor XII bind to negatively charged surfaces, presumably surfaces that are produced during injury and involve the vascular basement membrane. However, highly purified forms of collagen added to plasma did not activate factor XII as might be expected as collagen is an important interstitial molecule (Griffin et al., 1975; Fujikawa et al., 1980a). Surfaces play an essential role in the contact system during the coagulation and fibrinolytic cascade reactions as they allow the molecules involved to assemble and to act upon one another. For example, the result of the factors binding to the surface during contact activation is a burst of enzymatic activity to produce fibrin or plasmin in a matter of seconds.

Factor XII is the first zymogen to be activated in the contact activation mechanism (Griffin, 1981) although it is still unknown how the first molecule is activated. The protein becomes more susceptible to cleavage when bound to a surface (Griffin, 1978), and it is possible that small amounts of \(\alpha\)-factor XIIa in plasma could be responsible for the initial reaction (Heimark et al., 1980; Rosing et al., 1985). Therefore, according to the model, an active form of factor XII cleaves surface-bound factor XII molecules in a slow enzymatic reaction which is followed by the activation of surface-bound prekallikrein generating kallikrein (Griffin, 1981). The initiation reactions are now greatly enhanced since prekallikrein and
Figure 3: Schematic Diagram of the Surface-Dependent Contact Activation Mechanism

Schematic diagram of the surface-dependent reactions involved in contact phase activation of various plasma systems including the coagulation and fibrinolysis cascades (from Griffin and Cochrane, 1979). PK: Prekallikrein; Kal: Kallikrein; HMWKgn: HMW kininogen; HF: Hageman Factor (factor XII).
factor XII can undergo reciprocal proteolysis (Figure 3). Kallikrein, binding weakly to HMW kininogen, (Thompson et al., 1983) dissociates from the cofactor and activates many other factor XII molecules (Tans et al., 1986) as well as generating bradykinin from HMW kininogen. Alpha-factor XIIa displays greater activity towards surface-bound factor XI than to prekallikrein (Griffin, 1981) and so FXI is preferentially hydrolyzed, resulting in a cascade of enzymatic reactions. Factor XIA does not dissociate readily from HMW kininogen (Bouma et al., 1983) and is probably inhibited by its plasma inhibitor (α1-protease inhibitor) (Scott et al., 1982) in its complex form. Kallikrein cleaves factor XIIa a second time and generates β-factor XIIa which is no longer surface bound and enters plasma. The lifetime of both activated factor XII and prekallikrein in plasma is very short since they are inhibited rapidly by the C1-inhibitor (Schreiber et al., 1973).

B. PHYSIOLOGICAL FUNCTIONS OF THE CONTACT SYSTEM

1. Blood Coagulation

Haemostasis is the spontaneous arrest of blood flow during injury to a blood vessel (Guyton, 1977). Survival of the organism with a closed circulatory system such as in the vertebrates, depends on rapid initiation at the site of injury resulting in blood coagulation (Davie et al., 1979; Jackson and Nemerson, 1980). Platelets circulate in plasma and adhere to the subendothelial walls of torn vessels forming a platelet plug. Circulating molecules, the coagulation proteases, respond to platelet aggregation with the subsequent deposition of fibrin at the site of injury (Guyton, 1977). Many of the proteins involved are inactive zymogen forms of serine proteases and were discovered in patients with abnormal thromboplastin times resulting in a tendency to bleed (Bloom, 1981). At least 14 proteins
found in plasma and tissue are required in blood coagulation as well as phospholipid and $\text{Ca}^{2+}$ (Jackson and Nemerson, 1980; Burgess and Esnouf, 1985). The participating molecules interact with one another in a cascade effect (Fig. 1) whereby one molecule activates many others resulting in the conversion of a large number of fibrinogen molecules to fibrin (Davie and Ratnoff, 1964; MacFarlane, 1964). Fibrin is cross-linked and forms an insoluble fibrin clot which reinforces the platelet plug and prevents loss of fluid.

The order in which the coagulation factors are activated depends on which factors are involved when the coagulation cascade is initiated (Davie et al., 1979; Jackson and Nemerson, 1980). There are two enzymatic pathways, the intrinsic and the extrinsic pathway, which converge at a common step involving the activation of factor X. Figure 1 depicts the reactions in the coagulation cascade described in mammals. The intrinsic pathway is initiated by factor XII activation of prekallikrein and factor XI in the presence of a negatively charged surface and high molecular weight (HMW) kininogen (Griffin, 1981), as described previously. Initiation of the extrinsic pathway requires tissue factor released from tissue during injury. Tissue factor is a lipoprotein and acts as a cofactor which accelerates the activation of factor X by factor VIIa (Davie et al., 1979; Zur and Nemerson, 1981). The intrinsic pathway (Davie and Ratnoff, 1964) derives its activation molecules strictly from plasma whereas the extrinsic pathway (Zur and Nemerson, 1981) utilizes proteins from tissue extract; however, both pathways are usually activated upon injury to a blood vessel. Figure 1 also shows that the extrinsic pathway can be initiated by the kallikrein/factor XIIa systems by activating factor VII (Kisiel et al., 1977; Radcliffe et al., 1977; Seligsohn et al., 1979). Coagulation initiation via the extrinsic pathway has also been demonstrated in vitro by cold dependent activation of factor VII (Seligsohn et al. 1978).
The blood coagulation cascade is an elegant illustration of regulation of a complicated biological process through the action of a single type of enzymatic reaction—polypeptide chain cleavage by a serine protease (Davie et al., 1979; Jackson and Nemerson, 1980). The proteins involved are inactive zymogen forms of serine proteases and upon activation will recognize their specific zymogen substrate in the cascade and activate it by limited proteolysis at a specific site(s) and always after an arginine residue (Davie et al., 1979; Jackson and Nemerson, 1980). For example, coagulation factors XII, XI, X, IX, VII and prothrombin are partially cleaved to produce active serine proteases (XIIa, XIa, Xa, IXa and thrombin) shown in Figure 1. Cofactors such as V, VIII, HMW kininogen or tissue factor are required in some of these proteolytic reactions in addition to phospholipid and Ca\(^{2+}\) for the vitamin K-dependent coagulation proteins (factors VII, IX, X and prothrombin) (Suttie, 1985). Specific glutamic acid residues found at the amino-terminal regions of these coagulation proteins are converted to \(\gamma\)-carboxyglutamic acid (Gla) by a vitamin K-dependent carboxylase. The Gla residues are required in the Ca\(^{2+}\) bridging of the Gla-containing proteins to phospholipid membranes prior to their activation (Suttie, 1985). The culmination of all these reactions in the cascade is the polymerization of fibrin molecules (Doolittle, 1984) to reinforce the platelet plug. The fibrin network is strengthened by the formation of covalent cross links between monomers by the transglutaminase factor XIIIa which effectively stops the loss of blood (Curtis, 1981). The activated serine proteases are rapidly inhibited by plasma inhibitors and by feedback inhibition loops (Davie et al., 1979; Jackson and Nemerson, 1980) which will prevent the proteases from acting at any other site.

Although factor XII and kallikrein are central components of the surface activation mechanism in blood coagulation, individuals with either plasma deficiency are
assymptomatic. However, plasma coagulation times are greatly reduced and during extreme trauma or surgical procedures, the bleeding disorder may become life-threatening (Ratnoff and Saito, 1979; Griffin, 1981). It is reasonable that the coagulation system has evolved in such a way that it is activated by more than one mechanism since it is crucial that upon injury, the flow of blood must be arrested.

2. Fibrinolysis

In mammals the dissolution of fibrin clots requires the activation of the fibrinolytic enzymatic system (Collen, 1979). Plasminogen is the major precursor which is activated upon cleavage by a serine protease called plasminogen activator (Wiman, 1978). The product, plasmin, is a trypsin-like serine protease that cleaves 50-60 bonds in fibrin (Doolittle, 1981). Plasmin exhibits broad substrate specificity and therefore is inhibited rapidly in plasma by $\alpha_2$-antiplasmin in a 1:1 stoichiometric complex (Wiman and Collen, 1977). Within the complex the active site serine residue of plasmin interacts with the carbonyl group of a leucine residue of the inhibitor to form a covalent bond (Wiman and Collen, 1977). The rate of the inhibition reaction is one of the fastest described for protein-protein interactions and is one order of magnitude greater than the inhibition rate of reaction for trypsin by its inhibitors (Wiman and Collen, 1978). The action of the fibrinolytic enzymes is not limited to the dissolution of blood clots but also function in other biological phenomena such as tissue repair (Astrup, 1978), malignant transformation (Reich, 1975), macrophage function (Reich, 1975), ovulation (Strickland, 1978) and embryo implantation (Strickland, 1978).
a. Plasminogen and Its Activators

Human plasminogen is a single chain glycoprotein ($M_r$ 90,000) of which the entire amino acid sequence has been determined (Sottrup-Jensen et al., 1978a,b). Two forms of the molecule are found in plasma, Glu-plasminogen and Lys-plasminogen. Lys-plasminogen is a proteolytic degradation form of Glu-plasminogen. Lys-plasminogen is the most common form isolated from plasma and is activated by plasminogen activators by limited proteolysis between Arg$_{560}$-Val$_{561}$. The resulting plasmin molecule consists of a heavy chain and a light chain held together by a disulfide bond. The heavy chain has five characteristic kringle domains which are responsible for the binding of plasmin to fibrin. The light chain has the active site which has the His, Asp and Ser of the catalytic triad found in all serine proteases. Activation of Glu-plasminogen is much slower than Lys-plasminogen but the final product is the same, Lys-plasmin. Activation of plasminogen by the plasminogen activators is greatly enhanced in the presence of fibrin.

Plasminogen activation is a complicated process which is not yet clearly understood. The protein may be activated along three different pathways: an intrinsic pathway in which all the components are found in plasma; an extrinsic pathway in which the enzymes required originate from tissues and, an exogenous pathway in which therapeutic substances such as urokinase and streptokinase are involved (Collen, 1979; Gaffney, 1981; Mullertz, 1984).

Intrinsic fibrinolysis is activated in the presence of factor XII, prekallikrein and HMW kininogen and a negatively charged surface (Collen, 1979). The mechanism of activation is not known, and its physiological relevance is unclear. However, plasma deficient in factor XII shows a greatly reduced plasmin generation time (Ratnoff, 1969; Meier et al., 1977).
Physiologically, a few factor XII deficient individuals have shown a tendency towards deep venous thrombosis and pulmonary embolism (see Ratnoff and Saito, 1979; Cochrane and Griffin, 1982).

Plasminogen activators are serine proteases which are present in many different tissues, organs and secretions and have various functions (Collen, 1980; Bachmann and Kruithof, 1984). Tissue-type plasminogen activator (tPA) (M, 70,000) (Pennica et al., 1983) is a single-chain glycoprotein of 530 amino acids. It is not clear whether tissue-type plasminogen activator is synthesized as a zymogen form of the serine protease since both single-chain and double-chain forms are active and are found in plasma (Bachmann and Kruithof, 1984). The double-chain molecule is produced by limited proteolysis of the bond between Arg27-Ile27, by plasmin, plasma kallikrein or by factor Xa (Wallen et al., 1983; Ichinose et al., 1984). In the activation of plasminogen, the enzyme is active only when bound to fibrin and not when bound to fibrinogen. Because of its high affinity for fibrin (dissociation constant, 150 nM) and considering the high concentration of plasma fibrinogen (5 to 10 μM), tPA is almost quantitatively bound to the clot during blood coagulation. In the presence of fibrin, the catalytic activity of t-PA towards Glu-plasminogen increases more than 500-fold, making the activator a very specific and potent fibrinolytic effector molecule.

Urokinase plasminogen activator (M, 53,000) has been isolated from urine and kidney tissue and is a single-chain glycoprotein consisting of 411 amino acid residues (Salerno et al., 1984; Riccio et al., 1985). The physiological importance of this plasminogen activator has been associated with the role of plasminogen activators in tissue degradation and remodeling although it is not known how urokinase functions (see Nagamine et al., 1985; Riccio et al., 1985). The protein is synthesized as a zymogen form of a serine protease and
is activated by partial proteolysis generating a two-chain molecule (Wun et al., 1982).

3. The Inflammatory Response

The inflammatory response to vascular injury is the process whereby the blood flow is increased by vasodilatation, vascular permeability is increased, leucocytes are summoned by chemical attractants (chemotaxis) and fibrin is deposited and removed (Bennett and Ogston, 1981; Wright, 1982). Closer examination of these reactions reveals that the multiple plasma systems, coagulation, fibrinolysis, complement and kinin generation, become intimately linked (Pesce and Dosekun, 1983; Sundsmo and Fair, 1983; Colman, 1984). Vasodilatation is caused by a number of agents including bradykinin which is a plasma product from the contact activation mechanism and results from cleavage of HMW kininogen by kallikrein. Other vasodilators are found in plasma and platelets and interact with collagen, thrombin or components of the complement system. Permeability of vascular tissue has been associated with activation of factor XII, with kinins generated from HMW kininogen and with small peptides derived from the complement systems (C3a and C5a) that stimulate the release of vasoactive amines from mast cells. Neutrophils (polymorphonuclear leukocytes) are elicited from plasma to the site of injury through the action of chemotaxis (Wright, 1982). Kallikrein and tissue plasminogen activator are chemotactic for neutrophils as well as the complement components, C3, C5 and C5b67 (Bennett and Ogston, 1981). As part of the host defense mechanism, neutrophils penetrate the vascular site, undergo degranulation and remove foreign material by phagocytosis (opsonized microorganisms, cellular debris or immunoglobulin and complement fixed to a substrate which is not susceptible to phagocytosis) (Wright, 1982). Extracellular release of chemical substances such as proteases and lysosomal constituents by neutrophils causes
tissue damage and inflammation (Wright, 1982).

The role of kallikrein and factor XIIa in the neutrophil aggregation and degranulation reactions has been described (Schapira et al., 1982; Wachtfogel et al., 1983; Wachtfogel et al., 1986). The involvement of factor XIIa was measured by elastase and lactoferrin released from neutrophil granules in vitro (Wachtfogel et al., 1986). This release was stimulated by increasing amounts of FXIIa. Furthermore, Plow (1982) showed that neutrophil release of elastase was associated with blood coagulation. Pertinent to these events is that elastase is a major fibrinolytic protease that dissolves clots during the inflammatory process (Plow, 1980). The possible role of contact activation in the inflammatory response mechanism is clearly demonstrated in the neutrophil reactions and the release of elastase. Components of the complement system may also be acted upon by the contact activation proteases since it has been demonstrated that rabbit kallikrein generated chemotactic activity for neutrophils from rabbit C5 (Wiggins et al., 1981).

Dissolution of the fibrin deposited at sites of inflamed tissue is required in the resolution phase of the inflammatory process (Bennett and Ogston, 1981). As described above, elastase released from neutrophils is one mechanism which is independent of the fibrinolytic system that results in the removal of fibrin. However, plasmin is also often available at these sites of injuries (Bennett and Ogston, 1981). Plasminogen is probably activated by plasminogen activators released from endothelial cells. The plasminogen activators can be activated via the contact activation system. Factor XIIa and kallikrein would be available since coagulation would have been initiated earlier on at the damaged tissue site to produce the fibrin clot.
4. Conversion of Prorenin to Renin

Angiotensin II is a hypertensive agent required in the regulation of blood pressure. Angiotensinogen is synthesized as an inactive precursor which is partially hydrolyzed by renin to produce Angiotensin I, a less active form of Angiotensin II (Derks et al., 1976). A dipeptidase-converting enzyme, found in pulmonary endothelial cells, removes a dipeptide from Angiotensin I producing Angiotensin II (Erdos, 1979). The enzyme also destroys bradykinin in a similar reaction by cleaving the carboxy-terminal dipeptide, Phe-Arg (Erdos, 1979). Renin also circulates in plasma as an inactive zymogen. Prorenin can be activated in vitro by partial proteolysis by plasmin, trypsin, urinary kallikrein or plasma kallikrein (Cochrane and Griffin, 1982). It has been suggested that factor XIIa does not cleave prorenin directly but becomes necessary in its activation of plasma prekallikrein (Bennett and Ogston, 1981; Cochrane and Griffin, 1982). The importance of the contact system activation mechanism in vivo in the prorenin conversion reaction is questionable although plasma deficient in factor XIIa or kallikrein exhibit abnormal surface dependent activation of prorenin (Derks et al., 1979; Sealy et al., 1979).

C. THE STRUCTURE OF EUKARYOTIC STRUCTURAL GENES

1. Promoters

Expression of genes transcribed by RNA polymerase II is regulated by regions in the genome which are usually 5' to the coding sequence and are not transcribed (reviewed in Hansen and Sharp, 1984; Dynan and Tjian, 1985). These regions are important in regulating transcription levels, in determining specific sites of transcript initiation and in
determining rates of gene transcription. These various functions of the 5' region are accomplished by recognition of sequence motifs by RNA polymerase II and by important regulatory proteins, most of which are unknown (Hansen and Sharp, 1984). The composite promoter consists of three regions. The first is the initiation region, spanning from nucleotide -50 to +10 where nucleotide +1 is the point of transcription initiation, and including the "TATA" and cap sites. The second region is the immediate upstream sequence extending from nucleotides -50 to -110. The third region consists of the enhancer (activator or potentiator) sequences whose positions vary widely (Hansen and Sharp, 1984).

The best described and most commonly found features of 5' regions in genes are the "TATA" and "CAAT" sequences that are located approximately 30 bp and 80 bp respectively, from the start of the coding sequence in the gene (Breathnach and Chambon, 1981; Hansen and Sharp, 1984). The role of the "TATA" sequence is to allow transcription to proceed at approximately 25 bp downstream at a single site and in some cases to enhance transcription rates (Hansen and Sharp, 1984). The "CAAT" sequences are upstream regulatory elements initially found in globin genes where they act by increasing transcription rates (Benoist et al., 1980; Hansen and Sharp, 1984). However, sequence elements that enhance transcription can vary depending on the gene and not all genes have CAAT elements (Hansen and Sharp, 1984). For example, the presence of inducible enhancer sequences have been described for a number of genes (Kessel and Khoury, 1983) including the heat shock genes (Lindquist, 1986), mouse metallothionein I gene (Brinster et al., 1982), the human interferon-β gene (Kessel and Khoury, 1983) and the human interferon-α gene (Ragg and Weissmann, 1983). Inducible genes are responsive to heat, high metal concentrations, hormones and other agents (Kessel and Khoury, 1983).
2. Introns and Exons

RNA Polymerase II transcribed genes and some tRNA genes are found in the genome as mosaics of coding (exon) and noncoding (intron) regions. The discovery of heterogeneous nuclear RNA lead to the realization that transcripts entering the cytoplasm that were polyadenylated were in reduced forms of this nuclear RNA (Breathnach and Chambon, 1981). Introns were described (Breathnach et al., 1977) and the splicing apparatus revealed (Abelson, 1979) which elucidated the heterogeneous nuclear RNA processing mechanism in the nucleus (see Lewin, 1983). This splicing mechanism for RNA Polymerase II depends on recognition of conserved sequences at the donor and recipient sites of the introns by small nuclear RNAs and the subsequent cleavage by splicing proteins (Lerner and Steitz, 1979). The sizes of introns are not conserved between members of gene families and vary greatly. The minimum size of a eukaryotic intron is about 80 bp which may be due to the constraints caused by the splicing mechanism (Wieringa et al., 1984).

3. Transcription and Processing

Newly transcribed hnRNAs are processed in the nucleus in order to direct their export from the nucleus to the cytoplasm and to allow correct protein translation to occur at the appropriate start site in the mRNA. Sites of initiation (Baker and Ziff, 1981; Manley, 1983) and termination (Birnstiel et al., 1985) of transcription in genes are poorly defined and are subject to variation amongst eukaryotic genes. Transcription of most genes begins with a purine but cytosine or uridine have also been found (Baker and Ziff, 1981; Manley, 1983). Nontranslated 5' ends of mRNAs are usually 20 to 80 bp in length in 70% of mRNA 5' sequences compiled by Kozak (1984). The processing reactions include (i) 5' capping, (ii)
removal of part of the 3' untranslated region and subsequent addition of a poly A+ tract (approximately 200 bases in length) and (iii) removal of the the introns. Capping at the 5' end is important for efficient splicing of introns (Grabowski et al., 1985) and in translation (Shatkin, 1985). Polyadenylation occurs after part of the 3' untranslated region has been removed by a nuclease (Breathnach and Chambon, 1981) approximately 20 bp downstream of a conserved AAUAAA site (Proudfoot and Brownlee, 1976). This sequence is required for nuclease activity but is not required in the polyadenylation reaction (Montell et al., 1983). After the introns have been excised the mature transcript can pass through the nuclear matrix and enter the cytoplasm where it is translated on the ribosome to produce the polypeptide chain (Nevins, 1983).

D. MECHANISMS OF GENE EVOLUTION

1. Gene Duplication

Gene families have arisen by the mechanism of gene duplication (Li, 1983) resulting in proteins with common structure and function (Edgell et al., 1983). If the time of divergence between two related genes is large, then the genes begin to evolve differently and may assume new functions such in the lysozyme-a lactalbumin family (Hall et al., 1982) and in the immunoglobulin superfamily (Hood et al., 1985). The serine protease family of proteins provides a good example of gene duplication events with subsequent mutations during divergence. The resulting proteins maintain some amino acid sequence identity such that they retain the basic structure for their enzymatic reactions as serine proteases (Hewett-Emmett et al.; 1981; Patthy, 1985).
2. Gene Fusions

Variety in the genes can result from the fusion of one gene, or parts thereof, with another unrelated gene to give a new kind of protein (Doolittle, 1985). Polymorphism in the clotting genes has been attributed partially to this mechanism (Doolittle, 1985; Neurath, 1985) during gene duplication events (Calos and Miller, 1980).

3. Exon Shuffling

Complex proteins are usually encoded by genes made up of introns and exons (see above). It has been suggested (Gilbert, 1978; Blake, 1978; 1983; Rogers, 1985) that the proteins consist of functional and structural domains that can be identified in related proteins. In addition, introns demarcate these coding regions as separate exons in eukaryotic genes. The importance of this concept in evolution is that these exons can be inserted by an exon shuffling mechanism into other genes within the genome (Gilbert, 1978; Blake, 1978; 1983; Rogers, 1985). In this model, protein domains on separate exons can be exchanged between genes and may confer new structure and possibly new function to the recipient molecule.

E. EVOLUTION OF AMINO ACID AND DNA SEQUENCE

1. Molecular Clock

A phylogeny of species can be established by determining the rate of evolution of a conserved protein family within different species (Zuckerkandl and Pauling, 1965; Wilson et al., 1977; Li et al., 1985). The validity of this argument is dependent on random mutation
of DNA occurring at a constant rate which manifests itself as a constant rate of amino acid substitution (Zuckerkandl and Pauling, 1965). This constant rate of change of protein sequence is called the "molecular clock" and when established for a particular protein, divergence times between species can be assessed. Sometimes, protein evolution is not constant due to selective pressures on organisms (Wilson et al., 1977). Also, it has been shown (Wilson et al., 1977; Li et al., 1985) that regions within a protein diverge at different rates. This allows the detection of protein domains that may be more functionally important to the organism than other regions of the molecule.

2. Intron Sliding

Members of protein families exhibit regions of variation in the protein that would have arisen not from amino acid point mutations nor from insertion of domains from other proteins but from intron sequence within the gene (Craik et al., 1983). The mechanism by which small amino acid regions have been inserted into proteins is called intron/exon sliding, and if the junction has shifted in such a way as to maintain the correct translational reading frame, then it is possible that the new protein sequence could be retained through natural selection. These small inserts are often found on the surface of molecules where tertiary structural modifications are most easily tolerated, and they are usually not greater than 13 amino acids in length (Craik et al., 1983). Examples of intron sliding have been found in the dihydrofolate reductase and serine protease gene families (Craik et al., 1983).
F. PROTEIN DOMAINS OF THE COAGULATION AND PLASMINOGEN ACTIVATOR GENES

1. Trypsin-Like Family of Serine Proteases

The trypsin-like family of serine proteases has been conserved since the time of divergence of prokaryotes and eukaryotes (Neurath, 1984). These enzymes exhibit diverse functions such as digestion, coagulation, fibrinolysis, complement activation, neuropeptide processing and fertilization of germ cells (Neurath and Walsh, 1976). The serine protease domains of the trypsin-like enzymes share approximately 40% amino acid sequence identity and have been assumed to share a common mechanism of proteolysis (Young et al., 1978; Hewett-Emmett et al., 1981).

The trypsinogen, chymotrypsinogen and proelastase proteins differ from the coagulation and fibrinolysis enzymes in their amino-terminal domains. The digestive enzymes all have short amino-terminal extensions which are released upon activation by proteolysis. However, the coagulation and fibrinolysis proteins have very large amino-terminal fragments that consist of multiple putative domains within these proteases (Figure 4) (Jackson and Nemerson, 1980; Zur and Nemerson, 1981; Doolittle, 1985). Protein homologies were detected by alignment of common amino acid residues, especially the cysteine residues which suggest conserved tertiary structure in disulfide bridging patterns. These regions may have important roles in regulation and selective activation on and by other serine proteases of the cascade (Jackson and Nemerson, 1980).

The vitamin K-dependent proteins (see above) have a common "Gla-rich" domain which is
Figure 4: Amino Acid Sequence Homologies in Coagulation Factor Zymogens

Comparison of the structures of coagulation and fibrinolytic zymogens to trypsinogen. The solid bar represents the catalytic region in the proteases, the cross hatched region represents the Gla region, K represents the kringles, E represents regions homologous to epidermal growth factor precursor, 1 represent regions homologous to the type I homology of fibronectin, and A represents the homologous regions found in factor XI and prekallikrein. The lengths of the bars are approximately proportional to the lengths of the polypeptide chains. Arrows represent the locations of peptide bonds that are cleaved during the activation of the zymogens. Solid lines below the proteins represent disulfide bridges and do not necessarily represent their true locations.
contiguous with a prepro leader sequence (see Fung et al., 1985). These proteins include prothrombin (MacGillivray and Davie, 1984; Degen et al., 1983), factor VII (Hagen et al., 1986), factor IX (Kurachi and Davie, 1982; Jaye et al., 1983), factor X (Fung et al., 1984, 1985; Leytus et al., 1986) and protein C (Long et al., 1984; Foster and Davie, 1984; Beckmann et al., 1985) and protein S (Dahlback et al., 1986; Lundwall et al., 1987; Hoskins et al., 1987). The amino acid sequence of a seventh vitamin K-dependent protein called protein Z has recently been determined (Hojrup et al., 1985) though the role of this protein is not known.

A "kringle" structure that consists of 80 amino acid residues with six characteristic cysteine residues (Magnusson et al., 1975) is found in plasminogen (Sottrup-Jensen et al., 1978), tissue-type plasminogen activator (Pennica et al., 1983), urokinase plasminogen activator (Verde et al., 1984) as well as in prothrombin (Magnusson et al., 1975). The function of the kringle is not known but plasminogen kringle 4 may be associated with binding to fibrin molecules.

Another region found in both blood coagulation and fibrinolysis cascade zymogens is the epidermal growth factor like domain (Blomquist et al., 1984; Doolittle et al., 1984). This region has been identified in factor VII (Hagen et al., 1986), factor IX (Kurachi and Davie, 1982; Jaye et al., 1983) factor X (Fung et al., 1984, 1985; Leytus et al., 1984) protein C (Long et al., 1984; Foster and Davie, 1984; Beckmann et al., 1985), protein Z (Hojrup et al., 1985) and the plasminogen activators (Pennica et al., 1983; Verde et al., 1984). The epidermal growth factor domain is present in the 19 kdal fragment of vaccinia virus and several other proteins (Blomquist et al., 1984; Doolittle et al., 1984). In many of the EGF-like regions, a specific aspartic acid or asparagine residue is converted to
\( \beta \)-hydroxyaspartic acid or \( \beta \)-hydroxyasparagine, respectively.

Tissue-type plasminogen activator has a region that is homologous to the type I homology found in fibronectin (Ny et al., 1983; Freizner-Degen et al., 1985). This region has fibrin binding activity in fibronectin (Petersen et al., 1983; Yamada, 1983) and bears approximately 40% amino acid sequence identity with tissue-type plasminogen activator (Ny et al., 1984; Friezner-Degen et al., 1985).

G. EVOLUTION OF THE SERINE PROTEASES

Analysis of the blood coagulation and fibrinolysis families of serine proteases aids in determining the evolutionary process of the genes and how they diversified. Furthermore, by comparing these genes between species it is possible to determine a time scale within which these evolutionary events had occurred. Patthy (1985) provides a model utilizing the exon shuffling mechanism to describe how these families of proteins arrived at their present day form and from this type of speculation we may be able to determine the relationships of newly characterized blood coagulation and fibrinolysis proteins. With the help of biotechnology it has become possible to predict the amino acid sequence of many of these low abundance plasma and tissue proteins. For example cDNA sequences of blood coagulation proteins prothrombin (MacGillivray et al., 1980; Degen et al., 1983; MacGillivray and Davie, 1984; MacGillivray et al., 1986; Jorgensen et al., 1987), factor VII (Hagen et al., 1986), factor IX (Kurachi and Davie, 1982; Jaye et al., 1983; McGraw et al., 1986) factor X (Fung et al., 1984, 1985; Leytus et al., 1984; 1986) protein C (Long et al., 1984; Foster and Davie, 1984; Beckmann et al., 1985) and protein S (Dahlback et al., 1986; Lundwall et al., 1987; Hoskins et al., 1987) have been determined. Furthermore, the cDNA sequences of
the fibrinolytic proteins plasminogen (Malinowski et al., 1984), tissue-type plasminogen activator (Pennica et al., 1983) and urokinase (Verde et al., 1984) have been determined.

From the cDNA sequences it is possible to predict the amino acid sequences of the zymogens as well as to predict any regions which are cleaved by the cell prior to secretion such as the prepro peptides and the signal peptides. Therefore, this kind of technology has been extremely beneficial in the complete characterization of proteins. However, these techniques do not provide secondary structure data such as disulfide bridging patterns nor the identification of glycosylated residues.

**H. OBJECTIVES**

The contact system of activation of the various plasma systems described above may be an important physiological phenomenon required for maintaining homeostasis in the organism. Central to this mechanism is the glycoprotein, factor XII, which possesses the surface-binding property that is critical in the initiation reaction. The physical and biochemical means by which factor XII plays a role in the contact activation may be inferred from the protein sequence. Recombinant DNA technology will be used to obtain cDNA sequences from which the amino acid sequence can be predicted to circumvent the expensive and time consuming processes of protein purification and sequencing using the standard Edman degradation protein sequencing method. The cDNA can also be used as a probe for obtaining the factor XII gene.

The evolutionary history of factor XII will be determined from the gene structure since the family of serine proteases is well characterized and it may be possible to gain a better
understanding of the evolution of this old and complex family of genes. Many proteins with common function belong to the same family and it is possible that FXII is more closely related to the surface activating proteases, prekallikrein or factor XI, or it may even have regions of protein homology related to the binding domain of HMW kininogen that had been acquired through an exon shuffling mechanism in the genome. Clearly, from the determination of both the amino acid sequence and the gene structure for human factor XII, greater insight will be gained into the function and evolution of this protein.
II. MATERIALS AND METHODS

A. MATERIALS

Agarose, acrylamide, bisacrylamide, urea, ammonium persulphate, Sephadex G-25 and TEMED were from Bio-Rad Laboratories. Oligo (dT) cellulose (type 7) was from P.L. Biochemicals. Yeast extract, casamino acids, bacto-tryptone and bacto-agar were Difco grade from the Grand Island Biological Company. NZamine type A was from Sheffield Products. Nitrocellulose sheets and circles (82 and 132 mm) (BA-85) of 0.45 μm pore size were from Millipore or Schleicher and Schuell. 32P-labeled deoxyribonucleotide triphosphates (3000 Ci/mmol; 1 Ci = 3.7 x 10^{10} Bq) were from Amersham. Phenol (British Drug Houses Ltd.) was redistilled before use (boiling point 180°C) and frozen in portions at -20°C. Deoxy- and dideoxy-ribonucleotide triphosphates and the M13 sequencing primer (heptadecanucleotide) were from PL-Pharmacia. Isopropyl-β-D-thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-indolylβ-D-galactopyranoside (X-Gal), ethidium bromide, yeast transfer RNA (tRNA), ampicillin, tetracycline, chloramphenicol, ribonuclease A and deoxyribonuclease I were from Sigma. Cesium chloride was from Cabot Berylco Ltd. Ultragel AcA54 was from LKB. Most of the oligodeoxyribonucleotides including some of the M13 sequencing primer used, were synthesized on an Applied Biosystems 890A DNA Synthesizer (by Tom Atkinson, Dept. of Biochemistry) and purified by denaturing polyacrylamide gel electrophoresis prior to use (Atkinson and Smith, 1984). One oligonucleotide pool was synthesized and purified by Dr. M. Zoller at Cold Spring Harbor Laboratories, New York. All other chemicals were of reagent grade or better and were purchased from either Sigma Chemical Co., Fisher Scientific, or British Drug Houses Ltd.
Materials and Methods / 40

Restriction endonucleases, T4 DNA ligase, T4 DNA Polymerase, T4 polynucleotide kinase and bovine serum albumin (nuclease free) were from New England Biolabs, Bethesda Research Laboratories or PL-Pharmacia. Nuclease S1 was from Boehringer-Mannheim and avian myoblastosis virus reverse transcriptase was from Promega Biotec. DNA polymerase I and DNA polymerase I Klenow fragment were from Boehringer-Mannheim or PL-Pharmacia. Adult human liver was obtained from kidney donor patients. Liver samples were cut into 0.5 cm slices, rinsed in sterile PBS (10 mM sodium phosphate pH7.5, 150 mM NaCl), frozen in liquid nitrogen and stored at -70°C. Human genomic DNA was isolated by Dr. Colin Hay (Dept. of Biochemistry) and Escherichia coli rRNA was provided by Dr. P. Dennis (Dept. of Biochemistry).

B. STRAINS, VECTORS AND MEDIA

1. Bacterial Strains

*E. coli* MC1061 (araD 139, Δ(ara, leu) 7697, ΔlacX74, galU-, gal K-, hsr -, hsm +, str A) (Casadaban and Cohen, 1980) was host for the liver cDNA library cloned into the plasmid pKT218 containing cloned liver cDNA (Prochownik et al., 1983). *E. coli* C600 (F-, thi-1, thr-1, leuB6, lacY1, tonA21, supE44,λ-) (Maniatis et al., 1982) was host for human genomic DNA cloned into the λCharon28 vector (Blattner et al., 1977). *E. coli* JM83 (ara, Δlacpro, strA, thi-, φ80, lacZΔM15) (Vieira and Messing, 1982) was host for the transformation of pUC13 vector (Vieira and Messing, 1982) and was obtained from Dr. Jim Stone, Dept. of Microbiology, UBC. *E. coli* JM103 (Δlacpro, supE, thi-, strA, sbcB15, endA, hsdR' F', traD36, proAB, lacIQ, lacZΔM15 (Messing, 1983) was host for transformation of M13 mp 8, 9, 18 and 19 vectors (Messing, 1983). *E. coli* RY1088
Materials and Methods / 41

[ΔlacU169, supE, supF, hsdR-, hsdM+, metB, trpR, tonA21, proC::Tn5(pMC9), pMC9 is pBR322-lacIQ] (Young and Davis, 1983a,b) was host for screening and isolation of DNA cloned into the λgt11 vector (Young and Davis, 1983a,b).

2. Vectors

The M13 mp 8, 9, 18 and 19 vectors (Messing, 1983) were used for DNA sequence analysis. Small fragments of cloned DNA were subcloned into the vector pUC13 (Vieira and Messing, 1982), obtained from Dr. Mark Zoller, (Dept. of Biochemistry, University of British Columbia) for restriction endonuclease mapping and hybridization probes.

3. Media

The medium for growth of *E. coli* hosts for λ clones was NZCYM (Maniatis *et al.*, 1982), which is 10g NZamine type A, 2g MgCl₂, 5g NaCl, 5g yeast extract, 1g casamino acids per litre and the pH was adjusted to 7.5 with NaOH. Phage λ libraries were plated on NZCYM-agar plates (1.5% w/v) plates with an overlay of NZCYM-agarose (0.7% w/v) when screening or NZCYM-agar (0.7% w/v) when titering phage libraries or stocks. Phage were diluted or stored in SM buffer (5.8g NaCl, 2g MgSO₄·7H₂O, 50 ml 1M Tris-HCl pH7.5, 5 ml 2% gelatin per liter). Luria broth (5g yeast extract, 10g bacto-tryptone and 10g NaCl per liter) (Maniatis *et al.*, 1982) was used as medium for growth of *E. coli* JM83 containing the pUC13 plasmid derivative. For selection of *E. coli* JM83 colonies transformed with pUC derivatives containing cloned DNA, cultures were plated on LB-agar (1.5% w/v) plates supplemented with 50 ug/ml ampicillin. A liver cDNA library in *E. coli* MC1061/pKT218 vector screened on LB-agar (1.5% w/v) plates supplemented with tetracycline (12.5 ug/ml).
All DNA cloned into M13 vectors and transformed into JM103 was amplified in YT medium (Maniatis et al., 1982), which is 5g yeast extract, 8g bacto-tryptone and 5g NaCl per liter. Phage M13 transformants were plated on YT-agar (1.5% w/v) plates overlayed with YT-agar (0.7% w/v). E. coli JM103 was maintained on minimal medium plates (Messing, 1983), which were made up as follows: 3g of agar in 160 ml H₂O (sterilized and cooled to 55°C) was mixed with 40 ml 5X salts (2.1g K₂HPO₄, 0.9g KH₂PO₄, 0.2g (NH₄)₂SO₄, 0.1g NaCitrate.7H₂O per 40 ml), 2 ml 20% glucose, 0.2 ml 20% MgSO₄·7H₂O and 0.1 ml 10 mg/ml thiamine. Each of these solutions was sterilized separately by autoclaving except for thiamine which was filter sterilized. Large-scale cultures (more than 500 ml) of E. coli JM83 or MC1061 containing pUC or pKT218 plasmid vectors, respectively, were grown in Luria broth in the presence of the appropriate antibiotic. These cultures were further treated with chloramphenicol (250 ug/ml) in the mid-log phase of growth.

C. GEL ELECTROPHORESIS

1. Neutral Agarose Gels

Electrophoresis grade agarose was boiled in dH₂O (added according to the percentage w/v agarose gel required) for 10 minutes, cooled to 50°C and the appropriate volume of 50X TAE buffer (2M Tris-OH, 1M glacial acetic acid, 0.1M EDTA, pH7.2) was added to a final dilution of 1X TAE (Maniatis et al., 1982). To stain nucleic acids in the gel, a volume of ethidium bromide solution (10 mg/ml) was added to the cooled gel before pouring so that the final concentration of ethidium bromide was 0.5 ug/ml. The gel was poured into a mold and was allowed to set for 1 hour. DNA samples were adjusted to 1X loading buffer with a 10 times concentrated stock (in 30% ficoll, 0.2% xylene cyanol, 0.2% bromophenol blue) before
loading. Electrophoresis was carried out in 1X TAE buffer at 1-3 volts/cm after which the DNA was visualized by irradiation of UV light (260 nm or 360 nm).

2. Formaldehyde Agarose Gels

RNA fragments were separated in agarose gels prepared with a denaturing agent such as formaldehyde (Maniatis et al., 1982) as follows. An agarose solution (usually 0.6% w/v final concentration) was boiled and cooled to 60°C. The gel solution was adjusted to 2.2M formaldehyde (pH less than 4.0) and 1X running buffer from a 5X running buffer stock (0.2M MOPS pH7.0, 50mM sodium acetate, 5mM EDTA) and poured into a mini-gel mold. RNA samples were prepared in a 20μl loading volume consisting of 4.5μg RNA (up to 20μg), 2.0μl 5X running buffer, 3.5μl formaldehyde (pH4.0) and 10.0μl formamide and heated for 20 minutes at 50°C. The samples were electrophoresed for 1.5 hours at 15V/cm, and lanes containing ribosomal markers were excised and prepared for staining while the region of the gel containing mRNA was treated for a Northern blot transfer (below). The gel with the marker lanes was washed with sterile dH₂O for 2 hours with 5 changes of water. The gel was soaked in 0.1M ammonium acetate for 1 hour with one change of buffer followed by a final solution of 0.1M ammonium acetate, 0.1M β-mercaptoethanol and 0.5μg/ml ethidium bromide for 1 to 3 hours. Prior to visualization with UV light, the gel was destained for 2 hours in the same buffer but without ethidium bromide.

3. Neutral Polyacrylamide Gels

A 30% (w/v) stock solution of acrylamide (29:1 acrylamide:bisacrylamide) in H₂O was prepared and diluted according to the percent gel required. A 10X TBE buffer stock
solution (0.89M Tris-OH, 0.89M boric acid, 25 mM EDTA, pH 8.3) was added so that the gel solution was 1X TBE buffer and the solution was degassed using a water aspirator. Ammonium persulphate (0.066% w/v) and TEMED (0.04% w/v) were added to initiate polymerization, the gel was poured and was allowed to polymerize for 1 hour. A 10X stock of loading buffer was added to the DNA samples which were then adjusted to 0.3% ficol, 0.002% bromophenol blue, 0.002% xylene cyanol. The gel was subject to electrophoresis 6V/cm in 1X TBE buffer. DNA was visualized by staining the gel with a solution of ethidium bromide (1µg/ml) in 1X TBE buffer followed by irradiation under UV light (260nm).

4. Denaturing Polyacrylamide Gels

A 40% (w/v) stock solution of acrylamide (38:2 acrylamide:bisacrylamide) was prepared, and an appropriate volume was added to dH$_2$O, 10XTBE buffer (for a final concentration of 1XTBE buffer) and urea (for a final concentration of 8.3M) to attain the correct percentage of gel required. The urea was dissolved in the solution in a 37°C water bath, and then the solution was degassed. Polymerization was initiated with the addition of ammonium persulphate and TEMED to final concentrations of 0.066% (w/v) and 0.024% (w/v), respectively. The gel was poured and allowed to polymerize for 2 hours. The DNA samples were applied after heating at 90°C in loading buffer (98% deionized formamide, 10mM EDTA, 0.02% xylene cyanol, 0.02% bromophenol blue) for 3 minutes. DNA fragments separated in denaturing gels were visualized by autoradiography after drying under vacuum with a Bio-Rad gel drier at 80°C for 20-30 minutes, and exposing to Kodak XK-1 film, with or without intensifying screens (Dupont Lightning Plus).
D. ISOLATION OF DNA

1. Small-scale Isolation of Plasmid DNA

Overnight cultures of *E. coli* containing recombinant plasmids were grown in Luria broth or YT medium in the presence of the appropriate antibiotic at 37°C. Plasmid DNA was isolated from 1.5 ml of the culture according to the method of Birnboim and Doly (1979) and described in Maniatis *et al.* (1982). The bacterial culture was centrifuged for 30 seconds in an Eppendorf microfuge and the supernatant was discarded. The pellet was resuspended in 100μl of a solution (at 4°C) containing 50 mM glucose, 10mM EDTA, 25 mM Tris-HCl, pH 8.0, and 4 mg/ml lysozyme and incubated at room temperature for 5 minutes. A lysis solution (200μl of 0.2% NaOH, 1% SDS) was added and incubated on ice for 5 minutes. The SDS was precipitated by the addition of 150μl of a potassium acetate solution (60 ml 5M potassium acetate, 11.5 ml glacial acetic acid, 28.5 ml dH₂O, pH 4.8) followed by incubation at 4°C for 5 minutes. Cellular debris and potassium dodecylsulfate were removed by centrifugation in an Eppendorf centrifuge for 10 minutes at 4°C. The supernatant (350μl) was removed carefully and extracted with an equal volume of phenol:chloroform (1:1, v:v). The phases were separated by centrifugation for 2 minutes in an Eppendorf centrifuge, and two volumes of 95% ethanol were added to the supernatant. The mixture was incubated at room temperature for 5 minutes. The DNA was recovered by centrifugation in an Eppendorf centrifuge for 5 minutes. The pellet was washed with 1 ml of 70% ethanol, recentrifuged and air dried. The plasmid DNA pellet was resuspended in 25μl TE buffer (10mM Tris-HCl, pH 8.0, 1 mM EDTA).
2. Large-scale Isolation of Plasmid DNA

Large volumes (500 ml) of Luria broth were inoculated with 3 ml of a fresh overnight culture of *E. coli* containing plasmid DNA (antibiotics were added depending on the host resistance) and grown at 37°C with vigorous shaking until the OD600nm of the culture was 0.5. At this time 250 mg chloramphenicol was added, and the culture was shaken for 12-16 hours at 37°C. Cells were collected by centrifugation at 5 krpm for 10 minutes in a GS-3 rotor (Sorvall) at 4°C. The supernatant was discarded, and the cells were frozen at -20°C. The cells were resuspended on ice in a solution of 25% (w/v) sucrose, 50 mM Tris-HCl, pH 8.0 and 1.5 mg/ml lysozyme so that the final volume of the cell suspension was 5.0 ml. The suspension was transferred to a 25 ml Oakridge tube (Beckman) and kept on ice. The cells were lysed by the addition of 10 ml of a 0.2N NaOH, 1% SDS solution and incubated on ice for 10 minutes. The SDS and cell debris were precipitated with 7.5 ml of a 3M potassium acetate solution pH4.8 and incubated on ice for 10 minutes. (The potassium acetate solution was prepared from 60 ml of 5M potassium acetate to which were added 11.5 ml glacial acetic acid and 28.5 ml dH$_2$O.) The lysis mixture was centrifuged at 35 krpm, 4°C for 30 minutes in a Ti60 rotor (Beckman). The supernatant was decanted into a 50 ml centrifuge tube, 0.6 volumes of isopropanol was added, and the DNA was precipitated at room temperature for 15 minutes. The plasmid DNA was collected by centrifugation at 9 krpm for 30 minutes at room temperature in an HB-4 rotor (Sorvall). The pellet was resuspended in TE buffer, and solid CsCl was added for a final concentration of 1 g CsCl/ml plasmid DNA solution. The solution was adjusted to a final volume of 13.5 ml to which was added 300 μl of a solution of ethidium bromide (10 mg/ml). The solution was placed into a Beckman heat sealable tube and was centrifuged in a Ti70.1 rotor (Beckman) for 20 hours at 20°C at 60 krpm. Plasmid DNA was observed with UV light irradiation (360 nm) as a
band midway in the tube. The band was isolated by puncturing the tube with an 18 gauge syringe, and the ethidium bromide was removed from the DNA solution by multiple extractions with water saturated isopropanol. When the DNA solution was no longer pink in color, the extraction was repeated two more times, and the solution was diluted 4 times with dH₂O. The DNA was precipitated by the addition of 3M sodium acetate (to a final concentration of 0.2M) and two volumes of 95% ethanol. The suspension was left at -20°C overnight. The precipitated DNA was recovered by centrifugation at 9 krpm and 4°C for 30 minutes in an HB-4 rotor.

3. Small-scale Isolation of λ DNA

A fresh culture of host cells was grown overnight in NZCYM medium containing 0.02% maltose at 37°C with shaking. A plug of agar containing a single phage plaque was added to 100μl of the culture, and the phage were preabsorbed to the host cells for 15 minutes at 37°C. The culture was transferred to 20 ml of prewarmed NZCYM and shaken vigorously at 37°C until total cell lysis was observed. Chloroform (100μl) was added, and the culture was shaken for a further 5 minutes. Cell debris was removed by centrifugation at 5 krpm and 4°C for 5 minutes in an HB-4 rotor. The supernatant was transferred to a 50ml centrifuge tube which contained 6ml of 50% PEG (polyethylene glycol 6000, Carbowax 8000) and 3ml of 5M NaCl, and the phage were precipitated at 4°C overnight. The phage particles were collected by centrifugation at 5 krpm and 4°C for 15 minutes in an HB-4 rotor, and the supernatant was completely removed. The phage pellet was resuspended in 500μl of DNAsel buffer (50mM Tris-HCl pH7.5, 5mM MgCl₂, 0.5mM CaCl₂) which contained 5μg DNAsel and 50μg RNaseA. The reaction mixture was incubated at 37°C for 30 minutes, and then the cellular debris removed by centrifugation in an Eppendorf
Materials and Methods / 48

centrifuge at 4°C for 5 minutes. The supernatant was adjusted to 1% SDS, 5mM EDTA, and then 100μg of proteinase K (predigested for 1 hour at 65°C) was added and incubated at 65°C for 1 hour. The DNA solution was extracted twice with phenol:chloroform (1:1), and the DNA was precipitated by the addition of sodium acetate to 0.3M and 2 volumes of 95% ethanol followed by incubation at -20°C overnight. The precipitated DNA was collected by centrifugation at 9 kRPM and 4°C for 15 minutes in an HB-4 rotor. The pellet was washed with 1 ml 70% ethanol, air dried and resuspended in 50μl TE buffer.

4. Large-scale Isolation of λ DNA

The large-scale isolation of λDNA has been described in Maniatis et al. (1982). A fresh overnight culture of host bacteria was grown in NZCYM containing 0.02% maltose, and the OD600nm was taken. A volume of culture was centrifuged (at 3 kRPM for 5 minutes in an HB-4 rotor at 4°C) which corresponded to 2 x 10^9 bacteria (1 OD600nm = 8 x 10^8 cells/ml). The bacteria were resuspended in 1 ml SM medium, and 2 x 10^7 phage were added. The phage were preabsorbed to the bacteria at 37°C for 15 minutes after which they were transferred to 450 ml of prewarmed NZCYM medium and shaken vigorously until total cell lysis was observed. The culture was adjusted to a final concentration of 1% chloroform and 1M NaCl and shaken for a further 10 minutes. The debris was removed by centrifugation in a GS-3 rotor at 5 kRPM and 4°C for 10 minutes. DNaseI and RNaseA (50μg each) were added to the phage supernatant and incubated at 4°C for 1 hour. The phage particles were precipitated overnight at 4°C by the addition of 33g PEG. Phage were collected by centrifugation in a GS-3 rotor at 5 kRPM and 4°C for 10 minutes. The pellet was resuspended in SM for a final volume of 5 ml after which the suspension was extracted with chloroform. The aqueous supernatant was adjusted to a final volume of 13.5 ml with
SM buffer containing 0.75 g CsCl/ml SM buffer and transferred to a Beckman heat sealable tube. The phage were subjected to isopycnic centrifugation in a Ti70.1 rotor for 16-18 hours at room temperature and 60 krpm. A blue band midway in the gradient was observed by shining a light on the gradient. The band was extracted with an 18 gauge needle, and the CsCl solution containing λphage was adjusted to 0.02% gelatin and then transferred to a dialysis bag. The CsCl solution was exchanged with a solution of 10mM Tris-HCl pH8.0, 25mM NaCl, 10mM MgCl₂ by dialyzing two times against a 1000-fold volume excess of buffer. Extraction of DNA from the CsCl purified phage particles was performed as described for the small-scale preparation of λphage DNA.

5. Isolation of Single-stranded Plasmid DNA

DNA sequence was obtained from purified (Messing, 1983) single-stranded M13 DNA. Agar plugs containing M13 phage plaques representing DNA cloned into the M13mp8,9,18 or 19 vectors were removed from the agar plate, transferred to 2ml YT and grown overnight with shaking at 37°C. Bacterial cells from 1.5 ml of the overnight culture were removed by centrifugation in an Eppendorf centrifuge, and 650μl of the supernatant was transferred to a tube containing 300μl 40% PEG and 300μl 5M NaCl. The phage were allowed to precipitate for 15 minutes at room temperature and were collected by centrifugation in an Eppendorf centrifuge for 5 minutes. The pellet was resuspended in 200μl of low salt buffer (20mM Tris-HCl pH7.5, 20mM NaCl, 1mM EDTA). The phage were extracted with 200μl of phenol (redistilled and saturated in TE buffer) followed by two extractions with an equal volume of phenol/chloroform (1:1, v:v). The DNA in the aqueous phase was precipitated overnight at -20°C in a final concentration of 0.3M sodium acetate with 2 volumes of 95% ethanol. The DNA was collected by centrifugation, and the pellet
was resuspended in 50\(\mu\)l 0.3M sodium acetate and reprecipitated with 2 volumes of 95% ethanol overnight at -20°C. The final DNA pellet was resuspended in 40\(\mu\)l low salt buffer.

**E. ISOLATION OF POLY A+ RNA**

All glassware was baked at 180°C overnight, and pipet tips, centrifuge and Eppendorf tubes and solutions were autoclaved. Human liver RNA was isolated by the method of Chirgwin *et al.* (1979). Frozen liver (-70°C) was homogenized using a Polytron homogenizer in an appropriate volume of buffer [7.5M guanidine hydrochloride (GuHCl), 25mM sodium citrate pH7.0 and 0.1M DTT] so that 10ml/g tissue was obtained. A 10%(w/v) N-lauryl sarcosine solution was added to a final concentration 0.5% (w/v), and the insoluble matter was removed by centrifugation at 5 krpm for 30 minutes in an HB-4 rotor at 4°C. The RNA solution was adjusted to 33% ethanol, and RNA was precipitated overnight at -20°C. RNA was collected by centrifugation at 5 krpm in an SS-34 rotor at 0°C for 30 minutes, and the pellet resuspended in half of the starting volume of GuHCl buffer. The removal of insoluble material by low speed centrifugation, precipitation of RNA and resuspension in a half volume of GuHCl buffer was repeated two more times. The final RNA pellet was resuspended in a small volume of dH\(_2\)O, and the concentration of RNA was determined from absorption at 260nm [1OD260nm = 40 mg/ml (Davis *et al.*, 1980)]. RNA was stored in dH\(_2\)O at -70°C.

PolyA+ mRNA was isolated by chromatography on a column of oligo-dT cellulose (Edmonds *et al.*, 1971; Aviv and Leder, 1972). The solution of total RNA was adjusted to a final concentration of 0.4M sodium acetate pH7.5, 0.1% SDS and applied to the column equilibrated in column buffer (0.4M sodium acetate pH7.5, 0.1% SDS and 1 mM EDTA).
The unbound RNA fraction was reapplied three times. The column was washed until the OD260nm of the eluate was below 0.05. PolyA+ RNA was eluted from the column with 1 mM EDTA, 0.1% SDS. The OD260nm of each fraction was determined and those containing RNA were pooled. The solution of RNA was adjusted to 0.3M sodium acetate pH4.8 to which two volumes of 95% ethanol were added, and the RNA was precipitated overnight at -20°C. RNA was resuspended in dH2O at a high concentration and stored at -70°C.

F. Oligonucleotide Synthesis and Purification

Oligonucleotides were synthesized (by Tom Atkinson, Dept. of Biochemistry) with an Applied Biosystems 380A DNA Synthesizer (Atkinson and Smith, 1984). Two pools of mixed heptadecadeoxyribonucleotides were used as hybridization probes in the screening of the human cDNA liver library. Pool I (5'-dTCRAAYTGRTGNCCCCA-3', where R represents both dG and dA, Y represents dT and dC, and N represents dG, dA, dT and dC) was complementary to the mRNA sequence coding for amino acid residues 133-138 of β-factor XIIa (Fujikawa and McMullen, 1983) (Fig.5). Pool II (5'-dCCYTGRCANGCYTCNGT-3') was complementary to the mRNA coding for residues 182-188 of β-factor XIIa. Two oligonucleotides were synthesized which correspond to parts of the cDNA sequence (5'-CGAAAGTGTTGACTCCA-3' and 5'-GGCCAAAAGGTCTTGGAAA-3') and used to probe for specific regions in the gene and for nuclease S1 mapping and Northern blot analysis.

The synthetic products required purification by polyacrylamide gel electrophoresis and passage through Sep-Pac columns (Atkinson and Smith, 1984). Dried pellets of the newly
synthesized oligonucleotides were resuspended in 90μl dH₂O. A 10μl volume was transferred to 20μl deionized formamide and heated at 90°C for 3 minutes. The oligonucleotide sample and a sample of dye-formamide [xylene cyanol (mobility equivalent to 45 bases) and bromophenol blue (mobility equivalent to 12 bases) (Maniatis et al., 1982)] were applied to a 20% polyacrylamide-8M urea gel (42 cm) and electrophoresed at 1500V until the bromophenol blue had migrated two-thirds of the gel length. The gel was transferred to Saran Wrap and placed over a thin layer chromatography (TLC) plate, and the single-stranded DNA was visualized with UV irradiation (260nm). The band corresponding to the correct length of oligonucleotide was excised and eluted in 1.5 ml of buffer containing 0.5M ammonium acetate, 10mM magnesium acetate by incubation at 37°C overnight. The solution was vortexed vigorously and centrifuged for 5 minutes in an Eppendorf centrifuge and the supernatant removed. The debris was resuspended in 0.5ml of the elution buffer, vortexed, centrifuged and the supernatant pooled with the 1.5ml obtained previously. A Sep-Pac was attached to a 5ml syringe and was washed with 10ml HPLC grade acetonitrile followed by 10ml dH₂O. A new syringe was used for each solution. The sample was loaded onto the small column and pushed through while collecting the eluate in Eppendorf tubes. The column, loaded with DNA, was washed with 5ml dH₂O and 1ml fractions of the wash solution were collected. To elute the oligonucleotide from the SepPac column, 6ml of 60% methanol (or 20% acetonitrile) were pushed through the column, and 1 ml fractions collected. The DNA eluted in the first three fractions was detected by absorbance at 260nm. The concentration was determined from the OD260nm (1OD = 33μg). Fractions containing significant amounts of DNA were dried in a Speed vac at 80°C for 3 hours. The pellet was resuspended in the appropriate volume of dH₂O and was stored at -20°C.
G. LABELING OF DNA

1. Labeling With $^{32}$P by Nick Translation

DNA for use as hybridization probes was labeled with $^{32}$P by nick translation (Maniatis et al., 1975). DNA (100-500 ng) was labeled in 50μl of 50 mM Tris-HCl pH 7.5, 5 mM MgCl$_2$, 0.05 mg/ml BSA, 10 mM β-mercaptoethanol, 20 μM dGTP, 20 μM TTP, 1.4 μM dATP, 1.4 μM dCTP, 70 μCi α-$^{32}$P dATP, 70 μCi $^{32}$P dCTP, 0.2 mM CaCl$_2$, 50 pg DNaseI and 20 units E. coli DNA polymerase I (Kornberg). The reaction mixture was incubated at room temperature for 30 minutes and then terminated by the addition of three volumes of 1% SDS-10 mM EDTA, 25 μg tRNA and by heating at 68°C for 10 minutes. The unincorporated isotope was separated from the labeled DNA by chromatography on a column of Ultrogel AcA54 (5 ml) equilibrated in 10 mM Tris-HCl pH 7.5, 200 mM NaCl and 0.25 mM EDTA. Labeled DNA was eluted in the same buffer and collected in 200μl fractions from the column. Single-stranded labeled DNA used as hybridization probes was obtained by treating the labeled duplex DNA with a tenth of a volume 1 M NaOH at 68°C for 10 minutes followed by the addition of a tenth of a volume 1.5 M NaH$_2$PO$_4$ (untitrated) just prior to use.

2. Labeling with T4 Polynucleotide Kinase

Oligonucleotides were 5'-end labeled with T4 polynucleotide kinase (Chaconas and van de Sande, 1980) in a 20μl reaction volume containing 33 pmoles oligonucleotide, 2.5μCi/μl γ-$^{32}$P ATP, 0.1 M Tris-OH pH 8.0, 0.005 M dithiothreitol, 0.01 M MgCl$_2$ and 4 units T4 polynucleotide kinase. The reaction was incubated at 37°C for 45 minutes and stopped.
with 0.005M EDTA (pH8.0) and heated at 68°C for 10 minutes. Unincorporated γ-32P ATP was removed by chromatography on a column of Sephadex G-25 (superfine) equilibrated and eluted with 5 mM EDTA (pH 8.0). The excluded fraction was added directly to the hybridization mixture. The specific activities obtained were usually approximately 10^6 cpm/pmole.

**H. SCREENING A PLASMID LIBRARY**

An adult human liver cDNA library (Prochownik et al., 1983) was generously provided by Dr. S. H. Orkin (Children's Hospital Medical Center, Boston). This library consists of double-stranded cDNA (>500 bp in length) cloned into the PstI site of pKT218 by homopolymeric dG:dC tailing. Initially, 2 x 10^5 colonies of the cDNA library were plated directly onto ten cellulose nitrate filters (132mm diameter) placed on Luria broth plates containing tetracycline (12.5 mg/ml). The colonies were grown at 37°C until they reached a diameter of 1-2mm from which two sets of replica filters were prepared (Hanahan and Meselson, 1980; Fung et al., 1984). The lawn of bacteria was oriented by stabbing with an 18 gauge needle through the filter and the agar. The filters containing colonies were removed and another filter disc was placed on top of the first filter, and the orientation marks were placed on the new filter. The plaque lift procedure was repeated with a new filter disc. The replica filters were transferred to fresh LB-tetracycline plates, and the colonies were grown at 37°C until they were 3-4mm in diameter after which the filters were transferred to LB-chloramphenicol (250μg/ml) plates and incubated overnight at 37°C. The master filters were stored on LB-tetracyline plates at 4°C. Colonies were lysed on the filters by soaking the filters twice for 20 minutes on Whatman 3MM paper wetted in a solution of 0.5M NaOH, 1.5M NaCl. The replica discs were neutralized in a similar manner
Materials and Methods / 55

in a solution of 1M Tris-HCl pH7.5 for 20 minutes followed by treatment with 0.5M Tris-HCl pH7.5, 1.5M NaCl for 20 minutes. The nitrocellulose filters discs were air dried and baked at 68°C for 16-18 hours.

The filters were hybridized and washed according to conditions described in Fung et al. (1984). Prior to hybridization, the filter discs were washed three times in 6XSSC (1XSSC is 0.15M NaCl, 0.015M sodium citrate pH7.0) to remove cell debris and prehybridized in 6XSSC, 2X Denhardt's solution (1X Denhardt's is 0.02% BSA, 0.02% ficol, 0.02% polyvinylpyrrollidone) at 68°C for at least 6 hours. One set of replica filters was hybridized with the \( ^{32} \text{P} \)-labeled Pool I oligonucleotides, and the other set of filters was hybridized to the \( ^{32} \text{P} \)-labeled Pool II oligonucleotides. Each hybridization solution contained 6XSSC, 2X Denhardt's solution, 0.2% SDS and 50-100 x 10^4 cpm \( ^{32} \text{P} \)-labeled oligonucleotide. The filters were hybridized for 18 hours at 37°C and then washed in 6XSSC at room temperature for 5 minutes, 37°C for 5 minutes and then two times at 42°C for 5 minutes. Positive clones were detected by autoradiography at -70°C with an intensifier screen. Colonies that hybridized to both pools of oligonucleotides were rescreened at lower colony density until all the colonies plated were positive.

The cDNA library was also screened using a \( ^{32} \text{P} \)-nick translated hybridization probe specific for FXII. Preparation of replica nitrocellulose filters and their prehybridization conditions were the same as that for oligonucleotides but the hybridization conditions were changed. Replica filters were hybridized in a solution of 6XSSC, 2X Denhardt's solution, 0.5% SDS and 50 x 10^4 cpm at 68°C for 16-18 hours. The washing conditions for the specific probe were 4XSSC, 0.2%SDS at room temperature for 10 minutes followed by a wash in the same buffer at 68°C. More stringent conditions for washing can be used with
specific hybridization probes and so the filters were further washed at 1XSSC, 0.2%SDS for 1 hour at 68°C with three changes of buffer.

I. SCREENING λPHAGE LIBRARIES

Two λphage libraries were screened for FXII containing sequences according to the methods described in Benton and Davis (1977). The first was a human genomic phage library kindly provided by Dr. P. Leder (Harvard University). This library consists of a partial Sau3A digest of human placental DNA ligated into the BamHI site of Charon 28. One million recombinant phage were screened by in situ hybridization. Phage were plated on E. coli strain C600 at a density of 20,000 pfu per 15 cm diameter petri dish and grown at 37°C. Duplicate plaque lifts were obtained using nitrocellulose filters. The filters were placed on the plates when the phage were pin-point size and were not touching one another. The nitrocellulose disc was oriented by stabbing the filter and the agar using a syringe needle. The master plates were placed at 4°C. The filters were transferred to a fresh plate of medium (NZYCM) and grown overnight at 37°C. The phage were lysed on the filters in 0.5M NaOH, 1.5M NaCl for 5 minutes and neutralized in 1M Tris-HCl, pH 7.5 for 5 minutes followed by 0.5M Tris-HCl, pH7.5, 1.5M NaCl for 5 minutes. The filters were then washed with gentle rubbing to remove debris in 3XSSC and allowed to air dry. The filters were baked at 68°C overnight and were prepared for hybridization with specific cDNA probes for FXII. The nitrocellulose discs were prehybridized in 6XSSC, 2X Denhardt’s solution and 0.2%SDS at 68°C for 6 hours. A hybridization solution, consisting of the same buffer and 50-100 x 10^4 cpm of a FXII hybridization probe of FXII cDNA that was labeled by ^32P by nick translation, was prepared and the filters incubated at 68°C overnight. The filters were washed at room temperature in 4XSSC, 0.5%SDS followed by two washes for
Materials and Methods / 57

30 minutes at 68°C in 2XSSC, 0.5%SDS. Positive phage were identified by autoradiography at -70°C with intensifier screens. Positive phage found on both replica filters were replated and rescreened until all the phage hybridized with the probe.

A second library (Koschinsky et al., 1986) was screened and consisted of human liver cDNA fragments (less than 500 bp in size) cloned into the EcoRI site of λgt11 (Huynh et al., 1984). A total of $0.5 \times 10^6$ phage were plated on *E. coli* strain RY1088 at 20,000 phage per plate and incubated at 37°C. The screening procedure was the same as that described above for the λphage human genomic DNA library. The hybridization probe used was a $^{32}$P-nick translated 280 bp fragment of cDNA representing sequence encoding the N-terminal region of FXII.

**J. DNA SUBCLONING**

1. Production of DNA Fragments for Ligation

DNA fragments to be subcloned were generated either by restriction endonuclease digestion or by randomly shearing large pieces of DNA by sonication (Deininger, 1983). Restriction endonuclease digestions were performed by using the conditions recommended by the suppliers. Restriction fragments were electrophoresed through agarose or acrylamide gels, and those fragments that were to be cloned were excised and recovered from the gel by electroelution (Maniatis et al., 1982).

Random fragments of plasmid DNA were produced by sonication using a Heat Systems Sonifier at output level 2. The DNA (10-100μg) was sheared in 500μl of 0.5M NaCl, 0.1M
Tris-HCl, pH7.4, 10mM EDTA with 5 pulses of 5 seconds. The solution was cooled on ice and vortexed between pulses. The resulting DNA fragments were fractionated on a 5% polyacrylamide gel, and fragments (300-500 bp in size) were recovered by electroelution. The DNA solution (450µl) was removed from the dialysis bag and adjusted to 0.3M sodium acetate. Two volumes of 95% ethanol were added, and the DNA was precipitated by incubation at -20°C overnight. The fragments were collected by centrifugation in an Eppendorf centrifuge, and the pellet was resuspended in 25µl dH₂O. The ends were made blunt-ended in a 50µl reaction of 33mM Tris-acetate, pH7.8, 66mM potassium acetate, 10mM magnesium acetate, 100mg/ml BSA, 0.2mM each deoxynucleotide triphosphate and 6 units T4 DNA polymerase. The reaction conditions varied depending on the conditions recommended by the supplier. The repaired fragments were extracted with phenol and precipitated in 0.3M sodium acetate by the addition of two volumes of 95% ethanol and incubation at -20°C overnight. The DNA was recovered by centrifugation in an Eppendorf centrifuge for 5 minutes, and the pellet was resuspended in 10µl TE buffer.

2. Ligation of DNA into pUC13 or M13 Vectors

Fragments of DNA were ligated into the vectors, pUC13 or M13, in a reaction volume of 15µl consisting of DNA, 66mM Tris-HCl, pH7.5, 5mM MgCl₂, 5mM dithiothreiotol, 0.4mM ATP and T4 DNA ligase [1 unit for blunt-ended ligations and 0.1 unit for sticky-ended ligations (Maniatis et al., 1982)]. The amounts of insert and vector DNA used in the ligation reactions depended on the vector: for pUC13, 200ng insert DNA and 100ng vector DNA were used while for M13, 10ng sticky-ended insert DNA and 10 ng vector DNA were used, and 50ng blunt-ended insert DNA and 10ng vector DNA were used. Ligation of sticky-ended DNA was for 4 hours at 16°C whereas ligation reactions of blunt-ended DNA
were incubated at 16°C overnight. The ligation reactions were stored at -20°C.

3. Transformation of DNA into Bacteria

Host bacterial cells were made competent for DNA transformation using calcium chloride (Messing, 1983). *E. coli* JM83 or JM103 were grown in 50 ml LB or YT, respectively, at 37°C with vigorous shaking until the culture reached an OD600nm of 0.2-0.4. The cells were collected by centrifugation for 5 minutes in an HB-4 rotor at 2.5 krpm at 4°C, and the pellet was resuspended in ice-cold 50mM CaCl₂ at half the original volume of culture. The cells were incubated on ice for 40 minutes and then recovered by centrifugation as before. The calcium chloride treated cells were resuspended in one-tenth the original volume of ice cold CaCl₂ solution and incubated at 4°C for 2-18 hours. Aliquots (0.3ml) of competent cells were transferred to glass tubes (13x100mm) placed on ice. Volumes of 2-3 μl of the ligation mixture containing either M13 or pUC13 vectors were used to transform JM103 and JM83 competent cells, respectively. Transformation proceeded for 40 minutes on ice and then for 2 minutes at 42°C. JM103 transformations were plated immediately with the addition of 10μl 100 mM IPTG, 35-50μl X-Gal (2% w/v in dimethylformamide), 0.2ml host cells and 3ml soft YT agar and spread on YT plates. JM83 transformed cells were rescued with the addition of 0.7ml LB followed by shaking at 37°C for one hour. The cells were spread in 100μl aliquots with 50μl X-Gal (2% w/v) onto LB-ampicillin (50μg/ml) plates. The JM103 and JM83 plated transformants were incubated overnight at 37°C. Phage plaques or colonies which were clear or white, respectively, contained cloned DNA whereas plaques or colonies transformed with religated vector without insert DNA appeared blue on the plates.
K. Blot Hybridizations

1. Southern Blot Analysis of Genomic and Cloned DNA

DNA was transferred to nitrocellulose filter sheets from agarose gels essentially as described by Southern (1975). DNA samples (1 to 10 μg) were digested by restriction endonucleases and were separated by electrophoresis through agarose gels. Single-stranded nicks were introduced into the DNA by irradiation using 260nm ultraviolet light for 30 seconds. The gel was soaked in 0.5N NaOH, 0.6M NaCl for 45 minutes to denature the DNA and then the gel was neutralized by soaking twice in a solution of 1M Tris-HCl pH7.5, 0.6M NaCl for 30 minutes. The gel was placed on Whatman 3MM paper wicks which extended into the transfer buffer (10XSSC). A piece of nitrocellulose filter was cut to the size of the gel and soaked for 10 minutes in dH₂O and then 10XSSC. The filter was placed on the gel followed by two sheets of Whatman 3MM paper and stacks of paper towels which were cut smaller than the nitrocellulose sheet. The DNA was transferred to the filter paper for 16 hours for plasmid or lambda DNA but 36-48 hours for genomic DNA. After transfer, the paper towels and 3MM paper were removed and the lane origins were marked on the nitrocellulose filter. The filter was removed and washed in 3XSSC, air dried and baked at 68°C for 6 hours.

Hybridization of ³²P-labeled DNA to genomic DNA Southern blots was performed according to Kan and Dozy (1978) whereas Southern blots of plasmid or lambda DNA were hybridized as described in Maniatis et al. (1982). The nitrocellulose filter sheet containing genomic DNA was wetted in 3XSSC and prehybridized for 16 hours at 37°C in 10 ml of a solution containing 50% formamide, 6XSSC, 1mM EDTA, 0.1% SDS, 10mM Tris-HCl pH7.5, 10X
Denhardt's solution, 0.05% sodium pyrophosphate, 100μg/ml denatured herring sperm DNA, 25μg polyA<sup>+</sup> and 50 μg/ml tRNA. The hybridization solution was identical to that for prehybridization except for the addition of 20-50 x 10<sup>6</sup> cpm of denatured 32P-labeled DNA. Hybridization was allowed to proceed for 48 hours at 37°C, and then the blot was washed for 1 hour at room temperature in 2XSSC, 1X Denhardt's solution followed by two washes at 50°C for 45 minutes in 0.1XSSC, 0.1% SDS. The blot was rinsed twice at room temperature in 0.1XSSC, 0.1% SDS followed by four rinses at room temperature in 0.1XSSC. After air drying the hybridized blot was exposed to Kodak XK-1 film with an intensifier screen for 1 to 7 days at -70°C.

The hybridization conditions used for plasmid and lambda DNA containing inserted fragments of DNA were similar to those described in the screening of the λphage libraries. The baked filter was wetted in 3XSSC and prehybridized for 4 hours in 6XSSC, 2X Denhardt's solution, 0.2% SDS at 68°C. An identical solution was prepared which also contained 1 to 10 x 10<sup>6</sup> cpm of denatured 32P-labeled DNA was exchanged for the prehybridization solution. The blot was hybridized for 16 hours at 68°C followed by washing in 4XSSC, 0.5% SDS for 10 minutes at room temperature. Subsequent washes were in 1XSSC, 0.5% SDS at 68°C. The hybridized blots were air dried and exposed to Kodak film 2-16 hours at room temperature.

2. Northern Blot Analysis

RNA was separated by electrophoresis in formaldehyde agarose gels and then transferred to nitrocellulose filter sheets (Maniatis et al., 1982). The RNA was denatured in the presence of formaldehyde (Lehrach et al., 1977) which was present in the gel and in the
loading solution. After electrophoresis, the gel was treated with a solution of 50mM NaOH for 45 minutes and then neutralized in 0.1M Tris-HCl pH7.5 for 45 minutes. The gel was soaked in a final solution of 20XSSC for 45 minutes and then placed on Whatman 3MM filter paper wicks extending into a solution of transfer buffer (10XSSC). The nitrocellulose filter paper was wetted in dH₂O and 10XSSC, and the RNA was transferred to the paper for 16 hours as described above for Southern blotting. Specific mRNA species were detected by hybridization and washing as described for Southern blots of genomic DNA (see above).

L. DNA SEQUENCE ANALYSIS

1. Screening M13 Clones

DNA fragments to be sequenced by the chain termination method (Sanger et al., 1977) were ligated into the M13 sequencing vectors (Messing et al., 1981; Messing, 1983) as described above. To identify the M13 phage which contain the desired insert, the M13 plaques were screened by in situ hybridization (Benton and Davis, 1977). The M13 phage were plated on 1.5%(w/v) YT agar plates at a density of approximately 500 plaques/plate. Nitrocellulose filter discs (85mm) were placed over the plaques, the filter oriented with an 18 gauge needle and then the filter was transferred to a solution of 0.5N NaOH, 1.5M NaCl for 5 minutes to lyse the phage particles and denature the DNA on the filter. The filters were neutralized in 1M Tris-HCl pH7.5 for 5 minutes followed by a solution of 0.5M Tris-HCl pH7.5, 1.5M NaCl for 5 minutes. The filters were air dried and baked for 6 hours at 68°C. The nitrocellulose filter discs were washed with 3XSSC for 10 minutes at room temperature and then prehybridized in 6XSSC, 2X Denhardt's solution and 0.2% SDS for 4 hours at 68°C. The hybridization solution was the same as the prehybridization solution but also contained
Materials and Methods / 63

1 to 50 x 10⁶ cpm of denatured ³²P-labeled DNA to identify phage containing the DNA to be sequenced. The filters were incubated at 68°C for 16 hours and then washed in 4XSSC, 0.5% SDS at room temperature for 10 minutes. The excess probe was removed by further washes at 68°C in 1XSSC, 0.5% SDS with multiple exchanges of the washing buffer. The filters were dried and exposed to Kodak-XK1 film at -70°C for 16 hours with an intensifier screen. Selected phage plaques were transferred to YT media, diluted approximately 10⁴ times and replated. M13 phage containing insert DNA from these plates were grown and their single-stranded DNA isolated.

2. DNA Sequence Analysis

The nucleotide sequence of single stranded DNA isolated from M13 clones was determined by the chain termination method (Sanger et al., 1977) as modified for phage M13 templates (Messing et al., 1981). Table I shows the concentrations of the deoxy- and dideoxy-nucleotides used in the sequencing reactions. An M13 primer (1µl of a solution with 0.03 OD₂₆₀nm/ml) having the nucleotide sequence of 5'-GTAAAACGACGGCCAG-3' was hybridized to single stranded template DNA (in 4 µl) in a total reaction volume of 8µl that also contained 1µl dH₂O and 2µl 10X Hin buffer (600 mM NaCl, 100mM Tris-HCl pH7.5, 70mM MgCl₂). The hybridization mixture was incubated at 68°C for 10 minutes and then was allowed to cool slowly to room temperature followed by the addition of 1µl of 15µM dATP and 1.5µl ³²P-α-dATP (10 µCi/µl). Aliquots consisting of 2.0µl of the hybridization mixture were dispensed to four tubes containing 1.5µl of the appropriate deoxy/dideoxy-nucleotide mixture (Table I). The reactions were started with the addition of 0.4 units DNA polymerase I Klenow fragment and incubated at room temperature for 15 minutes. A chase solution was added (1µl of 0.5mM dATP) and the reaction mixtures were
Table 1: DNA Sequencing Mixes.

The numbers in the table refer to the μM concentrations of the dideoxy and deoxy-ribonucleotide triphosphates used in the sequencing mixes for M13 DNA sequencing. The values used were those determined by Dr. Joan McPherson, Dept. of Botany, UBC.

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>d/ddG</th>
<th>d/ddA</th>
<th>d/ddT</th>
<th>d/ddC</th>
</tr>
</thead>
<tbody>
<tr>
<td>dG</td>
<td>7.9</td>
<td>109.4</td>
<td>158.7</td>
<td>157.9</td>
</tr>
<tr>
<td>dT</td>
<td>157.6</td>
<td>109.4</td>
<td>7.9</td>
<td>157.9</td>
</tr>
<tr>
<td>dC</td>
<td>157.4</td>
<td>109.4</td>
<td>158.7</td>
<td>10.5</td>
</tr>
<tr>
<td>ddG</td>
<td>157.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ddA</td>
<td></td>
<td>116.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ddT</td>
<td></td>
<td></td>
<td>550.3</td>
<td></td>
</tr>
<tr>
<td>ddC</td>
<td></td>
<td></td>
<td></td>
<td>191.6</td>
</tr>
</tbody>
</table>
incubated for 20 minutes at room temperature followed by the addition of 5μl of a stop-dye mixture (98% formamide, 10 mM EDTA pH8.0, 0.02% xylene cyanol, 0.02% bromophenol blue). The extended products were denatured by heating to 92°C for three minutes, and 1-2μl of each reaction mixture was analyzed on 6% and 8% polyacrylamide-8M urea gels in 1XTBE buffer at 1W/cm. After electrophoresis, the gels were dried and exposed to Kodak XK-1 film overnight at room temperature.

3. Computer Analysis of DNA Sequence Data

Analysis of the DNA sequence data obtained from the sequencing gels was assisted by the computer programs of Staden (1982) and Delaney (1982).

M. MAPPING THE END OF A MRNA TRANSCRIPT

1. Primer Extension Analysis

A hybridization primer was prepared as follows. Single stranded DNA of a fragment representing the 5' end of the FXII mRNA was subcloned M13mp19 and was made double stranded and radioactive with DNA polymerase I-Klenow fragment in the following reaction. The M13 primer was hybridized to the single stranded DNA template at 68°C for 10 minutes in a mixture consisting of 2.5μl M13 primer, 2.5μl DNA, 2.5μl 10X HincII buffer (1M NaCl, 0.1M Tris-HCl pH7.4, 0.07M MgCl₂, 0.01M dithiothreitol) and 7.5μl dH₂O. The reaction was cooled slowly to room temperature and then adjusted with 2μl 15μM dATP, 3μl 3²P-α-dATP (10μCi/μl), 2μl 0.5mM dGTP, dCTP and TTP and 1.25 units of Klenow fragment (2.5μl). The extension reaction was incubated for 5 minutes at room temperature and transferred to 68°C for 10 minutes. The reaction was cooled to 37°C, and
20 units of the restriction endonuclease *Hinc* II was added and the DNA digested for 20 minutes. An equal volume of stop-dye mixture (98% deionized formamide, 10 mM EDTA, pH 7.5, 0.02% bromophenol blue, 0.02% xylene cyanol) was added and the DNA was denatured at 90°C for 5 minutes. The DNA fragments were subjected to denaturing gel electrophoresis through a 6% polyacrylamide-8 M urea gel. After autoradiography for 5 minutes, the gel slice containing the primer DNA (180 bp) was excised and the DNA recovered by electroelution. The DNA was precipitated with ethanol and resuspended in 6.5μl of a buffer containing 40 mM Pipes pH 6.7, 1 mM EDTA and 0.4 M NaCl (Kingston et al., 1984). The mixture was added to 3μl polyA+ mRNA from human liver (18μg) and 30μl deionized formamide. Another reaction was prepared in the same manner but without mRNA. The reaction mixtures were heated at 80°C for 15 minutes and then transferred to a 30°C water bath for 16 hours (Kingston et al., 1984). The nucleic acids were precipitated with 140μl of 0.3M sodium acetate (pH 5.0) and 400μl of 95% ethanol. The pellets, containing approximately 60,000 cpm, were resuspended in 50μl of AMV reverse transcriptase buffer (50 mM Tris-OH pH 8.3, 20 mM KCl, 10 mM MgCl₂, 5 mM dithiothreitol) (Kingston et al., 1984) to which was added 30 units of AMV reverse transcriptase. The reactions proceeded at 37°C for 1 hour and then the nucleic acids were precipitated with 5μl of 3M sodium acetate and 125 ul of 95% ethanol. The pellets were resuspended in 5μl of load-dye mix and denatured at 90°C for 5 minutes. The primer-extended product was analyzed on a 6% polyacrylamide-8 M urea gel. DNA sequencing reactions of the template from which the original primer was obtained were used as size markers. Following electrophoresis, the gel was dried and exposed to Kodak XK-1 film overnight at room temperature and then 3 to 14 days at -70°C with an intensifier screen.
2. Nuclease S1 Mapping

A hybridization probe was prepared in the same manner as that described above. The single stranded primer, however, was obtained from an M13 subclone of the 450 bp HpaII fragment that represents the 5' end of the mRNA transcript and 5' most flanking sequences. An oligonucleotide primer (5'-CGAAAGTGTTGACTCCA-3') was synthesized on an Applied Biosystems 380A DNA Synthesizer (Atkinson and Smith, 1984) and is complementary to a region within the 450 bp HpaII fragment. The preparation of the long $^{32}$P-labeled primer was the same as that described for the primer extension reaction above. To generate the correct hybridization probe for nuclease S1 mapping, the Klenow $^{32}$P-labeled DNA was digested with EcoRI and the single stranded primer was isolated from a 6% polyacrylamide-8M urea gel as described above. Hybridization of this primer (210 bp in length) to polyA+ mRNA was as described for the primer extension reaction. Another identical reaction was prepared except that dH$_2$O was added instead of polyA+ mRNA. After hybridization, the nucleic acids were precipitated by the addition of 140μl of 0.3M sodium acetate and 400μl of 95% ethanol followed by incubation at -70°C for 20 minutes. The nucleic acids were collected by centrifugation at 4°C for 10 minutes in an Eppendorf centrifuge. To the pellets were added 270μl nuclease S1 buffer (270μl of 0.28 M NaCl, 50mM sodium acetate, 4.5mM ZnSO$_4$), 10μg denatured herring DNA and 10 to 30 units of nuclease S1. The reaction mixture was incubated at 37°C for 15 minutes to 1 hour. The digested DNA was precipitated with 10μg tRNA and 600μl 95% ethanol and incubated at -70°C for 15 minutes. The DNA was recovered by centrifugation and the pellet resuspended in 2.5μl of 1mM Tris-HCl pH 7.5, 0.6mM NaCl, 0.1mM MgCl$_2$ and 2.5μl of stop-dye mix. The sample was heated for 3 minutes at 90°C and analyzed on a 6% polyacrylamide-8M urea gel together with the DNA sequencing reactions of the template.
from which the original probe was prepared. The nuclease S1 reaction on the same hybridization mixture was repeated with the addition of 10μg tRNA instead of 10μg of denatured herring DNA and 200 units of enzyme. The gel was dried and exposed to film as described for the primer extension reaction above.
III. RESULTS

A. CHARACTERIZATION OF THE HUMAN FXII CDNA

1. Isolation of Factor XII cDNA Clones

A human liver cDNA library (provided by Dr. Stuart Orkin at Children’s Hospital, Harvard University) was screened to isolate the human FXII cDNA. Two hundred thousand bacterial colonies of the cDNA library were screened by in situ hybridization with two mixtures of synthetic oligonucleotides that encoded two regions of β-factor XIIa as predicted from the known amino acid sequence (Fig. 5). Although each of the oligonucleotide mixtures hybridized to about 60 different colonies, only two colonies hybridized to both (Fig. 6). After rescreening at lower colony density, plasmid DNA was isolated from the two positives that were designated pcHXII-11 and pcHXII-14 (Fig. 7). Subsequent sequence analysis revealed that these plasmids contained DNA coding for the known amino acid sequence of β-factor XIIa. Because pcHXII-11 and pcHXII-14 contained only short cDNA inserts, $2 \times 10^5$ colonies of the cDNA library were rescreened with the [$^{32}$P]-labeled PsII insert of pcHXII-11 as a hybridization probe. Thirty-six positive clones of various lengths were obtained; plasmids pcHXII-17 and pcHXII-501 contained the longest cDNA inserts and were studied further. Restriction endonuclease mapping of the four positive plasmids showed that they contained overlapping DNA (Fig. 7) except that pcHXII-17 appeared to contain a deletion (Fig. 7, open dashed bar). This was confirmed by subsequent DNA sequence analysis (see next section).
Flow Chart indicating the regions of the β-factor XIIa amino acid sequence (Fujikawa and McMullen, 1983) from which the oligonucleotide sequences were predicted; mixtures of synthetic heptodecaderoxyribonucleotides were used as hybridization probes in screening the liver cDNA library. The underlined numbers represent the two pools of oligonucleotides used in the double screening technique (Figure 6). Oligo 1 and Oligo 2 sequences represent amino acids 537 to 542 (Thr to Gly) and 486 to 491 (Trp to Glu), respectively, in the β-factor XIIa amino acid sequence (Fig. 6).
**β-factor XIIa**

\[ \text{H}_2\text{N} - \text{AsnGlyProLeuSerCys} - \text{GlyGlnArg} \]

\[ \text{ValValGly} - \text{Cys} - \text{TrpGlyHisGlnPheGlu} - \text{ThrAspAlaCysGlnGly} - \text{(Ser)} - \text{ThrSerVal-COOH} \]

**mRNA Sequence**

Complementary Oligo Sequence

\[ 5'-\text{TC}^A\text{AA}^C\text{TG}^A\text{GCCC}^-\text{CCCA} - 3' \]

32 Possible Sequences

Complementary Oligo Sequence

\[ 5'-\text{CC}^T\text{GTG}^A\text{CATGC}^A\text{TCA}^G^-\text{TCTGT} - 3' \]

128 Possible Sequences
Figure 6: Double Oligonucleotide Screen For Factor XII cDNA

A liver cDNA library was plated at a high density (100,000 colonies/filter) and duplicate filters were hybridized with either $^{32}$P-labeled oligo-1 or oligo-2 pools of synthetic heptadecadeoxyribonucleotides (upper discs). The overlapping positive colony (arrows) was rescreened at a lower density (200 colonies/filter) in duplicate and hybridized with the same oligonucleotide mixtures (lower discs). Positive colonies which overlapped (arrows) were selected and their DNA sequence determined.
Figure 7: Restriction Enzyme Map and Sequencing Strategy for Factor XII cDNA

The bars below the restriction map represent cDNA clones; pcHXII-11, pcHXII-14, pcHXII-17 and pcHXII-501 and include regions coding for the signal peptide (dotted bar), the heavy chain of α-factor XIIa (solid bars) and the 3’ untranslated region (slashed bar). The open dashed line in pcHXII-17 represents a deletion (see text for details). Arrows below the cDNA clones represent the sequencing strategy used, where each arrow represents the M13 clone. The extent of sequencing is indicated by the length of arrows. DNA sequence determined on the coding strand is shown by an arrow pointing left; sequence determined on the non-coding strand is shown by an arrow pointing right. Thick arrows represent DNA sequences obtained from randomly sheared cDNA fragments and thin arrows represent sequence obtained from restriction enzyme fragments. The scale at the bottom represents nucleotides in kilobase pairs.
2. **Sequence Analysis of Human Factor XII cDNA Clones**

The strategy used to determine the nucleotide sequence of the four clones, pcHXII-11, 14, 17 and 501 is shown in Figure 7. The thick arrows represent the sequence determined from M13 phage containing randomly sheared DNA. Although more than 90% of the nucleotide sequence was determined using this strategy, few random clones were isolated that overlapped the *AvaI* site or the 5' end of factor XII cDNA (Fig. 7). To complete the nucleotide sequence analysis, plasmid pcHXII-501 (Fig. C) was digested with *PstI* and the 1200 bp and 280 bp fragments were isolated by polyacrylamide gel electrophoresis. The 280 bp fragment was cloned directly into the *PstI* site of M13mp8. The 1200 bp fragment was digested with both *AvaI* and *SmaI*, and the ends were made blunt-ended using the Klenow fragment of *E. coli* DNA polymerase I (Smith *et al.*, 1979). The resulting *PstI/AvaI* and *AvaI/SmaI* fragments were isolated by polyacrylamide gel electrophoresis and subsequently ligated into the *SmaI* site of M13mp18. The DNA sequence analysis of these clones is shown in Fig. 7 (thin arrows).

The nucleotide sequences of the four factor XII cDNA clones were compiled into a consensus sequence (Fig. 8) using the computer program designed by Staden (1982). Each nucleotide in the sequence was determined an average of 6.6 times and 84% of the sequence was determined on both strands. In the regions of overlap, the four factor XII cDNA clones contained identical nucleotide sequences with one exception. In agreement with the restriction map of pcHXII-17 (Fig. 7), the DNA sequence analysis showed that this plasmid contains a deletion of 53 bp between nucleotides 1330 and 1383 (Fig. 8). The deleted region is flanked by an inverted repeat having the sequence: CTCGC(G/C)CGTC (nucleotides 1320-1329 and nucleotides 1383-1392, Fig. 8). This deletion alters the reading frame of
Figure 8: Nucleotide Sequence of Factor XII cDNA

The sequence was determined by analysis of the overlapping clones shown in Figure 7. The predicted amino acid sequence of human factor XII is shown above the DNA sequence. The oligonucleotide sequences used for screening the cDNA library are represented by boxes at nucleotides 1,474-1,490 and 1,627-1,643. The signal peptide is numbered backwards from the putative signal peptidase cleavage site (open arrow). The proteolytic cleavage sites that give rise to β-factor XIIa are shown by the heavy arrows (at amino acid residues 334, 343 and 353). The residues that make up the catalytic triad (His 393, Asp 442 and Ser 544) are underlined. The TGA stop codon is located at nucleotides 1,807-1,809 and followed by the 3' untranslated region (nucleotides 1,810-1,959) and a poly A+ tail. The poly A+ recognition site AATAAA is encoded by nucleotides 1,937-1,942. A carbohydrate attachment site (Fujikawa and McMullen, 1983) is indicated by the solid diamond.
factor XII mRNA and probably represents a cloning artifact.

3. Predicted Amino Acid Sequence of Human Factor XII

Translation of the consensus cDNA sequence using the standard genetic code results in a single open reading frame coding for human factor XII (Fig. 8). The derived amino acid sequence agrees well with those regions of human factor XII that have been determined by protein chemistry techniques. Nucleotides 19-78 encode the amino-terminal region of factor XII (Fujikawa et al., 1980b; Fujikawa and Davie, 1981). Since the amino-terminal residue of plasma factor XII is encoded by nucleotides 19-21 (Fig. 8), the factor XII cDNA sequence also encodes part of an amino-terminal extension (nucleotides 1-18, Fig. 8). Nucleotides 1021-1047 and 1078-1806 encode the light and heavy chains of β-factor XIIa as determined by Fujikawa and McMullen (1983). In the presence of an anionic surface, kallikrein cleaves factor XII at a position carboxy-terminal to Arg\textsubscript{335}, resulting in the formation of α-factor XIIa (Fujikawa and McMullen, 1983). Further proteolytic cleavages occur carboxy-terminal to Arg\textsubscript{334} and Arg\textsubscript{343}, and give rise to β-factor XIIa. During the formation of β-factor XIIa, residues 1-334 and 345-353 of factor XII are removed. The codon for the carboxy-terminal residue of the heavy chain of β-factor XIIa (Ser-596) is followed by a TGA stop codon (nucleotides 1807-1809, Fig. 8), a 3’ untranslated region of 150 nucleotides and a polyA\textsuperscript{+} tail. Nucleotides 1937-1942 contain the sequence AATAAA that is involved in the polyadenylation of mRNA (Proudfoot and Brownlee, 1976; Wickens and Stephenson, 1984). This sequence occurs 17 nucleotides upstream of the polyA\textsuperscript{+} tail.

The predicted amino acid sequence of factor XII corresponds to a mature protein of 596 amino acids. The amino acid composition of plasma factor XII was determined to be as
follows: Ala_{51}, Arg_{39}, Asn_{13}, Asp_{22}, 1/2-Cys_{40}, Gln_{37}, Glu_{31}, Gly_{48}, His_{27}, Ile_{9}, Leu_{53}, Lys_{19}, Met_{4}, Phe_{15}, Pro_{56}, Ser_{36}, Thr_{34}, Trp_{13}, Tyr_{19}, Val_{30}. From the predicted amino acid sequence, the molecular weight of plasma factor XII is 66,915 in the absence of carbohydrate and 80,427 with the addition of 16.8% carbohydrate (Fujikawa and Davie, 1981). Within experimental error, both the predicted amino acid composition and molecular weight of factor XII are in excellent agreement with those determined for the purified protein (Griffin and Cochrane, 1976; Fujikawa and Davie, 1981).

The complete amino acid sequence for human α-factor XIIa was also determined by McMullen and Fujikawa (1985). The amino acid sequence was determined by automated Edman degradation using peptides produced by chemical and enzymatic cleavages of factor XII. These results appeared in the literature after the completion of the cDNA sequence for factor XII described above. The two amino acid sequences were compared, and a serine residue at position 314 is found in place of a proline residue reported by McMullen and Fujikawa (1985).

4. 5' End of the Factor XII cDNA

The factor XII cDNA clone, pcHXII-501, did not extend to the initiator methionine residue and so another human liver cDNA library (prepared by W. Funk in Dr. MacGillivray’s laboratory) was screened with a $^{32}$P-labeled 280 bp PstI probe representing nucleotides 1 to 280 of pcHXII-501. The library was prepared by using oligonucleotides as random primers for reverse transcription of mRNA. Fragments were separated on a column of Agarose A 0.5M, and were cloned into the expression vector λgt11. A library containing small inserts (less than 500 bp) was screened for a clone that would extend further 5' to
pcHXII-501. A total of $1 \times 10^6 \lambda gt11$ recombinant phage were screened and two positive clones were isolated. The nucleotide sequences of the cDNA clones (pcHXII-WF7 and -WF9) were determined (Fig. 9). The two clones were identical and contained 220 bp inserts and extended a further 40 bp 5' to pcHXII-501. Translation of the sequence revealed an identical amino acid sequence to pcHXII-501 but extended the signal sequence by a further 13 amino acid residues. At this point, there is an ATG sequence in the same reading frame as the putative signal peptide. Therefore, the signal peptide sequence for factor XII probably consists of 19 amino acid residues with the characteristic hydrophobic core (Watson, 1984). Two other cDNA libraries were screened in order to obtain a longer cDNA clone which would contain the 5' untranslated end of the mRNA but no longer clones were isolated.

5. A 2.40 kb mRNA From Liver Codes For Human Factor XII

To determine the size of factor XII mRNA, human liver polyA+ RNA was electrophoresed on an agarose gel under denaturing conditions and transferred to nitrocellulose. The blot was hybridized to $[^{32}P]$labeled pcHXII-501, which contains 1959 bp of coding sequence. Autoradiography revealed a single band that was 2,400 ± 100 nucleotides in size (Fig. 10). The total length of mRNA sequenced from the cDNA clones, pcHXII-501 and pcHXII-WF7, is 2000 nucleotides. Since eukaryotic mRNAs usually contain polyA+ tails of 180-200 nucleotides (Perry, 1976) it seems that there is approximately 200 bases of mRNA sequence still to be determined.
A human liver cDNA library was screened with a $^{32}$P-labeled 5' fragment from pcHXII-501 (Figure 7) and a positive clone (pcHXII-WF7) was isolated. The nucleotide sequence of this clone was determined and is shown. The predicted amino acid sequence is given below the nucleotide sequence. The numbers correspond to the cDNA sequence given in Figure 8. The open arrow indicates the signal peptidase cleavage site and the diamond shows the initiator Methionine residue. Only 3 nucleotides of the 5' untranslated end were contained in the pcHXII-WF7 clone.
Figure 10: Size Determination of the Factor XII mRNA Using rRNA Markers

Human liver poly A+ RNA (10 μg) was electrophoresed in a denaturing agarose-formaldehyde gel and was transferred to nitrocellulose. The filter was hybridized with $^{32}$P-labeled pcHXII-501. The filter was exposed for 18 hours at -70°C with Lightning Plus intensifying screens (Dupont). The positions of RNA size standards are indicated, including mammalian 28 S (4.72kb) and 18 S (1.87 kb) rRNA and E. coli 23 S (2.90 kb) and 16 S (1.54 kb) rRNA (Noller, 1984).
B. THE HUMAN FACTOR XII GENE

1. Isolation and Characterization of Factor XII Genomic Clones

The factor XII gene was isolated from a human genomic DNA library (provided by Dr. P. Leder) which contains partial Sau3A fragments (10-20 kbp in length) inserted into the BamHI site of λCharon 28. One million phage from this library were screened with a human factor XII cDNA (pCHXII-501) as a hybridization probe. This cDNA contains DNA representing coding sequence for the complete zymogen of factor XII, 6 amino acids of the signal peptide and the 3' untranslated region (150 bp). Two recombinant phage designated λHFXII-27 and λHFXII-76 were purified to homogeneity by rescreening four times with the same hybridization probe.

DNA was isolated from large scale lysates of each of the two phage and analyzed by restriction enzyme digestion. Southern blot analysis indicated that the 5' and 3' ends of the cDNA were represented in the two phage. BamHI and PstI digests of the clones were hybridized with ³²P-labeled 280bp or 360 bp PstI fragments, representing the 5' and 3' end of the cDNA, respectively. Figure 11 shows that the 280bp PstI fragment which encodes exons 2,3 and part of 4 (see Fig. 12) hybridized to the 2.5 and 0.6 kbp PstI fragments in λHFXII-76 and only to the 0.6 kbp PstI fragment in λHFXII-27. The 360 bp PstI fragment representing exon 14 (see Fig. 12) hybridized to λHFXII-27 and not to λHFXII-76. The 2.0 kbp and the 0.3 kbp PstI fragments and a very large BamHI fragment in λHFXII-27 hybridized to the 3' probe. Therefore, λHFXII-76 contains the 5' but not the 3' end of the cDNA whereas λHFXII-27 contains the 3' end and not the 5' end of the cDNA.
Figure 11: 5' and 3' Southern Blot Analysis of FXII Genomic Clones

Two clones (λHFXII-27, -76) were isolated from a genomic λ phage library encoding the human factor XII gene. The DNA was cleaved by restriction endonucleases, the fragments were electrophoresed through neutral agarose gels and Southern (1975) blotted to nitrocellulose filter paper. Two sets of filters were prepared and one set was hybridized with a $^{32}$P-labeled 280 bp PstI (nucleotides 1-280, 5' end of the cDNA, Figure 9) or with a $^{32}$P-labeled 360 bp PstI fragment (nucleotide 1,680-3' end of the pcHFXII-501 insert, Figure 9). The blots were autoradiographed for 18 hours. Left side blot is hybridized with the 280 bp fragment: A, λHFXII-76 digested with PstI (Lane P) or Bam HI (Lane B); B, λHFXII-27 digested with PstI (Lane P) or Bam HI (Lane B). Right side blot is hybridized with the 360 bp fragment: C, λHFXII-76 digested with PstI (Lane P) or BamHI (Lane B); D, λHFXII-27 digested with PstI (Lane P) or Bam HI (Lane B).
A partial restriction enzyme map of the cloned genomic DNA for FXII was constructed from multiple restriction endonuclease digestions and Southern blot analysis of the fragments hybridized to various $^{32}$P-labeled cDNA probes. The map (Fig. 12) shows that $\lambda$HXII-27 and -76 contained DNA sequence which overlapped but were not identical. Together, the two phage span approximately 30 kb of genomic DNA, with the factor XII gene mapping to a 12 kbp region in the middle of this DNA (Fig. 12). A second, unamplified human genomic phage library (prepared by V. Geddes in Dr. MacGillivray's lab) was screened with $[^{32}P]$-labeled 280 bp PstI fragment of pcHXII-501. However, of $0.5 \times 10^4$ phage screened, no positive clones for factor XII were isolated.

2. Southern Blot Analysis of the Human FXII Gene

The BamHI, HindIII and PstI fragments in the restriction enzyme map correspond to those in the Southern blot (Fig. 13) of genomic DNA that was hybridized to $[^{32}P]$-labeled factor XII cDNA (pcHXII-501). Autoradiography revealed several hybridizing bands in each of the restriction enzyme digests (Fig. 13). For example, the lengths of the BamHI (greater than 9.5 and 7.6 kbp), the HindIII (9.5 and 5.8 kbp) and the PstI (2.5, 1.5, 1.0 and 0.6 kbp) fragments agree closely with the fragment lengths which contain exon sequence in the restriction enzyme map (Fig. 12). The 7.6 kbp BamHI bands observed in the genomic Southern blot compares to the 3' end of the gene including most of exon 14. The 3' BamHI site is not shown in the restriction enzyme map. Since no other bands were observed in the genomic Southern blot, it appears that the human genome contains a single factor XII gene.
Figure 12: Partial Restriction Map and Intron/Exon Organization of the FXII Gene.

A partial restriction enzyme map of the gene is given. Abbreviations used are: B-BamHI; Hp-HpaII; Bg-BglII; P-PstI; H-HindIII; Hc-HincII. Exons are shown in the bar above the restriction map as black boxes and are numbered 1 through 14. The two genomic phage clones, λHFXII-76 and λHFXII-27 are shown below the restriction enzyme map as open bars. Breaks in the bars indicate that the phage DNA extends approximately 5.0 and 13.0 kb 5' and 3' from the gene, respectively. The scale represents kilobase pairs.
Human liver DNA (10μg) was digested by various restriction enzymes and electrophoresed in a 1.0% agarose gel. After denaturation, the DNA was transferred to nitrocellulose and hybridized with $^{32}$P-labeled pcHXII-501. The filter was autoradiographed for 4 days at -70°C with intensifying screens. Lane M shows the positions of the $^{32}$P-labeled λ-HindIII DNA used as molecular weight markers. The human DNA was digested with EcoRI (Lane 1), HindIII (Lane 2), PstI (Lane 3), and BamHI (Lane 4).
3. Localization of Intron/Exon Junctions

To identify exon-containing DNA sequences, sonicated fragments of the genomic clone, \( \lambdaHXII-27 \), were subcloned into M13 vectors and transformed into \( E. coli \) JM103. The phage were plated and screened with \(^{32}\text{P}-\text{labeled pcHXII-501.} \) The DNA sequences were determined from purified single stranded DNA of the positive phage. The intron/exon junction sequences for exons 3 to 14 were obtained on both strands and are given in Table II. The positions of exons 1 and 2 in \( \lambdaHXII-76 \) were determined by restriction enzyme analysis. Small fragments containing the exons were subcloned and their DNA sequences determined on both strands (Table II). All introns follow the GT-AG rule (Breathnach and Chambon, 1981) except for the donor sequence of intron J which has a GC dinucleotide instead of GT. The consensus sequences surrounding splice junctions in RNA polymerase II transcribed genes (Mount, 1982) were also found to be in agreement with the intron/exon junctions of the FXII gene (Table III). The localization of exons by DNA sequence analysis and restriction enzyme mapping has shown that the FXII gene consists of 13 introns (A through M, Table IV) and 14 exons (1-14, Table IV) of which 12 are located within 4.2 kb of genomic DNA (Fig. 12).

4. Nucleotide Sequence of the Human Factor XII Gene

The partial DNA sequence for the human factor XII gene is shown in Fig. 14. A total of 4,735 bp of DNA sequence was determined of which 92% was obtained on both strands. Sequence data that was obtained on one strand only was determined at least twice and includes nucleotides -120 to -357 of the 5’ end of the gene. There are 357 and 195 bp of 5’ and 3’ flanking regions, respectively, given in Fig. 14. As determined by DNA sequence
Table 1. Nucleotide Sequence of Intron/Exon Junctions in the Factor XII Gene.

Exon sequence is shown in upper case; intron sequence in lower case. The codon phase refers to the position of the intron in the codon triplet. 0-intron occurs between codons, 1-intron occurs after the first nucleotide and 11-intron occurs after the second nucleotide in the codon.
### Donor Frequencies

<table>
<thead>
<tr>
<th></th>
<th>+4</th>
<th>+3</th>
<th>+2</th>
<th>+1</th>
<th>-1</th>
<th>-2</th>
<th>-3</th>
<th>-4</th>
<th>-5</th>
<th>-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>13</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>CON</td>
<td>N</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>R</td>
<td>A</td>
<td>G</td>
<td>T</td>
</tr>
</tbody>
</table>

### Acceptor Frequencies

|   | -20 | -19 | -18 | -17 | -16 | -15 | -14 | -13 | -12 | -11 | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | +1 | +2 | +3 | +4 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| G | 3   | 3   | 2   | 4   | 1   | 4   | 2   | 3   | 2   | 1   | 3   | 0   | 3   | 3   | 1   | 1   | 4   | 0   | 0   | 13  | 5   | 3   | 2   | 2   |
| A | 3   | 3   | 4   | 0   | 0   | 1   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 3   | 1   | 13  | 0   | 4   | 1   | 3   | 2   |     |
| T | 0   | 3   | 2   | 3   | 1   | 6   | 6   | 4   | 7   | 5   | 7   | 3   | 5   | 5   | 3   | 4   | 3   | 2   | 0   | 0   | 1   | 4   | 4   | 4   |
| C | 7   | 4   | 5   | 6   | 9   | 7   | 5   | 3   | 7   | 5   | 5   | 7   | 5   | 8   | 8   | 3   | 10  | 0   | 0   | 3   | 5   | 4   | 5   |
| CON | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | Y | A | G | G | N | N | N |

**Table III. Frequencies of Nucleotides at Intron/Exon Junctions.**

The frequencies of the different nucleotides at the intron/exon junctions of the human FXII gene are compared to the consensus (CON) of Mount (1982). Splice junctions are between -1 and +1.
Table IV. Size and Positions of Exons and Introns in the Factor XII Gene.

*Sizes of these introns were estimated from restriction enzyme analysis (Fig. 12). All other sizes were determined from the nucleotide sequence of the gene (Fig. 14).
The sequence was determined by the DNA sequence analysis of M13 clones. The predicted amino acid sequence of FXII is given above the nucleotide sequence. Five transcriptional start sites are indicated by triangles and circles between +1 and +33. Solid and open triangles indicate strong and weak primer extension products, respectively; solid and open circles represent strong and weak nuclease S1 products, respectively. A solid line indicating a possible TATA sequence is shown at position -26. The CAAT-like sequence at position -103 is shown as a double line. Inverted repeats are marked by arrows under the nucleotide sequence. An open diamond represents HpaII sites that demarcate the nuclease S1 primer. A closed diamond indicates the complemented 5' end of the 180 bp primer used in the extension reaction. The probable initiator methionine at amino acid -19 is indicated by a box. The sizes of the larger introns are approximated from restriction enzyme mapping (Fig. 13). The polyadenylylation signal AATAAA (nucleotides 11693-11698) is boxed. In the protein coding region, the cleavage site which gives rise to plasma FXII is denoted by a closed arrow, and the two sites of cleavage which produces β-FXIIa by kallikrein are denoted by an open arrow.
Results / 100

analysis of the 4.2 kb region described above and by sizing of introns A and B by restriction enzyme analysis the total length of the FXII gene is approximately 12 kb. Transcription initiation was determined by nuclease S1 mapping and primer extension analysis (described below) and is designated as nucleotide +1 in Figure 14. The nucleotide sequence GAC GCC precedes the ATG codon. This sequence agrees with the consensus sequence found prior to initiator methionine residues (Kozak, 1984). The conserved AATAAA (Proudfoot and Brownlee, 1976) in the 3' flanking sequence was identified at position 11693-11698. DNA sequence analysis showed that there is an AluI-like repetitive element (Schmid and Jelinek, 1982) at nucleotide -370 (data not shown) preceeding the 5' end of the gene for factor XII. Furthermore, two other AluI elements were identified in intron B by DNA sequence analysis although their positions in the gene were not determined.

The sizes of the exons and introns and their positions in the gene are given in Table IV. Exons range in size from 57 to 320 bp, and introns range in size from 80 to 4,500 bp. Exon 14 is the largest exon (320 bp) and includes the 3' untranslated region of the mRNA. The average length of the remaining exons is 132 bp which is similar to the average length of eukaryotic exons of 150 bp (Naora and Deacon, 1982). The predicted size of the mRNA is 2,036 bp which is in close agreement with the size of the mRNA determined by Northern blot analysis (2,400 +/- 100 bp, including a poly A+ tail). The predicted amino acid sequence of the gene coding sequence and the cDNA sequence were identical except for a proline,414 (CCT) in the gene compared to a serine residue (TCT) in the cDNA sequence.
5. Transcription Initiation Site of the Factor XII Gene

The site of transcription initiation was determined by primer extension and nuclease S1 mapping analysis of the 5' end of the gene. The primers used in these experiments are indicated in Fig. 14 and the reaction products are shown in Figs. 15 and 16. The nucleotide sequences of the templates from which the primers were prepared are also given in Figs. 15 and 16. The furthest extension product is a weak band seen with a long exposure of the gel (7 days) and corresponds to position +1 in Fig. 14. This extension product corresponds to the longest nuclease S1-resistant fragment. Other nuclease S1 and primer extension products were observed and could be aligned within 1 nucleotide of each other. These positions in the 5' end of the gene are: +14 and +22, represented by two strong bands in the primer extension reactions, and +27 and +33, represented by weak bands in the autoradiograms of both the primer extension and nuclease S1 reactions. Higher molecular weight bands were not observed even with prolonged exposures (two weeks) of the gels (Fig. 15). Since the nuclease S1 and primer extension products correspond to one another, it appears that there is no intron within the 5' untranslated region of the factor XII gene. The primer extension experiment was repeated three times resulting in the identical banding pattern seen in Fig. 15. Furthermore, longer (220 bp) and shorter (50 bp) primers were used in the extension reaction; the longer primer folded upon itself and could not prime RNA, and the shorter primer resulted in only a few nucleotides of extension. No extended product was observed when oligonucleotide primers (17 to 20 bp in length) were used in the primer extension reaction. In separate experiments, the primer extension and nuclease S1 reactions of the primer without mRNA showed no extended product and no protected fragments, respectively (data not shown).
Figure 15: Nuclease S1 Mapping of the 5' End of the FXII Gene.

A 450 bp *HpaiI* fragment containing all of exon 1 and 5' flanking sequence was used to obtain a 228 bp probe for the nuclease S1 mapping reaction. The probe begins at the position 90 and extends to position -120 (Fig. 15). The $^{32}$P-labeled fragment was hybridized to 18 μg Poly A+ mRNA and was digested with 200 units nuclease S1 for 1 hour at 37°C. The reaction was electrophoresed in a 6% polyacrylamide-8M urea gel along with a DNA sequencing ladder of the probe DNA (*Lanes GATC*). The gel was autoradiographed for 18 hours (*Lanes GATC*) and 3 days with an intensifier screen and -70°C (*Lane 1*). *Lanes GATC* are Sanger sequencing reactions of the probe DNA. *Lane 1* is the nuclease S1 reaction of poly A+ mRNA from liver. The wide arrows show the nuclease S1 digestion products and the narrow arrows, the exact nucleotide to which the nuclease S1 band corresponds.
Figure 16: Primer Extension Analysis of the 5' End of the FXII Gene.

A 180 bp $^{32}$P-labelled fragment of sequence complementary to FXII mRNA was hybridized to 18 $\mu$g of poly A+ mRNA from liver and extended from the HindIII site at position 96 bp (Fig.15). The reaction products were electrophoresed in a 6% polyacrylamide-8M urea gel along with Sanger sequencing reactions performed on the pcHXII-WF7 used in preparation of the primer. Lanes 1 and 2 are primer extension products of mRNA from liver reacted with reverse transcriptase at 37°C and 45°C for 1 hour. The gel was exposed for 18 hours (Lanes GATC) and for 7 days at -70°C with an intensifying screen (Lanes 1 and 2). Lanes GATC are the sequencing reactions of the primer DNA. P indicates the primer before extension and wide arrows indicate the extended products. Met indicates the ATG initiator methionine sequence. HcII is the HincII restriction site used to generate the single stranded primer used in the primer extension reaction. The numbers beside the sequencing gel correspond to the extended product sizes counted from the Met (ATG) (as "CAT" in the gel shown) sequence.
The position of the nuclease S1-protected product was verified further by hybridizing an oligonucleotide complementary to nucleotides -23 to -6 (Fig. 14) to a Northern blot (Fig. 17) of human liver polyA+ mRNA. This oligonucleotide did not hybridize to the mRNA although another oligonucleotide complementary to the region +89 to +106 did hybridize. Therefore, the sequence immediately upstream of the putative transcriptional start site does not contain coding sequence.

The DNA sequence of the 5' flanking region to the transcription initiation site was determined and was analyzed for the characteristic recognition sequences "TATA" and "CAAT". The closest sequences to these consensus sequences were a "CTATTTTC" at -26 and a "CCAAG" at -105. The SV40 enhancer core consensus TGG(A/T)(A/T)(A/T) was found at position -330 (ATGGTAGT). A similar sequence, ATGGTATG, was observed at position -123 in the promoter region of the albumin gene (Gorski et al., 1986).
Figure 17: Northern Blot Analysis of the 5' End of the FXII Gene

Human liver poly A⁺ mRNA (10 µg) was electrophoresed in a denaturing agarose-formaldehyde gel. The gel also contained BamHI/HindIII digestion product of a subcloned fragment (pB/H76-5.5) of λHXII-76 representing the 5.5 BamHI/HindIII fragment at its 5' end (see Fig. 14). The nucleic acids were transferred to nitrocellulose and hybridized with two ³²P-labeled oligonucleotides ("a" and "b") representing -23 to -6 (oligo "a") and +89 to +106 (oligo "b") (Fig. 15) of the gene. The filter was autoradiographed for 24 hours at -70°C with an intensifying screen. *Lane A*, pB/H76-5.5 digested with BamHI/HindIII and hybridized with the "b" oligo; *Lane B*, poly A⁺ mRNA hybridized with oligo "b"; *Lane C*, polyA⁺ mRNA hybridized with oligo "a"; *Lane D*, pB/H76-5.5 digested with BamHI/HindIII and hybridized with the "a" oligo. The numbers indicate the lengths of the λDNA-HindIII marker fragments.
IV. DISCUSSION

A. CHARACTERIZATION OF THE HUMAN FACTOR XII CDNA

1. Characterization of Factor XII cDNA Clones

A factor XII cDNA of 1959 bp was isolated from a cDNA library prepared from human liver mRNA. This cDNA (pcHXII-501) codes for part of a leader peptide followed by the entire coding region of plasma factor XII, a TGA stop codon, a 150 bp non-coding region at the 3' end and a poly A+ tail of at least 75 bp. A second human liver cDNA library was screened and two identical clones (pcWFHXII-WF7,-WF9) were isolated that contained cDNA sequence coding for the signal peptide region, the putative initiator methionine residue of factor XII and three nucleotides of 5' untranslated end of the message.

The cDNA sequence data have enabled us to locate the kallikrein cleavage sites (solid arrows, Fig. 8) that produce α- and β-factor XIIa during zymogen activation (Cochrane et al., 1973; Revak et al., 1977; Dunn et al., 1982). The three cleavage sites occur carboxy-terminal to arginine residues in the sequences: Leu-Thr-Arg, Gly-Gln-Arg and Met-Thr-Arg. Because kallikrein does not recognize a unique sequence, however, the locations of other kallikrein cleavage sites in the Mr 52,000 fragment of α-factor XIIa cannot be determined from the cDNA sequence data.
2. Factor XII Is Synthesized as a Precursor

Factor XII is a secreted protein and would be expected to contain a signal peptide which functions in the transport of the protein across the rough endoplasmic reticulum membrane (Blobel et al., 1979). The cDNA sequence predicts that factor XII is synthesized as a precursor containing an amino-terminal extension of at least 19 amino acid residues (encoded by nucleotides 1-18, Fig.D and nucleotides 3-40, Fig.9). Cleavage of a Ser-Ile bond (encoded by nucleotides 16-21, Fig.8) in the precursor would give rise to plasma factor XII. This bond cleavage is consistent with the specificity of signal peptidase (Watson, 1984).

Several clotting factors require vitamin K for their biosynthesis, including prothrombin, factor X, factor IX, factor VII and protein C (Jackson and Nemerson, 1980; Suttie, 1985). Each of these proteins undergoes a post-translational modification of 10-12 glutamic acid residues by a vitamin K-dependent carboxylase. The resulting γ-carboxyglutamate residues bind calcium ions. In contrast to factor XII, the vitamin K-dependent blood clotting factors are synthesized as prepro-proteins (Kurachi et al., 1982; Degen et al., 1983; Jaye et al., 1983; Long et al., 1984; MacGillivray and Davie, 1984; Fung et al., 1984; 1985). In each of these proteins, formation of the plasma protein results from the cleavage of a bond carboxy-terminal to an arginyl residue. Since signal peptidase does not cleave such bonds (Strauss et al., 1978; Gordon et al., 1983) it has been proposed that the leader peptides of these proteins may be cleaved by signal peptidase to form proproteins, and a second protease(s) then converts the proproteins to the forms found in plasma. Other plasma proteins are also synthesized as prepro-proteins, including albumin (Straus et al., 1977; MacGillivray et al., 1979; Lawn et al., 1981), apolipoprotein A-II (Gordon et al., 1983) and PAs (Pennica et al., 1983; Freizner-Degen et al., 1986; Riccio et al., 1985). The pro-proteins
of prothrombin, factor X, factor IX, and protein C share a high level of sequence homology (Fung et al., 1985) that is not found in the pro-regions of albumin, apolipoprotein A-II or plasminogen activators. The pro region has now been shown to be important for carboxylation of the vitamin K-dependent proteins (Jorgensen et al., 1987).

**B. HOMOLOGIES WITH OTHER PROTEINS**

Comparison of residues 1-276 of factor XII with protein sequences in the National Biomedical Research Foundation (Washington, D.C.) Protein Data Base reveals extensive sequence homology with both fibronectin and plasminogen activators. Based on these sequence homologies, a schematic representation of the structure of factor XII is shown in Fig. 18. This structure is based on five different regions that share homology with regions of both fibronectin and plasminogen activators. In addition, factor XII contains a proline-rich region that does not share sequence homology with any known protein sequence. Each of these six regions of factor XII are discussed in detail in the following sections.

**1. Amino-Terminal End of the Zymogen Form of Factor XII**

The amino terminal region of factor XII (region NH$_2$-A, Fig.18) contains the signal peptide amino acid sequence with its characteristic hydrophobic core (Blobel et al., 1979). The signal peptide sequence is followed by a region that is rich in charged amino acids (amino acids 1-20, Fig. 8). There are 9/19 amino acids that are charged at neutral pH, five of which have a positive charge and four of them appear clustered near the middle of the domain. This region is a candidate for the anionic surface binding property of factor XII
Figure 18: Proposed Model for Factor XII

The model was based on amino acid sequence homologies with tPA (Pennica et al., 1983; Ny et al., 1984) and bovine fibronectin (Petersen et al., 1983). Thin arrows demarcate the proposed protein structural domains found in human factor XII. Regions A-B represents a fibronectin type II homology; B-C and D-E are growth factor-like regions; C-D region is a fibronectin type I homology; region E-F is a kringle structure and F-G is a proline rich region whose structure is undefined, and G-COOH represents the catalytic region (see text for details). Disulfide bonds have been inserted according to their arrangement in the homologous domains of other proteins; the disulfide bond arrangement of the fibronectin type II region has been determined by Patthy et al. (1984). The disulfide bonds in the catalytic region were based on the proposed model of β-factor XIIa (see Discussion). Kallikrein cleavage sites that give rise to β-factor XIIa are indicated by the heavy arrows.
although both positively and negatively charged amino acids are represented.

2. Fibronectin Homology

The next region of factor XII (region A-B, Fig. 18) shares sequence homology with the type II homology regions of fibronectin. The type II homologies are each composed of about 60 residues including four half-cysteine residues (Petersen et al., 1983), and share approximately 54% sequence identity with each other (31/57 residues in corresponding positions are identical). Residues 13-69 of factor XII share 39% sequence identity (22/57 residues are identical) and 40% sequence identity (23/57 residues are identical) with the two fibronectin sequences, respectively, including the four half-cysteine residues (Fig. 19-A). The type II homologies are present in a Mr 45,000 fragment of fibronectin that binds to collagen (see Yamada, 1983; Hermans, 1985); by analogy, the region of factor XII that shares sequence homology to the type II regions may be responsible for the putative collagen-binding properties of factor XII (Wilner et al., 1968). Evidence for collagen activation of factor XII has been strongly disputed, however, since purified collagen added to plasma does not initiate coagulation as expected for surface-dependent activation of the cascade (Griffin et al., 1975; Fujikawa et al., 1980a). The proposed disulfide-bridging pattern shown in the model for factor XII corresponds to that described by Patthy et al. (1984) for fibronectin and has a kringle-like structure in which the first and third and the second and fourth cysteine residues form disulfide bridges.

Separating the two growth factor-like regions of factor XII (described below) is a 43 amino acid peptide (regions C-D, Fig. 18) that shares limited sequence homology with the type I regions of fibronectin (Fig. 19-C). This domain in fibronectin is characterized by two
Figure 19: Homologies Between the Amino Acid Sequence of Factor XII, Bovine Fibronectin (Petersen et al., 1983) and Human tPA (Pennica et al., 1983)

Identical residues in corresponding positions are boxed; gaps have been inserted in the sequences to maximize homology. A: Comparison of Segment S3, residues 14-73 (type II-1), and residues 74-130 (type II-2) of fibronectin with residues 13-69 of factor XII (FXII).

B: Comparison of residues 51-84 of tissue-type plasminogen activator (tPA) and residues 79-111 (FXII-1) and residues 159-190 (FXII-2) of factor XII. Arrows denote glycine and cysteine residues that are invariant in this type of homology.

C: Comparison of segment S1, residues 17-57 (type I-1) and segment S1 residues 66-104 (type I-2) of fibronectin with residues 116-151 of factor XII (FXII).

D: Comparison of residues 87-173 of tissue-type plasminogen activator (tPA) with residues 193-276 of factor XII (FXII).
A  
Fibronectin Type II Homology:

Type II-1

TAVTQTYGNSN EPCVLPFTYNKTFYSCTTEGRQ DGHLCW

Type II-2

TVLVQTRGNNSNALCHFPPFLYNNHNYTDCTS EGRDRDNM KWC

FXII

KAEEHTVLTVTGEPCCHFFQYHRQLYHCKTHKGRPG PQ PW

Type II-1

STTSNYEODQQY SFC

Type II-2

GTQONYDADQQFGFC

FXII

ATTNFDQQDORWCYC

B  
Growth Factor Homology:

tPA

CSEPFCNNGGTCQALYFSDFVCQCPFGAFACKCC

FXII-2

CRTNPCCLHGGRRCLEVEGHR LCHCPVGY TGPFC

FXII-1

CSKHSPCQKGGTCVNMPSPGPHCLCPQHLTGNHC

C  
Fibronectin Type I Homology:

TYPE I-1

CKP G SYDNGKH YQIN QOWER TLYGSALVCTCUC G GSRGFC

TYPE I-2

CFDK TYNRY RGDY RER PKDS M IWDCTCIGAR GR IS C

FXII

CFEPQ LLLRFHKNIEI WYRT EQ AAV AR C QCKGPDAHCC

D  
Kringle Homology:

tPA

DTRATCYE DQGUISYRG TWSTAESGAECTN WNS S SALAQ KPYSGRR

FXII

DKASCYDGRGLSYRGLAR TLLSGAPCQP WASEATY R NVTAEQA

tPA

PDAR LG NH YCNRP DRSK PWC YVFKA G K S FSCT P A C

FXII

RNW GLGHAFCRNPD NDIRPWCFLN RDRLSW E YCDL AQ C
disulfide bonds giving a two loop structure that has been named a "finger" domain (Petersen et al., 1983). The functions of the finger domains in fibronectin may be varied and involve the fibrin, heparin and perhaps gelatin binding properties of the protein (see Yamada, 1983; Petersen and Skorstengaard, 1985).

Fibronectin is a macromolecular dimer of nearly identical polypeptide chains (Mr 220,000-250,000) which are held together at their carboxy termini by two disulfide bonds (Petersen and Skorstengaard, 1985). The protein is found inside the cell, in the extracellular matrix and in plasma (McDonagh et al., 1985). Plasma fibronectin is synthesized in the liver and is secreted. Fibronectin is also synthesized by megakaryocytes and found in platelets (Plow et al., 1979). A single gene codes for human fibronectin although three different proteins have been described which result from alternative splicing of the mRNA transcript (Schwarzauer et al., 1983). The protein is comprised of multiple structural and functional domains consisting of three repeated units, 12 with type I homology, two with type II homology and 15 with type III homology (Petersen and Skorstengaard, 1985). As determined by electron microscopy studies (Hirano et al., 1983) the chicken fibronectin gene consists of at least 48 exons with each being about 150 bp in length and correspond to the lengths of the type I and type II homologies.

The functions of fibronectin are varied due to its binding to many different biological substances, especially to those on cell surfaces (Petersen and Skorstengaard, 1985). Interactions with collagen, fibrin and cells implicates fibronectin in the process of wound repair (Clark and Colvin, 1985) either as a result of severe injury or in inflammation. After fibrin, fibronectin is the major protein found in the blood clot (4-5% of the total protein) and has been shown to be cross-linked to fibrin by factor XIIIa in vitro (Mosher, 1975).
Fibronectin also binds to collagen as well as being chemotactic for fibroblasts, neutrophils and macrophages. Therefore, it appears to be associated with various stages of tissue repair in organisms including the initial stage, cellular infiltration and the phagocytic removal of debris (McDonagh et al., 1985). Importantly, fibronectin may be a central component in cell migration, attachment and differentiation although formal proof is lacking (Clark and Colvin, 1985).

3. Epidermal Growth Factor Homology

Two regions of factor XII (regions B-C and D-E, Fig. 18) are homologous to an epidermal growth factor-like sequence that has been found in many proteins including the 19K protein from Vaccina virus, transforming growth factor type 1, tPA, and several clotting factors (Blomquist et al., 1984; Doolittle et al., 1984). In each of these proteins, there is a highly conserved region of 50 amino acids with nine invariant cysteine and glycine residues. After the insertion of three gaps in the factor XII sequences, all cysteine residues align perfectly (Fig. 19-B). The carboxy-terminal growth factor domain also contains the invariant glycine residues; in the amino-terminal domain, however, one invariant glycine residue has been replaced with a histidine.

4. Kringle Homology

Another type of homology found in factor XII is the kringle domain (region E-F, Fig. 18). Kringle domains are typically 80 amino acids in length and form three characteristic disulfide bonds. Kringles have been found in prothrombin (Magnusson et al., 1975), plasminogen (Sottrup-Jensen et al., 1978), tPA (Pennica et al., 1983) and urokinase
(Gunzler et al., 1982) and share approximately 40% sequence identity with each other including six invariant half-cysteine residues (see Jackson and Nemerson, 1980). An alignment of one of the kringle regions of tPA with factor XII is shown in Fig. 19-D. If three gaps are inserted into the factor XII sequence, the kringle shares 41% sequence identity (36 out of 87 residues are identical). The function of the kringle regions is unclear but it may be involved in binding factor XII to fibrinogen and fibrin as has been described for a tPA kringle (van Zonneveld et al., 1986). The presence of both a finger domain and a kringle suggests that the fibrin binding capacity of factor XII may be important in vivo.

5. Proline-Rich Domain

The kringle structure is followed by a region (residues 279-330, Fig 8; region F-G, Fig. 18) in which 33% of the residues are proline (17 out of 52). Despite the high proline content, this region does not share any sequence homology with other proline-rich proteins such as those from salivary glands (Ziemer et al., 1984). However, residues 299-308 of the proline-rich region of factor XII share sequence homology with residues 29-38 of calf thymus HMG-17 (Walker et al., 1977) in which eight out of ten residues in corresponding positions are identical. The significance of this homology and the function of the proline-rich region of factor XII are unclear.

6. Serine Protease Domain

Downstream of the proline-rich region in factor XII is the catalytic region (region G to the COOH-terminus, Fig. 18) that shares amino acid sequence homology with other serine proteases. An alignment of the amino acid sequences of the catalytic regions of β-factor
XIIa, plasminogen activators and the three pancreatic serine proteases is given in Figure 20. An analysis of the sequence homology between these enzymes is summarized in Table V. It is notable that the sequence identity between β-factor XIIa and the pancreatic enzymes (average value of 35%) is comparable to the identity present between the pancreatic enzymes themselves (average value of 42%).

A computer generated structural model of β-factor XIIa was developed by Dr. Brayer and G.V. Louie (Department of Biochemistry, University of British Columbia) (see Cool et al., 1985), based on the known three dimensional structures of the pancreatic enzymes. The model provided structural information about factor XII which may affect its catalytic activity. For example, it is known that all 14 cysteine residues of β-factor XIIa are involved in disulfide bridging (Fujikawa and McMullen, 1983). From the sequence alignment of β-factor XIIa (Figure 20), it is apparent that four disulfide bridges (involving residues 42-58, 136-201, 168-182 and 191-220) are conserved in all three pancreatic enzymes, and a further homologous disulfide bridge is shared with chymotrypsin (residues 1-122). The disulfide bridging pattern of the remaining four cysteine residues of β-factor XIIa cannot be determined from primary structure alignments alone. Within the tertiary structure model of β-factor XIIa, however, pairing of these remaining cysteines residues were proposed and are Cys-50 to Cys-111 and Cys-77 to Cys-80. This accounts for the seven possible disulfide bonds in the serine protease domain.

Also evident from the tertiary structural model of β-factor XIIa are the positions of major insertions that occur in the amino acid sequence of β-factor XIIa [residues 61 A-C, 109 A-E and 205 A-C (Figure 20)]. All three of these insertions occur towards the end of existing surface β-loops in the pancreatic enzymes; two insertions are positioned on the forward
Figure 20: Amino Acid Sequence Alignment of the Catalytic Regions of β-Factor XIIa (FXII), Bovine Trypsin (TRYP), Porcine Elastase (ELAS), bovine chymotrypsin (CHYM) and Human tPA

The sequences of trypsin, elastase, chymotrypsin and tPA are taken from Marquart et al. (1983), Sawyer et al. (1978), Cohen et al. (1981) and Pennica et al. (1983), respectively. The numbering system used is based on that of chymotrypsinogen A (Hartley and Kauffman, 1966), with insertions in the sequences of related enzymes denoted by letters (36A, 36B, etc.). Deletions are indicated by dashed lines. Identical residues in corresponding positions of all five proteins are boxed. The standard single letter code for amino acids is used.
### Table V. Comparative Analysis of the Amino Acid Sequence Alignment Between the Catalytic Regions of β-Factor Xlla and the Pancreatic Serine Proteases.

The total number of residues in each protein is given in parentheses below its name in the heading. For each pair of proteins aligned in Figure 20, the matrix contains the number of identical residues followed by the overall sequence homology (given as a percentage of the smaller protein) in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Chymotrypsin</th>
<th>Elastase</th>
<th>Trypsin</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Factor Xlla</td>
<td>86 (35%)</td>
<td>79 (33%)</td>
<td>84 (38%)</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>—</td>
<td>94 (41%)</td>
<td>101 (45%)</td>
</tr>
<tr>
<td>Elastase</td>
<td>—</td>
<td>—</td>
<td>87 (39%)</td>
</tr>
</tbody>
</table>

Chymotrypsin (243)  Elastase (240)  Trypsin (223)
active site face of the enzyme (residues 61 A-C and 109 A-E), and the other is positioned towards the back surface of the enzyme (residues 205 A-C). Thus, the insertions observed for β-factor XIIa occur at sites easily accommodated within the overall structural framework of the pancreatic enzymes.

The insertion of five residues at position 109 is found in a region in other proteases which shows no amino acid sequence variability amongst them and includes the pancreatic proteases, thrombin, factor Xa and factor IXa (Furie et al., 1982). Interestingly, this insert is also found in tPA (Figure 20). Since the insertion occurs on the surface of the enzyme it is most probably related to the ability of β-factor XIIa to recognize the physiological molecular surfaces with which it must interact. It is notable that the carbohydrate moiety of β-factor XIIa (bound to Asn-74; Fujikawa and McMullen, 1983) is positioned near this insertion and according to the tertiary structural model, is on the substrate binding surface of the enzyme. Specialized surface features such as these and the variable regions already discussed provide a mechanism by which complex cascade pathways like those involved in blood coagulation and fibrinolysis can occur by the specific recognition of substrate molecules (Jackson and Nemerson; Nemerson and Furie, 1980).

7. Plasminogen Activator Homology

From these sequence comparisons, it can be seen that factor XII shares remarkable sequence homology with tPA (Ny et al., 1984), a protease that converts plasminogen to plasmin in the fibrinolytic pathway (see Collen, 1980). The complete amino acid sequence of human tPA has been predicted from the cDNA sequence (Pennica et al., 1983; Fisher et al., 1985) and shows that tPA consists of a finger domain, a growth factor-like domain, two
kringle structures and the catalytic serine protease domain. This organization is similar to regions C to the COOH-terminus of factor XII except that one of the tPA kringles is replaced by the proline-rich region (Fig. 18). However, the tPA transcript contains a leader sequence with not only the signal peptide sequence but also a pro-region which is removed prior to secretion (Ny et al., 1984). No corresponding pro-region is found in factor XII but it is present in other proteases involved in coagulation such as, prothrombin, factor X, factor IX and protein C (discussed above).

8. Physiological Significance of the Homologous Protein Domains

Complex proteins such as the coagulation and fibrinolysis serine proteases and fibronectin consist of multiple structural and functional domains which have been assembled during evolution to produce the molecules found today. When a function has been defined for a particular domain such the epidermal growth factor homology, it is tempting to attribute an EGF-like function to the acceptor protein. This argument is flawed, however, since proteins with homologous structures can possess different functions [for example, hemoglobin, a protein in red blood cells and myoglobin (Blanchetot et al., 1983) have homologous domains; or, a-crystallin which is an eye lens protein, is similar to a Drosophila heat shock protein (Doolittle, 1985)]. Evidence for functionality of a protein domain within a particular protein will have to be obtained directly. For example, functional assays have been performed on the kringle domains of tPA (van Zonneveld et al., 1986). These studies showed that the kringle binds to fibrin which would be important in the degradation of fibrin by plasmin (see Collen, 1980). However, similar studies have not been performed on the putative functional domains of FXII.
C. A PERSPECTIVE ON THE ROLE OF THE CONTACT SYSTEM IN HOMEOSTASIS

Our understanding of the complex process of wound repair and the regenerative process (reviewed in Clark and Colvin, 1985) is becoming clearer as the components of the plasma systems involved are being defined. The plasma systems, blood coagulation, fibrinolysis and inflammation have already been described in detail (see "Introduction"). Other plasma systems involved in wound repair include the complement system and the immunoglobulin molecules and are described elsewhere (Bennett and Ogston, 1981; Pesce and Dosekun, 1983). Also important to the healing process is the regeneration of new tissue at the site of injury. Epithelium cells at the wound margin are stimulated to elongate and divide and migrate to cover the thick mat of fibrin and fibronectin. Blood vessels sprout branches and migrate to the wound bed and also begin to divide. Fibroblasts and macrophages are summoned to the wound and newly synthesized glycosaminoglycans (GAG), fibronectin and collagen, probably supplied by the fibroblasts are used to form the new extracellular matrix. Central to these events is fibronectin which binds to fibrin and fibronectin. Blood vessels sprout branches and migrate to the wound bed and also begin to divide. Fibroblasts and macrophages are summoned to the wound and newly synthesized glycosaminoglycans (GAG), fibronectin and collagen, probably supplied by the fibroblasts are used to form the new extracellular matrix. Central to these events is fibronectin which binds to fibrin, GAG, collagen, can stimulate cell migration (epithelium) and is chemotactic for neutrophils and macrophages.

The convergence of the plasma systems involved in wound repair and the regenerative process occurs as a result of the assembly of the essential components at the site of injury. Plasma and tissue factors initiate the coagulation cascade producing a network of fibrin. Trapped in this mesh are the molecules that are vital for repair. These include plasminogen activators, plasminogen, fibronectin and perhaps factor XII. The proteins are all related structurally although it is not known if each homologous domain exhibits similar function. The plasminogen activator, factor XII and plasminogen have homologous kringles (possible
fibrin binding property) and serine protease domains. Plasminogen activator, factor XII
and fibronectin are related through the type I homology (possible fibrin binding property).
Factor XII also contains a type II homology (possible collagen binding property) found in
fibronectin.

Interestingly, like fibronectin, factor XII appears central in the processes of wound repair
(see "Introduction") in that factor XII is important in the contact activation of coagulation,
fibrinolysis, inflammation and kinin generation through the activation of prekallikrein.
Also, the epidermal growth factor domains in factor XII may be preordial growth factors
which may still have stimulatory properties since at the site of injury, cell proliferation
takes place and it may be possible that these regions in factor XII (and tissue-type
plasminogen activator) may play a minor role. Factor XII and fibronectin are structurally
unalike except in the fibronectin type I and type II homologous domains found in factor XII.
It is possible that during the process of evolution, factor XII acquired these two domains (see
below), and since factor XII was also functional in wound repair, the newly acquired
domains may have assumed similar function. The amino acid sequence of the factor XII
protein has changed during the process of evolution but may still have retained essential
amino acids required for the binding of collagen or fibrin. Finally, this evolutionary
strategy may also be applied to the other proteins important in wound repair (plasminogen
and tissue-type plasminogen activator) which have also acquired the homologous domains
(kringles and type I homology).

It should be emphasized that there is only circumstantial evidence to implicate fibronectin in
the wound repair and the regenerative processes. In addition, factor XII and prekallikrein
are not essential for wound repair since individuals deficient in either of these proteins are
assymptomatic. However, it is likely that the plasma systems are interrelated in such a way that if an initiation factor is not present then the reaction will still occur via another, perhaps less efficient, mechanism. Secondary activation mechanisms of any one of these plasma systems involved in wound repair are crucial since inactivation of the system would probably be lethal to the organism.

An understanding of the mechanism of gene evolution may elucidate the importance and relevance of the exchange of protein domains within a family and between families of proteins. This study of the factor XII gene structure has provided more data concerning the evolution of the serine proteases and their structural and functional relationships.

D. DNA SEQUENCE ANALYSIS OF THE HUMAN FACTOR XII GENE

Two recombinant phage containing the entire factor XII gene were isolated from a human genomic phage library. The gene is approximately 12 kb in length and consists of 13 introns and 14 exons. Restriction enzyme mapping and DNA sequence analysis were used to localize all exon sequence for factor XII mRNA in the gene (Fig. 12). An AATAAA sequence is often found approximately 20 bp 5' to the polyadenylation site in eukaryotic genes (Birnstiel et al., 1985). This sequence was located at nucleotides 11693-11699 in the FXII gene. A consensus sequence CAYTG is another possible recognition site downstream of the polyadenylation site (Berget, 1984). A similar sequence was found at position 11703 as CTTTG. All the splice junction sequences agree with the consensus sequence of Mount (1982) except for the splice donor of intron J which has a GC instead of a GT dinucleotide. However, other genes have been described with similar junction sequences (Wieringa et al., 1984; Dush et al., 1985; Irwin, 1986) and Aebi et al. (1986) have shown that GC is a
reasonable splice donor sequence which is spliced accurately in vivo.

An unusual feature of the FXII gene is that it has many very small introns such that exons 3 through 14 are contained in a 4.2 kbp fragment of genomic DNA with three introns less than 100 bp in length. However, due to the presence of the first two large introns, the total length of the gene is 12 kb which is an expected size when compared to other genes with similar lengths of coding sequence (Naora and Deacon, 1982).

**E. ANALYSIS OF THE PROMOTER REGION IN THE HUMAN FXII GENE**

Promoters of many genes have a CCAAT or a TATA element (reviewed in Corden et al., 1980; Breathnach and Chambon, 1981; Hansen and Sharp, 1984; Dynan and Tjian, 1985; Bucher and Trifonov, 1986) which are positioned at -70 to -80 bp and -27 to -34 bp from the transcriptional start site, respectively. The primer extension and nuclease S1 mapping results for the human FXII gene show that these elements are not present in the 5' end of the gene although a CCAAG sequence at -105 and a CTATTTC at -26 bp are present (see Table VI for frequency of nucleotides at each position).

1. The CAAT Element

These promoter elements are not essential for the expression of all RNA Polymerase II-transcribed genes. For example, the CCAAT region in the globin gene family is required for expression but there is no corresponding sequence in many genes expressed in the liver such as the blood clotting factors, retinol binding protein (Laurent et al., 1985) and α-1 antitrypsin (Table VI). Deletion studies on the α-1 antitrypsin (Ciliberto et al., 1985) and
Consensus (Corden et al., 1980)

(-33 to -27) (-27 to -21)

T(37) A(39) T(37) A(35) A(37) T(20)
A (3) T (1) A (4) T (3) T (18) T (4) A (14)
G (6)
C (1)

\[ a -1 \text{ Antitrypsin} \]
(Ciliberto et al., 1985) -26 TTAAATA

Bovine Prothrombin
(Irwin, 1986) -29 CATTAAC

Factor VIII
(Gitschier et al., 1984) -28 TAAAAAG

Factor IX
(Yoshitake et al., 1985) -44 TAAATAC

Protein C
(Plutzky et al., 1986) -34 CAAATAT

E-\(a\) Immune Response
(Mathis et al., 1983) -30 TTATAAT

Factor XII
-26 CTATTTC

Table VI. Summary of TATA-like Elements Found in Liver Specific Genes, Coagulation Factor Genes and the E-\(a\) Immune Response Gene.

The consensus sequence for TATA-like sequences was determined from 40 eukaryotic Polymerase II transcribed genes and were summarized in Corden et al. The numbers above the sequence elements describe the position in the gene from the transcriptional start sites in where these elements are commonly found.
albumin (Gorski et al., 1986) promoter regions have shown that both genes have strong promoters upstream of CAAT or TATA regions. Albumin has a CCAAT element at -89 bp, but it was shown that deletion of this region reduced the transcription rate by only 10% whereas sequences further 5' reduced the rate by 85%. In addition, promotion of both the albumin and α-1 antitrypsin genes required liver specific factors since the promoters did not function significantly when the deletion mutants were expressed in tissues that did not normally produce the gene products. Therefore, it appears that there are strong promoters in these liver genes that are independent of the CAAT regions. Although similar studies have not been described in the literature for clotting factor promoters, it is possible that the genes which are highly expressed (prothrombin and antithrombin III) probably will also have non-CAAT-like promoter sequences recognized by liver specific factors.

2. The TATA Box

The blood clotting factor genes as well as many others have atypical or no canonical TATA elements (Table VI). Other genes such as the housekeeping genes [3-hydroxy-3-methylglutaryl CoA (Reynolds et al., 1984); dihydrofolate reductase (Crouse et al., 1982); hypoxanthine-guanine phosphoribosyl transferase (Melton et al., 1984), and adenosine deaminase (Valerio et al., 1985)], Thy-1 (Giguere et al., 1985), EGF receptor (Ishii et al., 1985) and viral promoters (Baker et al., 1979; Ghosh et al., 1979) do not have TATA elements but do have G+C rich regions 5' to the transcriptional start site. These genes also often have Sp1 binding sequences (GGGCGG) (reviewed in Dynan and Tjian, 1985). Another class of non-TATA like promoters includes those which have no apparent DNA features that are particularly distinguishable. These genes are human β-tubulin (Lee et al., 1983), T-cell receptor/T3 complex (van den Elsen et al., 1986) and perhaps factor XII.
In most of the cases listed, all the mRNA transcripts appear to have heterogeneous 5' ends suggesting that the major role of the TATA sequence in these genes is to direct the position of the 5' end of the transcript.

Variability in the nature and function of the TATA sequence is demonstrated in *in vivo* deletion studies. The TATA sequences in the SV40 (Benoist and Chambon, 1980) and sea urchin H2A (Grosschedl *et al.*, 1980) gene promoters showed that the TATA sequence was required for precision in transcript initiation but did not affect transcription rates. Sequences around the mRNA capping site may also contribute to the determination of the transcription initiation site as illustrated in a mutant rabbit β globin gene which lacks a proper TATA sequence (Charnay *et al.*, 1985; Dierks *et al.*, 1983) and yet the majority of mRNA transcripts contained the correct 5' end. On the other hand, both rate and placement of transcript initiation required TATA sequence elements in expression of the thymidine kinase and globin genes (McKnight and Kingsbury, 1982; Charnay *et al.*, 1985; reviewed in Hansen and Sharp, 1984). Certainly the lack of TATA or presence of weaker TATA-like sequences (Table VI) is exceptional as most eukaryotic RNA Polymerase II genes do have the consensus sequence (Breathnach and Chambon, 1981). However, many genes function normally without the TATA sequence and probably possess other factors pertinent to transcription initiation and regulation.

Until further studies are performed with the FXII 5' flanking region, the identity of the regulatory sequence(s) of the factor XII gene remains unknown. The nuclease S1 and primer extension products for FXII strongly suggest that FXII is lacking a typical TATA sequence possibly resulting in transcription initiation at more than one site. However, analysis of the 5' flanking regions of the genes listed in Table VI and those discussed above
indicates that many genes are lacking or have atypical TATA sequences. The absence of true CAAT or TATA sequences may be a means of maintaining a low level of transcription for those genes coding for low abundant mRNAs (Bucher and Trifonov, 1986). Genes which require stronger promotion but do not have typical promoter sequences (for example, α-1 antitrypsin and prothrombin, Table VI) may possess unique promoter elements such as those recognized by liver specific factors (Ciliberto et al., 1985; Gorski et al., 1986) resulting in high levels of gene expression.

**F. PUTATIVE PROTEIN DOMAINS ARE FOUND ON SEPARATE EXONS**

The positions of the introns in the factor XII gene divide the coding region into discrete protein modular units (demarcated by arrows in Fig. 18) which share amino acid sequence identity with similar regions in tPA (Ny et al., 1984) and fibronectin (Peterson et al., 1983). Comparisons of the FXII amino acid sequence with those of tPA and fibronectin show that there is only approximately 30-40% (Figure 19) sequence identity but the positions of half-cysteine residues in each unit are conserved.

The schematic representation of the FXII coding sequence, shown in Fig. 21, was based on sequence homologies found in tPA and fibronectin. The lengths of each exon are given in Table IV. A putative signal peptide sequence (Blobel et al., 1979) at the N-terminus of the zymogen is located on the first exon (NH2-A, Fig. 21) followed by a region of unknown homology (A-B, Fig. 21) that is encoded by the second exon. A "type II" homology with apparent collagen binding properties in fibronectin (Petersen et al., 1983; Yamada, 1983) is represented by exons three and four (B-D, Fig. 21). Next is the EGF-like domain (D-E, Fig. 21) on exon five (Blomquist et al., 1984; Doolittle et al., 1984) followed by the "type I"
Figure 21: Schematic Diagram of the FXII Protein Coding Domains and the Positions of Introns.

The standard one-letter code of each amino acid is given in the open circles. The solid black bars, indicating possible disulfide bonding between cysteine residues is based on protein homologies of other proteins (Magnusson et al., 1975; Yamada, 1983; Blomquist et al., 1984; Patthy et al., 1984; Cool et al., 1985). The arrows A - M indicate the positions of introns in the coding sequence. Ile at position 1 indicates the first residue of the zymogen and is marked by a curved arrow. The short, closed arrows show cleavage activation sites by kallikrein which produce β-factor XIIa.
homology, or, the fibrin finger in fibronectin (E-F, Fig. 21) (Petersen et al., 1983) found on exon six. Exon seven encodes another EGF-like domain (F-G, Fig. 21) preceding a kringle structure and a proline rich region (G-I, Fig. 21) on exons eight and nine.

FXII activation requires cleavage by kallikrein, carboxy terminal to two Arg residues (thick arrows in Fig. 21), generating one form of the serine protease, $\beta$-FXIIa (I-COOH, Fig. 21). The serine protease consists of five exons with the catalytic triad residues, His, Asp and Ser separated by introns. Exon 14 is the largest (320 bp) and encodes 55 amino acid residues of the serine protease and 150 bp of the 3' untranslated end of the mRNA.

Although the protein domains described above may be functional in other proteins, it is not yet known whether similar functions can be attributed to FXII or tPA. It has been speculated that exons may represent functional building blocks in a protein and that these elements can be exchanged between existing genes and generate new kinds of proteins (Gilbert, 1978; Blake, 1978). Patthy (1985) has suggested that the proteins involved in coagulation and fibrinolysis may have been altered through this action of exon shuffling to account for their present day structures.

G. EVOLUTION OF THE FACTOR XII GENE

The number and position of introns can be used to describe the evolutionary relatedness of proteins (Craik et al., 1983) since proteins with conserved structure and function have common gene organizations and have probably arisen through gene duplication events (Patthy, 1985; Rogers, 1985). The urokinase and tPA genes have been described (Ny et al., 1984; Riccio, et al., 1985; Friezner Degen et al., 1986) and are a family of serine proteases.
separate from the clotting factors (Patthy, 1985; Rogers, 1985; Irwin, 1986). It was anticipated that since FXII and tPA had many regions of protein homology which were organized in the same manner in both proteins and that half-cysteine residues were conserved, then the number and position of introns within their genes would likely be similar. The positions of introns of the plasminogen activators were compared to those of FXII and were found to be more conserved in the carboxy terminal half of the molecules than in the amino terminal region of the proteins (Figs. 22 and 23).

1. Evolution of the Amino Terminal Half of Factor XII

Protein coding sequences of the amino terminal regions of the plasminogen activators and FXII are interrupted by introns in identical codon reading phases although there is considerable variation in the coding sequence itself. For example, there are only two exons (those encoding the EGF-like domain and the kringle) which are common to all three proteins but the lengths of each exon is conserved and varies by only 10 bp. These exons, whose functions have been described above, have similar sequence identity and conserved half-cysteine residues (Fig. 22). The fibronectin type I homology is found amino terminal to the EGF-like domain in FXII and tPA but not in human urokinase. The fibronectin homology appears in the FXII and tPA genes on separate exons of almost identical length and with conserved half-cysteine residues.

Further variation in gene organization was found in the connecting regions in which there is a kringle in tPA (85 amino acids), a proline-rich region in FXII (55 amino acids), but no corresponding region in urokinase. The porcine urokinase protein has a 27 bp insert at this position relative to the human urokinase protein (Nagamine et al., 1984). Analysis of the
Figure 22: Comparison of the Exon Organization in the N-Terminal Chains of Human FXII, tPA (Ny et al., 1984) and uPA (Riccio et al., 1985).

Exons are represented by open bars; 5' untranslated regions are represented by solid bars and signal or pre-regions by dotted bars. The sizes of exons are drawn to scale. Triplet codon phase in which the exon is interrupted by the intron is designated as 0, I or II. The sizes of the introns are not to scale. The vertical lines below the open bars represent positions of cysteine residues in the exons. The lettering in the open boxes refer to the type homology encoded by the exon: FnII, type II homology in fibronectin; EGF, epidermal growth factor; FF, type I homology in fibronectin, the "fibrin finger"; K, kringle; and P, proline-rich region. The scale represents 50 bp.
gene shows that the intron/exon junction has shifted downstream 27 bp keeping the protein in the correct reading frame (Nagamine et al., 1985). The insert is not long enough to be a result of an inserted exon with subsequent loss of an intron in the porcine gene (Blake, 1983).

Kringles are encoded by two exons (reviewed in Patthy, 1985), interrupted by an intron in the II codon reading frame. The second exon does not usually contain more than about 30 amino acids of coding sequence in the plasminogen activators, plasminogen or prothrombin (Patthy, 1985). Unlike these genes, the FXII gene contains a 55 amino acid proline-rich region attached to the second exon of the kringle. However, the length of the first exon of the FXII kringle and the placement of the middle intron are almost identical to the plasminogen activator kringles and therefore, the kringle in the FXII gene was probably present prior to the time of FXII and plasminogen activator divergence. It is possible that, by exon shuffling or another mechanism of gene evolution, FXII acquired a proline-rich exon between the second exon of the kringle and the first exon of the serine protease domain with subsequent loss of an intron. Other genes such as collagen, insulin and actin exhibit loss of introns when their intron/exon gene organizations are compared to other members in their respective families (see Blake, 1983).

Sequences amino terminal to the fibronectin type I homology domain in the plasminogen activators and FXII share no common homologies although codon phase interruption is conserved (Fig. 22). The FXII gene has two exons which bear amino acid sequence homology to the type II collagen binding domain in fibronectin and an EGF domain. These domains are not present in the plasminogen activators. However, the plasminogen activators have a prepro region which is cleaved prior to secretion, generating the single...
chain form of the plasminogen activators found in plasma (Ny et al., 1984). This region is found in the gene as part of the second exon which also contains 26 bp of the 5' untranslated end of the transcript and nucleotide sequence for the signal peptide domain. The prepro region is interrupted by an intron in the plasminogen activators and shares the third exon with amino acid sequence of the zymogens. This third exon is very small (27 nucleotides) in urokinase and (45 nucleotides) in tPA, and is flanked by O and I triplet codon intron insertions in both genes. The second exon of FXII encodes the first 19 amino acids of the zymogen, which follows the signal peptide domain and is flanked by introns in the 0 and I reading frames, similar to the third exon of the plasminogen activator genes. Since it is not known whether exons are lost or are acquired through mechanisms such as exon shuffling in the genome, it is difficult to assess the actual structure of the ancestral gene prior to the divergence of the plasminogen activator and FXII genes.

The untranslated 5' end of the plasminogen activators and FXII genes are also different. There is an intron in the plasminogen activators which is not found in FXII. Urokinase and tPA differ also in that the intron is inserted in different reading frames in the genes and the lengths of the first exons vary. The first exon of FXII contains coding sequence for the 5' untranslated end (47 bp) and the signal peptide sequence. Interestingly, the first intron in the FXII gene is positioned immediately following the last residue of the signal peptide in the O triplet codon reading frame. The 5' untranslated region of the gene is subject to more variation than the coding regions as observed in the plasminogen activators themselves. Furthermore, the clotting factors, human FIX (Yoshitake et al., 1985), bovine prothrombin (Irwin, 1986) and human protein C (Foster et al., 1985; Plutzky et al., 1986) which belong to the same family of serine proteases (Rogers, 1985; Irwin, 1986) also exhibit variation in that there is an intron in the 5' untranslated region of the protein C gene but not in the
2. Evolution of the Serine Protease Domain of Factor XII

The serine protease domains of FXII, tPA and uPA consist of five exons, four of which are flanked by introns that have inserted into the same triplet codon reading frame in all three genes (Fig. 23). This organization was compared to other serine proteases (Fig. 23) in order to determine the evolutionary history of the FXII gene. Serine proteases can be divided into five different families of genes based on their intron/exon gene organization (Rogers, 1985; Irwin, 1986). Three of these families are shown in Figure 23 and are: first, the pancreatic, kallikrein, nerve growth factors, plasminogen activators and factor XII family; second, the factor IX and protein C family; and third (but closely related to the FIX and Protein C family), the prothrombin family. Two other distantly related families belonging to the serine protease superfamily of proteins (not shown in the figure) are represented by haptoglobin and complement factor B (Rogers, 1985). Haptoglobin has lost its serine protease activity although it has retained its activation by proteolysis property common to the serine proteases. Also, the haptoglobin gene does not have introns in the equivalent domain (Doolittle, 1985; Rogers, 1985). Complement factor B has serine protease activity upon activation by proteolysis and has two more introns compared to the FXII/PA family of proteins.

Although there are differences, FXII is closely related to the plasminogen activator, pancreatic, nerve growth factor and kallikrein family of serine proteases since each of these genes contain introns 3' to codons for both the active site histidine and aspartate and a 5' intron to the codon for the active site serine (Fig. 23). However, the placement of all four
Figure 23: Comparison of the Serine Protease Exon Organization in the Trypsinogen (Craik et al., 1984), Chymotrypsinogen (Bell et al., 1984), Proelastase (Swift et al., 1984), Kallikrein (Mason et al., 1983; van Leeuwen, 1986), α and γ Subunits of Nerve Growth Factor (Evans and Richards, 1985), Tissue-type Plasminogen Activator (Ny et al., 1984; Degen et al., 1986), Urokinase (Nagamine et al., 1984; Riccio et al., 1985), Factor IX (Anson et al., 1984; Yoshitake et al., 1985), Protein C (Foster et al., 1985; Plutzky et al., 1986), Prothrombin (Irwin, 1986) and Factor XII genes.

Exons are shown as open boxes; the scale at the bottom represents 100 bp. Triplet codon phase in which the intron is positioned is represented as 0, I or II. The activation sites of the zymogens are shown by an arrow (the gamma subunit of nerve growth factor is not activated this way). The codons for the histidine, aspartic acid and serine residues comprising the catalytic triad are indicated by H, D and S, respectively. The 3' untranslated ends of factor IX, tPA and urokinase have been abbreviated and are 1935, 914 and 1119 bp in length, respectively. Dotted bars represent exons coding for signal peptides and slashed bars represent 3' untranslated regions of the mRNAs.
introns in almost identical locations and in the same codon phase in the plasminogen activators and FXII genes suggests that these genes may be more closely related to one another than to the other serine proteases of this family. Figure 23 also shows that the intron/exon gene organization of FXII and the plasminogen activators is unlike that found in the FIX and protein C family and the prothrombin gene.

The evolution of the serine protease family of genes has been intensively studied as the amino acid sequence and tertiary structure of many of the proteins involved have been determined as well as their respective intron/exon gene organizations. In addition, these proteins represent a very old family of genes, present in both prokaryotic and eukaryotic species. The presence of the blood proteins probably arose in organisms when the vertebrates evolved (Doolittle, 1984). At this time a primordial protease gene was amplified by gene duplication events and the family was generated (Young et al., 1978; Doolittle, 1984a; Neurath, 1984; Doolittle, 1985; Patthy, 1985).

Important to gene evolution is the emergence of introns. It is not known and is heavily debated as to whether introns invaded coding sequences or introns from ancient genes were lost during the evolution of gene families (Blake, 1983; Gilbert, 1979; Rogers, 1985). Analysis of the serine proteases in Figure 23 shows that the introns could have been present prior to gene duplication and subsequently lost. Thus, the older the protein, the fewer introns which are present. For example, the most distantly related protein is haptoglobin which has no introns and may have been present in invertebrate species prior to the clotting factors. Also, invertebrates generally have fewer introns (Gilbert, 1985; Gilbert et al., 1986) due to loss of introns or less insertion. The role of introns is not known but may be important in providing raw genetic material (as in intron/exon sliding events observed in the
porcine gene and described below) or they allow recombination to occur between genes but not within coding sequences so that important protein domains are not interrupted (Gilbert, 1979). Evidence for insertion of introns or loss of introns will be provided as we gain more information about the intron/exon organization of related proteins in lower species and a time scale of intron loss or gain in a protein family can be assigned.

3. Intron/Exon Sliding

The mechanism of intron/exon junction sliding could explain any small variation in lengths of exons found in homologous domains (Craik et al., 1983; Nagamine et al., 1985) in the plasminogen activators and FXII (Fig. 23). According to the tertiary structural model proposed for β-FXIIa (Cool et al., 1985), there are three insertions in the serine protease domain which occur on the surface of the molecule. Two of these inserts (also found in tPA) appear at or close to a junction (positions Arg₃₉₄ to Ala₄₀₀ and Glu₄₅₀ to Gly₄₅₄, Fig. 14) and are at positions 61 and 109 of the serine protease domain (Figure 20). Craik et al. (1983) proposed that intron/exon sliding, which generates new coding sequence occurs on the surfaces of molecules. Such modifications in β-factor XIIa also map to the surface of the molecule (Cool et al., 1985). However, the third insertion at position 205 is 13 amino acids away from the junction (position Ala₅₅₅ to Arg₅₅₇, Fig. 14) and could be acquired through other mechanisms. These length variations could confer new function to a molecule since the insert at 109 is near an Asn-74 residue to which a carbohydrate moiety is bound (Fujikawa and McMullen, 1983) and, according to the tertiary structural model, the insert is on the substrate binding face of the enzyme.
**H. A MODEL FOR THE EVOLUTION OF THE HUMAN FACTOR XII GENE**

According to the model of evolution of the serine proteases proposed by Platthy (1985) and based on the amino acid sequence homology in the serine protease domain, FXII and the plasminogen activators evolved separately from the coagulation family of proteins (Figure 24). The model also illustrates that the plasminogen activator family diverged from the clotting members after the EGF(A) domains were acquired but before the kringles. It is interesting that after this divergence, both the FXII and the clotting factor genes were invaded by another EGF domain of the B type but the plasminogen activators were not, unless this exon was lost during evolution. The uPA and tPA/FXII genes separated at some time prior to the acquisition of the fibrin finger domain in tPA and FXII. Since the amino terminal region of the plasminogen activators and FXII proteins are remarkably different in terms of their structure and function, it is very difficult to apply the model of exon shuffling completely to the evolution of the FXII gene as it relates to either the plasminogen activator or the family of coagulation proteins. Perhaps by some recombination event, all or parts of the amino terminal region of an ancient gene (represented by the 5' untranslated end, signal peptide sequence, fibronectin type II homology and possibly the EGF domains in FXII) were spliced onto or fused with a primordial plasminogen activator gene (Doolittle, 1985) followed by the insertion of exon domains as described in the exon shuffling model. A recombination event involving gene fusion has also been proposed for the generation of the FXI and prekallikrein genes (Chung et al., 1986). The model suggests that the amino terminal chain (heavy chain consisting of the four tandem repeat units) was fused with the serine protease domain of a FIX related ancestral gene (Chung et al., 1986) producing FXI or prekallikrein. Analysis of plasminogen activators and clotting factors of lower vertebrates and invertebrates may verify this model and provide a time scale to the events.
Figure 24: A Model for the Evolution of the Plasminogen Activators and FXII.

Rectangles with S represent the serine protease domain. Triangles with $G_B$ or $G_A$ represent the epidermal growth factor type B or type A homology, respectively. Diamonds with I or II represent fibronectin homologies with the fibrin finger or collagen properties, respectively. Circles with a K indicates a kringle domain or with an N, the amino terminal domains for the plasminogen activators. Circles with an X represents the unknown homology in FXII with possible surface binding properties. A rectangle with N indicates the factor XII amino terminus or with Pro, the proline-rich domain. (see text for details)
CONCLUDING REMARKS

The complete amino acid sequence of human factor XII, including the signal peptide sequence, has been predicted from the DNA sequence of the cDNA and the gene. The human gene has been identified and the intron/exon gene organization and the 5' and 3' flanking regions determined. Complete characterization of the human factor XII protein and its gene has expanded our understanding of the role of factor XII in the contact system. The presence of putative functional domains which bear homology with fibronectin and plasminogen activators suggests that the overlapping functions in wound repair for all these proteins may be a result of these common protein domains. The major plasma systems involved in wound repair, coagulation, fibrinolysis, complement and regeneration require the components of the contact system (factor XII, prekallikrein, HMW kininogen), plasminogen activator and fibronectin. Since the components involved have overlapping functions (at least in vitro) they may be the means through which the plasma systems are interrelated and homeostasis maintained during injury.

Analysis of the gene for factor XII shows that it is a member of the plasminogen activator gene family. Residual function may still be present in the modern factor XII molecule so that it can act, perhaps weakly, in plasma systems other than coagulation. The complexity of these serine protease and fibronectin molecules seems to depend on the acquisition of multiple protein domains. Whether this mechanism is merely an evolutionary strategy to rapidly increase sequence variation in proteins or whether it is a strategy to confer new function as well on a protein remains to be determined for human factor XII.
V. LITERATURE CITED


Bovine Factor XII (Hageman Factor). Biochemistry, 16; 2270-2278.


USA 83, 5086-5090.


Complementary Deoxyribonucleic Acid Coding for Human and Bovine Plasmingen. Biochemistry 23; 4243-4250.


Literature Cited / 174


USA 80,137-141.


1056-1065.


Sequence Coding for Human Coagulation Factor XII (Hageman Factor). Nucl. Acids Res. 14, 3146.


Young, C.L., Barker, W.C., Tomaselli, C.M., and Dayhoff, M.O. (1978). Serine Proteases, in Atlas of Protein Structure (Dayhoff, M.O. Ed.) vol. 5 (suppl. 3), National Biomedical Research Foundation, Silver Spring, Maryland, pp. 73-93.


