CHARACTERIZATION OF THROMBIN cDNAs RANGING FROM MAMMALS TO CYCLOSTOMES: structural analysis and evolution of prothrombin in vertebrates.

by

David Karl Banfield

B. Sc (Hon.), Simon Fraser University, 1986.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy in

THE FACULTY OF MEDICINE

Department of Biochemistry

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1991

(c) David K. Banfield, 1991
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of BIOCHEMISTRY

The University of British Columbia
Vancouver, Canada

Date [9.9.9]
ABSTRACT

The cDNA sequence of the B-chain of thrombin has been determined from nine vertebrate species (rat, mouse, rabbit, chicken, gekko, newt, rainbow trout, sturgeon, and hagfish). The amino acid sequence identities vary from 96.5% (rat vs. mouse) to 62.6% (newt vs hagfish). Of the 240 amino acids spanned in all the species compared, there is identity at 110 (45.8%) positions. When conservative changes are included, the amino acid similarity increases to 75%. The most conserved portions of the B-chain are the active-site residues and adjacent amino acids, the B-loop, and the primary substrate binding region. In addition, the Arg-Gly-Asp motif is conserved in 9 of the eleven species compared and the chemotactic/growth factor domain is well conserved in all of the eleven species compared. The least conserved regions of the B-chain correspond to surface loops including the thrombomodulin binding-site and one of the hirudin binding regions. The extent of the amino acid sequence similarity and the conservation of many of the functional/structural motifs suggests that in addition to their role in coagulation, vertebrate thrombins may also play an important role in the general mechanisms of wound repair.

In addition, the complete cDNA sequences of chicken and hagfish prothrombin has been determined. The sequences predict that prothrombin from both species is synthesized as a prepro-protein consisting of a putative Gla domain, two kringles and a two chain protease domain. Like the mammalian prothrombin cDNAs, both chicken and hagfish prothrombin have either AGT or AGC encoding their active-site serine. Chicken and hagfish prothrombin share 51.6% amino acid sequence identity (313/627 residues) and 61.7% similarity. Hagfish prothrombin contains a 19 amino acid residue insertion between the Gla domain and the first kringle. This insertion in hagfish prothrombin may form a 16
amino acid residue disulphide loop. Both chicken and hagfish prothrombin are structurally very similar to human, bovine, rat, and mouse prothrombin and all six species share 41% amino acid sequence identity and 67.5% amino acid sequence similarity. Amino acid sequence alignments of human, bovine, rat, mouse, chicken, and hagfish prothrombin suggest that the thrombin B-chain and the Gla domain are the regions most essential for the common function of vertebrate prothrombins. The structural composition of hagfish prothrombin implies that vertebrate prothrombins acquired the propeptide, Gla domain, and both kringles prior to the divergence of cyclostomes from the main vertebrate lineage (400 MYA). Attempts to identify an invertebrate prothrombin homologue were unsuccessful, supporting the notion that the hemostatic processes in invertebrates and vertebrates may be quite dissimilar.
# TABLE OF CONTENTS

**ABSTRACT** ........................................................................................................... ii

**Table of Contents** ............................................................................................... iv

**List of Tables** ........................................................................................................ viii

**List of Figures** ......................................................................................................... ix

**LIST OF ABBREVIATIONS** ................................................................................... xi

**Acknowledgements** ............................................................................................... xiv

**I. INTRODUCTION** ................................................................................................. 1

A. An Overview of Hemostasis ................................................................................... 1

1. The Endothelium ...................................................................................................... 1

2. Vasoconstriction ..................................................................................................... 3

3. The Formation of the Platelet Plug. ...................................................................... 3

4. Blood Coagulation ................................................................................................. 6

5. Fibrin Formation ..................................................................................................... 10

6. Fibrinolysis ............................................................................................................. 13

B. The Evolution of Hemostatic Mechanisms. ............................................................. 14

1. Coagulation in Non-mammalian Vertebrates ...................................................... 15

2. Coagulation in Invertebrates ............................................................................... 16

C. Prothrombin ............................................................................................................ 20

1. Biosynthesis and Post-translation Modifications ............................................... 20

2. Structure of the cDNA and Gene for Prothrombin ........................................... 21

3. The Propeptide ..................................................................................................... 23

4. The γ-carboxyglutamic acid Containing Region .............................................. 24

5. The Kringle Domain ............................................................................................ 25

6. The Aromatic Amino Acid Stack ....................................................................... 26

7. The Protease Region ............................................................................................. 26

   a. Non-enzymatic Functions of Thrombin. ......................................................... 27

   b. Three Dimensional Structure of Thrombin. .................................................... 28

   c. Thrombin Receptor ......................................................................................... 32

D. Families of Serine Protease Zymogens ................................................................. 33
E. Homologous Domains Found Within the Coagulation and Fibrinolytic Proteins .................................................................34
  1. Homologous Domains Within the Activation Peptide of Prothrombin .................................................................34
  2. Homologous Domains Found in Coagulation and Fibrinolytic Proteins Other Than Prothrombin ...........................................34
F. The Evolution of Serine Proteases Involved in Coagulation and Fibrinolysis .................................................................37
  1. Molecular Mechanisms Involved in the Evolution of Coagulation and Fibrinolytic Proteins ...........................................39
  2. The Unique Features of the Serine Codon .........................................................................................................................40
  3. Relationships Among the Coagulation and Fibrinolytic Serine Proteases ........................................................................41
  4. Objectives of the Present Study ......................................................................................................................................44
II. MATERIALS AND METHODS ................................................................................................................................................46
  A. MATERIALS .......................................................................................................................................................................46
  B. STRAINS, VECTORS, AND MEDIA ..................................................................................................................................48
    1. Bacterial Strains and Vectors ....................................................................................................................................48
    2. Media ................................................................................................................................................................................48
  C. COLLECTION OF ANIMAL SPECIMENS .................................................................................................................................49
    1. Collection of Hagfish Specimens ................................................................................................................................49
    2. Collection of Plasma from Hagfish .................................................................................................................................49
    3. Hagfish Liver Collection ..................................................................................................................................................50
  D. GEL ELECTROPHORESIS ......................................................................................................................................................50
    1. Non-Denaturing agarose gels and Southern Blots ..............................................................................................................50
    2. Denaturing Agarose Gels and Northern Blots ....................................................................................................................51
    3. Denaturing Polyacrylamide Gels .......................................................................................................................................52
    4. SDS-Polyacrylamide Gels ................................................................................................................................................52
  E. PURIFICATION OF HAGFISH PROTHROMBIN .....................................................................................................................53
  F. ISOLATION OF RNA ............................................................................................................................................................53
    1. Isolation of Total Cellular RNA .......................................................................................................................................53
2. Isolation of Poly A+ RNA .......................................................... 54

G. ISOLATION OF DNA .............................................................. 55
1. Isolation of Plasmid DNA .......................................................... 55
2. Phage DNA Isolation ................................................................. 56
3. Genomic DNA Isolation ......................................................... 57

H. THE POLYMERASE CHAIN REACTION .................................. 57
1. Preparation of Single-Stranded cDNA ...................................... 57
2. Selection of Primers for use in the Polymerase Chain Reaction .... 58
3. Reaction Conditions for the Polymerase Chain Reaction ........... 61

I. DNA SUBCLONING ............................................................... 63
1. Preparation and Isolation of DNA fragments for Subcloning ....... 63
2. Ligation and Transformation of DNA into Bacteria .................... 63

J. RADIOACTIVE LABELING OF DNA ......................................... 64
1. Klenow Labeling ................................................................. 64

K. DNA SEQUENCE ANALYSIS .................................................. 65
1. DNA Sequence Determination from Plasmid DNA Templates ...... 65
2. Preparation of Unidirectional Deletions with Exonuclease III ..... 66
3. DNA Sequence Determination from PCR Generated Templates ... 69
4. DNA Sequence Data Analysis ................................................ 70

L. PREPARATION OF AND SCREENING OF cDNA LIBRARIES ......... 71
1. cDNA Synthesis ............................................................... 71
2. Linker Addition and Size Fractionation ................................... 71
3. In Vitro Packaging and Library Titration ................................ 72
4. Plating l Phage Libraries ..................................................... 72
5. Screening of l Phage Filters and Plaque Purification ................. 73

III. RESULTS ............................................................................. 74
A. Isolation of the 5'end of the Chicken Prothrombin cDNA ......... 74
1. Northern Blot Analysis of Chicken Prothrombin mRNA .......... 74
2. Isolation and Sequence Determination of cDNA Clones for Chicken Prothrombin ......................................................... 74

B. The Purification of Hagfish Prothrombin from Plasma ............. 84
C. Isolation of Prothrombin cDNA Fragments by the PCR ............ 88
1. Selection of Oligodeoxyribonucleotide Primers for use in the PCR ........88
3. Sequence Determination of Prothrombin cDNA fragments ............94
D. Isolation of a Hagfish Prothrombin cDNA fragment by the PCR ..........95
   1. Selection of oligodeoxyribonucleotide Primers for use in the PCR ....95
   2. Sequence Determination of Hagfish Prothrombin cDNA Fragments ...96
E. Isolation of cDNA Clones for Hagfish Prothrombin ..........................96
   1. Analysis of the Hagfish Prothrombin mRNA .................................96
   2. Screening Hagfish cDNA libraries ............................................96
   3. Isolation of Longer Hagfish Prothrombin cDNAs ............................97
F. Identification of Prothrombin Homologous Sequences in Invertebrates ...97
   1. Selection of Primers for use in the PCR ..................................97
   2. Amplification of ssoDNA from Amphioxus, Tunicate, and Sea Star . 102
G. Amino Acid Sequence Alignments of Vertebrate Thrombin B Chains ....102
H. DNA Sequence Alignments of Vertebrate Thrombin B Chains .............105
IV. DISCUSSION .................................................................................113
   A. Analysis of Vertebrate Prothrombin B chains ...............................113
      1. Comparisons of the B Chain Amino Acid Sequence From Eleven
         Vertebrate Species ..................................................................113
   B. Analysis of Vertebrate Prothrombins .........................................124
      1. Characterization of cDNAs For Chicken and Hagfish
         Prothrombin ........................................................................124
      2. Comparison of Vertebrate Prothrombin Amino Acid Sequences ....126
   C. The Evolution of Prothrombin in Vertebrates ...............................136
      1. Structural Organization of Vertebrate Prothrombin ....................136
      2. Evolution of Vertebrate Thrombin B Chains ..............................137
   D. Identification of Additional Coagulation Factors in Vertebrates .......138
IV. appendix ......................................................................................141
V. REFERENCES ..................................................................................150
LIST OF TABLES

Table 1. An Inferred Order of Appearance for Some Blood-clotting and Fibrinolytic Proteins ................................................................. 45
Table 2. Oligodeoxyribonucleotide Primers used for DNA Sequence Determination and the Polymerase Chain Reaction .................................. 59
Table 3. Differences in Amino Acid Sequence Assignments for Chicken Prothrombin ................................................................. 83
Table 4. Primer Mismatches Compatible and Incompatible with Obtaining Sequence ................................................................. 92
Table 5. Length of Thrombin B-chain Coding and 3’ Untranslated (UTS) Sequences (nts) for the Nine Species ......................................................... 112
Table 6. Percent Amino Acid Sequence Identity Among Vertebrate Thrombin B-Chains ................................................................. 114
Table 7. Percent Amino Acid Identities of Prothrombin From Six Species ................................................................. 133
Table 8. Amino Acid Sequence Identity and Similarities for the Domains of Prothrombin ................................................................. 134
LIST OF FIGURES

Figure 1. Thromboresistant Properties of Endothelium .................................................. 2
Figure 2. The Blood Coagulation Cascade ............................................................................. 7
Figure 3. Schematic Diagram of Fibrinogen and Fibrin Polymer Formation ..................... 11
Figure 4. Phylogenetic Scheme for the Main Groups in the Animal Kingdom .................... 17
Figure 5. Schematic Organization of the Prothrombin Gene, Prothrombin, and Thrombin .......................................................... 22
Figure 6. Computer Generated Model of the Structure of the B Chain of Human Thrombin ............................................................... 29
Figure 7. The Position of the B-Loop and Trp 486 Loop in the Model of the Human Thrombin B Chain ........................................................................... 30
Figure 8. Amino Acid Sequence Homologies in Coagulation Factor Zymogens ............. 35
Figure 9. Exon-Intron Structures of the Genes Encoding the Blood Coagulation Serine Proteases ............................................................................. 38
Figure 10. Phylogenetic Trees Showing the Relationships Among Serine Proteases ......... 42
Figure 11. Scheme for the Amplification of sscDNA ............................................................ 60
Figure 12. Scheme for Unidirectional Deletions using Exonuclease III and S1 Nuclease ................................................................................. 67
Figure 13. Northern Blot Analysis of Total Cellular RNA from Chicken Liver ............... 76
Figure 14. Restriction Map and cDNA Cloning Strategy for Chicken Prothrombin ......... 78
Figure 15. The cDNA Sequence and Predicted Amino Acid Sequence of Chicken Prothrombin ............................................................................. 79
Figure 16. Flow Diagram for Hagfish Prothrombin Isolation ............................................ 85
Figure 17. SDS-PAGE Analysis of Hagfish Prothrombin .................................................. 86
Figure 18. The PCR strategy for the Amplification of cDNA Fragments of the B-chain of Prothrombin ............................................................................. 89
Figure 19. Analysis of the PCR Amplification of Vertebrate Prothrombin B Chains...... 90
Figure 20. Restriction Map and cDNA Cloning Strategy for Hagfish Prothrombin ......... 98
Figure 21. The cDNA Sequence and Predicted Amino Acid Sequence of Hagfish Prothrombin ............................................................................. 99
Figure 22. Amino Acid Sequence Alignment of Vertebrate Thrombin
LIST OF ABBREVIATIONS

A  absorbance
Amp  ampicillin
ADP  adenosine diphosphate
ATP  adenosine triphosphate
AT III  anti-thrombin III
bis  N,N'-methylenebisacrylamide
bp(s)  basepair(s)
BSA  bovine serum albumin
cDNA  complementary deoxyribonucleic acid
Ci  Curie
Co  original concentration
cpm  counts per minute
DEAE  diethylaminoethyl
DMSO  dimethyl sulfoxide
DNA  deoxyribonucleic acid
DNase I  deoxyribonuclease I
dNTPs  deoxyribonucleotide triphosphates
DTT  dithiothreitol
E. coli  Escherichia coli
EDTA  ethylenediaminetetraacetic acid
EGF  epidermal growth factor
EtBr  ethidium bromide
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exo III</td>
<td>exonuclease III</td>
</tr>
<tr>
<td>Gla</td>
<td>γ-carboxyl glutamic acid</td>
</tr>
<tr>
<td>Hfl</td>
<td>high frequency lysogene</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>kbp(s)</td>
<td>kilo base pair(s)</td>
</tr>
<tr>
<td>Klenow</td>
<td><em>E.coli</em> DNA polymerase fragment I</td>
</tr>
<tr>
<td>LB</td>
<td>Luria broth</td>
</tr>
<tr>
<td>MeHgOH</td>
<td>methylmercuric hydroxide</td>
</tr>
<tr>
<td>MMLV-RT</td>
<td>Moloney Murine Leukemia Virus reverse transcriptase</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-[N-morpholino]propanesulfonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>MYA</td>
<td>millions of years ago</td>
</tr>
<tr>
<td>nt (s)</td>
<td>nucleotide(s)</td>
</tr>
<tr>
<td>OAc</td>
<td>acetate</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PFU</td>
<td>plaque forming unit</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>RGD</td>
<td>Arg-Gly-Asp</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SSC</td>
<td>standard saline citrate</td>
</tr>
<tr>
<td>sscDNA</td>
<td>single-stranded complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>TAE</td>
<td>tris acetate EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>TM</td>
<td>thrombomodulin</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer ribonucleic acid</td>
</tr>
<tr>
<td>UTS</td>
<td>untranslated sequence</td>
</tr>
<tr>
<td>UV</td>
<td>ultra-violet</td>
</tr>
<tr>
<td>vWF</td>
<td>von Willebrand factor</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to thank my supervisor Ross MacGillivray, for providing the space and the opportunity for me to pursue this project. I would also like to thank members of the MacGillivray lab both past and present for their patience. Thanks also to Tammy Massey, for providing reagents and ordering skills beyond the call of duty. I thank Kelley Thomas for his much appreciated help during the early PCR experiments. Thanks to Roger Graham and Blair Main for the recreational distractions. Thanks also to Mike Murphy and Terry Lo for their help with the computer graphics presented in this thesis.

A final and special thanks to Marilyn, whose patience, support, and encouragement kept me going. Thanks for riding the roller coaster with me.
I. INTRODUCTION

A. AN OVERVIEW OF HEMOSTASIS

In vertebrates, a closed circulatory system is essential for a number of physiological processes including: the transport of nutrients, the removal of waste products, the immune response, and the regulation of the hormonal response. Blood normally circulates through endothelium-lined vessels without appreciable fluid loss. Hemostasis is the process by which blood volume and blood flow are maintained by arresting blood loss from damaged blood vessels. In mammals, four interrelated processes interact to stop blood loss and repair damage in response to injury. These four processes are: vasoconstriction; the formation of a platelet plug; blood coagulation and fibrinolysis. The end result of the first three processes is the formation of a stable blood clot which mechanically impedes the flow of blood from the injured vessel to reduce blood loss. Fibrinolysis is involved in clot dissolution, a process that occurs during the repair of the damaged tissue.

1. The Endothelium.

Endothelial cells line the luminal surface of blood vessels and modulate vascular perfusion, permeability, and maintain blood fluidity. Vascular endothelial cells are largely responsible for the prevention of abnormal blood clot formation in blood vessels (see Figure 1). When the endothelial lining of a blood vessel is damaged, platelets adhere to the uncoated subendothelial extracellular matrix. The initial contact of a platelet through a few portions of the disrupted membrane is followed by platelet spreading and then interaction of that platelet with other platelets to form a platelet aggregate or plug (see below).
Figure 1. Thromboresistant Properties of Endothelium.

Endothelial cells synthesize prostaglandin I2 (PGI2), thrombomodulin, heparin, and plasminogen activators all of which inhibit hemostasis. (from Colman et al., 1987).
2. Vasoconstriction.

When vessels are injured they contract; this immediate and intense vasoconstriction temporarily reduces blood loss from the site of injury. Vasoconstriction in response to direct stimulation is generally transient. However, additional vasoconstrictors such as platelet-derived thromboxane A2, serotonin and ADP secreted from the activated platelets may contribute to a more prolonged effect (Mustard et al., 1987), helping to reduce blood loss during the initial stages of platelet plug formation and coagulation (see below).

3. The Formation of the Platelet Plug.

As mentioned above, platelets do not adhere to normal vascular endothelial cells (see Figure 1) but in areas of endothelial disruption they readily adhere to integral glycoprotein components of the subendothelial connective tissue. It is this process which initiates the formation of the platelet plug. To adhere to the region of vascular injury, or to one another to form a closely packed aggregate, platelets require receptors for the multimeric plasma protein, von Willebrand factor (vWF) and fibrinogen (Weksler, 1987). While these receptors are normally present on the platelet membrane, they do not interact with adhesive macromolecules in the surrounding plasma unless the platelets are first activated. In addition to platelet binding to vWF and fibrinogen, platelets may also bind to the vascular basement membrane proteins, laminin and fibronectin, further facilitating platelet localization to the site of injury during the early stages of hemostasis. Three principal effects occur upon platelet activation: secretion of the contents of intracellular granules releasing ADP, fibrinogen, vWF, thromboxane, and factor V; exposure of latent receptors for vWF and fibrinogen; and changes in the lipid structure of the platelet surface.
membrane leading to the acceleration of plasma coagulation via the formation of the prothrombinase complex (Holmsen, 1987.; Weksler, 1987.; Colman et al., 1987).

While the exact in vivo activation mechanisms are not known, they most likely include ADP, thrombin, thromboxane A2, and or collagen (Colman et al., 1987). The first two stages of hemostatic plug formation are platelet contact and platelet spreading. After platelets make contact with the site of vascular injury and spread, they begin to interact with one another to form an aggregate. Immediately after their exposure to activating agents, the platelets change shape from flattened disks to spheres with multiple projecting pseudopods. This change in shape is a result of the polymerization of actin and myosin monomers to form actin and myosin microfilaments. The triggering event for platelets to aggregate in vitro is the ADP catalyzed exposure of binding sites for fibrinogen on the platelet membrane (Colman and Walsh, 1987). In the course of activation, platelets develop receptors for specific plasma clotting factors such as activated factor V. Activated factor V secreted by the platelet or absorbed from plasma may serve as a binding site for factor Xa. This surface-orientated complex provides a powerful catalytic environment for the conversion of prothrombin to thrombin. There are at least two ways of stimulating platelet secretion in vivo (Holmsen, 1987) including: thromboxane A2 which stimulates vasoconstriction and the release of calcium across intracellular membranes; and an increase in cytoplasmic calcium resulting from platelet stimulation by exogenous ADP. vWF is one of at least six components of the extracellular matrix that can potentially bind to platelets. The binding of vWF to platelets is mediated by R-G-D receptors (Weksler, 1987).

The activation of platelets and the processes involved in platelet aggregation are clearly complex. Platelet activation is regulated largely by several features of intact endothelial cells (see Figure 1). The thromboresistance of normal vascular endothelial cells
is contributed by several properties (Colman et al., 1987) which are summarized in Figure 1. Endothelial cells synthesize prostaglandin I₂ (PGI₂), a powerful inhibitor of platelet aggregation (Weksler et al., 1977). In addition to synthesizing PGI₂, endothelial cells synthesize thrombomodulin and heparin sulphate (Esmon and Owen, 1981). Thrombin loses its ability to activate coagulation factors after binding thrombomodulin, but has enhanced ability to bind protein C. Protein C in tum inactivates factors VIIIa and Va and may stimulate the release of plasminogen activator (Esmon et al., 1982). Thrombin also binds to receptors on the surface of vascular endothelial cells, where it is inactivated by anti-thrombin III. The inactivation of thrombin by anti-thrombin III is accelerated in the presence of the glycoaminoglycan, heparin sulphate (Colman et al., 1987). Thrombin may also stimulate endothelial cell synthesis of PGI₂ (MacIntyre et al., 1978). Exposure of structures in the vessel wall can lead to activation of coagulation, activation of complement, and platelet adhesion to the wall. These vesicles include dense bodies (which contain ADP and calcium), α granules (which contain fibrinogen, vWF, Factor V, and high molecular kininogen), and lysosomes. Platelet activation and its effects are modulated by a number of regulatory substances including cyclic AMP (cAMP). Platelets bind to collagen components of the subendothelial connective tissue. The activation of platelets is regulated in part by cAMP and is a calcium dependent process. A number of calcium dependent activities have been identified in platelets including: adhesion, platelet shape change, aggregation, secretion of intracellular vesicular contents, clot retraction, activation of cyclooxygenase, and inhibition of adenylate cyclase (Holmsen, 1987). cAMP stimulates a calcium/magnesium ATPase-dependent pump that leads to a reduction in cytoplasmic calcium, thus offsetting the effects of platelet agonists (Holmsen, 1987).
Simultaneous with platelet plug formation, exposure of the injured vessel wall and activated platelets to the plasma clotting factors initiates blood coagulation. This process involves a sequential enzymatic activation of a family of serine protease proenzymes by limited proteolysis.


Blood coagulation involves the sequential enzymatic activation of a number of serine protease proenzymes (zymogens) resulting in the formation of the cross-linked polymer of fibrin molecules (Jackson and Nemerson, 1980). This process can be subdivided into two steps: the formation of thrombin, and the deposition of the fibrin clot in the platelet mass. Two pathways (or models) for the initiation of blood coagulation have been described (MacFarlane, 1964; Davie and Ratnoff, 1964): the intrinsic and the extrinsic pathways (see Figure 2). Both pathways converge in the activation of factor X. Activated factor X (Xa) converts prothrombin to thrombin. The initiation of the intrinsic pathway (contact activation pathway) involves four proteins: factor XII, factor XI, prekallikrein, and HMWK (Griffen, 1981) see Figure 2. The protease responsible for the first proteolytic cleavage necessary for the initiation of the intrinsic pathway has not been identified (Jackson and Nemerson, 1980.). However, the pathway is most likely initiated by the reciprocal activation of factor XII and prekallikrein when blood is exposed to a negatively charged surface (Rosenberg, 1987). Subendothelial collagen exposed after vessel damage may provide the negatively charged surface required for this reaction (Colman et al., 1987). While the initial steps of the contact activation pathway may be of minor physiological significance the latter steps (including factors IX, X, and VII) are of great physiological significance as indicated by the severe effects of deficiency of these proteins in hemophilia.
Figure 2. The Blood Coagulation Cascade.

Outline of the mammalian blood coagulation cascade with the intrinsic pathway and extrinsic pathway converging at the activation of factor X, and ending with the formation of the insoluble fibrin clot.
Anionic surface  
Kininogen  
Prekallikrein

Factor XII

Factor XI

Factor IX

Factor X

Factor IXa

Factor Xa

Factor Va

Prothrombin

Factor VIIIa

Factor VIIa

Factor VII

Tissue factor

Factor X

Factor Xa

Factor V

Thrombin

Fibrinogen

Ca++

Fibrin
patients (Rosenberg, 1987). The extrinsic pathway is initiated by the binding of factor VII to tissue factor. Tissue factor, a membrane lipoprotein, is found primarily in perivascular tissue and is made accessible to blood when the vascular endothelium is disrupted (Colman et al., 1987). The factor VII/tissue factor complex possesses weak enzymatic activity and is able to activate factors IX and X. Activated factor IX and X subsequently activate factor VII to VIIa which has greatly enhanced enzymatic activity towards the activation of factor IX and X. Both the intrinsic and extrinsic pathways result in the formation of platelet bound factor Xa. Factor Xa in the presence of factor Va, calcium ions and phospholipid forms a complex on the activated platelet surface. This complex binds to prothrombin, generating thrombin by limited proteolysis. Thrombin releases fibrinopeptides A and B from fibrinogen by limited proteolysis promoting the polymerization of fibrin monomers. Thrombin also activates factor XIII, a transglutaminase, which cross-links polymerizing fibrin molecules. In addition, thrombin also activates platelets, factor V, factor VIII, platelets, and plays a central role in the regulation of the coagulation cascade (Jackson and Nemerson, 1980.; Rosenberg, 1987.; Colman et al., 1987).

Thrombin regulates the coagulation cascade through its interaction with thrombomodulin on the surface of endothelial cells. The interaction of thrombin with thrombomodulin alters the substrate specificity of thrombin such that it activates protein C. Activated protein C anchors to the cell surface by binding to membrane bound protein S. The protein C/protein S complex converts factor Va and factor VIIIa to their inactive forms by limited proteolysis (Owen, 1987). The concentration of thrombin is regulated by anti-thrombin III (ATIII), which forms a 1:1 stoichiometric complex with thrombin. ATIII together with protein C plays a crucial role in the regulation of the coagulation cascade.
through the modulation of thrombin activity (Jackson and Nemerson, 1980.; Colman et al.,
1987).

The existence of a cascade or waterfall pathway of blood coagulation allows for a rapid amplification of the response to injury essential for hemostasis. Each activated zymogen (proenzyme) is able to activate catalytically a large number of proenzymes in the proceeding step of the cascade (see Figure 2). The large number of protease inhibitors found in plasma (Harpel, 1987) and the multiple steps of the coagulation cascade provide a large number of opportunities to regulate the process (Jackson and Nemerson, 1980.; Colman et al., 1987.; Rosenberg, 1987). The rigorous regulation of the clotting cascade prevents coagulation beyond the site of injury and allows termination of coagulation once the platelet plug has formed.

5. Fibrin Formation.

The formation of fibrin strands represents a second phase in hemostasis (the first being the formation of the platelet plug). The blood clot is formed by the polymerization of fibrin monomers into a network which incorporates the platelet plug, thrombin and other proteins, and cells into a mechanical plug to prevent fluid loss. The precursor of fibrin is fibrinogen, a large glycoprotein of Mr 340,000 present in high concentrations in both plasma and platelet granules (Colman et al., 1987). Fibrinogen is comprised of six polypeptide chains: 2 Aα, 2Bβ, and 2 γ chains (Doolittle, 1984). Thrombin cleaves four peptide bonds in each monomer, one in each of the Aα and Bβ chains, releasing fibrinopeptides A and B, and the fibrin monomer (Doolittle, 1984). Progressive lengthening of the polymer chain occurs by half-overlap, side-to-side approximation of fibrin monomers, and the two-strand protofibrils interact laterally to form long, thin fibrin
**Figure 3. Schematic Diagram of Fibrinogen and Fibrin Polymer Formation.**

Schematic diagram of fibrinogen and thrombin-induced fibrin monomer and polymer formation with factor XIIIa induced fibrin cross-linking. The cross-hatched lines represent coiled-coils between central (small circles) and terminal domains (large circles). The small solid lines represent crosslink sites induced by factor XIIIa between γ chains of two contiguous terminal domains. The fibrinopeptides are indicated as small vertical lines connected to the central domain and are absent from fibrinogen monomers following thrombin action. (From Colman et al., 1987).
Fibrinogen → Thrombin → Fibrin Dimer → Thrombin → Factor XIIIa → Fibrin Trimer → Fibrin Polymer (Protofibril) → Two Protofibrils → Fibrin Mesh
strands or short broad sheets of fibrin (see Figure 3). The fibrin network is further
strengthened by the formation of covalent cross links between monomers by activated
factor XIII (Doolittle, 1984). Factor XIII is activated by limited proteolysis by thrombin
(see Figure 2). Factor XIIIa catalyzes the formation of covalent isopeptide bonds between
lysine and glutamine residues. In mature forms the fibrin fiber contains approximately 100
protofibrils, with a somewhat random pattern of branching that links the fibers together
(see figure 3). The fibrin mesh binds the platelets together and contributes to their
attachment to the vessel wall (Doolittle, 1984.; Colman and Walsh, 1987).

6. Fibrinolysis.

Several mechanisms for controlling and localizing hemostasis exist. These
mechanisms include: proteolytic feedback by thrombin, inhibition by plasma proteins
(ATIII), activation of inhibitory enzymes (protein C), and fibrinolysis (Colman et al.,
1987). The fibrinolytic system is the principal effector of clot removal and controls the
enzymatic degradation of fibrin (Colman et al., 1987). Fibrinolysis resembles the
coagulation cascade in that it involves zymogen to enzyme conversions, feedback
potentiation and inhibition (Colman et al., 1987). The principal enzymes capable of
digesting fibrin are plasmin and leukocyte-derived proteases (Francis and Marder, 1987).
The conversion of plasminogen to plasmin is accomplished by several plasminogen
activators (Colman et al., 1987.; Francis and Marder, 1987). In blood the conversion of
plasminogen to plasmin is accomplished by tissue-type plasminogen activator (tPA),
umokinase, and via the activation of the contact phase of coagulation (factor XII,
prekallikrein, high molecular weight kininogen, and factor XI); such a mechanism is called
the factor XII-dependent fibrinolytic pathway (Colman et al., 1987). The activation of the
fibrinolytic system is controlled by protease inhibitors. These inhibitors include inhibitors of tPA and urokinase, and α-antiplasmin (Francis and Marder, 1987).

B. THE EVOLUTION OF HEMOSTATIC MECHANISMS.

In mammals complex mechanisms have evolved to limit blood loss following vascular damage (see above). Analogous responses to injury have been observed in non-mammalian vertebrates as well as invertebrates, but are far less characterized than in mammals. These responses include the contraction of tissues, cellular aggregation, and the gelation of body fluids (Ratnoff, 1987). The evidence available to date suggests that these responses have evolved independently in vertebrates and invertebrates. With few exceptions, the clotting mechanism among different mammals closely resembles that of humans. The majority of the variation is restricted to the relative concentrations of the various clotting factors in the plasma of different mammals. For example, marsupials and non-human primates have relatively high levels of factor XII (Lewis and Doyle, 1964; Hampton and Mathews, 1966; Saito and Ratnoff, 1979). In addition, variations in the circulating levels of factors V, VIII, and prekallikrein have also been reported in different mammalian species (Ratnoff, 1987). Furthermore, bats appear to have little or no factor XI and relatively lower levels of a number of additional clotting factors including prothrombin, factor VII, and factor XII (Lewis, 1972). Finally, both toothed and baleen cetaceans (whales) are unique among mammals in their apparent functional deficiency in factor XII (Robinson et al., 1969; Ratnoff et al., 1976). However, the commonly used techniques in these studies are the mixing of plasmas or partially purified clotting factors from the various species being compared, and the use of standard clotting assays. This approach does not take species specific variability into account and likely results in false information regarding
the absence, relative activities, or relative abundance of the various components of the clotting cascade. Amino acid sequence variability, particularly on the surface of proteins, can cause changes in the ability of one species homologous protein to activate the others by altering protein protein interactions. For example, bovine thrombin readily cleaves fibrinogen in mammalian plasmas whereas fibrinogen from fish, amphibians, reptiles and birds is a poorer substrate for bovine thrombin such that these plasmas clot more slowly (Doolittle and Surgenor, 1962).

1. Coagulation in Non-mammalian Vertebrates

Unlike mammals, where the anucleated platelet plays a central role in hemostasis (see above), in non-mammalian vertebrates the circulating cells that participate in hemostasis are the nucleated thrombocytes (Ratnoff, 1987). Thrombocytes are spindle-shaped motile cells and like platelets, their cytoplasm generally contain granules. The origin of thrombocytes is apparently different from that of platelets, which arise from fragmentation of the cytoplasm of megakaryocytes (Rebuck, 1971). Like platelets, thrombocytes appear to adhere to sites of vascular injury as well as to collagen, and may be functionally analogous to platelets in some species. For example, thrombocytes (like platelets) accelerate the generation of thrombin as well as support clot retraction (Belamarich, 1976; Doolittle and Surgenor, 1962; Kimura, 1969).

Conversion of fibrinogen to fibrin by a thrombin-like enzyme is the basis of blood clot formation in all vertebrates (Doolittle, 1984). In many vertebrate classes, the existence of other coagulation factors has not been investigated in detail (Ratnoff, 1987). The chicken appears to have the best characterized coagulation system of the non-mammalian
vertebrates (Didisheim et al., 1959; Walz et al., 1975). In chicken, most of the mammalian clotting factors have been identified including a partially characterized prothrombin (Didisheim, 1959; Walz et al., 1974; Walz et al., 1975; Irwin, 1986). However, chickens appear to be deficient in factor XII (Ratnoff, 1987). Evidence for an extrinsic pathway of thrombin generation has been described in all vertebrate species, including the lamprey. Prothrombin has been partially purified from the lamprey (Doolittle et al., 1962; Doolittle, 1965) and activated lamprey prothrombin is able to coagulate bovine fibrinogen (Doolittle et al., 1962). However, studies using factor deficient plasmas from man have failed to detect factors VII and X in some amphibian species as wells as factors X and V in some species of bony fish (Ratnoff, 1987). Evidence of an intrinsic pathway of blood coagulation has been described in some species of bony fish (Doolittle and Surgenor, 1962). The evidence of an intrinsic pathway in other vertebrates is inconclusive and based largely on the use of factor deficient plasma from man (Ratnoff, 1987).

The reported absence of a number of the coagulation factors in non-mammalian vertebrates is not consistent with the available molecular data (Doolittle and Feng, 1987). This apparent contradiction is likely the result of species specific variability (as discussed above).

2. Coagulation in Invertebrates.
Blood coagulation is not limited to vertebrates. Hemostasis of some type has been observed in many phyla including coelenterates, echinoderms, molluscs, annelids, and arthropods (reviewed by Ratnoff, 1987) see Figure 4. In many of these invertebrate
Figure 4. Phylogenetic Scheme for the Main Groups in the Animal Kingdom.
species, hemostasis is the result of the aggregation of blood cells at the site of injury (analogous to the formation of the platelet aggregate in mammals).

The simplest of the metazoa, the Porifera (sponges) are composed of loose aggregates of cells. These animals consist of a series of cavities and chambers through which water circulates by the beating of flagellae. Porifera respond to injury by contraction of their gastrual and dermal layers, which seals the wound (Ratnoff, 1987).

The coelentrates (jelly fish, corals) are more structured than the porifera, and seal wounds by the aggregation of surrounding cells. Although the coelomic fluid does not clot, a mucus-like material helps to seal the wound (Gregoire and Tagnon, 1962; Ratnoff, 1987).

Animals with bilateral symmetry can be divided into two general groups: the Protostomes and the Deuterostomes (see Figure 4). This classification is based in part on the origin of the mouth (Meglitish, 1972). Among the Protostomia are the platyhelminths, molluscs, annelids, and arthropods. The more primitive Protostomia (platyhelminths) lack anything resembling a circulatory system. These animals respond to injury by muscular contraction and the secretion of mucus, although the wound seal in planaria appears to have a fibrin-like structure (Needleham, 1972).

Phyla such as the mollusca, annelida, and arthropoda (see Figure 4) have a vascular system consisting of a heart, blood vessels, and a circulating fluid (hemolymph) (Belamarich, 1976). In some species a phenomena resembling blood clotting has been observed, whereby the blood rapidly gels to form a gelatinous coagulum (Ratnoff, 1987). In annelids (earthworms), both cellular aggregation and blood clotting have been observed (Hellibrunn, 1961; Meglitisch, 1972; Ratnoff, 1987).
Arthropods (insects, crusteans) have been more extensively studied than the lower forms of protostomes (Gregoire and Tagnon, 1962; Ratnoff, 1987). Hemostasis in arthropods is accomplished by the break up of hemocytes on contact with foreign surfaces and the formation of a coagulum (Gregoire and Tagnon, 1962; Ratnoff, 1987). The best characterized invertebrate coagulation proteins are coagulogen from the horseshoe crab (Levin and Bang, 1964), and the "fibrinogen" molecule from the spiny lobster (Fuller and Doolittle, 1971a,b). Despite its name, the horseshoe crab (*Limulus polyphemus*) is not a true crab, but a member of a more primitive subphylum, the Chelicerata (Belamarich, 1976). The coagulogen is released from amoebocytes (the sole circulating cells) when blood is shed or exposed to endotoxin (Levin and Bang, 1968). Endotoxin appears to activate an amoebocyte-derived enzyme with the properties of a serine protease (Torano et al., 1984). The complete amino acid sequence of the precursor of coagulogen has been determined (Miyata et al., 1983; Cheng et al., 1986) and has no similarity to vertebrate fibrinogens. Despite the lack of structural similarity to vertebrate fibrinogens, the coagulogen of *Limulus*, is clotted by thrombin (Fumarola et al., 1975). Unlike fibrinogen, the coagulogen can also be clotted by trypsin (Solum, 1973). The hemolymph of the spiny lobster contains a protein that coagulates on exposure to clot-promoting agents found in hemocytes and in muscle extracts (Fuller and Doolittle, 1971a,b). In this animal, clot formation is caused by the polymerization of a plasma "fibrinogen" (unlike vertebrate fibrinogen; Fuller and Doolittle, 1971a) by a calcium dependent transglutaminase (Fuller and Doolittle, 1971b). A similar transglutaminase from the hemolymph of sand crabs (Madaras et al., 1981) is able to cross-link fibrin in an analogous fashion to factor XIIIa, and vice versa (Ghidalia et al., 1982).
Among the Deuterostomes are echinoderms, hemichordates, protochordates, and chordates (which include the vertebrates). The echinoderms such as the sea stars, sea urchins, and sea cucumbers, have a rudimentary vascular system. However, echinoderms do have an elaborate water-vascular system which is used for locomotion. Loss of body fluid from this system is controlled by constriction of the body wall, the formation of aggregates of phagocytic cells, or by the extrusion of proteins that polymerize into a meshwork of fibers (Boolootain and Giese, 1959; Ratnoff, 1987). In some species these processes may also require calcium (Ratnoff, 1987). A fibrinogen homologous to vertebrate fibrinogen has not been identified in invertebrate plasma. However, Xu and Doolittle have identified two cDNAs that are clearly related to vertebrate fibrinogen β and γ chains (Xu and Doolittle, 1990) in the hepatic caecae of the sea cucumber (*Parastichopus parvimensis*).

C. PROTHROMBIN

1. **Biosynthesis and Post-translation Modifications**

Prothrombin is the circulating zymogen of thrombin, the serine protease responsible for the limited proteolysis of fibrinogen to produce fibrin (Davie et al., 1979, Jackson and Nemerson, 1980). The amino acid sequences for bovine and human prothrombin have been determined (Magnusson *et al.*, 1975; Butkowski *et al.*, 1977; Thompson *et al.*, 1977; Walz *et al.*, 1977; Seeger, 1979). Both bovine and human plasma prothrombin are glycoproteins of Mr~70,000 (Davie *et al.*, 1979). The major site of prothrombin synthesis is the liver (Anderson and Bornhart, 1964) where it represents approximately 1% and 0.1% of the translation products of bovine (MacGillivray *et al.*, 1979) and human (Besmond et
Prothrombin biosynthesis requires vitamin K, a cofactor for a microsomal carboxylase that is responsible for the post translational modification of glutamic acid residues to γ-carboxyglutamic acid (Gla) (Suttie, 1985; Furie and Furie, 1990; Wu et al., 1991). Circulating human prothrombin contains 8% carbohydrate by weight and 579 amino acids including 10 Gla residues at its amino-terminal end (Walz et al., 1977; Butkowski et al., 1977). Circulating bovine prothrombin contains 10% carbohydrate by weight, and is comprised of 582 amino acids including 10 Gla residues. Besides the conserved amino-terminal Gla domain, both human and bovine prothrombin contain two kringles and a serine protease domain (Magnussen et al., 1975; Walz et al., 1977; Butkowski et al., 1977).

2. Structure of the cDNA and Gene for Prothrombin

The sequence of bovine (MacGillivray and Davie, 1984), human (Degen et al., 1983), mouse (Degen et al., 1990) and rat (Dihanich and Monard, 1990) cDNAs have been determined. The precursor forms of prothrombin in all four species have an identical domain organization. In each of the four species prothrombin is synthesized as a prepro-protein with an amino-terminal leader sequence that contains the signal peptide required for secretion, and a peptide that appears to direct the vitamin K-dependent carboxylation of prothrombin (see below). The mRNAs for each of the four species are approximately the same length (2,000 nts) and include short segments of 5' untranslated sequence (UTS; 20-30 nts) and 3'UTS (80-150 nts).

The structure of the human (Degen and Davie, 1987) and bovine (Irwin et al., 1988) prothrombin genes have been determined. The human prothrombin gene is
Figure 5. Schematic Organization of the Prothrombin Gene, Prothrombin, and Thrombin.

Panel A: Schematic representation of the structure of the human prothrombin gene. The shaded boxes represent exons and the exon numbers are indicated above.

Panel B: Schematic representation of prothrombin. The arrows represent the factor Xa cleavage sites. Panel C: Schematic representation of thrombin.
approximately 21 kbp in length and comprises 14 exons interrupted by 13 introns (Degen and Davie, 1987). The bovine prothrombin gene (Irwin et al., 1988) is approximately 15 kbp in length and comprises 14 exons interrupted by 13 introns. The human and bovine genes appear to be organized in a similar manner with regard to the number and placement of intervening sequences. The introns separating exons IV and V, VI and VII, and XII and XIII, in the human gene are significantly larger than their bovine counterparts (Degen and Davie, 198; Irwin et al. 1988). A schematic representation of the prothrombin along with the structure of the prothrombin gene is shown in Figure 5.

3. The Propeptide.

The leader peptide of human, bovine, rat and mouse prothrombin is cleaved at an Arg-Ala bond prior to secretion from the liver (Magnussen et al., 1975; Walz et al., 1977; Degen et al., 1983; MacGillivray and Davie., 1984; Degen et al., 1990; Dihanich and Monard, 1990). The site cleaved to produce plasma prothrombin is more similar to a pro-peptide cleavage sequence such as prepro-albumin (Steiner et al., 1980) than a signal peptidase cleavage sequence. On the basis of the cDNA sequences of the bovine and human prothrombin cDNAs, it was suggested that prothrombin is synthesized as a prepro-protein and contains both a pre (signal) and a pro-peptide in the leader sequence. Similar prepro-leader peptides have been found in other vitamin K-dependent coagulation factors (see Section E.1). The observation of incomplete carboxylation associated with a propeptide point mutation, the sequence homology of this domain in Gla containing proteins and the absence of Gla residues in a factor IX with its propeptide deleted (Diugurd et al., 1986; Pan and Price, 1985; Jorgensen et al., 1987) led to the proposal of a role for
the propeptide is designating proteins for γ-carboxylation. In addition to directing proteins for γ-carboxylation, the propeptide also functions as a recognition element for the carboxylase (Jorgensen et al., 1987; Foster et al., 1987; Price et al., 1987). Recent support for the dual function of the propeptide comes from studies using a synthetic 59 amino acid peptide representing residues -18 to 41 of factor IX (Wu et al., 1990; Wu et al., 1991). The 59 amino acid peptide includes the entire propeptide and Gla containing region of factor IX (Wu et al., 1990) which was found to be an excellent substrate for in vitro γ-carboxylation. This peptide was later used to purify the vitamin K-dependent carboxylase from bovine microsomes (Wu et al., 1991).

Although the propeptidase has not been identified, the specificity for Arg suggests it may be a serine protease. The propeptidase may be related to a family of subtilisin serine proteases of which Kex2 (Mizuno et al., 1988; Fuller et al., 1989a; Fuller et al., 1989b), furin (Bresnahan et al., 1990; van de Ven et al., 1990; van den Ouweland et al., 1990), and PC (Smeekens and Steiner, 1990) propeptidases are members.

4. The γ-carboxyglutamic acid Containing Region

The γ-carboxyglutamic containing region or Gla domain is located at the amino terminal end of prothrombin. The Gla domain contains 10 γ-carboxylated glutamic acid residues (Magnussen et al., 1975). Two or three of these Gla residues bind a single calcium ion and form a noncovalent intramolecular bridge between regions of the polypeptide backbone (Furie et al., 1979). The formation of this intramolecular bridge stabilizes the tertiary structure of this region of the protein (Tai et al., 1984). Upon binding calcium, prothrombin undergoes conformational changes that expose a membrane-binding site (Nelsestuen, 1976; Prendergast and Mann, 1977; Borowski et al., 1986). Whether
some of the Gla residues directly bind to membrane surfaces through calcium bridges, (Mann et al., 1982) or calcium binding to the Gla residues alters the structure of adjacent domains exposing a membrane-binding surface (Borowski et al., 1986; Soriano-Garcia et al., 1989) remains unclear. In addition, the calcium-Gla residue interaction is essential for the efficient activation of prothrombin (Jackson, 1981). Further support for the role of the Gla domain in the activation of coagulation factors comes from studies on factor IX deficiencies. To date, several factor IX deficiencies have been identified which result from mutations at specific γ-carboxylated Glu residues (Hamguchi et al., 1991; Chen et al., 1987; Wang et al., 1990). The majority of these mutations result in severe bleeding tendencies presumably resulting from structural changes in the Gla domain and subsequent decreased activation potential.

5. The Kringle Domain

Following the Gla region are the structures known as kringles (Magnussen et al., 1975). Kringles are composed of about eighty amino acid residues containing six invariant cysteine residues which form three internal disulphide bridges (Magnussen et al., 1975): see Figure 5. The folding of the kringle is characterized by the close contacts between the sulphur atoms of two of the disulphide bridges which form a cluster in the center of the kringle (Park and Tulinsky, 1986; Soriano-Garcia et al., 1989). While the function(s) of the kringles remain uncertain, the second kringle of prothrombin has been reported to bind factor Va (Esmon and Jackson, 1974), the essential protein cofactor in prothrombin activation (see above).
6. The Aromatic Amino Acid Stack.

Between the Gla region and the first kringle is a short segment known as the aromatic amino acid stack (Furie and Furie, 1988). This region of prothrombin contains the sequence Phe-Trp-X-X-Tyr. The aromatic amino acid residues have their side chains stacked into a ring cluster which stabilizes the protein structure (Park and Tulinsky, 1986; Soriano-Garcia et al., 1989). Although hydrophobic, this region is orientated toward the surface where it may play a role in macromolecular assembly (Park and Tulinsky, 1986; Soriano-Garcia et al., 1989; Furie and Furie, 1988).

7. The Protease Region

The carboxy-terminal half of prothrombin contains the serine protease catalytic domain, thrombin (Magnussen et al., 1975). Thrombin is generated from its zymogen, prothrombin, by the limited proteolytic action of factor Xa in the presence of factor Va, calcium ions, and phospholipid (Rosenberg, 1987). Factor Xa cleaves the polypeptide chain in two places releasing the amino terminal activation peptide (Gla and kringles domains) from the thrombin domain (see Figure 5). In addition to its proteolytic action on fibrinogen, thrombin also activates/inactivates other coagulation factors such as factor XIII, factors V and VIII, and protein C. The anticoagulant activity of thrombin is regulated through the interaction of thrombin with thrombomodulin (Rosenberg, 1987; Jackman et al., 1986). Thrombomodulin, an endothelial cell membrane protein related in structure to the low density lipoprotein receptor (Jackman et al., 1987), binds to thrombin via an exposed surface loop (Susuki et al., 1990). The thrombin/thrombomodulin complex
activates protein C which in the presence of protein S then degrades factors Va and VIIIa (Rosenberg, 1987).

The enzymatic activity of thrombin is regulated by the endogenous protein inhibitor antithrombin III (Rosenberg, 1987). Thrombin is also inhibited by protein inhibitors from non-plasma sources such as hirudin from the European medicinal leech *Hirudo medicinalis* (Markwardt 1970). The substrate specificity of thrombin is similar to that of trypsin, cleaving after basic amino acid residues (Fenton, 1981). However, thrombin has a more restricted substrate range than trypsin. The molecular basis of this restricted substrate specificity is still unresolved but may be the result of interactions of the substrates with secondary binding sites distant from the active site (Fenton and Bing, 1986; Fenton 1988).

Thrombin consists of two polypeptide chains joined by a disulfide bond. The A chain (49 amino acids in bovine and 36 in human) has no known function. The B chain (259 amino acids in both human and bovine) is structurally very similar to other serine proteases (Fenton and Bing, 1986; Fenton 1988). When the amino acid sequence of the thrombin B chain is aligned with that of bovine chymotrypsin, amino acid sequence insertions in the B chain are found at exon junctions (Furie et al., 1982). The crystal structure of human oxthrombin has placed these amino acid insertions on the surface of the protein (Bode et al., 1990).

*a. Non-enzymatic Functions of Thrombin.*

In addition to the role in hemostasis and thrombosis, thrombin is a potent stimulator of tissue plasminogen activator release from endothelial cells (Levin et al., 1986). At basal concentrations, various forms of thrombin are chemotactic for monocytes and neutrophils (Bar-Shavit et al., 19; Bizios et al., 1986; Morin et al., 1990). This activity is blocked by
binding with hirudin or antithrombin III (Bar-Shavit et al., 1985). For monocytes, the chemotactic activity has been localized in the thrombin domain to residues 334-399 of prothrombin (Bar-Shavit et al., 1984). In addition to chemotactic activity, this polypeptide sequence has been found to induce differentiation of certain macrophage lines (Bar-Shavit et al., 1986). Thrombin also contains an RGD tripeptide analogous to the adhesion site in adhesive proteins such as laminin, fibronectin, and fibrinogen and is a possible site through which thrombin binds to receptors (Bar-Shavit et al., 1991). The fact that thrombin possesses these additional bioregulatory and growth stimulating activities suggests it may also play an important role in the wound healing process as well as fibrinolysis.

b. Three Dimensional Structure of Thrombin.
Crystal structures have now been described for human α-thrombin (Bode et al., 1990) as well as the thrombin/hirudin complex (Grutter et al., 1990; Rydel et al., 1990). Thrombin is an almost spherical molecule (dimensions: $45 \times 45 \times 50 \, \text{Å}^3$), and the A- and B- chain are not organized in separate domains (Bode et al., 1990) see Figure 6. Compared with other serine proteinases whose three dimensional structures are known, thrombin has a deeper and more prominent groove making up the active-site cleft (Bode et al., 1990, Yue et al., 1991) see Figure 7. Three of the four disulphide bridges are topologically and conformationally identical to those found in chymotrypsin (Bode et al., 1990). The B-chain is organized in two interacting 6-stranded barrel-like domains covered by turn structures and four helical regions (Bode et al., 1990). The B-chain displays a similar polypeptide fold to the other known serine protease structures with about half of the approximately 200 spatially equivalent residues are identical or similar in character (Bode et al., 1990). Relative to the structure of chymotrypsin, α-thrombin exhibits a number of
Figure 6. Computer Generated Model of the Structure of the B Chain of Human Thrombin.

Spaced filled representation of a model of the B chain of human thrombin. The active-site groove runs from left to right through the center of the molecule. The active-site residues are numbered as in human prothrombin (Degen et al., 1983) His 363 (blue), Asp 419 (red) is mostly buried, and Ser 525 (orange). The position of the active-site residues are indicated with arrows. The coordinates for the human thrombin model were obtained by digitizing the stereo image produced from the crystal structure (Bode et al., 1990) but is a fair representation of the structure (Yue et al., 1991).
Figure 7. The Position of the B-Loop and Trp 486 Loop in the Model of the Human Thrombin B Chain.

Panel A: View into the active-site cleft of human thrombin. The loops projecting into the active-site groove are Tyr 367 - Phe 374 (top center, in red) and Lys 465 - Lys 474 (bottom right, in red).
Figure 7. cont'd.

Panel B: Same as panel A except the model has been rotated $90^\circ$ to emphasize the projection of Trp 373 into the active-site cleft. The location of the active-site groove is indicated by the arrow.
deviating segments, all of which are located at the surface (Bode et al., 1990). The active-site residues and the substrate binding site are integrated into a deep narrow groove. The rims of the groove are made up of loop segments project into the active-site cleft (see Figure 7). The rims of the active-site cleft are primarily lined by hydrophobic groups, whereas the base is mainly lined with polar/charged side chains (Bode et al., 1990). In agreement with previous suggestions (Fenton and Bing, 1986; Fenton, 1988), the substrate specificity of thrombin is mainly determined by its insertion loops (Bode et al., 1990). The most striking feature of the B-chain structure is its unique loop composed of Tyr 367 - Phe 374 (B loop). The B loop together with the loop segment around Trp 468, shapes, narrows and deepens the active-site cleft (Bode et al., 1990). The most exposed part of the B loop, Tyr 370 - Pro 371- Pro 372- Trp 373, is compact and rigid (see Figure 7). Sterical hindrance by this segment is one of the presumed reasons for the narrow substrate specificity of thrombin (Bode et al., 1990).

c. Thrombin Receptor

A thrombin receptor has recently been identified (Vu et al., 1991). The receptor is a member of the 7 transmembrane domain receptor family, the mRNA for which has been found in human platelets and vascular endothelial cells. The thrombin receptor appears to be activated by a novel mechanism. The thrombin receptor contains a putative thrombin cleavage site at its amino-terminal end. On binding to the receptor, thrombin presumably cleaves the amino-terminal extension to create a new receptor amino-terminus that functions as a ligand and activates the receptor (Vu et al., 1991).
D. FAMILIES OF SERINE PROTEASE ZYMOGENS.

Serine proteases play a role in a diverse number of physiological processes including blood coagulation, fibrinolysis, digestion, complement activation, neuropeptide processing, and post-translational processing (Neurath and Walsh, 1976; Brenner, 1988; Mizuno et al., 1988; Fuller et al., 1989a; Fuller et al., 1989b; Bresnahan et al., 1990; van de Ven et al., 1990; van den Ouweland et al., 1990; Smeekens and Steiner, 1990). Two families of serine proteases have been identified which share a similar mechanism; the subtilisin-like proteases and the trypsin-like proteases (Neurath, 1984). The subtilisin-like proteases, while sharing the same catalytic mechanism (including the catalytic triad), do not share amino acid or three dimensional structure homology with the trypsin-like proteases. The trypsin-like proteases appear to be a larger family of proteins and have been identified in a wider distribution of animals than the subtilisin-like proteases. Both types of proteases have been found in eukaryotes as well as prokaryotes (Brenner, 1988; Mizuno et al., 1988; Fuller et al., 1989a; Fuller et al., 1989b; Bresnahan et al., 1990; van de Ven et al., 1990; van den Ouweland et al., 1990; Smeekens and Steiner, 1990). All of the trypsin-like serine proteases share amino acid sequence homology. However, the blood coagulation factors and many other serine proteases differ from trypsinogen in possessing long amino-terminal non-catalytic segments. In general, the function of these non-catalytic segments is to mediate binding of the protease or their zymogens to other macromolecules and through these interactions regulate the cascades of fibrinolysis and coagulation (Patthy, 1985).
E. HOMOLOGOUS DOMAINS FOUND WITHIN THE COAGULATION AND FIBRINOLYTIC PROTEINS

1. Homologous Domains Within the Activation Peptide of Prothrombin.

When the amino-terminal extensions of many of serine proteases involved in coagulation and fibrinolysis are aligned, several homologous domains are observed (Hewitt-Emmett et al., 1981; Furie and Furie, 1988). Prothrombin contains two kringle structures. Kringles have also been identified factor XII, plasminogen, tissue plasminogen activator (tPA), and urokinase (Furie and Furie, 1988). Kringles have also been found in apolipoprotein a (McLean et al., 1987). Prothrombin also contains a Gla domain, found in other vitamin K-dependent coagulation factors including factor VII, factor IX, factor X, and protein C, all of which also contain a prepro leader peptide (Furie and Furie 1988). Figure 8 illustrates some of the homologies within the coagulation and fibrinolytic serine protease zymogens (Hewett-Emmett et al., 1981). The prepro leader and Gla domain is also found in the bone matrix protein osteocalcin (Price et al, 1987). Gla containing peptides/proteins have also been identified in the venom of the predatory cone snail Conus (Olivera et al., 1990), as well as in hermatypic corals (Hamilton et al., 1982).

2. Homologous Domains Found in Coagulation and Fibrinolytic Proteins Other Than Prothrombin.

Additional domains are found in coagulation and fibrinolytic proteins other than prothrombin. A region with epidermal growth factor (EGF) homology has been identified
Figure 8. Amino Acid Sequence Homologies in Coagulation Factor Zymogens.

Comparison of the structures of coagulation and fibrinolytic zymogens to trypsin. The solid bar represents the protease domain, the cross-hatched region represents the activation peptide, and the dark shaded region represents the Gla domain. K represents the kringles, E represents regions homologous to epidermal growth factor precursor, A represents the apple domain, and 1 and 2 represent regions homologous to the type I and type II homologies of fibronectin. Arrows represent the locations of peptide bonds that are cleaved during activation of the zymogens. Solid lines below the proteins represent disulphide bridges but do not necessarily represent their true locations. The lengths of the bars are approximately proportional to the lengths of the polypeptide chains. (from Irwin, 1986; and Hewitt-Emmett, 1981).
in a number of proteins including factor VII, factor IX, factor X, factor XII, protein C, and tissue plasminogen activator (see Figure 8). These domains are found not only in fibrinolytic and coagulation proteins but also in non-serine proteases such as the low density lipoprotein receptor (LDL; Sudhoff et al., 1983) as well as the Notch-1 protein of Drosophila (Wharton et al., 1985) and the homeotic protein lin-12 of C. elegans (Greenwald, 1985). In addition, regions homologous to fibronectin type I and type II are found in factor XII. The type II fibronectin homology is also found in tPA (Furie and Furie, 1988): see Figure 8. The amino terminal extensions of factor XI and prekallikrein share four homologous apple domains (see Figure 8; McMullen et al., 1991;). Among the non-serine protease coagulation proteins factors V and VIII contain share homologous domains (Kane and Davie, 1988): in addition both factors V and VIII share amino acid sequence identity with the copper binding protein ceruloplasmin (Koschinsky et al., 1986). Amino acid sequence identities to factors V, VIII and EGF have also been identified in a mouse mammary epithelial cell surface protein (Stubbs et al., 1990).

The blood coagulation and fibrinolytic proteins are a prime example of a family of proteins with diverse functional properties but common structural elements (Hewitt-Emmett et al., 1981; Furie and Furie, 1988).

F. THE EVOLUTION OF SERINE PROTEASES INVOLVED IN COAGULATION AND FIBRINOLYSIS.

The plasma proteins involved in hemostasis can be separated into three functional groups with interrelated but distinct physiologic roles. One group consists of the blood coagulation proteins, which promote clot formation; the second consists of the proteins that modulate and localize clot formation to regions of tissue injury; the third group comprises
Figure 9. Exon-Intron Structures of the Genes Encoding the Blood Coagulation Serine Proteases.

Exons are shown schematically (see key) and are approximately drawn to scale. Introns are indicated by lines connecting the exons and are not drawn to scale. (from Furie and Furie, 1988).
the fibrinolytic proteins which dissolve fibrin clots as a late component of the healing process.

The structure and organization of the genes coding for the blood coagulation proteins emphasize that the evolution of new function occurs via gene duplication, gene modification, and exon shuffling (Patthy, 1985; Doolittle and Feng, 1987). Figure 9 shows the exon-intron structures of a number of genes encoding blood coagulation serine proteases.


The primary mechanism involved in the evolution of coagulation and fibrinolytic proteins is the shuffling of exons or modules (Patthy, 1985; Doolittle and Feng, 1987; Doolittle, 1987). There are two types of exon shuffling: exon duplication and exon insertion. Exon duplication refers to the duplication of one or more exons of a gene and exon insertion is the process by which structural or functional domains are inserted into or exchanged between genes (Li and Graur, 1991). The shuffling of exons (or modules) between proteins showing sequence homology may be facilitated by mispairing and double-cross over, or by gene conversion (Patthy, 1985). The insertion of an exon from one gene into another may ultimately result in the formation of mosaic or chimeric proteins. Exon shuffling is not without its limitations. For an exon to be inserted into an intron of a gene without causing a frameshift, the reading frame of the gene must be respected. Irrespective of the mechanisms responsible for the shuffling of exons (modules), the introns separating the pre-existing modules likely increase the probability of recombination outside the coding regions and thus aid in the reassortment of the modules (Patthy, 1985). Viruses may also serve as a vehicle for the transfer of exons between genes (Patthy, 1985). Both exon duplication and exon insertion have been used in the evolution of the proteins...
involved in coagulation and fibrinolysis (Patthy, 1985; Doolittle and Feng, 1987; Furie and Furie, 1988).

2. The Unique Features of the Serine Codon.

Several amino acids have multiple codon representation, but serine is unique in that its two sets of codons TCN and AGY (N = T,C,A,G; Y = T,C) cannot be converted into one another by a single nucleotide substitution. Thus, the only way to change from one serine codon set to another is via a non-serine intermediate. The advent of molecular biology techniques have made available a large number of cDNA and gene sequences for serine proteases and other enzymes which use serine as part of their catalytic mechanism (Brenner, 1988). Proteins which use serine in this way include the serine proteases (trypsin-, and subtilisin-like), alkaline phosphatases, esterases, lipases, and β-lactamases (Brenner, 1988). Within the family of serine proteases, both the AGY and TCN codons are represented. This is true of a number of other enzymes including the β-lactamases, alkaline phosphatases, esterases and lipases. Only the subtilisin-like serine protease family appears to use a single serine codon set (TCN, Brenner, 1988). It has been suggested that the presence of both the TCN and AGY codon sets in serine proteases proves that there have been at least two different lines of decent for the active site sequences of the serine proteases (Brenner, 1988). Furthermore, the simplest pathway for this convergent evolution was by the divergence of each line from a precursor that was itself catalytically active and had much the same sequence (Brenner, 1988). The codons that connect the two serines are ACN for threonine and TGY for cysteine. According to Brenner, the precursor of all serine proteases may have been cysteine proteases (Brenner, 1988).
An alternative explanation has been suggested by Irwin (1988). Vertebrate serine proteases have been grouped into five classes based on the gene structure, amino acid sequence, and type of active-site serine codon (Irwin et al., 1988). The AGY proteases fall into two clusters within this vertebrate serine protease phylogeny, with representatives in two classes (Irwin, 1988). The larger class includes the vitamin K-dependent coagulation factors, and the second class includes plasminogen and apolipoprotein a (Irwin, 1988).

Molecular phylogenies of the vertebrate serine proteases have been produced (Doolittle and Feng, 1987; Irwin, 1988) none of which support Brenner's suggestion that the TCN and AGY proteases represent separate lines of decent (Irwin, 1988). Rather, these phylogenies suggest that the TCN serine codon evolved on two separate occasions to an AGY codon. This occurred once along the lineage leading to the vitamin K-dependent coagulation factors, and once along the lineage leading to plasminogen and apolipoprotein a presumably via a non-serine intermediate (see Figure 10, Panel A; Irwin, 1988). In one case, a descendant of an intermediate non-serine protease (haptoglobin) has been identified (Kurosky et al., 1980).

3. Relationships Among the Coagulation and Fibrinolytic Serine Proteases.

A number of molecular phylogenies of vertebrate serine proteases have been reported (Hewitt-Emmett, 1981; Patthy, 1985; Doolittle and Feng, 1987; Irwin, 1988). These molecular phylogenies were constructed using progressive amino acid sequence alignments, gene structures, and the class of serine codon used (see Figure 10). Molecular phylogenies such as those presented in Figure 10 are useful for a number of reasons. Molecular phylogenies can be used to infer relationships among proteins (Figure 10 panels A, B, and C) as well as provide information regarding the mechanism(s) by which some
**Figure 10.** Phylogenetic Trees Showing the Relationships Among Serine Proteases.

**Panel A:** Phylogenetic tree relating selected vertebrate serine proteases and related proteins by codon usage and exon-intron structure. Gene classes indicate genes which share exon-intron structure, uncharacterized genes are indicated by question marks, X denotes the loss of a serine codon and an open circle the gain of a new serine codon (from Irwin, 1988). **Panel B:** Phylogenetic tree reconciling phylogenetic trees of individual domains. P represents the protease domain, C represents the Gla domain, K represents the kringle domain, F represents a finger domain, Ga represents an A-type growth factor domain, Gb represents a B-type growth factor domain. PT represents prothrombin, PC represents protein C, IX and X represent factors IX and X respectively, u-PA represents urokinase, t-PA represents plasminogen activator, and PL represents plasminogen. The solid circles represent an internal gene duplication. The vertical axis has a time dimension, and the arrows indicate the relative time of acquisition of the various modules. The large arrow stretching from plasminogen to prothrombin represents lateral gene transfer of kringle modules (from Patthy, 1985). AGY represents the type of serine codon found at the active-site serine. **Panel C:** Phylogenetic tree showing the relationship of sequences from 14 serine proteases. The tree is based on progressive alignments of amino acid sequence data (from Doolittle and Feng, 1987). AGY represents the type of serine codon found at the active-site serine.
proteins may have evolved from others (Figure 10 panels A, and B). The major limitation of this approach is the size and scope of the data set. Panels A and C of Figure 10 are based on amino acid sequence alignments of the protease domain (Doolittle and Feng, 1987; Irwin, 1988). Panel B is a reconciling phylogeny based on phylogenies of individual domains including the protease domain, kringle, Gla, and EGF domains (Patthy, 1985). The results of all three molecular phylogenies taken together suggest that (1) the addition of the Gla/propeptide domains to the vitamin K-dependent clotting protein preceded the addition of the krings to prothrombin (Patthy, 1987; Doolittle and Feng, 1987), (2) the krings of prothrombin are related to the krings of plasminogen, and (3) the prothrombin krings were derived from plasminogen on two occasions; the first transfer likely occurred prior to the multiplication of plasminogen krings and the second occurred at or about the time of the plasminogen kringle multiplication and divergence (Patthy, 1985). Furthermore, by extrapolation from amino acid sequence comparisons it is possible to infer the order of appearance for some of the proteins involved in hemostasis (Doolittle and Feng, 1987) see Table 1.

4. Objectives of the Present Study

The elucidation of the sequences of many of the proteins involved in hemostasis in mammals has allowed inferences to be made about the time, order and mechanisms by which these proteins were introduced during evolution. Molecular phylogenies (like those presented in Figure 10) predict specific intermediate states in the evolution of the proteases of blood coagulation and fibrinolysis. The validity of these predictions can be examined by determining the structures of components of the blood coagulation and fibrinolytic pathways in lower vertebrates and invertebrates. In addition to assigning a time scale to the
Table 1. An Inferred Order of Appearance for Some Blood-clotting and Fibrinolytic Proteins\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Factor</th>
<th>MYA\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibrinogen</td>
<td>600</td>
</tr>
<tr>
<td>Prothrombin</td>
<td></td>
</tr>
<tr>
<td>Tissue Factor</td>
<td></td>
</tr>
<tr>
<td>Plasminogen</td>
<td></td>
</tr>
<tr>
<td>Factor XII</td>
<td>500</td>
</tr>
<tr>
<td>Factor X</td>
<td></td>
</tr>
<tr>
<td>Factor VII</td>
<td></td>
</tr>
<tr>
<td>Factor V</td>
<td></td>
</tr>
<tr>
<td>Urokinase</td>
<td>450</td>
</tr>
<tr>
<td>Factor IX</td>
<td></td>
</tr>
<tr>
<td>Factor VIII</td>
<td></td>
</tr>
<tr>
<td>Factor XI</td>
<td></td>
</tr>
<tr>
<td>Prekallikrein</td>
<td>200</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Adapted from Doolittle and Feng., 1987.

\textsuperscript{b} MYA indicates millions of years ago.

predicted events, comparisons of amino acid sequences of blood clotting and fibrinolytic proteins from a number of different species may help to identify regions of structural / functional importance. To examine the evolution of coagulation in vertebrates, cDNAs for chicken and hagfish prothrombin have been characterized. In addition, portions of prothrombin corresponding to the B chain of thrombin have been characterized from a number of different vertebrate species. The results of these experiments and their significance to the evolution of prothrombin and coagulation in vertebrates are discussed in the following sections.
II. MATERIALS AND METHODS

A. MATERIALS

Yeast extract, casamino acids, bacto-tryptone, and bacto-agar were purchased from Difco Laboratories. Deoxy-, dideoxyribonucleotides, random hexadeoxyribonucleotides (p(dN₆)), oligo(dT) (type 7) cellulose, Sephacryl S300, DEAE- Sephadex CL6B and Sepharose CL-4B were purchased from Pharmacia. Acrylamide, bisacrylamide, urea, ammonium persulphate, and TEMED (N',N',N',N'- tetramethylethylenediamine) were purchased from Bio-Rad Laboratories. Guanidinium isothiocyanate, bovine serum albumin (BSA, pentax fraction 5) and phenol was purchased from BRL. Phenol was equilibrated with either buffer or water prior to use. Nitrocellulose discs (82 and 132 mm) and Nytran membranes (0.45 um) and were purchased from Schleicher and Schuell. α-[³²P]-dATP and α-[thio-³⁵S]-dATP were purchased from New England Nuclear.

Oligodeoxyribonucleotides were synthesized on either an Applied Biosystems 380A or 391A DNA Synthesizer, and purified by reverse phase chromatography on Sep-Pak C18 cartridges as described by Atkinson and Smith (1984). Bacteriophage arms (λgt11 and λgt10), λ phage packaging extracts (Gigapack Gold) as well as the E.coli host strains used to propagate λ phage were purchased from Stratagene. Reagents for the majority of cDNA syntheses were purchased from Invitrogen. Ribonuclease A (RNase A), deoxyribonuclease I (DNase I), dimethysulfoxide (DMSO), 3-(morpholino)propanesulphonic acid (MOPS), dithiothreitol (DTT) and lysozyme were purchased from Sigma Chemicals. Isopropyl-β-D-thiogalactopyranoside (IPTG), and 5-bromo-4-chloro-3-
indolyl-β-D-galactopyranoside (XGAL) were purchased from 5 Prime -> 3 Prime Inc. Methymercuric hydroxide (MeHgOH) was purchased from Alpha Catalog Chemicals. Ultrafree-MC 30,000 NMWL filter units were purchased from Millipore. Centricon 10, and 30 microconcentrators were purchased from Amicon. All other chemicals were of reagent grade or better and were purchased from either Sigma Chemicals, Fisher Scientific, or BDH.

Restriction endonucleases, T4 DNA ligase, T7 polymerase, *E.coli* DNA polymerase fragment I (Klenow), and S1 nuclease were purchased from either BRL or Pharmacia. Sequenase version 1.0 or 2.0 was purchased from USB. Exonuclease III (ExoIII) was purchased from New England Biolabs. *Taq* polymerase was purchased from Perkin-Elmer-Cetus. Moloney Murine Leukemia Virus reverse transcriptase (MMLV-RT) was purchased from BRL.

Liver samples from rat (*Ratus norwegicus*), mouse (*Mus musculus*), rabbit (*Oryctolagus cuniculus*), chicken (*Gallus gallus*), gekko (*Gekko gekko*), rainbow trout (*Onchorincus mykiss*), sturgeon (*Acipenser transmontanus*), and xenopus (*Xenopus leavis*) were generously provided by colleagues at the University of British Columbia and Simon Fraser University. Hagfish (*Eptatretus stouti*) were collected in Bamfield Inlet, British Columbia or purchased from Seacology Inc., Vancouver, British Columbia. Japanese firebelly newts (*Cynops pyrogastor*) were purchased from Jungle Land "B's Friendly Pet Store", Vancouver, British Columbia. Dogfish (*Squalis acanthias*), ratfish (*Hydrolagus collei*), sea stars (*Pisaster ochraceus*), and tunicates (*Pyura hauvster*) were purchased from Seacology Inc., Vancouver, British Columbia. Amphioxus (*Branchiostoma californiense*) were purchased from Pacific Bio-Marine Labs., Inc., Venice
Beach, California. Opossum (*Didelphis virginiana*) liver was obtained from W. Kelley Thomas, Department of Biochemistry, U. C. Berkeley, California.

Kodak X-Omat and Kodak XAR film was used for autoradiography. Intensifying screens (Lighting Plus) were purchased from Dupont.

**B. STRAINS, VECTORS, AND MEDIA**

1. **Bacterial Strains and Vectors**

   *E. coli* C600 and *E. coli* C600 HflA were the hosts for screening and isolation of DNA from clones in λgt10. *E. coli* Y1088, *E. coli* Y1089, and *E. coli* Y1090 (Young and Davis, 1983) were the hosts for screening and isolation of DNA from clones in λgt11. *E. coli* JM 83 (Messing, 1983) and *E. coli* DH 5α and *E. coli* DH 5αF (Hanahan, 1983) were the hosts for transformation and DNA isolation from clones in pUC 18, pUC19, (Vierra and Messing, 1982, Yanisch-Perron et al., 1985).

2. **Media**

   The medium for growth and screening of λ clones and bacterial hosts was NZCYM (Maniatis et al., 1982). Phage libraries were screened by plating phage on NZCYM agar (1.5% w/v) plates with overlay of NZCYM agarose (0.8% w/v). The medium for growth and selection of plasmid containing bacteria was Luria broth (Maniatis et al., 1982). For the selection of plasmid containing bacteria, clones were plated on LB-agar (1.5% w/v) plates supplemented with 100ug/mL AMP, 25 ug/mL IPTG, and 50 ug/mL XGAL.
Plasmid containing bacterial cultures were grown in LB supplemented with 100ug/mL AMP.

C. COLLECTION OF ANIMAL SPECIMENS

1. Collection of Hagfish Specimens

Pacific hagfish (*Eptatretus stouti*) were collected from the Broken Rocks region of Bamfield Inlet, Bamfield, British Columbia. Animal, blood and liver collection were accomplished using the facilities of the Bamfield Marine Station. Hagfish traps were prepared by placing 250-500 g of frozen fish in punctured 30 liter aluminium containers. The containers with secured lids attached, were lowered to the sea floor, and retrieved after 2 hours. Hagfish were stored in large tanks with a continuous flow of unfiltered sea water.

2. Collection of Plasma from Hagfish

Prior to blood collection, hagfish were sedated with tricaine methanesulfonate (0.4g/L sea water) in small plastic tubs. Sedated hagfish were removed from sea water and impaled through the head with a nail attached to a wooden plank. The plank was placed upright to allow blood to drain and collect in the large sinus at the ventral end of the animal. The pooling of blood in the sinus was expedited by gently stroking the length of the animal in a downward motion. The stroking action also facilitated the removal of excess slime. Blood was removed by inserting an 18 gauge needle attached to a 10mL syringe into the ventral sinus. An average of 3-5 mL of blood was removed from each of approximately 150 hagfish. The blood was mixed with 0.15 volumes of 80 mM citric acid, 70 mM
trisodium citrate, 110 mM glucose to prevent the blood from coagulating. Blood cells were removed by low speed centrifugation, and the light blue plasma was stored at -70°C.

3. Hagfish Liver Collection

Animals were sedated, bled and their livers removed following dissection. Liver samples were immediately frozen in liquid nitrogen, and kept at -70°C until required.

D. GEL ELECTROPHORESIS

1. Non-Denaturing agarose gels and Southern Blots

DNA fragments were separated according to size by electrophoresis in agarose or polyacrylamide gels. The buffer for native agarose gel electrophoresis was 1x TAE, 0.02μg/mL ethidium bromide (EtBr) (Maniatis et al., 1982). DNA samples in 1x loading dye (3% ficoll, 0.02% xylene cyanol, 0.02% bromophenol blue) were separated by electrophoresis at 5-10 volts/cm. DNA fragments were visualized by irradiation under UV light.

DNA fragments were transferred to Nytran or Nitrocellulose filters (Southern blots) as described by Maniatis et al. (1982). The filters were equilibrated in 0.5 mL/cm² of a solution consisting of 6x SSC, 0.5% SDS, 5x Denhardt's solution (1x Denhardt's solution: 0.02% BSA, 0.02% ficol, 0.02% polyvinylpyrolidine) at 55-68°C for 1-12 hours. Hybridizations were carried out in 0.2mL/cm² of the same solution at 55-68°C for > 3x Cot₁/₂ (Maniatis et al., 1982). The filters were washed sequentially in 6x SSC, 0.1% SDS, 1x SSC, 0.1% SDS, and 0.1x SSC, 0.1% SDS at the hybridization temperature.
After air drying, filters were exposed to Kodak XK or XAR films with intensifying screens for 12-24 hours at -70°C.

2. Denaturing Agarose Gels and Northern Blots

Total cellular and poly A+ RNA were separated by electrophoresis on formaldehyde agarose gels in 1x MOPS (20mM MOPS, 5mM NaOAc, 1 mM EDTA) as described by Davis et al., 1985. RNA samples were denatured by boiling for 2 min in loading buffer (50 % formamide, 25 % formaldehyde, 1x MOPS, 0.02% bromophenol blue, 0.8% glycerol) and separated by electrophoresis at 5 volts/cm. Agarose gels were prepared in 1x MOPS, 0.66 M formaldehyde, 0.6 ug/mL EtBr, and the RNA visualized by irradiation under UV light.

All buffers for Northern blot analysis were autoclaved prior to use. Following electrophoresis, gels were photographed, and soaked in 10x SSC to remove the formaldehyde. The RNA was transferred to Nytran filters by capillary action using 20x SSC for 8-12 hours. After transfer, the RNA was covalently cross-linked to the filters with UV light as described by the manufacturer. Filters were either stored at room temperature after drying or used immediately. Specific mRNAs were detected with 32P-labeled probes (see Radioactive Labeling of DNA). Filters were equilibrated in a buffer consisting of 50% formamide (deionized), 6x SSC, 5x Denhardts, and 0.5 % SDS (0.5 mL/cm²) at 42°C for 1-2 hours. Hybridization reactions were performed in 0.2 mL/cm² of buffer with the addition of denatured labeled probe to at least 1x10^6 cpm/mL. Hybridization was for >3x Cot₁/₂ (4-12 hours) at 42°C. After hybridization the filters were removed and washed as described for Southern blots.
3. Denaturing Polyacrylamide Gels

Fragments from DNA sequencing reactions were separated on 0.4 mm thick denaturing polyacrylamide gels. Gels were prepared in 0.5x TBE (Maniatis et al., 1982), 8.3 M urea and the concentration of acrylamide varied by adjusting the volume of the stock solution (38:2 acrylamide :bisacrylamide) and water added. Polymerization was initiated by the addition of ammonium persulphate and TEMED to final concentrations of 0.066% (w/v) and 0.024% (w/v) respectively. DNA fragments were separated by electrophoresis in 0.5x TBE at 50 watts constant power for 2-3 hours. Gels were dried under vacuum and exposed to either Kodak X-Omat or XAR films.

4. SDS-Polyacrylamide Gels

Proteins were separated according to size by electrophoresis in polyacrylamide gels containing SDS (SDS-PAGE). The running buffer for electrophoresis of protein samples was 0.025 M Tris-glycine pH 8.9, 0.1% SDS. The 5% acrylamide stacking gels were prepared in 125 mM Tris pH, 6.8, 0.1% SDS. Resolving gels (8-12%) were prepared in 6.25mM Tris glycine pH 8.9, 0.025% SDS, 1% glycerol by adjusting the volume of acrylamide stock solution (30:0.8 acrylamide:bisacrylamide) and water. Protein samples were boiled prior to electrophoresis in 1x loading dye (8% glycerol, 3% SDS, 0.02% bromophenol blue, +/- 0.72 M 2-mercaptoethanol, 0.125x stacking gel buffer) for 2 min. Proteins were visualized by staining with Coomasie brilliant blue (0.2% w/v, 25% methanol, 1.22 M glacial acetic acid).
E. PURIFICATION OF HAGFISH PROTHROMBIN

Hagfish prothrombin was partially purified using procedures adapted from Allen et al. (1977) and Mann (1976). The purification scheme involved chromatography on DEAE-Sephadex, precipitation with barium citrate, and ammonium sulphate fractionation. During the course of the purification process fractions were analyzed by SDS-PAGE and assayed for thrombin-like activity using venom from *Echaris caritinatis* (Echarin), and the chromogenic substrate S2238 (Ciba Geigy).

F. ISOLATION OF RNA

1. Isolation of Total Cellular RNA

Total cellular RNA was prepared by the method of Chomczynski and Sacchi (1987) using sterile disposable plasticware and freshly autoclaved solutions. One gram portions of fresh or previously frozen liver samples were homogenized with a Polytron in a sterile 50 mL Falcon tube containing 10.0 mL of 4 M guanidinium isothiocyanate, 25 mM sodium citrate, pH 7.0; 0.5% sodium lauryl sarcosyl, 0.2 M 2-mercaptoethanol (solution D). Sequentially, 1.0 mL of 2 M sodium acetate, pH 4, 10 mL of phenol (water saturated), and 2 mL of chloroform/isoamyl alcohol mixture (49:1) were added to the homogenate. The suspension was mixed thoroughly by inversion after the addition of each reagent. The final suspension was shaken vigorously for 20 sec and incubated on ice for 20 min. Samples were centrifuged at 5,000xg for 20 min at 4°C. After centrifugation, the aqueous phase was carefully removed with a disposable pipet and transferred to a sterile 50mL Falcon
tube. The RNA was precipitated by the addition of isopropanol to 50% and incubation at -20°C for 2-12 hours. RNA was collected by centrifugation (20 min at 5000xg). The RNA pellet was resuspended in one third the volume of solution D and insoluble material removed by centrifugation (5 min at 5,000xg). RNA was precipitated as before. RNA preparations were stored in solution D/isopropanol (1:1) at -20°C prior to use. The concentration of RNA was determined by assuming that a 1mg/mL solution of RNA had an OD_{260 nm} of 20.

2. Isolation of Poly A⁺ RNA

Poly A⁺ RNA was isolated by chromatography on oligo(dT) cellulose using either the Fast Track mRNA Isolation Kit (Invitrogen) or as described by Badley et al. (1988). Oligo(dT) cellulose was prepared by rehydrating 0.2 g of lyophilized cellulose per 10 g tissue in 10-20 fold excess volume of 10mM Tris, pH7.5. The hydrated oligo(dT) cellulose was collected by centrifugation, and the pellet resuspended in 10-20 fold excess volume of 10mM Tris, pH7.5; 500mM NaCl. The centrifugation step was repeated and the oligo(dT) cellulose resuspended a second time. The oligo(dT) cellulose was centrifuged once more and the pellet stored at room temperature under 0.5-1.0 mL of 10mM Tris, pH7.5; 500mM NaCl. The RNA was collected from the second isopropanol precipitation (see Isolation of Total Cellular RNA) by centrifugation and resuspended in 500 mM NaCl, 200 mM Tris, pH7.5; 100 mM MgCl₂, and 2% SDS (binding buffer). Insoluble material was removed by centrifugation. The RNA was added to the rehydrated equilibrated oligo(dT) cellulose and incubated at room temperature with intermittent shaking for 45-60 min. The oligo(dT) cellulose was collected by centrifugation and the supernatant poured off. The oligo(dT) cellulose was resuspended in fresh binding buffer and resuspended
three more times. The oligo(dT) cellulose was poured into a sterile 5mL syringe the base
of which had been plugged with sterile glass wool. The packed oligo(dT) cellulose was
washed with binding buffer until the A$_{260}$ of the eluate was less than 0.05. The column
was allowed to run dry and the poly A$^+$ eluted with 0.5mL fractions of 10mM Tris,
pH7.5. The RNA containing fractions were pooled and precipitated by the addition of 0.1
volumes of 3 M NaOAc, pH 5.2, 2 volumes of 95% ethanol and incubation at -20°C.
RNA was resuspended in sterile water and stored at -70°C.

G. ISOLATION OF DNA

1. Isolation of Plasmid DNA

Small amounts of plasmid DNA for DNA sequence determination and restriction
mapping were prepared by a modified boiling lysis procedure (Gatermann et al., 1988).
Bacterial cells (E. coli DH 5α) were harvested by centrifugation (5,000xg at 4°C for 10
min) from 1 - 5mL of an overnight culture of the clone of interest. The bacterial cell pellet
was resuspended in 200µL of STET buffer (8% sucrose, 5% Triton X-500, 50 mM EDTA,
0.5 mg/mL lysozyme) and placed in a boiling waterbath for 2 min. Bacterial cellular debris
was collected by centrifugation at 12,000xg for 10 min at room temperature and removed
with a sterile toothpick. The DNA was precipitated by the addition of 200 µL of
isopropanol. The DNA pellet was collected by centrifugation at 12,000xg for 20 min at
room temperature, washed with cold 75% ethanol (-20°C) and the residual ethanol
removed with a flame drawn pipet. DNA pellets were resuspended in sterile dH$_2$O and
stored at 4°C until needed.
Larger scale plasmid DNA preparations were prepared using a modification of the alkaline lysis procedure (Maniatis et al., 1982) and further purified using pZ523 columns (5 Prime -> 3 Prime Inc.) as described by the manufacturer. The concentration of DNA was determined by assuming that a 1mg/mL solution of DNA had an OD\textsubscript{260 nm} of 20.

2. Phage DNA Isolation

For 20 mL λ phage DNA preparations, 100 μL of an overnight culture of host bacterial cells was added to 1-3x10\textsuperscript{6} pfu of λ phage. The mixture was incubated at 37°C for 10 min to allow the λ phage to adsorb to the cells. This mixture was used to inoculate 20 mL of pre-warmed NZCYM broth and the culture incubated at 37°C until lysis. Chloroform was added (0.5 mL) and the incubation at 37°C continued for 5 min. Bacterial debris was removed by centrifugation at 10,000xg for 10 min. The λ phage were precipitated by the addition of 0.3 volumes of polyethylene glycol 6000 (PEG), 0.15 volumes of 5M NaCl, and incubation overnight at 4°C. λ phage were collected by centrifugation at 10,000xg for 15 min at 4°C. The λ phage were resuspended in 0.5 mL of DNase I buffer (50mM Tris, pH7.5, 5 mM MgCl\textsubscript{2}, 0.5 mM CaCl\textsubscript{2}) to which 5 μL of DNase I (1 mg/mL) and 5 μL RNase A (5 mg/mL) were added, and the solution was incubated at 37°C for 30 min. Debris was removed by centrifugation at 12,000xg for 5 min and the solution transferred to a clean 1.5 mL microcentrifuge tube containing 50 μL of 10 % SDS, 5 μL of 0.5 M EDTA, and 20 μL of Proteinase K (5 mg/mL). The solution was incubated at 68°C for 60 min. Following the incubation at 68°C, the solution was extracted once with 400 μL of TE (10 mM Tris, 1 mM EDTA, pH 8.0) equilibrated phenol, once with 400 μL of phenol/chloroform (1:1), and once with 400 μL of chloroform. The λ phage DNA was precipitated by the addition of 0.1 volume of 3 M NaOAc, pH 5.2 and an
equal volume of isopropanol, and incubation at -70°C for 20 min. The λ phage DNA was collected by centrifugation at 12,000xg for 20 min and washed with cold 75% ethanol. The last traces of ethanol were removed, and the DNA resuspended in 50 μL of TE.

3. Genomic DNA Isolation

Liver tissue was ground to a fine powder in liquid nitrogen with a mortar and pestle. The liver powder was dissolved in 50 mM Tris, pH 8.5; 100 mM EDTA, 200 mM NaCl, 0.5 mg/mL proteinase K, 0.2% SDS (10 mL/g tissue). RNase A (DNase free) was added to a concentration of 100 μg/mL, and the sample digested overnight at 50°C. The DNA solution was extracted gently three times with an equal volume of phenol equilibrated in TE, and twice with an equal volume of chloroform. Following each extraction the sample was centrifuged at 3,000g at room temperature for 10 min. The aqueous phase was removed with a Pasteur pipet, and transferred to a sterile 50 mL Falcon tube. The DNA was precipitated by the addition of 0.1 volumes of 3 M NaOAc, pH 5.2 and 0.8 volumes of isopropanol. The genomic DNA precipitate was removed with a Pasteur pipet and transferred to a Falcon tube containing 70% ethanol. The DNA was washed twice with 75% ethanol and the DNA collected by centrifugation. The final DNA pellet was dissolved in TE, 10 μg/mL RNase A (DNase free) and adjusted to a final concentration of 1 mg/mL.

H. THE POLYMERASE CHAIN REACTION

1. Preparation of Single-Stranded cDNA

Single-stranded cDNA (sscDNA) was prepared from total RNA with MMLV-RT using either an anchored oligodeoxyribonucleotide (T17,XSP, see Table 2), random
hexadeoxyribonucleotides (p(dN<sub>6</sub>)), or specific oligodeoxyribonucleotides (see below and Table 2) as first-strand synthesis primers (see Figure 11). First-strand synthesis reactions were performed in 50 µL volumes using the reverse transcriptase buffer provided by the manufacturer. Total cellular RNA (5-10 µg) was denatured in 20 mM methylmercuric hydroxide (MeHgOH) for 10 min at room temperature. The denatured RNA was subsequently frozen on dry ice. After freezing, a reverse transcriptase cocktail consisting of 200 µM of each of the four deoxyribonucleotide triphosphates, 1-2 µg of deoxyribonucleotide primer, 200 mM DTT, and 20 units of human placental ribonuclease inhibitor (BRL) was added to the frozen RNA/MeHgOH. The mixture was thawed rapidly, placed on wet ice and 200-600 units of MMLV-RT was added. Reactions were incubated at 37°C for one hour, diluted to 200 µL with sterile distilled water and stored at -20°C.

2. Selection of Primers for use in the Polymerase Chain Reaction

A list of all the oligodeoxyribonucleotide primers used in this study are presented in Table 2. For details regarding oligodeoxyribonucleotide primer selection for the PCR see the Results section.
Table 2. Oligodeoxyribonucleotide primers used for DNA sequence determination and the polymerase chain reaction.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ 3</td>
<td>5'-pGCGGCCGCACCTGCAATG-3'</td>
</tr>
<tr>
<td>EQ 4</td>
<td>5'-CGCCGGCGTGGACGTTACTTAA-5'</td>
</tr>
<tr>
<td>RACE 36</td>
<td>5'-CGAGCATGCGTCGACAGGCATTTTTTTTTTTTTTTTTT-3'</td>
</tr>
<tr>
<td>T1'XSP</td>
<td>5'-ACACTGCAGGAGCTCTAGATTTTTTTTTTTTTTTTTT-3'</td>
</tr>
<tr>
<td>Th 3</td>
<td>5'-GAGCTGCTGTGGGCGCAGCCTCATTAG-3'</td>
</tr>
<tr>
<td>Th 4</td>
<td>5'-GGCTTTGTAACCAGCAGCAACAT-3'</td>
</tr>
<tr>
<td>Th 7</td>
<td>5'-AGGCGACCTTGGGAGGTATG-3'</td>
</tr>
<tr>
<td>Th 10</td>
<td>5'-AAGGCGGTGTGACTGGAATGGGG-3'</td>
</tr>
<tr>
<td>Ser 1</td>
<td>5'-ACAAAAGCTTG/AGGICC/CC/C/GTT/C/AA/GTC/CC-3'</td>
</tr>
<tr>
<td>His 1</td>
<td>5'-ACAGAATTCTGGGGTIG/CT/AC/ICIG/CIG/CIG/CAC/TTG-3'</td>
</tr>
<tr>
<td>FII 1</td>
<td>5'-GCCGCNTGGTGGC/AGA/AC/TTGGG-3'</td>
</tr>
<tr>
<td>FII 2</td>
<td>5'-TCACANCTTCNCCC/ANNA/3'</td>
</tr>
<tr>
<td>B-Loop</td>
<td>5'-ACAGAATTCTGCT/C/TIC/T/IT/AT/CCC/CTTG-3'</td>
</tr>
<tr>
<td>FX 1</td>
<td>5'-GGTGAA/GTGG/CCCNTGGCA/GGG-3'</td>
</tr>
<tr>
<td>FX 2</td>
<td>5'-CATCCT/CTC/NNCC/CAA/GTNAC-3'</td>
</tr>
<tr>
<td>FIX 1</td>
<td>5'-GGTCAA/GTAT/CC/CT/ATGGCAGGT-3'</td>
</tr>
<tr>
<td>FIX 2</td>
<td>5'-TCTCC/AGCCAG/TAAT/ATTNCC-3'</td>
</tr>
<tr>
<td>FIX 3</td>
<td>5'-GC/AC/GCAC/TTCC/TC/NC/CC/AGC/ACT-3'</td>
</tr>
<tr>
<td>SWG</td>
<td>5'-ACAAAGTCC/CTGG/CICCCCAIG/CT/AC/ICIC/TIC/ATIC-3'</td>
</tr>
<tr>
<td>XVII</td>
<td>5'-ACAGAATTCTGGCAGC/AA/AGG/ICAA/GTGG/CCC/ITGG-3'</td>
</tr>
</tbody>
</table>

Nucleotides separated by the / symbol indicate that the oligodeoxyribonucleotide was synthesized with two nucleotides present at this position.

N indicates that all four oligodeoxyribonucleotides were included in the synthesis at this position.

I indicates that inosine was used in the synthesis at this position.
Figure 11. Scheme for the Amplification of sscDNA

A sample of RNA was mixed with an anchored oligo(dT) primer such as RACE 36 or T17 XSP (see Table 2) in buffer, dNTPS, and reverse transcriptase (see Materials and Methods). Following sscDNA synthesis, cDNA fragments were amplified by the PCR using either two cDNA specific primers (A), or one cDNA specific primer and the anchored oligo(dT) primer (B).
cDNA Synthesis

(anchored oligo dT)
mRNA + TTTT + Reverse Transcriptase + dNTPs

Template for the PCR

Amplification of Internal cDNA Fragments

Amplification of 3' end of cDNA
3. Reaction Conditions for the Polymerase Chain Reaction

The polymerase chain reaction (PCR) was performed essentially as described by Saiki et al. (1985) using a Perkin Elmer Cetus DNA Thermocycler. Two different buffers were used for DNA amplification, buffer A (67 mM Tris, pH 8.8, 16.6 mM, (NH₄)₂SO₄, 10 mM 2-mercaptoethanol, 1.0 mM MgSO₄) or buffer B (10 mM Tris, pH 8.5, 0.05% Tween-20, 0.05% Nonidet P-40, 1.0 mM MgCl₂). Samples for the PCR were prepared in 50 μL volumes containing either buffer A or buffer B, 200 μM dNTPs (from a stock solution 5 mM in dGTP, dCTP, dTTP, dATP, and 20 mM MgCl₂), 5-10 ng of ssDNA, 20 pmol of each of two oligodeoxyribonucleotide primers, and 1 unit of Taq polymerase. A typical PCR cycle consisted of 10 sec at 94°C, 30 sec at 62°C, and 1.0 min at 72°C for a total of 35 cycles; the 72°C incubation step was extended to 10 min for the final cycle. On completion of the PCR cycles, 10 μL of the 50 μL reaction volume was removed for analysis. Products of the PCR reaction were visualized on 1.0% agarose gels stained with EtBr. Annealing temperatures for oligodeoxyribonucleotide primers used in the PCR were estimated with the following equation:

\[ T_A = (4°C(GC)+2°C(ATT)) - 5°C \] (Maniatis et al., 1982)

Where \( T_A \) is the estimated annealing temperature, and I is deoxyriboinosine. The temperatures derived from this equation were found to be reliable for oligodeoxyribonucleotide primers 20 to 30 nucleotides in length.
I. DNA SUBCLONING

1. Preparation and Isolation of DNA fragments for Subcloning

DNA fragments for ligation into pUC vectors were produced either by restriction endonuclease digestion or by the PCR (see below). DNA fragments generated by the PCR not digested with restriction endonucleases were made blunt-ended with the Klenow fragment of *E. coli* DNA polymerase I (Maniatis et al., 1982). Following amplification the PCR samples were extracted with an equal volume of phenol/chloroform, and precipitated by the addition of 0.1 volumes of 3 M NaOAc, pH 5.2; 2 volumes of 95% ethanol and incubation at -70°C for 10 min. The DNA precipitate was collected by centrifugation at 12,000xg for 20 min, and the pellet resuspended in 1x Klenow buffer (10 mM Tris, pH 8; 10 mM MgCl$_2$, 50 mM NaCl, 1mM DTT). DNA samples were incubated at 37°C for 5 min with one unit of Klenow. The samples were adjusted to 500 uM dNTPs and incubated at 37°C for a further 10 min. The DNA fragments were separated by electrophoresis on 1 - 2% low melting point agarose gels in the presence of EtBr. The DNA fragments were visualized by irradiation under UV light, and recovered from gel slices with *GeneClean* (Bio 101) as described by the manufacturer.

2. Ligation and Transformation of DNA into Bacteria

For restriction endonuclease mapping and DNA sequence determination, DNA fragments were ligated into pUC vectors. Ligations were carried out with 10-20 ng of restriction endonuclease digested vector DNA and 25-50 ng of insert DNA in 20 - 30 µL volumes of a buffer consisting of 50 mM Tris, pH 7.6; 10 mM MgCl$_2$, 1 mM ATP, 1 mM
DTT and 5 % (w/v) PEG 8000. One unit of T4 DNA ligase was added to each sample and the ligation allowed to proceed for 4-12 hours at room temperature. If not used immediately, ligations were stored at -20°C. After ligation, 1-5 μL of the ligation mixture was used to transform 50 μL of competent *E. coli* DH5α. The cells and DNA were mixed by gentle tapping and incubated on ice for 30 min. The transformed bacterial cells were incubated at 37°C for 30-60 sec, and then placed on ice for 2 min. The volume of the bacterial suspension was increased to 1.0 mL by the addition of NZCYM and placed in a shaking incubator for 15-20 min. Transformed bacteria were selected by plating 50-100 μL of the bacterial culture on LB agar plates, supplemented with 100 μg/mL AMP, 25 μg/mL IPTG, and 50 μg/mL XGAL.

Transformation competent bacteria were prepared by harvesting bacteria from liquid cultures when the OD$_{600}$ reached 0.5-0.6. Bacteria were collected by centrifugation at 5,000xg for 10 min, the pellet resuspended in ice cold 50 mM CaCl$_2$ and incubated on ice for 30 min. The bacteria were recentrifuged and the pellet resuspended in 50 mM CaCl$_2$, 15 % glycerol at 1/20th of the original culture volume and stored at -70°C. Competent cells were thawed on ice just prior to use.

**J. RADIOACTIVE LABELING OF DNA**

1. **Klenow Labeling**

DNA fragments were labeled by a modification of the method described by Feinberg and Vogelstein (1983). DNA fragments (25-50 ng) were mixed with 1.4 μL of random hexadeoxyribonucleotides (p(dN)$_6$), 90 OD$_{260}$ units/mL in dH$_2$O) in a final volume of 8 μL. The DNA sample was denatured by boiling for 5 min, and cooled on ice. The
labeling reactions were performed in 25 μL volumes in 200 mM HEPES (N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid), pH 6.6; 25 uM dCTP, 25 uM dGTP, 25 uM dTTP (in 250 mM Tris, pH 8.0; 25 mM MgCl₂, 50 mM 2-mercaptoethanol), 0.4 ug/mL BSA, 50 μCi α-[³²P]-dATP (3000 Ci/mM), using 5 units of Klenow. The reaction was incubated at 37°C for 30-60 min, and unincorporated nucleotides removed by spun chromatography (Maniatis, et al., 1982) using TE equilibrated Sephacryl S300. Prior to use, labeled DNA fragments were denatured by boiling in 50% formamide for 2 min.

K. DNA SEQUENCE ANALYSIS

1. DNA Sequence Determination from Plasmid DNA Templates.

   DNA sequence was determined by the chain termination method (Sanger et al., 1977) using double-stranded DNA templates in pUC vectors (Gatermann et al.,1988) and T7 DNA polymerase, or Sequenase. Sequencing reactions were carried out using the dideoxy- and deoxyribonucleotide concentrations recommended by the manufacturer (see below). Sequencing reactions were performed by denaturing plasmid DNA (3-5ug) in 200 mM NaOH, 2mM EDTA in presence of 2 pmoles of sequencing primer (M13 forward, M13 reverse, or sequence specific see Table 2) in boiling water for 2 min. The denatured plasmid DNA was precipitated by the addition of 0.1 volumes of 3 M NaOAc, pH 5.2; and 2 volumes of cold 95% ethanol (-20°C) followed by gentle mixing. The DNA precipitate was collected by centrifugation at 12,000g for 20 min at room temperature. The DNA pellet was washed with cold 75% ethanol (-20°C) and the last traces of ethanol removed with a flame drawn pasteur pipet. The DNA pellets were resuspended in 2uL of 5x reaction buffer (200 mM Tris, pH 7.5; 100 mM MgCl₂, 250 mM NaCl) and 8uL of dH₂O.
The samples were mixed (Vortex shaker) and incubated at 37°C for 10 min. The DNA samples were removed from the water bath and 1 μL of 100 mM DTT, 2 μL of a 1/8 dilution of T7 polymerase (or Sequenase), 2 μL of labeling mix (9 uM dATP, 9 uM dTTP, and 9 uM dGTP) and 0.5 μL of α-[thio-35S]-dATP (3000 Ci/mmol) added to each sample. Sample contents were mixed by brief centrifugation and incubated at room temperature for 2-5 min. Following this incubation, 3.5 μL aliquots of each sample were mixed with 2.5 μL of the ddGTP, ddATP, ddTTP, and ddCTP termination mixes (80 uM dGTP, 80 uM dATP, 80 uM dTTP, 80 uM dCTP, 8 uM ddNTP) and incubated at 37°C for 5-10 min. Reactions were stopped by the addition of 4 μL of 98% formamide, 10 mM EDTA, 0.02% xylene cyanol, 0.02% bromophenol blue. Sequencing reactions were stored for up to one week at -20°C.

2. Preparation of Unidirectional Deletions with Exonuclease III

Unidirectional deletions were prepared from DNA fragments greater than 1.0 kilobase pairs (kbp) with Exonuclease III (Exo III) (Henikoff, 1987). For up to 25 individual aliquots, 10 μg of insert containing plasmid DNA was used. The DNA was digested to completion with two restriction enzymes, leaving a 4 base 3' protrusion protecting the vector, and a 5' protrusion adjacent to the insert (see Figure 12). The digested DNA was extracted with an equal volume of phenol/chloroform and precipitated by the addition of 0.1 volumes of 3M NaOAc, pH 5.2; 2 volumes of ethanol, and incubated at -70°C for 10 min. The DNA precipitate was collected by centrifugation at 12,000xg for 20 min, washed with 75% ethanol, and the pellet resuspended in 60 μL of 1x Exo III buffer (66 mM Tris, pH 8.0; 0.66 mM MgCl₂). The DNA sample was pre-incubated at 37°C for 5 min and mixed with 500 units of Exo III and the incubation
Figure 12. Scheme for Unidirectional Deletions using Exonuclease III and S1 Nuclease.

The cross-hashed box represents the region of DNA to be deleted, and the solid box represents the DNA sequencing primer site. A and B represent the restriction endonuclease sites which leave a 5' and 3' protrusion respectively. After the initiation of the digestion reaction aliquots are removed at timed intervals, treated with S1 nuclease to remove 3' protrusions generated by the action of Exo III, the ends made blunt with Klenow, and ligated together with T4 DNA ligase.
Plasmid with insert DNA

A B primer sequence

Restrict with A, B

5' Protrusion

Exo III digestion
Remove timed aliquots

S1 nuclease
Klenow
T4 DNA ligase

Transform *E. coli*
Template Preparation
DNA Sequence Determination
continued at 37°C. At 30 sec intervals 2.5 µL aliquots of the Exo III digestion reaction were removed and mixed with 7.5 µL of S1 nuclease mix (30 mM KOAc, pH 4.6, 25 mM NaCl, 0.5 % glycerol, 1 mM ZnSO₄, 2.5 units of S1 nuclease) on ice. After all the desired aliquots had been removed and added to the S1 mix, the samples were incubated at room temperature for 30 min. The S1 digestion was stopped by the addition of 1 µL of 300mM Tris, 50 mM EDTA, pH 8.0, and incubation at 70°C for 10 min. Following S1 nuclease digestion, 2 µL aliquots from each time point were removed and analyzed by agarose gel electrophoresis. The ends of the Exo III digested DNA were made blunt ended by the addition of 1 µL 20 mM Tris, pH 8.0; 100 mM MgCl₂, 0.1 units of Klenow, and incubation at 37°C for 2 min, followed by the addition of 1 µL of 125 uM dNTPs, and incubation at 37°C for a further 5 min. The ends of the Exo III digested samples were ligated together by the addition of 40 µL of 82.5 mM Tris, pH 7.6; 8.25 mM MgCl₂, 1.25 mM DTT, 12.5 ug/mL BSA, 125 uM spermidine, 25 uM ATP, 1 unit of T4 DNA ligase and incubation at room temperature for 4-12 hours. An aliquot of the ligation reaction (2 uL) was used to transform competent bacterial cells (as described previously).

3. DNA Sequence Determination from PCR Generated Templates.

DNA samples were amplified from sscDNA as described above. An aliquot of the amplification reaction was separated by electrophoresis on a 1.5 % low melting point agarose gel, and the DNA fragment(s) visualized by irradiation with UV light. DNA fragments were removed in small agarose plugs (using a disposable micropipet tip) and diluted in 200 µL of dH₂O. Sequencing templates (ssDNA) were prepared using the unbalanced oligodeoxyribonucleotide primer method described by Gyllensten and Erlich (1988). A aliquot of the gel purified, diluted PCR product was amplified using buffer A
(see above) and a oligodeoxyribonucleotide primer ratio of 1:1/100. The PCR cycle parameters were different from those used in balanced amplification reactions, and performed in 25 μL volumes. A typical amplification cycle consisted of 92°C for 45 sec, 60°C for 1 min, ramp from 60°C to 72°C for 1 min, 72°C for 2 min and was repeated 35 times. An aliquot (5 μL) of the ssPCR reaction was analyzed by agarose gel electrophoresis, and the remaining sample concentrated and the excess oligodeoxyribonucleotide primers, and dNTPs removed using 30K NMWL ultrafree-MC filter units. The volume of the concentrated, desalted ssDNA was adjusted to 7-10 μL. The ssDNA (7μL) was annealed with 2 pmoles of the limiting oligodeoxyribonucleotide primer, and 2 μL of 5x buffer (200 mM Tris, pH 7.5; 100 mM MgCl$_2$, 250 mM NaCl) in a final volume of 10 μL by incubation at 68°C for 2 min. The DNA samples were removed from the water bath and 1 μL of 100 mM DTT, 2 μL of a 1/8 dilution of T7 polymerase (or Sequenase), 2 μL of labeling mix (0.75 uM dATP, 0.75 uM dTTP, and 0.75 uM dGTP; for fragments 200-500 bases in length, or 0.375 uM dATP, 0.375 uM dTTP, and 0.375 uM dGTP; for fragments less than 250 bases in length) and 0.5 μL of α-[thio-$^{35}$S]-dATP (3000 Ci/mmol) added to each sample. Sample contents were mixed and incubated at room temperature for 2-5 min. The DNA sequencing reactions were completed as described previously.

4. DNA Sequence Data Analysis

DNA sequence data was analyzed using the Delaney Sequence Analysis Program Version PC-2.2 (Delaney Software Ltd), ESEE (Cabot and Beckenbach, 1989), and PCGENE (Intelligenetics).
L. PREPARATION AND SCREENING OF CDNA LIBRARIES

1. cDNA Synthesis

Oligo(dT)-primed, randomly primed, and specifically primed cDNA libraries were prepared using a modification of the method described by Gubler and Hoffman (1983) and cDNA synthesis kits purchased from Invitrogen (The Copy Kit). The specifically primed cDNA library from chicken liver poly A⁺ RNA was prepared by priming first-strand synthesis with a unique complementary oligodeoxyribonucleotide 38 nucleotides in length (5'-CTGCAGTAGTTCTCAGTGAGGTCAGGATAAATGGAGGC-3'). The specifically primed cDNA library from hagfish liver poly A⁺ RNA was prepared by priming first-strand synthesis with a unique complementary oligodeoxyribonucleotide 40 nucleotides in length (5'GAGCATTCCCTGAGGCCGCTTCCTGTACCATGACTTTG C-3'). The products of the cDNA synthesis reactions were labeled by the addition of 5 μCi of α-[32P]-dATP (3000 Ci/mM) in the first-strand reaction. The cDNA synthesis kits were used as per the manufacturers instructions.

2. Linker Addition and Size Fractionation

Following synthesis, cDNA fragments were made blunt-ended with Klenow fragment and EcoRI/NotI linkers attached by ligation with T4 DNA ligase. The EcoRI/NotI adaptors were prepared by mixing equimolar amounts of the oligodeoxyribonucleotides EQ3 and EQ4 (see Table 2) in TE, heating to 68°C for 10 min, and gradual cooling to room temperature. Linkers were ligated to cDNA fragments in 10uL of 50 mM Tris, pH7.5; 7mM MgCl₂, 1 mM DTT, containing 250-500 ng of blunt-ended
cDNA, 0.5 mM ATP, 30 ng of linker, and 1 unit of T4 DNA ligase. The linker ligated
cDNA fragments were size fractionated by gel filtration chromatography on Sepharose
CL4B. The Sepharose CL4B was equilibrated in 10 mM Tris, pH 7.5, 100 mM NaCl,
1 mM EDTA, 20 μg/mL tRNA and packed in 2 mL disposable pipets. The cDNA samples
in TE were adjusted to 0.8 % bromophenol blue, applied to the column, and 100 μL
fractions collected in 10 mM Tris, pH 7.5, 100 mM NaCl, 1 mM EDTA. The total
radioactivity per fraction was determined by scintillation counting and the fractions
containing the highest number of cpms were pooled. The volume of the pooled fractions
was reduced to 2-5 μL using Ultrafree-MC 30K filtration units. The cDNA (25-50 ng)
was ligated to non-dephosphorylated λgt10 (or λgt 11) arms in 5 μL volumes consisting of
50 mM Tris, pH7.5; 7 mM MgCl₂, 1 mM DTT, 1 mM ATP, 0.5 units of T4 DNA ligase
and incubation at room temperature for 1 hour followed by incubation at 4°C for 12 hours.

3. In Vitro Packaging and Library Titration

An aliquot of the ligation reaction (3.5 μL) was packaged into phage particles using
Giga Pack Gold packaging extracts (Stratagene) as described by the manufacturer. The λ
phage libraries were stored at 4°C in 500 μL of SM (100 mM NaCl, 8 mM Mg(SO)₄, 50
mM Tris, pH7.5, 0.01 % gelatin) containing a drop of chloroform.

4. Plating λ Phage Libraries

The titres of λ phage libraries were determined by plating 1 μL, 0.1 μL, and 0.01
μL of the packaged library on either E. coli C600, E. coli C600 HflA⁻ (λgt 10) or E. coli
Y1088 (λgt 11). For primary screens, λ phage were plated at a density of 35-40,000
pfu/150 mm plate. Plates were incubated at 37°C until the λ phage plaques were just
visible but not touching each other and placed at 4°C for one hour. Replicas of the plaques were transferred to nitrocellulose discs (150 mm) and incubated inverted on NZCYM plates at 37°C overnight. For screens other than this high density screen, this amplification step was omitted. The λ phage DNA on the nitrocellulose discs was denatured, neutralized and immobilized as described by Maniatis et al. (1982).

5. Screening of λ Phage Filters and Plaque Purification

The λ phage containing the cDNAs of interest were detected by hybridization to 32P-labeled probes as described for Southern blots (see above). The λ phage identified in this way were purified by successive platings at lower phage densities until 100% of plated phage reacted with the hybridization probe. The λ phage DNA was purified as described above.
III. RESULTS

A. ISOLATION OF THE 5'END OF THE CHICKEN PROTHROMBIN CDNA

1. Northern Blot Analysis of Chicken Prothrombin mRNA

The size of the mRNA for chicken prothrombin was determined by separating denatured chicken liver RNA on formaldehyde agarose gels, and immobilizing the RNA on Nytran membranes. When these blots were hybridized with $^{32}$P-labeled chicken prothrombin cDNA probes, two mRNAs were detected (see Figure 13). These mRNAs are approximately 2200 and 3200 nucleotides in length. The signal intensity of the two mRNAs on the Northern blot suggests that the 2200 nucleotide transcript is the more abundant of the two (> 90% message) see Figure 13.

2. Isolation and Sequence Determination of cDNA Clones for Chicken Prothrombin

A number of cDNAs for chicken prothrombin have been isolated and their sequence determined in a previous study (Irwin, 1986). The total length of the chicken prothrombin cDNA determined was 2561 nucleotides, and included the 3' untranslated sequence of 1145 nucleotides (UTS), and 1416 nucleotides of coding sequence (472 amino acids). From the predicted amino acid sequence, it was determined that circulating prothrombin consisted of a two chain protease domain and two Kringles (Irwin, 1986). The DNA sequence of the 5' end of the transcript was required to determine if chicken prothrombin (like mammalian prothrombin) was synthesized as a prepro-protein containing a Gla...
domain and if the propeptide sequence was homologous to the propeptide sequences of other vitamin K-dependent proteins.

To isolate the remaining 5' cDNA sequence of chicken prothrombin, a randomly primed cDNA library was prepared from chicken liver poly A⁺ RNA and screened with a ³²P-labeled BamHI-PstI fragment isolated from λCII201 (Irwin, 1986; see Figure 14). From the initial 250,000 λ phage screened, 7 positives were identified; one of these (λCIIrpl) contained a 1.3 kilobase pair (kbp) insert extending an additional 330 nucleotides 5' of λCII201. The remaining 5' cDNA sequence was obtained by screening a specifically primed cDNA library. From the initial 25,000 recombinant λ phage screened, 5 positives were identified and the DNA sequence of their inserts determined. One of these, λCII38;1 contained a 510 bp insert. From the sequence of λCII38;1, 157 amino acids of chicken prothrombin were predicted. The complete sequence of the chicken prothrombin cDNA is presented in Figure 15, and a restriction map and cDNA cloning strategy are presented in Figure 14. The sequence presented in Figure 15 indicates that chicken prothrombin has a very similar structure to that of the mammalian prothrombins. The chicken prothrombin cDNA predicts an open reading frame of 607 amino acids consisting of a 22 amino acid leader sequence, a 21 amino acid propeptide, and 564 amino acids of plasma prothrombin with a predicted molecular weight of 64,130 (for the unglycosylated protein). The amino acid sequence predicted from λCII38;1 differs from the amino acid sequence determined for this region of chicken prothrombin in two positions. The cDNA sequence predicts an Ala residue at position 68 and a Trp residue at position 90. The amino acid sequence determined by protein sequencing assigned Val and Tyr residues to these positions 68 and 90 respectively (Walz, 1978), see Table 3.
Figure 13. Northern Blot Analysis of Total Cellular RNA from Chicken Liver.

Sample preparation and protocols are as described in the Materials and Methods. 

Panel A: 1.5% denaturing agarose gel of chicken liver RNA. Lane M, RNA size ladder (RNA ladder, BRL) the sizes of the RNA fragments are as indicated; Lane 2, 10 ug of total cellular RNA from chicken liver. The RNA from the agarose gel in Panel A was transferred to Nytran and hybridized with $^{32}$P- labeled chicken prothrombin cDNA probe see Panel B. Panel B: Northern blot analysis of chicken liver RNA. Lane 1, 10ug of total cellular RNA from chicken liver; M, RNA ladder, BRL. The sizes of the marker fragments are as indicated. The arrows denote the position of the chicken prothrombin transcripts and there respective sizes (3.2, 3200 nts; 2.2, 2200 nts).
Figure 14. Restriction map and cDNA cloning strategy for chicken prothrombin.

The length of the combined cDNA sequences is represented by the thick box and the positions of restriction endonuclease sites are indicated above. The solid portion of the box represents the coding region, the open portion the 5' UTS, and the shaded portion the 3'UTS. The arrow indicates the position and primer orientation of the cDNA synthesis primer (not drawn to scale). The cross-hatched boxes below the restriction map denote the length and position of hybridization probes (see text for details). The positions, names and lengths of isolated cDNA fragments are indicated below the restriction map. The scale is indicated above the restriction sites.
Figure 15. The cDNA Sequence and Predicted Amino Acid Sequence of Chicken Prothrombin.
The arrows indicate the positions of the putative signal peptidase cleavage sites, the position of the putative propeptidase cleavage site (a.a -21,-20; -1,+1 respectively) and the positions of the putative factor Xa cleavage sites. The Asn residues of potential glycosylation sites are boxed, and the active site His, Asp, and Ser residues are in bold text. The RGD peptide sequences are in bold type. The polyadenylation consensus sequences AATAAA are in bold type and the polyadenylation addition site of the shorter transcript is indicated by an arrow. The position and length of the inverted repeats in the 5' UTS are indicated by arrows above the sequence.
2400 CCAACAAAATACCATCTGTCTTTCAAAAGTTAGGTCAATCTTATATGCAACTTACCTG
2475 AGGTACAGAAATAGGGGGTTGGAATAGATGAAGTAACACGCACTTGCTAGACAGTAACTTTCTATATGGCC
2550 AAGTACTGCTGGGACTTTGAAGTATATTTTCCACTGCGATAACTAGATTCAGAAGGAAGCACTTTCAGTACACAGAA
2625 TTTTCAGGCTTTCCAAAGGCGAGCATGGTACGTCTGACCTTTTGTCTTTTATATGGAAGTTGATTCACC
2700 CTPAAAAGAACCCCACTTTCCATGAACCTCTTCCACAGCAAGACAGAAGATCAATATAGTATATATAGACTGGC
2775 CAATCTGTTCAGACCCATTTTCTCTCAACTAAGGGGTGGAAAGCTATTTTGCTCCCCAAAACATCATTC
2850 TCAAATCCCTCATCCCCACACGGCAATCAACTTACAGAAACAAGCAATAGACAAAAGTTGCTCTGCTTAAATG
2925 GAGTATTAAGGAAGATGACTTTGAAAGAGATGCTAGAGAAACTATCCAAAATTTGTTGGAATAAAACAGTTAT
3000 TAATC
Table 3. Differences in Amino Acid Sequence Assignments for Chicken Prothrombin.

<table>
<thead>
<tr>
<th>Amino Acid Residue</th>
<th>Protein Sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>cDNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>90</td>
<td>Tyr</td>
<td>Trp</td>
</tr>
<tr>
<td>168</td>
<td>Glu</td>
<td>Lys</td>
</tr>
<tr>
<td>310</td>
<td>Glu</td>
<td>His</td>
</tr>
<tr>
<td>326</td>
<td>Phe</td>
<td>Lys</td>
</tr>
<tr>
<td>391</td>
<td>unassigned</td>
<td>Trp</td>
</tr>
</tbody>
</table>

<sup>a</sup> Protein sequence was determined as described in Walz et al., 1978
B. THE PURIFICATION OF HAGFISH PROTHROMBIN FROM PLASMA

To initiate the isolation of hagfish prothrombin, hagfish plasma was collected and preliminary purification studies begun. A summary of the purification scheme used in the partial purification of hagfish prothrombin is presented in Figure 16. The advent of the polymerase chain reaction eliminated the need to purify hagfish prothrombin for the purpose of cloning the cDNA. However, some biochemical features of hagfish plasma prothrombin were inferred from the results of partial purification studies.

Hagfish plasma contains a protein that when treated with the snake venom Echarin, is capable of cleaving the chromogenic substrate S2238. This protein activity is precipitated with barium salts. When the peak S2238 esterase activity fractions from the second DEAE chromatography are analyzed by SDS-PAGE, one of the major protein components has an Mr~ 90-95,000 (see Figure 17; lane 3). Incubation of peak protein fractions from the second DEAE chromatography with Echarin results in the apparent cleavage of the 90-95 kD protein (see Figure 18; lane 2). Although no N-terminal amino acid sequence was determined for this protein, and the apparent MW of unglycosylated hagfish plasma prothrombin determined from the cDNA sequence is somewhat lower (see below), these observations support the 90-95 kD protein as a candidate for hagfish plasma prothrombin.
Figure 16. Flow Diagram for Hagfish Prothrombin Isolation.

Citrated Blood

Cells → Centrifugation

Citrated Plasma
1) dialyse in 25 mM sodium pyrophosphate, pH 7.5, 0.38 g trisodium citrate/100 mL
2) Chromatography on DEAE-Sephadex

Flow through (discard)

Citrated Plasma
250 mM eluate
1) echarin/S2238
2) pool fractions with peak activity

1) dialyse into 10 mM trisodium citrate, 0.9% NaCl, pH 7.5
2) add BaCl2 to 74 mM
3) centrifuge

Supernatant (discard)

Barium citrate precipitate
1) dialyse into 0.2 M EDTA
2) add saturated ammonium sulphate to 40% saturation
3) centrifugation

Precipitate (discard)

Supernatant
Add saturated ammonium sulphate to 70% saturation

Precipitate
1) dissolve in 50 mM Tris, pH 7.5, 100 mM NaCl
2) DEAE-Sephadex chromatography

Supernatant (discard)

250 mM eluate → SDS-PAGE
Figure 17. SDS-PAGE Analysis of Hagfish Prothrombin.

Analysis of partially purified fractions of hagfish prothrombin. Lane 1, protein molecular weight markers (BRL) from top to bottom 220 kD, 97.4 kD, 68 kD, 43 kD, 29 kD, 18.4 kD, 14.3 kD; Lane 2, fraction from the second DEAE-Sephadex chromatography treated with Echarin; Lane 3, same as lane 2 but without Echarin treatment; Lane 4, Echarin; Lane 5, aliquot of the reconstituted 70% ammonium sulphate fraction treated with Echarin; Lane 6, same as lane 5 but with no Echarin treatment. Echarin treatment prior to electrophoresis: all samples with or without Echarin were incubated at 37°C for 10 min in 50mM Tris, pH7.5, 100mM NaCl. The reactions were stopped by the addition of loading dye and boiling for 2 min. The arrow indicates the position of the putative hagfish plasma prothrombin.
C. ISOLATION OF PROTHROMBIN CDNA FRAGMENTS BY THE PCR

I. Selection of Oligodeoxyribonucleotide Primers for use in the PCR

The oligodeoxyribonucleotide primers used in the polymerase reaction were designed following alignment of the cDNA sequences of human prothrombin (Degen et al., 1983), bovine prothrombin (MacGillivray and Davie, 1984), and chicken prothrombin (Irwin, 1986). The oligodeoxyribonucleotide primers synthesized following this alignment were designated Thx (where x = 3-10) and are presented in Table 2. These oligodeoxyribonucleotide primers were used in combination with primers designed to amplify the 3' end of cDNAs (Frohman et al., 1988) see Table 2 (primers T17XSP and RACE 1) and Figure 19. Amplification of sscDNA from rat, mouse, rabbit, gekko, newt, rainbow trout, and sturgeon with primers Th7 and T17XSP (or RACE 36) produced a fragment of approximately 900 base pairs (bp). Amplification of sscDNA with oligodeoxyribonucleotide primers Th7 and Th4 produced a fragment of approximately 600 bp in length, and amplification with oligodeoxyribonucleotide primers Th10 and T17XSP yielded a fragment approximately 600 bp in length (see Figure 20 Panel A). The oligodeoxyribonucleotide primers listed in Table 2 failed to amplify homologous cDNA fragments from xenopus, dogfish, or hagfish sscDNAs. The primer Th3 used in combination with any of the other Th primers listed in Table 2 failed to amplify homologous cDNA fragments from newt. The cDNA sequence of a number of the Th priming sites and their efficacy in the PCR are presented in Table 4. From Table 4 it appears that both the position of primer/template mismatch as well as the number of
Figure 18. The PCR strategy for the Amplification of cDNA fragments of the B-chain of prothrombin.

The solid line represents the 3' portion of the mRNA of vertebrate prothrombin. The positions and names of the oligodeoxyribonucleotide primers used to amplify portions of the B-chain of vertebrate prothrombin are shown above the mRNA, and are not drawn to scale. The sequence of the oligodeoxyribonucleotide primers are given in Table 2.
Figure 19. Analysis of the PCR Amplification of Vertebrate Prothrombin B Chains.

Panel A: 1% agarose gel stained with EtBr. Lane 1 shows the product produced following amplification of sturgeon sscDNA with oligodeoxyribonucleotide primers Th7 and T17_{XSP}, lane 2 shows the product produced following amplification of sscDNA with oligodeoxyribonucleotide primers Th7 and Th4, and lane 3 shows the product produced following amplification with oligodeoxyribonucleotide primers Th10 and T17_{XSP} (see text for details). The lane designated M contains DNA size markers, the lengths of which are indicated. Panel B: 1% agarose gel stained with EtBr. Amplification of sscDNA from five different vertebrate species using the oligodeoxyribonucleotide primers His 1 and Ser 1 (see Table 2). Lane, 1 Hagfish; Lane, 2 Dogfish; Lane, 3 Sturgeon; Lane, 4 Newt; Lane, 5 Mouse.
Table 4. Primer Mismatches Compatible and Incompatible with Obtaining Sequence.

The sequence of the oligodeoxyribonucleotide primers are shown below each line in the primer sequence column. Sequence identities are indicated with a period (.) and sequence differences are as indicated. The + symbol indicates product was obtained in the PCR using all of the compatible primers shown in Figure 18. The - symbol indicates no product was obtained in the PCR using any of the compatible primers shown in Figure 18.
<table>
<thead>
<tr>
<th>Species</th>
<th>Primer Sequence (5' - 3')</th>
<th>Amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Th 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>GAGCTGCTGTGGGGCCAGCCTC</td>
<td>T...A...</td>
</tr>
<tr>
<td>Rat</td>
<td></td>
<td>A...T...</td>
</tr>
<tr>
<td>Rabbit</td>
<td></td>
<td>C.C...T...</td>
</tr>
<tr>
<td>Chicken</td>
<td></td>
<td>T...</td>
</tr>
<tr>
<td>Gekko</td>
<td>..CT...T...</td>
<td></td>
</tr>
<tr>
<td>Newt</td>
<td>..CA...A.A.A.A...T...A...</td>
<td></td>
</tr>
<tr>
<td>Rainbow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sturgeon</td>
<td>..T...A...</td>
<td></td>
</tr>
<tr>
<td>Hagfish</td>
<td>.GAA...C...T...A.TT.G...</td>
<td></td>
</tr>
<tr>
<td><strong>Th 4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>GGCTTGTAACCACAGCACAAGAC</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>AC...A.G...</td>
<td></td>
</tr>
<tr>
<td>Rabbit</td>
<td>AC...A.G...</td>
<td></td>
</tr>
<tr>
<td>Chicken</td>
<td>..AC...G...A...</td>
<td></td>
</tr>
<tr>
<td>Gekko</td>
<td>..AC...C..T...</td>
<td></td>
</tr>
<tr>
<td>Newt</td>
<td>..T...A...T...</td>
<td></td>
</tr>
<tr>
<td>Rainbow</td>
<td>..T...A.G...</td>
<td></td>
</tr>
<tr>
<td>Sturgeon</td>
<td>..GC...A.T...T...</td>
<td></td>
</tr>
<tr>
<td>Hagfish</td>
<td>..TGAA...T...A...</td>
<td></td>
</tr>
<tr>
<td><strong>Th 10</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>AAGGCGGTGTGACTGGCTGGGG</td>
<td>A...</td>
</tr>
<tr>
<td>Rat</td>
<td>.A...</td>
<td></td>
</tr>
<tr>
<td>Rabbit</td>
<td>.A...G...A...</td>
<td></td>
</tr>
<tr>
<td>Chicken</td>
<td>.A...T...</td>
<td></td>
</tr>
<tr>
<td>Gekko</td>
<td>.A...A.T...T...</td>
<td></td>
</tr>
<tr>
<td>Newt</td>
<td>.A...T...T.C...</td>
<td></td>
</tr>
<tr>
<td>Rainbow</td>
<td>.A...T...</td>
<td></td>
</tr>
<tr>
<td>Sturgeon</td>
<td>..T...</td>
<td></td>
</tr>
<tr>
<td>Hagfish</td>
<td>.A.T.A.T...</td>
<td></td>
</tr>
</tbody>
</table>
mismatches affect the ability of a given oligodeoxyribonucleotide to act as a primer for Taq polymerase.

3. Sequence Determination of Prothrombin cDNA fragments

Initially, prothrombin cDNA fragments were identified by determining the sequence of the amplified cDNA fragments directly from the PCR reaction, (Gyllensten and Erlich, 1988; see Materials and Methods) and alignment of the DNA and predicted amino acid sequences with human prothrombin (Degen et al., 1983), bovine prothrombin (MacGillivray and Davie, 1984) and chicken prothrombin (Irwin, 1986). The identification of prothrombin cDNAs was based on the presence of the tetrapeptide sequence Tyr-Pro-Pro-Trp in the predicted amino acid sequence of cDNAs. This tetrapeptide sequence represents a portion of the B loop, unique to the B chain of thrombin (Furie et al., 1982). The complete DNA sequence was determined following the cloning of cDNA fragments into pUC vectors (see Materials and Methods).

When the DNA sequence determination was completed, the portion of the B chain corresponding to residues 344-579 (human prothrombin numbering) of prothrombin as well as the 3' UTS had been determined for the rat, mouse, rabbit, gekko, newt, rainbow trout, and sturgeon. The cDNA sequences, predicted amino acid sequences, and restriction maps for all seven species are presented in the Appendix, panels A-G. With the exception of the mouse and rat B-chains, the 3' UTS regions show no significant nucleotide sequence identity among any of the species compared excluding the polyadenylation consensus sequence AATAAA. While this study was in progress the cDNA sequences of rat (Dihanich and Monard, 1990) and mouse (Degen et al., 1990) prothrombin were published. The sequences determined by PCR in this study are identical to these published
sequences. These results suggest that the error rate for *Taq* polymerase may be lower than previously reported (Tindall and Kunkel, 1988). Furthermore, when PCR generated cDNA fragments were re-amplified, for a total of 70 cycles, the number of nucleotide substitutions observed was less than 1/500 bases sequenced (results not shown).

**D. ISOLATION OF A HAGFISH PROTHROMBIN CDNA FRAGMENT BY THE PCR.**

1. **Selection of oligodeoxyribonucleotide Primers for use in the PCR**

   As discussed above, the *Th* oligodeoxyribonucleotide primers listed in Table 2 failed to amplify cDNA fragments from xenopus, dogfish or hagfish *sscDNA*. A cDNA fragment of hagfish prothrombin was identified following amplification of hagfish *sscDNA* with degenerate oligodeoxyribonucleotide primers containing deoxyriboinosine in the third position of 4 fold degenerate codons. The degenerate oligodeoxyribonucleotide primers used in the amplification were adapted from Sakanari et al. (1989) and are based on the amino acid consensus sequence of the active-site regions of trypsin-like serine proteases. These oligodeoxyribonucleotide primers (Ser 1 and His 1; see Table 2) were selected following alignment of amino acid sequences surrounding the active-site serine and aspartic acid residues of a number of trypsin-like serine proteases (Sakanari et al., 1989). Amplification of *sscDNA* using primers Ser 1 and His 1 produces a fragment approximately 600 bp in length in all the species tested (see Figure 20, panel B).
2. Sequence Determination of Hagfish Prothrombin cDNA Fragments.

As the products of amplification with Ser1 and His1 contain more than one type of serine protease cDNA (Sakanari et al., 1989), the products of the PCR were cloned in pUC vectors (see Material and Methods) and their DNA sequence determined. A fragment of the hagfish prothrombin cDNA was identified by searching the predicted amino acid sequences of several clones for the presence of the B loop sequence (Tyr-Pro-Pro-Trp, see above).

E. ISOLATION OF CDNA CLONES FOR HAGFISH PROTHROMBIN

1. Analysis of the Hagfish Prothrombin mRNA

The size of the hagfish mRNA for prothrombin was determined by Northern blot analysis. When immobilized hagfish poly A+ RNA was hybridized with a 32P-labeled cDNA fragment isolated by the PCR, a single transcript of about 2600 nucleotides was detected (data not shown).

2. Screening Hagfish cDNA libraries

Initially, cDNA clones were isolated using the hagfish prothrombin cDNA fragment generated by the PCR as a hybridization probe (see above). From the initial 250,000 oligo dT primed hagfish clones screened, three positive λ phage were plaque purified. One of the three positive phage, λ ESII2, contained a 1.4 kbp insert. The sequence of λ ESII2 was determined. Analysis of the predicted amino acid sequence of λ ESII2 revealed an open reading frame of approximately 0.95 kbp encoding the entire thrombin B-chain and
partial A-chain sequences. In addition to the protease region this clone also contained approximately 0.4 kbp of 3' untranslated sequence (UTS).

3. Isolation of Longer Hagfish Prothrombin cDNAs

In order to obtain the remainder of the hagfish prothrombin cDNA, a specifically primed library was prepared. The 5' end of the hagfish prothrombin cDNA was obtained by screening a specifically primed cDNA library using the 5' end of pESII2 as a hybridization probe (see Figure 21). A total of 10 positive λ phage were plaque purified, and DNA was prepared. One of the 10 positive phage, λESII40;3 contained a 1.2 kbp insert. The DNA sequence of λESII40;3 was determined. The combined sequences of AESII and λESII40;3 yield 2.4 kbp of contiguous cDNA sequence consisting of 30 bp of 5' UTS, an open reading frame of 1881 bp representing 627 amino acids of hagfish prothrombin, and 407 bp of 3' UTS. The combined sequence of λESII2 and λESII40;3 is presented in Figure 22. The restriction map of the hagfish prothrombin cDNAs, and the cloning strategy used to isolate these cDNAs is presented in Figure 21.

F. IDENTIFICATION OF PROTHROMBIN HOMOLOGOUS SEQUENCES IN INVERTEBRATES

1. Selection of Primers for use in the PCR

To examine invertebrate species for the presence of a prothrombin homologous cDNA, 3 degenerate oligodeoxyribonucleotide primers were prepared. All three primers chosen are located in the thrombin B chain and were based on the consensus sequence of vertebrate thrombin cDNA sequences. One primer sequence was based on the amino acid
Figure 20. Restriction map and cDNA cloning strategy for hagfish prothrombin.

The length of the combined cDNA sequences is indicated by the thick box and the positions of restriction endonuclease sites are indicated above. The solid portion of the box represents the coding region, the open portion the 5' UTS, and the shaded portion the 3'UTS. The arrow indicates the position and primer orientation of the cDNA synthesis primer. The cross-hatched boxes below the thick box denote the lengths and positions of hybridization probes. The positions, and names of the isolated cDNA fragments are indicated below the restriction map. The scale is indicated above the restriction sites.
Figure 21. The cDNA Sequence and Predicted Amino Acid Sequence of Hagfish Prothrombin.

The open arrows indicate the positions of the putative signal peptidase cleavage site, and the position of the putative propeptidase cleavage site (a.a -18,-17; -1,+1 respectively) and the solid arrows indicate the positions of the putative factor Xa cleavage sites. The Asn residues of potential glycosylation sites are boxed, and the active site His, Asp , and Ser residues and are in bold text. The RGD peptide sequences are in bold type. The boxed amino acid and cDNA sequence represents the unique insertion sequence in hagfish prothrombin. The polyadenylation consensus sequence AATAAA is in bold type and the polyadenylation addition site indicated by a solid triangle. The position and length of the direct repeats in the 3' UTS are indicated by arrows.
sequence of the B loop (B-loop, see Table 2), and the other 2 were based on amino acid
sequences from the N-terminal and C-terminal regions of the thrombin B chain (FII1 and
FII2, see Table 2).

2. Amplification of sscDNA from Amphioxus, Tunicate, and Sea Star.

The B loop primer was used in combination with the His1 primer described above.
Amplification of sscDNA samples from amphioxus, sea star and tunicate with the B loop/
His1 primer combination failed to produce any product. Amplification was successful
from vertebrate sscDNA samples prepared from liver RNA (data not shown). Similarly,
amplification of invertebrate sscDNA samples with the FII1/FII2 primer combinations
failed to produce any product, but were successful in the amplification of vertebrate
sscDNA samples. The Ser1/His1 primer combination was successful in amplifying cDNA
fragments from sea star and tunicate. However, due to the inherent difficulty of identifying
possible prothrombin homologues in the apparent absence of the B-loop sequence, DNA
sequence determination of these fragments was not pursued.

G. AMINO ACID SEQUENCE ALIGNMENTS OF VERTEBRATE THROMBIN B
CHAINS.

The predicted amino acid sequences of the rat, mouse, rabbit, gekko, newt,
rainbow trout, sturgeon, and hagfish thrombin B chains are aligned with the sequences of
the human (Degen et al., 1983), bovine (MacGillivray and Davie, 1984), and chicken
(Irwin, 1986) thrombin B chains in Figure 23. The length of this portion of the B chain is
relatively invariant; only a single amino acid insertion is present in some species at position
Figure 22. Amino acid sequence alignment of vertebrate thrombin B-chain sequences.

All of the aligned amino acid sequences were predicted from the translation products of the cDNA sequences. Amino acid sequence numbering is based on human prothrombin (Degen et al., 1983). Residues identical to the human sequence are indicated by a dash. The (*) above the human sequence identifies identical amino acid residues in all eleven species. Amino acid substitutions are as indicated and the deletion at position 472 in some species is marked by an *. The active-site His, Asp, and Ser residues are marked by the narrow boxes. Arrows above the human sequence indicate the positions of Arg 382, Arg 418, and Gly 558 (see section IV A.1). For descriptions of residues bounded by boxes A, B, C, D and E, see section IV A.1. The regions marked by dashed lines and designated CR or VR refer to positions representing either conserved (CR) or variable (VR) regions as described by Furie et al. (1982).
472 (Figure 23, box D). The amino acid residue at position 472 is absent in chicken, newt, gekko, rainbow trout, sturgeon, and hagfish thrombin. In addition to the variability at position 472, there is variability in the composition and length of the C-terminal end. The amino acid sequence surrounding the active-site residues His-363, Asp-419, and Ser-525 are conserved in all 11 species (see Figure 23). All of the eight tryptophan and seven cysteine residues seen in the human B chain sequence are conserved in each of the species examined. Ten of the twelve prolines are also conserved. When one aligns the 240 amino acids of the B-chain from the eleven different species, there is amino acid sequence variation at 130 positions (54%). The majority of amino acid sequence variation can be attributed to conservative changes and/or changes occurring in the variable regions (VR, see Figure 23.).

H. DNA SEQUENCE ALIGNMENTS OF VERTEBRATE THROMBIN B CHAINS.

The DNA sequences of the rat, mouse, rabbit, gekko, newt, rainbow trout, sturgeon, and hagfish thrombin B chains are aligned with the DNA sequences of the human (Degen et al., 1983), bovine (MacGillivray and Davie, 1984), and chicken (Irwin, 1986) thrombin B chain sequences in Figure 24. The sequences are compared with the human thrombin B chain cDNA sequence. The majority of nucleotide substitutions are found in the third position of codons and appear to be distributed randomly throughout the B chain. In addition, a large number of codons have two or more changes and in many instances there are codons with no sequence identity with the human cDNA sequence. When the frequencies of nucleotide substitutions are plotted, several features are evident (see Figure 25). With few exceptions the frequency of nucleotide substitutions at the third positions of
Figure 23. The DNA Sequence Alignment of Vertebrate Thrombin B Chains with Human Prothrombin.

The DNA sequence of the B chain of thrombin from bovine (MacGillivray and Davie, 1984), mouse, rat, rabbit, chicken (Irwin, 1986), gekko, newt, rainbow trout, sturgeon, and hagfish are compared to the human sequence (Degen et al., 1983). The amino acid sequence for human prothrombin is shown above the nucleotide sequence of human prothrombin. The active site His, Asp, and Ser residues are in bold text. The regions marked by dashed lines and designated CR or VR refer to positions representing either conserved (CR) or variable (VR) regions as described by Furie et al. (1982).
<table>
<thead>
<tr>
<th>Animal</th>
<th>Species</th>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Sequence 3</th>
<th>Sequence 4</th>
<th>Sequence 5</th>
<th>Sequence 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>GAC</td>
<td>AGC GGG GGA CCC TTT</td>
<td>AGG GGG GGA CCC TTT</td>
<td>AGG GGG GGA CCC TTT</td>
<td>AGG GGG GGA CCC TTT</td>
<td>AGG GGG GGA CCC TTT</td>
<td>AGG GGG GGA CCC TTT</td>
</tr>
<tr>
<td>Bovine</td>
<td>C</td>
<td>C C C</td>
<td>C C C</td>
<td>C C C</td>
<td>C C C</td>
<td>C C C</td>
<td>C C C</td>
</tr>
<tr>
<td>Mouse</td>
<td>A</td>
<td>A T T</td>
<td>A T T</td>
<td>A T T</td>
<td>A T T</td>
<td>A T T</td>
<td>A T T</td>
</tr>
<tr>
<td>Rat</td>
<td>T</td>
<td>T A C</td>
<td>T A C</td>
<td>T A C</td>
<td>T A C</td>
<td>T A C</td>
<td>T A C</td>
</tr>
<tr>
<td>Rabbit</td>
<td>G</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
</tr>
<tr>
<td>Chicken</td>
<td>C</td>
<td>C T A</td>
<td>C T A</td>
<td>C T A</td>
<td>C T A</td>
<td>C T A</td>
<td>C T A</td>
</tr>
<tr>
<td>Gekko</td>
<td>C</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
</tr>
<tr>
<td>Newt</td>
<td>T</td>
<td>T A C T</td>
<td>T A C T</td>
<td>T A C T</td>
<td>T A C T</td>
<td>T A C T</td>
<td>T A C T</td>
</tr>
<tr>
<td>Rainbow Tr</td>
<td>G</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
<td>G A T</td>
</tr>
<tr>
<td>Sturgeon</td>
<td>C</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
</tr>
<tr>
<td>Hagfish</td>
<td>C</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
<td>C T A T</td>
</tr>
</tbody>
</table>

---

| Human        | GAC     | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT |
| Bovine       | A       | A T T               | A T T               | A T T               | A T T               | A T T               | A T T               |
| Mouse        | A        | A T T               | A T T               | A T T               | A T T               | A T T               | A T T               |
| Rat          | T        | T A C               | T A C               | T A C               | T A C               | T A C               | T A C               |
| Rabbit       | G        | G A T               | G A T               | G A T               | G A T               | G A T               | G A T               |
| Chicken      | C        | C T A               | C T A               | C T A               | C T A               | C T A               | C T A               |
| Gekko        | C        | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             |
| Newt         | T        | T A C T             | T A C T             | T A C T             | T A C T             | T A C T             | T A C T             |
| Rainbow Tr   | G        | G A T               | G A T               | G A T               | G A T               | G A T               | G A T               |
| Sturgeon     | C        | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             |
| Hagfish      | C        | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             |

---

| Human        | GAC     | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT | AGT GGG GGA CCC TTT |
| Bovine       | A       | A T T               | A T T               | A T T               | A T T               | A T T               | A T T               |
| Mouse        | A        | A T T               | A T T               | A T T               | A T T               | A T T               | A T T               |
| Rat          | T        | T A C               | T A C               | T A C               | T A C               | T A C               | T A C               |
| Rabbit       | G        | G A T               | G A T               | G A T               | G A T               | G A T               | G A T               |
| Chicken      | C        | C T A               | C T A               | C T A               | C T A               | C T A               | C T A               |
| Gekko        | C        | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             |
| Newt         | T        | T A C T             | T A C T             | T A C T             | T A C T             | T A C T             | T A C T             |
| Rainbow Tr   | G        | G A T               | G A T               | G A T               | G A T               | G A T               | G A T               |
| Sturgeon     | C        | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             |
| Hagfish      | C        | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             | C T A T             |
Figure 24. Frequency of Nucleotide Substitutions in Vertebrate Thrombin B Chains.

The number of nucleotide substitutions is plotted against a moving window of 5 amino acids for the B chain of vertebrate thrombins. The number of nucleotide substitutions is based on comparisons with the human sequence. A represents the number of nucleotide substitutions at the first position of codons, B represents the number of nucleotide substitutions at the second position of codons, C represents the number of nucleotide substitutions at the third position of codons (where both transitions and transversions have been scored). I represents a block diagram of the CR and VR regions of the B chain, where the shaded regions represent CR regions and the open regions represent VR regions (Furie et al., 1982). II represents (from left to right) the B loop, Hirudin binding site, and the thrombomodulin binding domain.
codons appears to be higher in the CR vs the VR regions of the B chain (Furie et al., 1982) see Figure 23 and Figure 24. Furthermore, the frequency of substitutions at the first and second positions of codons is far lower than that at the third position, and occurs at the much higher frequency in the VR regions. Taken together these results suggest that the areas of the thrombin B chain corresponding to surface loops (VR regions) are evolving more rapidly than the regions of the thrombin B chain required for the core structural features of the protease. The sole exception appears to be the surface loop making up the B loop. The B loop of vertebrate thrombins is highly conserved at the amino acid level and this is reflected in the frequency and type of substitutions observed in this region (see Figures 23, 24, 25).

The length of the 3' UTS and B chain coding region is summarized in Table 5. While the length of the coding sequence is similar in all the species examined (see below for the exceptions), the length of the 3' untranslated sequence (UTS) is quite heterogeneous. The length of the 3' UTS varies from 82 nucleotides (nts) in the trout and sturgeon to 1145 nts in chicken, with the majority approximately 80-100 nts. (see Table 5)
Table 5. Length of thrombin B-chain coding and 3' untranslated (UTS) sequences (nts) for the nine species

<table>
<thead>
<tr>
<th>Species</th>
<th>coding sequence</th>
<th>3' UTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rat</td>
<td>705</td>
<td>115</td>
</tr>
<tr>
<td>Mouse</td>
<td>705</td>
<td>114</td>
</tr>
<tr>
<td>Rabbit</td>
<td>705</td>
<td>121</td>
</tr>
<tr>
<td>Chicken</td>
<td>702</td>
<td>154/1145</td>
</tr>
<tr>
<td>Gekko</td>
<td>705</td>
<td>224</td>
</tr>
<tr>
<td>Newt</td>
<td>705</td>
<td>108</td>
</tr>
<tr>
<td>Rainbow Tr.</td>
<td>717</td>
<td>82</td>
</tr>
<tr>
<td>Sturgeon</td>
<td>702</td>
<td>82</td>
</tr>
<tr>
<td>Hagfish</td>
<td>708</td>
<td>407</td>
</tr>
</tbody>
</table>

* Two transcripts are observed for chicken prothrombin.
IV. DISCUSSION

A. ANALYSIS OF VERTEBRATE PROTHROMBIN B CHAINS.

1. Comparisons of the B Chain Amino Acid Sequence From Eleven Vertebrate Species.

Comparisons of amino acid sequences of rat, mouse, rabbit, gekko, newt, rainbow trout, sturgeon and hagfish with previously characterized portions of human, bovine, and chicken prothrombin suggest that this portion of prothrombin is highly conserved throughout vertebrate evolution (see Table 6.). Amino acid sequence identities range from 96.5% (between mouse and rat) to 62.6% (between newt and hagfish). Thrombins from mammalian species share greater than 82% amino acid sequence identity. The percent identity is much lower among the non-mammalian species (66% average). Rainbow trout and sturgeon share 82.5% amino acid sequence identity; this is 10% more identity than either shares with any of the other nine species. This high degree of amino acid sequence identity may reflect a slower rate of nucleotide sequence change in fish or constraints on the B-chain through interactions with other proteins involved in hemostasis. The overall amino acid sequence identity in this region of prothrombin is 43.3% among the eleven vertebrates; when conservative changes are included, the overall amino acid sequence similarity increases to 74.6%.

The crystal structure for thrombin (Bode et al., 1989) demonstrates that the B-chain displays a similar polypeptide fold to chymotrypsinogen, with almost half of the approximately 200 spatially equivalent residues being similar or identical in character. The
Table 6. Percent Amino Acid Sequence Identity Among Vertebrate Thrombin B-chains *

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>M</th>
<th>Rbt</th>
<th>B</th>
<th>C</th>
<th>G</th>
<th>N</th>
<th>RTr</th>
<th>S</th>
<th>Hgf</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN</td>
<td>88.5</td>
<td>89.8</td>
<td>85.1</td>
<td>87.3</td>
<td>72.2</td>
<td>72.8</td>
<td>70.2</td>
<td>68.6</td>
<td>65.0</td>
<td>65.3</td>
</tr>
<tr>
<td>RAT</td>
<td>96.5</td>
<td>82.6</td>
<td>86.0</td>
<td>71.4</td>
<td>73.2</td>
<td>68.5</td>
<td>68.5</td>
<td>65.5</td>
<td>63.8</td>
<td></td>
</tr>
<tr>
<td>MOUSE</td>
<td>83.0</td>
<td>86.4</td>
<td>71.4</td>
<td>73.2</td>
<td>68.5</td>
<td>69.4</td>
<td>65.0</td>
<td>64.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RABBIT</td>
<td>82.6</td>
<td>73.1</td>
<td>71.9</td>
<td>68.9</td>
<td>66.4</td>
<td>64.1</td>
<td>64.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOVINE</td>
<td>72.2</td>
<td>72.3</td>
<td>68.5</td>
<td>68.8</td>
<td>65.0</td>
<td>64.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHICKEN</td>
<td>77.8</td>
<td>73.5</td>
<td>72.2</td>
<td>65.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEKKO</td>
<td>71.1</td>
<td>70.2</td>
<td>70.1</td>
<td>69.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEWT</td>
<td>69.4</td>
<td>69.2</td>
<td>62.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAINBOW TROUT</td>
<td>82.5</td>
<td>68.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STURGEON</td>
<td>66.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Percent identity data was generated with PALIGN (Intelligenetics) using the structure-genetic matrix, open gap=5; unit gap=50. When these data are plotted as a phylogeny, using a difference matrix, the results are consistent with the present evolutionary origin of birds from a reptilian ancestor.

R, rat; M, mouse; Rbt, rabbit; B, bovine; C, chicken; G, gekko; N, newt; RTr, rainbow trout; S, sturgeon; Hgf, hagfish.
specificity of thrombin for its substrates is likely determined by its insertion loops (Fenton, 1988; Furie et al., 1982; Bode et al., 1989). The locations of these insertion loops are indicated in Figure 23 and correspond to the VR regions defined by Furie et al. (Furie et al., 1982). A number of the loop segments are similar to those of chymotrypsin but contain additional residues which alter the overall three dimensional structure of the region (Bode et al., 1989). Several of the loop structures in the B chain of thrombin have been implicated in interactions with substrate (Figure 22 boxes A, C, and the first 4 residues of box D) and as sites of interaction with hirudin (Figure 22 box B) (Grutter et al., 1990; Rydel et al., 1990) and thrombomodulin (box D) (Susuki et al., 1990). It is interesting to note that the overall structural features of the B chain have been conserved throughout vertebrate evolution. However, with the exception of the B-loop, corresponding to residues 365-369 (box A in Figure 22.) and the loop highlighted by box C (Figure 22), all other loop structures have highly variable amino acid sequences. This variability in surface loop amino acid sequence may contribute to some of the species specific differences observed between thrombin and fibrinogen (Doolittle, 1965; Doolittle et al., 1962; Ratnoff, 1987) as well as between thrombin and thrombomodulin (Miyata et al., 1987).

The human thrombin B chain contains a region with growth factor and chemotactic activity (Bar-Shavit et al., 1984; Bar-Shavit et al., 1986). This chemotactic/growth factor domain corresponds to residues Leu 335 - Met 400 of prothrombin (see Figure 22). In addition to the chemotactic/growth factor activity attributed to this region, residues 335-400 also contains the B-loop (box A, Figure 22.), active-site histidine (His 363) and a hirudin-binding loop (box B, Figure 22). With the exception of the hirudin binding loop, this portion of thrombin contains very few amino acid sequence differences among the species examined. Analysis of the thrombin crystal structure has identified the B-loop (box A,
Figure 22) as forming an extended loop structure that restricts access to the active-site cleft and is likely responsible for the limited substrate specificity of thrombin (Bode et al., 1989). Due to the rigid-kinked structure of the B-loop, it has been suggested that this loop may also represent part of the active substructure of the chemotactic/growth factor domain (Fenton, 1986; Bode et al., 1989); see Figure 25. The degree of amino acid sequence similarity in this region (28/56 identical residues, 8/56 conservative changes) suggests that the chemotactic/growth factor activity attributed to this region in human thrombin may have a similar activity in other vertebrate thrombins. In addition to the role of the B-loop in substrate specificity and chemotactic/growth factor activity, this portion of the B chain contains a site for N-linked glycosylation (Asn 373, see Figure 22). The asparagine residue at position 373 is conserved in all of the eleven species. Removal of the carbohydrate moiety has no effect on thrombin activity in vitro (Bar-Shavit et al., 1984), suggesting that Asn 373 may be required more for its overall geometrical contribution rather than addition of carbohydrate moieties. Residues Lys 385 - Glu 396 (box B, Figure 22.) represent one of the surface contact loops of hirudin with human thrombin (Grutter et al., 1990; Rydel et al., 1990).

Interestingly, only 3 of the 12 residues in this region are identical or have conservative changes: Arg 388, Arg/Lys 390 (except hagfish and newt) Arg/Lys 393, and Lys 397 (except hagfish) see Figure 22, box B. Analysis of the crystal structure of the thrombin-hirudin complex reveals a number of possible electrostatic interactions between these residues and the acidic C-terminal residues of hirudin (Grutter et al., 1990). The location of the hirudin binding loop in the model of the B chain of human thrombin is shown in Figure 26.
Figure 25. The Location of the B Loop in the Model of the Human Thrombin B Chain.

The amino acid residues representing the B loop (Tyr 367-Phe 374) are shown in red. The active-site groove runs from left to right through the centre of the structure.
Figure 26. The Location of an Hirudin Binding Site in the Model of the Human Thrombin B Chain.

The amino acid representing the hirudin binding domain (Lys 385-Glu 396) are shown in orange. The active-site groove runs from left to right through the centre of the structure.
A thrombomodulin binding site has recently been localized in human thrombin (Suzuki et al., 1990). The binding site corresponds to residues Thr 468 - Ser 478 (box D, Figure 22) where the surface loop formed by Glu 467 - Tyr 470 is located at the entrance of the binding cleft (Figure 27). The interaction of thrombomodulin with thrombin has the effect of altering the substrate specificity of thrombin allowing thrombin to activate protein C. In this way, thrombomodulin changes thrombin from a procoagulant to an anticoagulant. The thrombomodulin binding region (Figure 22, box D) is highly variable with only 2 of the 12 amino acid residues being invariant (Trp 468 and Pro 477). The majority of the amino acid sequence changes in this region are nonconservative. In addition this is the only region of the B-chain to contain an amino acid deletion.

A recent study by Hirahara and co-workers (Hirahara et al., 1990) reported a species specific variability in the ability of thrombomodulin to act as an inhibitor of thrombin inactivation by antithrombin III. These investigators reported that human thrombomodulin acts as a weak competitive inhibitor of thrombin inactivation by antithrombin III while thrombomodulin purified from rabbit lung accelerates thrombin inactivation. Furthermore, they suggest that this difference is likely the result of species specific differences in the ability of thrombomodulin to bind directly to antithrombin III reflecting a species specific variation in the regulation of the coagulation cascade via protein C. An alternative explanation for the species variability in the inhibition of thrombin by antithrombin III is suggested by our data. The region corresponding to the thrombomodulin binding site varies significantly among the eleven species. The species specific phenomenon observed (Hirahara et al., 1990) may be a reflection of the ability of rabbit thrombomodulin to bind to this region of human thrombin. This decreased binding ability of rabbit thrombomodulin for human
Figure 27. The Location of the Thrombomodulin Binding Domain in the Model of the Human Thrombin B Chain.

The amino acids representing the thrombomodulin binding domain (Thr 468- Ser 478) are shown in orange. The thrombomodulin domain includes a portion of the Trp 486 loop shown in Figure 7.
thrombin would result in an overall increase in the inactivation of thrombin by antithrombin III as observed (Hirahara et al., 1990). At present there is little evidence that any thrombomodulin/thrombin interaction occurs in non-mammalian vertebrates if indeed any of these animals express a thrombomodulin-like molecule. Therefore the variation seen in the thrombomodulin binding region of non-mammalian vertebrates may be a reflection of genetic drift rather than species specific variability in the regulation of this part of the coagulation cascade.

Human thrombin contains an RGD tripeptide sequence at position 517-520 (box E, Figure 22.) analogous to the adhesion site in adhesive proteins such as laminin, fibronectin, and fibrinogen (Bar-Shavit et al., 1991). This region of thrombin has been shown to promote endothelial cell (EC) adhesion, spreading, and cytoskeletal reorganization potentially contributing to repair mechanisms and maintenance of the internal blood vessel lining (Bar-Shavit et al., 1991). Although the cell adhesion activity was greatest with a chemically modified form of thrombin (NO₂-α-thrombin), native thrombin incubated at 37°C was also found to promote EC adhesion. While a portion of the loop segment Tyr 509 to Gly 518 is exposed to the solvent, the tripeptide RGD is not (Bode et al., 1989). The RGD sequence is conserved in nine of the eleven species, supporting a possible in vivo role for the RGD sequence in vertebrate thrombins. The fact that this sequence is absent in rainbow trout and sturgeon is intriguing. It would be interesting to know if the RGD associated function is absent or replaced by some other mechanism in fishes.

The C-terminal region of thrombin is variable in both composition and length. Within the C-terminal ten residues of human thrombin, only Trp 569 and Lys 572 are conserved in all of the species compared. According to the crystal structure, the C-
terminus of human thrombin is exposed to the solvent (Bode et al., 1989); however, no interactions or functions have been assigned to this region.

A number of abnormal human prothrombins have been characterized and the molecular defects identified. Three of these are found in the B-chain. Prothrombin Quick I (Henriksen and Mann, 1988) is characterized by an Arg to Cys change at residue 382 (between boxes A and B, Figure 22.). The Arg residue is conserved in all eleven species. This amino acid has been identified from the 3-dimensional structure as one of the residues lining the long groove extending from the active site and may form part of the putative fibrinogen secondary binding site (Bode et al., 1989): see Figure 28. Substitution of Cys for Arg at this position probably disrupts thrombin/fibrinogen interactions. In Prothrombin Tokushima, the Arg at position 418 is replaced by a Trp (Miyata et al., 1987). This Arg residue is also conserved in all eleven species. Arg 418 is adjacent to box C (Figure 22.) which forms one of the surface loops projecting out from the active site cleft (Bode et al., 1989) see Figure 28. It may be the nature of the amino acid change at position 418 that leads to the decreased enzyme efficacy observed in Prothrombin Tokushima. A Gly to Val substitution at position 558 is found in Prothrombin Quick II (Henriksen and Mann, 1989). Gly 558 is adjacent to the Cys 551 to Tyr 557 loop segment (Bode et al., 1989) and is conserved in all eleven species. Substitution of Val for Gly appears to alter the primary substrate binding pocket (Henriksen and Mann, 1987). However, in the model of the human thrombin B chain (Yue et al., 1991), this Gly residue appears to be buried and is not exposed to the active-site groove. Thus, it may be that the Gly-Val substitution at position 558 alters the polypeptide fold in this region, disrupting interactions with substrate.
The location of the Quick I and Tokushima mutants in the B chain of thrombin are indicated by the labels and are coloured orange. An Arg - Cys change at position 382 results in the prothrombin Quick I mutation, and an Arg - Trp change at position 418 results in the prothrombin Tokushima mutation.
B. ANALYSIS OF VERTEBRATE PROTHROMBINS

1. Characterization of cDNAs For Chicken and Hagfish Prothrombin.

The predicted amino acid sequences of both chicken and hagfish prothrombin suggest they are synthesized as prepro-proteins consisting a signal peptide, a propeptide, Gla domain, two kringles, and a two chain protease domain. Thus, with few exceptions (see below) both chicken and hagfish prothrombin are structurally identical to mammalian prothrombins (Degen et al., 1983; MacGillivray and Davie, 1984; Dihanich and Monard, 1990; Degen et al., 1990).

Two transcripts were detected for chicken prothrombin which differ only in the length of their 3' UTS (see Figure 13). The largest chicken prothrombin transcript is the least abundant of the two and represents about 10% of the prothrombin message found in the liver. The 3' UTS of the chicken prothrombin cDNA contains a number of direct repeats, inverted repeats, and one hairpin loop structure. The hairpin loop (nts 2713-2731, see Figure 15) is found only in the 3.2kb transcript, and consists of a 7 base pair (bp) stem and a 5 bp loop. The 5' UTS of the chicken prothrombin cDNA contains a 14 bp inverted repeat (nts 1-29, see Figure 15).

In hagfish liver, RNA a single prothrombin transcript of 2.4 kb was detected on Northern blots (data not shown). From the sequence determined from the cDNA, the transcript contains a 450 nt 3' UTS region consisting of a number of direct repeat
sequences of varying lengths. The longest direct repeats are of 15 nts and 13 nts respectively (see Figure 22).

From Northern blot analysis and the frequency of positive phage in the cDNA libraries, we estimate that the fraction of mRNA encoding chicken prothrombin in liver is approximately the same as for the bovine and human transcripts (0.1% - 1%, MacGillivray and Davie, 1984). In contrast, the relative abundance of prothrombin mRNA in hagfish liver is at least an order of magnitude lower (0.01-0.05%). The decreased prothrombin mRNA levels observed in the hagfish may reflect lower circulating levels of prothrombin in plasma, a long half-life for the protein in plasma, or an inherent stability of the transcript. The presence of multiple tandem repeat sequences in the 3' UTS of the hagfish prothrombin mRNA may play a role in this mRNA stability.

There are no significant nucleotide sequence identities in the 5' UTS of the chicken prothrombin and hagfish prothrombin cDNAs. The 5' UTS of the chicken prothrombin cDNA does not conform to the Kozak consensus translation initiation sequence (CCACC, Kozak, 1984). The sequence matches the Kozak consensus at 2/5 positions (ACAGG). However, the ATG at nts 41-43 in the chicken prothrombin cDNA most likely encodes the initiator Met, as there are no in frame or out of frame Met codons preceding it. Furthermore, application of the -1, -3 rule (Von Heijne, 1986) identifies the putative signal peptide cleavage site as being between Ser -21 and His -20 (see Figure 15). The nucleotide sequence adjacent to Met-38 in hagfish prothrombin matches the Kozak consensus sequence at 2/5 positions (AAACA). Although the position of the signal peptidase cleavage site is less clear, application of the -1,-3 rule suggests that cleavage may occur between Ala -18 and Val -17 (see Figure 22).
The chicken prothrombin cDNA predicts an open reading frame of 607 amino acids and the cDNA of hagfish prothrombin predicts an open reading frame of 627 amino acids. Circulating chicken prothrombin is predicted to be 564 amino acids in length with a molecular weight of 64,130; circulating hagfish prothrombin is predicted to 589 amino acids in length with a molecular weight of 67,647. This difference in size is primarily the result of the 19 amino acid residue insertion in hagfish prothrombin in the region between the Gla domain and the first kringle (see below). The glycosylation status of the two prothrombin molecules is unknown at present.

2. Comparison of Vertebrate Prothrombin Amino Acid Sequences

The predicted amino acid sequences of chicken prothrombin and hagfish prothrombin are compared with human prothrombin (Degen et al., 1983), bovine prothrombin (MacGillivray and Davie, 1984), mouse prothrombin (Degen et al., 1990) and rat prothrombin (Dihanich and Monard, 1990) in Figure 30. The overall structure of prothrombin is identical in all six species, consisting of a signal peptide, a propeptide, a Gla domain, two kringles, and a two chain protease domain. Gaps and insertions have been placed to allow for maximum identity with the minimum number of deletions and/or insertions but with retention of common structural features. With the exception of insertions and deletions of one or a few amino acid residues, the only major difference in the primary structures of all six species is the presence of an additional 19 residues in hagfish prothrombin. The additional 19 amino acids in hagfish prothrombin are found in the region between the Gla domain and the first kringle (see Figure 30). A portion of this 19 amino acid insertion may form a 16 amino acid disulfide loop. The region preceding the 19 amino acid insertion in hagfish may also form a disulphide loop (a.a 47-60)
homologous to the 14 amino acid loop found in the other five species compared with respect to size and location. However, the amino acid sequence of the 14 amino acid residue loop is not very well conserved in hagfish prothrombin (see Figure 30). Furthermore, the absence of the Pro residue in the hagfish sequence suggests that the 3 dimensional structure of this loop may be different than in the other five species. In the human (Degen and Davie, 1988) and bovine (Irwin et al., 1988) prothrombin genes, the 14 residue disulfide loop and the region adjacent to it are found in separate exons. In the bovine prothrombin gene, exon 3 contains the sequence DAFWAKYT and exon 4 contains the 14 residue loop and adjacent sequence ACESARNPREKLNECLEG (see Figure 31). While the structure of the chicken and hagfish prothrombin genes have not been determined, the amino acid sequence of chicken prothrombin suggests the gene structure in this region is likely to be similar to that of the human and bovine genes. On the basis of the amino acid sequence alignment shown in Figure 30, the 19 amino acid sequence insertion in hagfish prothrombin corresponds to the junction of exons 3 and 4 of the bovine gene (Irwin et al., 1988) suggesting that this insertion may represent either an additional exon, or an intron sliding event in the hagfish gene (Craik et al., 1983) see Figure 31. The loop formed by the 19 amino acid residue insertion in hagfish prothrombin may be located on the surface of the protein as 5/16 amino acid residues are charged (see Figure 31).

There are 4 potential N-linked glycosylation sites in the sequence Asn-X-Thr/Ser at Asn 79, Asn 101, Asn 118, and Asn 360, in chicken prothrombin (see Figure 15), and in hagfish prothrombin at Asn 68, Asn 95, Asn 198, and Asn 381 (see Figure 22). The potential glycosylation site at Asn 68 is unique to hagfish prothrombin and is found in the 19 amino acid insertion sequence. The potential glycosylation sites at Asn 79 and Asn 95 (for chicken and hagfish prothrombins respectively) are homologous to first glycosylation
Figure 29. Alignment of Prothrombin Amino Acid Sequences.

Alignment of the amino acid sequence human prothrombin with rat, mouse, bovine, chicken, and hagfish prothrombin. The position and length of the various regions of prothrombin are indicated above the human amino acid sequence. The stars (*) above the human sequence indicate the position of identical residues in all six species. Dots (.) represent amino acid sequences identical to the human sequence, and dashes (-) indicate amino acid residue deletions / insertions. Amino acid sequence differences between human prothrombin and prothrombin from other species are as indicated. Amino acid sequence numbering is for human prothrombin (Degen et al, 1983).
site in kringle 1 of human, bovine, rat, and mouse prothrombin. The potential glycosylation site at amino acid 101 in chicken prothrombin is analogous to the second glycosylation site found in kringle 1 of human, bovine, rat and mouse prothrombin. The potential glycosylation site at Asn 118 in chicken prothrombin is not found in any of the other species compared. The potential glycosylation site at Asn 198 in hagfish prothrombin is located in kringle 2; no equivalent site is seen in any of the other species. The potential glycosylation sites at Asn 360 and Asn 381 (for chicken and hagfish prothrombin respectively) are analogous to the glycosylation sites found in the B-loop of human, bovine, mouse, and rat prothrombin.

The cell adhesion sequence RGD found in human, bovine, rat, and mouse prothrombin is conserved in chicken and hagfish prothrombin (Ruoslahti and Pierschbacher, 1986.; Bar-Shavit et al., 1990.) see Figure 23. In addition to the RGD sequence found in the protease domain, hagfish prothrombin has an additional RGD sequence in kringle 2 at amino acids 194 - 196 (see Figure 22). This RGD sequence is adjacent to a potential N-linked glycosylation site at Asn\textsubscript{198}^\textsuperscript{-Ile-Thr}^{200}. Although the crystal structure of kringle 2 has not been determined, the structure of kringle 1 has been solved at 2.8Å resolution (Park and Tulinsky, 1986). By analogy, the equivalent region of kringle 2 is most likely on the surface of the protein. The position of the RGD sequence in kringle 2 of hagfish prothrombin suggests that it may be able to directly interact with any putative receptor(s). The presence of an active RGD sequence at this position in hagfish prothrombin may reflect a variation in the activation of prothrombin in the hagfish.

Pairwise amino acid sequence identities have been determined for the precursor forms of prothrombin from six species and are presented in Table 7. The percent amino acid sequence identities vary from 93.5% (mouse vs. rat) to 50.7% (rat vs. hagfish). The
Figure 30. Potential Cysteine Loop Structures in Fragment 1 of Bovine and Hagfish Prothrombin.

Panel A, 14 amino acid cysteine loop of bovine fragment 1. The position of the intron/exon junctions for exons 3 and 4, and exons 4 and 5 are indicated by the arrows (Irwin et al., 1985). Amino acid sequence numbering is from Degen et al. (1983); Panel B, potential cysteine loops of hagfish fragment 1. The positions of the arrows indicate the putative amino acid sequence insertion with respect to the bovine prothrombin gene (see Panel A).
precursor forms of chicken and hagfish prothrombin share approximately 63% and 52% identity with the other five species (see Table 7). The precursor forms of prothrombin from the six species share identity at 266 positions (41.0%) and similarity at 172 positions (26.5%) for an overall similarity of 67.5%. The most conserved regions of prothrombin among the species compared are the Gla domain (55.9% identity) and the B-chain (54.4% identity) see Table 8. The A-chain shares the lowest amount of amino acid sequence identity (42%). The least conserved portions of prothrombin among the six species compared are the regions connecting the Gla domain and kringle 1, the region connecting kringle 1 and kringle 2, and the region connecting kringle 2 and the A-chain (see Figure 30). The amino acid sequence identities and similarities have been calculated across the six species compared and are presented in Table 8. The amino acid sequence identity/similarity scores suggest that the Gla domain and the thrombin B-chain are the regions most essential for the common function of vertebrate prothrombins (see Table 8). Kringle 1 may play a less essential role, and the connecting regions likely only function to separate the different domains.

Unlike the bovine prothrombin mRNA whose base composition is somewhat GC rich (MacGillivray and Davie, 1984), the base composition of the chicken prothrombin mRNA (30.1% A, 21.4% U, 25.7% G, and 22.6% C) and the hagfish prothrombin mRNA (27.1% A, 22.7% U, 28.4% G, and 21.6% C) are somewhat AG rich. The slight increase in AG content is reflected in the frequency of codons using G in the first position (31.9% in chicken and 32.7% in hagfish) and A in the second (35.6% in chicken and 34% in hagfish). In the chicken prothrombin cDNA 26% of all codons are either AAN or GAN (where N=A,U,C,G), 82% of which encode charged amino acid residues. In chicken prothrombin, 142/608 (23.5%) of amino acids are charged. In the hagfish prothrombin
Table 7. Percent Amino Acid Identities of Prothrombin From Six Species

<table>
<thead>
<tr>
<th></th>
<th>Bovine</th>
<th>Rat</th>
<th>Mouse</th>
<th>Chicken</th>
<th>Hagfish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>81.5%</td>
<td>79.6%</td>
<td>81.2%</td>
<td>62.1%</td>
<td>53.4%</td>
</tr>
<tr>
<td>Bovine</td>
<td>- 76.8%</td>
<td>78.3%</td>
<td>93.5%</td>
<td>63.6%</td>
<td>51.5%</td>
</tr>
<tr>
<td>Rat</td>
<td>-</td>
<td>-</td>
<td>63.3%</td>
<td>51.5%</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>63.9%</td>
<td>52.1%</td>
</tr>
<tr>
<td>Chicken</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>51.6%</td>
</tr>
</tbody>
</table>

a Amino acid sequences were aligned with PALIGN (Intelligenetics) using a unit gap=5, and open gap=10.
Table 8. Amino Acid Sequence Identity and Similarities\textsuperscript{a} for the Domains of Prothrombin\textsuperscript{b}.

<table>
<thead>
<tr>
<th>Region Compared</th>
<th>Identical Amino Acids</th>
<th>Similar Amino Acids</th>
<th>Overall Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propeptide</td>
<td>10/20 (50.0%)</td>
<td>3/20 (15.0%)</td>
<td>65.0%</td>
</tr>
<tr>
<td>Gla</td>
<td>19/34 (55.9%)</td>
<td>8/34 (23.5%)</td>
<td>79.4%</td>
</tr>
<tr>
<td>Kringle 1</td>
<td>34/79 (43.0%)</td>
<td>23/79 (29.1%)</td>
<td>72.1%</td>
</tr>
<tr>
<td>Kringle 2</td>
<td>28/80 (35.0%)</td>
<td>17/80 (21.3%)</td>
<td>56.3%</td>
</tr>
<tr>
<td>A Chain</td>
<td>21/50 (42.0%)</td>
<td>12/50 (24.0%)</td>
<td>66.0%</td>
</tr>
<tr>
<td>B Chain</td>
<td>129/237 (54.4%)</td>
<td>74/237 (31.2%)</td>
<td>85.6%</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Percent identities and similarities were determined with Clustal (Intelligenetics) using an open gap=10 and a unit gap=10. The following groups of residues were defined as similar: A,S,T., D,E., N,Q., R,K., I,L,M, V., and F,Y,W.

\textsuperscript{b}Values are based on the simultaneous alignments of human, bovine, mouse, rat, chicken, and hagfish prothrombin.
cDNA 23.8% of all codons are either AAN or GAN, 78% of which encode charged amino acids. In hagfish prothrombin 168/626 (26.8%) of amino acids are charged. Unlike the mRNAs for human and bovine prothrombins all the Glu residues in the Gla domain of chicken and hagfish prothrombin are not encoded by GAG. In chicken prothrombin 7/10 glutamic acid residues are encoded by GAG and in hagfish prothrombin 5/10 are encoded by GAG. The difference in codon frequency in this region may reflect the codon usage bias in the human and bovine genes where 80% of glutamic acids are encoded by GAG (MacGillivray and Davie, 1984.; Degen et al. 1983). In the chicken and hagfish cDNAs both Glu codons are used with equal frequency. Although the codon usage for Glu acid residues varies among vertebrate prothrombins the number of Glu residues and their distribution in the Gla domain is conserved among the vertebrates compared.

Finally, the active site serine in all six species compared is encoded by either AGT or AGC. The presence of the AGY class of serine codon in the hagfish prothrombin cDNA implies that the events leading to the appearance of this distinct sub-group of serine proteases (Irwin, 1988., Irwin et al., 1988.) must have occurred prior to the divergence of cyclostomes from the vertebrate lineage (over 450 million years ago). The structural similarity of hagfish prothrombin with mammalian prothrombin indicates that the appearance of the propeptide and Gla domain as well as the acquisition of both kringle, all pre-dated the divergence of hagfish from the main vertebrate lineage.
C. THE EVOLUTION OF PROTHROMBIN IN VERTEBRATES

1. Structural Organization of Vertebrate Prothrombin

From the predicted amino acid sequences of chicken and hagfish prothrombin, it is apparent that vertebrate prothrombins share the same general domain organization, and that this organization was complete prior to the divergence of hagfish from the main vertebrate lineage > 400 MYA. Vertebrate prothrombin is synthesized as a prepro-protein consisting of a signal peptide, a propeptide, a Gla domain, two Kringles, and a two chain protease domain (see Figure 30). Amino acid sequence comparisons suggest that the Gla domain and the B chain of thrombin are the regions most essential for common function of vertebrate prothrombins (see Table 8). The presence of both a propeptide and Gla domain in hagfish prothrombin suggests that the vitamin K-dependent γ carboxylation in vertebrates is at least 400 million years old. While Gla residues have been found in invertebrates (Hamilton et al., 1982; Olivera et al., 1990) there is no evidence to suggest that the mechanism of γ carboxylation is analogous to the vitamin K-dependent process observed in vertebrates. Furthermore, the presence of the propeptide and Gla domain in hagfish prothrombin suggests that the post-translational endoproteolytic cleavage of the propeptide has co-evolved along with vitamin K-dependent γ carboxylation during the course of vertebrate evolution. While the protease responsible for the propeptide cleavage has not been identified it is likely an integral membrane protein located in a post-ER compartment related to the subtilisin family of endoproteases of which Kex 2 and furin are members (Mizuno et al., 1988; Fuller et al., 1989a; Fuller et al., 1989b; Bresnahan et al., 1990; van de Ven et al., 1990; van den Ouweland et al., 1990). Hagfish prothrombin, like
chicken and mammalian prothrombins contains two kringles. According to Patthy (1984) the transfer of the kringle domains to prothrombin occurred on two separate occasions. The N-terminal kringle is related to the ancestral kringle of plasminogen (prior to the internal gene duplication events in plasminogen) while the C-terminal kringle is more similar to kringle 2 of plasminogen than to any other kringles (Patthy, 1984) see Figure 10 Panel B. If this is true, the presence of two kringles in hagfish prothrombin suggests that the duplication of the primordial kringle of plasminogen must have occurred prior to the divergence of hagfish from the main vertebrate lineage (> 400 MYA). Thus, one would expect a hagfish plasminogen to contain at least two kringles.

2. Evolution of Vertebrate Thrombin B Chains

Thrombin cDNAs have been identified from a diverse group of vertebrates, supporting the notion that thrombin is the central procoagulant protease in vertebrate blood coagulation (Surgenor and Doolittle, 1962). The vertebrate thrombin B chains characterized share 75% amino acid sequence similarity (see Table 6). Regions of high amino acid sequence similarity include those surrounding the catalytic triad residues (see Figure 23). In addition, the amino acid residues which make up the chemotactic domain, and the B loop are also conserved (see Figure 23), suggesting that in addition to the role of thrombin in coagulation, it may also play a role in the wound healing process of vertebrates. The regions of thrombin which are the least conserved among the vertebrates compared correspond to surface loops (with the exception of the B loop, see Figure 23). Examination of the nucleotide sequence in these regions suggests that they may be evolving more rapidly than the core structural elements of the protease (see Figures 22, 25). This result is particularly intriguing in light of thrombin's role in hemostasis. Thrombin
interacts with a number of macromolecules including fibrinogen, thrombomodulin, and protein C. In addition, thrombin also binds to platelets and endothelial cell surfaces (presumably via the receptor, see Section C.7.c). With the possible exception of fibrinogen these interactions occur via amino acid residues on the surface of the protein (surface loops). The variability in amino acid sequence observed in the surface loop regions of thrombin may explain many of the species-specific interactions observed between thrombin and its substrates (Doolittle and Surgenor, 1962; Walz et al., 1975; Ratnoff, 1987; Hattori et al., 1990). The physiological significance of this phenomenon is unclear.

Attempts to isolate a thrombin homologous sequence from invertebrates have been unsuccessful to date. While the methods used in these studies do not conclusively prove that a thrombin-like molecule is not expressed in the invertebrate species examined, the apparent absence of fibrinogen in these species does support this observation (Doolittle, 1984; Xu and Doolittle, 1990). Furthermore, the little that is known about hemostasis in invertebrates strongly suggests that it bares little resemblance to the mechanisms involved in vertebrate hemostasis (see Section B.2).

D. IDENTIFICATION OF ADDITIONAL COAGULATION FACTORS IN VERTEBRATES

The structural organization of hagfish prothrombin suggests that vertebrate prothrombins are identical in their domain organization and that the events leading to this organization must have occurred > 400 MYA. However, the organization of hagfish prothrombin tells us little about the time scale over which the acquisition of the various domains occurred. Examining the structures of proteins which share domains with
prothrombin may provide insight into: the time scale over which these domain shuffling processes occurred, the complexity of the coagulation process in non-mammalian vertebrates, and structure/function relationships. Of particular interest are those coagulation proteins which comprise the latter stages of the mammalian coagulation cascade such as factors IX, X, and VII. At present little is known about the protease which activates prothrombin in lower vertebrates. Using a PCR based approach similar to the one(s) described in this dissertation it should be possible to identify these proteins in lower vertebrates. However, unlike the protease domain of prothrombin which contains the unique B loop sequence, the protease domains of factors IX, X, and VII are very similar in composition and length. However, it should be possible to identify these proteases as a subset of other trypsin-like serine proteases, by taking advantage of the sequence of the active-site serine codon. Like prothrombin, factors IX, X, and VII use the AGY class of serine codons at their active-site serine (see Section F.2). To date, the AGY codon has been identified in only a few vertebrate serine proteases including plasminogen and protein C. While the sequence of the active-site codon is essential to the identification of this subset of serine proteases this feature can not be exploited for use in PCR primers in the absence of additional unique sequence information. However, the unique nature of the serine codon can be exploited at other essential serine residues. The carboxy-terminal end of trypsin-like serine proteases contains the sequence Ser - Trp - Gly. Examination of the cDNA sequences of a number of serine proteases reveals that the prothrombin, factors IX, X, VII, protein C, and plasminogen use AGY at this serine. Using a combination of PCR primers based on the sequence of this region as well as additional conserved regions of serine proteases (see Table 2 FX 1-X/VII), a number of serine proteases containing AGY active-site serine codons have been identified in lower vertebrates. While the
characterization of these proteases is incomplete at this time, this approach should provide a means to identify additional coagulation and fibrinolytic proteases from lower vertebrates.
IV. APPENDIX
The cDNA sequences and partial restriction maps of a number of vertebrate prothrombin B chains.

The species name along with the restriction maps and partial cDNA sequences are presented in panels A thru G. **Restriction maps:** The length of the cDNA sequence is represented by the box. The solid portion of the box represents the coding sequence and the cross-hatched portion of the box represents the 3' UTS. **cDNA sequences:** The nucleotide sequence commences immediately following the 3' nucleotide of primer Th7 (see figure 18, and Table 2). The predicted amino acid sequence is shown above the DNA sequence, and is not numbered. However, the first amino acid in each species corresponds to Asn 344 (human prothrombin numbering) but the first nucleotide of Asn 344 is numbered as 1 and the length of the DNA sequence is indicated in the right margin. The active site residues His-373, Asp-419, and Ser-525 are indicated with bold type. The DNA sequence of the translational stop codons are indicated with bold type as are the polyadenylation consensus sequences AATAAA.
A Partial cDNA sequence and restriction map of rat prothrombin

Q E L L C G A S L I S D R W V L T A A H C I L Y P P W D K N
CAAGAGCTGCTGTTGAGCAAGCCAGCTTTACATCGATACGGGCTCTACGTGCGCCACATCTGCTAGCCACTGGGACAAAGAC
90
F T E N D L L V R I G K H S R T R Y E R N V E K I S M L E K
TTCATCTGAGAATGACCTCCATGGCAACACTCAGAACCAGAAAATGAGCCGGAATGGAACAAGATTCCATGGTGGAAAG
180
I Y I H P R Y N W R E N L D R D I A L L L K L K K P V P F S D
ATCTACATCCACCCAGATATAACTGGGCGGAACCTAGACCGTGACATTGCTCTGGAGGCAAGGAGCCAGTGCCCCCTCAGTGAC
270
Y I H P V CL P D K Q T V I S L L Q A G Y K G R V T G W G N
TACATTCAACCCTGCTGCTTGCCAGACACACAGACAGTAAACGTTTGCTCCAGGCTGGTTTATAAGGGCGGTTGAGGACGGCTGGGCAAC
360
L R E T W T T N I E I O Q P S V L Q V V N L P I V E R P V C
CTTCGGGAGACGGAGCAACACACATGACTACAGACACACCCACCGTTGCTGCAGTGGATATGGAACCTGCCCCATTTGAGCGCAATGG
450
AAGGCCCTCTACCCCGGGTAACCTACTACGAAACACGTGGTGCTCGCTGTTCAAGGTGAAGTAGAACACTCCAGGCAAGAGGCTGGTGCAGATGC
540
GACAGTGGGGACCTTTTGTGATGAGCCGCCCTCTAACAACCACCGCTGGTACCAATGGTTATTGTCTCCTGGGTGAAAGGCTGGTGGACCGG
630
N G K Y G F Y T H V F R L K R W M Q K V I D Q H R *
AATGGAAATATGGCTCTACACCGCAGTGTTCCGCTGAAAAAGGTTGGATGCAAAATGCTACTGAGATGATGAGTACGATAGATAGATGAGTACGAT
720
GTGAGAGACCCACAGTATGGATCCTCATTTGCAAAAAATCAGAAAACCAGTCAAAATATTTTGTGTTTTGTTTCTGCTTCTACAATAAAA
810
GATTGTCACT
Partial cDNA sequence and restriction map of mouse prothrombin

QELLCGASLISDRWVTAAHCILYPPPWDKNCAAGAGCTGCTGTTGGGCGCAACCTTTATCAGTGACCGATGGGTCCTCACTGCTGCCCACTGCATTCTGTACCCACCCTGGGACAAGAAC
FEENDLLLVRIGKHSRTRYERNVEKISMLEK
TTCACTGAGAATGACCTCTGCGAGCATTTCCGAACCAGATATAGAGCGGAATGTTGAAGATCTGCTGTTACCTGCCCTACGTGAAC
IYVHPRYNWRNLDRIALLKLLKPKVPVFSD
ATCTACGTCCACCACATATAACTGCGGGGAGAACCTAGACCGCGATATCGCTCTGCTCAAGCTAAGAAGCTACCTGCTACCTGAC
YIHVPVCPLDPKQTVTSLRLRAGYKGRVRTGWGN
TATATCACCACCGTGTTTCGCCAGACAAGCAGATGATACCCAGGCTCTGCGAGGTGAACCTGCCCATTGTAGAGCGGCCAGTGTGC
KASTRIRITDINMCAGFKVNDTKRGDAECG
AAGGCCTCCACCGGATTTCAACTGACAAATGTCTCTGCTGCGCTCAAGTGTAATGCAAGACACCAAGGGAGAGATGCTTGTAAGGT
DSGGPFVMMKSPFNNRWYQMGIVSWEGECGDRA
GACAGTTGGAACCTTTTGTCATGAAAGACGCCCTAAAACAACGGCTGTATACAAAATTGGTATTGGTACCATGGGGTGAAGGATGTGC
KGYGFYTHVFRLKRWIQKVIDQFG*
AAGGGAAATACCCGCTTCTACAGCGATATGGTGTTCCGCCTGCTGAAAGTGTTGATACAGAAATGTCATTGCAATGGGAAGATGGTGATGCTGAGG
GTAAAAGCCTACATGGAATCCATGCTCGAAATCCACAGAAACCAAATACATTCTTTTGTTGTTTCTGCTTCAATGAAAGTG
ATTGTCACT

819
C Partial cDNA sequence and restriction map of rabbit prothrombin

QELLCAASLISDRWVTAAHCLLYPPWDKN
CAGGAGCTGCTGCGCCGCCACCTCATCAGCGACCGCTGGGTTCTCACTGCTGCCCACTGCCTCCTCTACCCGCCCTGGGACAAGAAC

FTVNDILVRIGKYARSRYERNMEEKISTLEK
TTCACCGTGAATGACATTTCTGTGGCCATCGGGCAATACGCCCGACAGATCAGACCAGGAACATGGAGAAGATCTCCACCCTGGAAAAG

IIIHPGYNWRNLDLMKLMKLPVAFSD
ATCATCATCCACCCCGGTACAACCTGGGACACATCCGGCCCTCATGATGAGCTCAAGAAGCCGGTTGCCTTCAGCGAC

YIHPVCLPDKQIVTSLLQAGHKGRVTGWGN
TACATCCACCCCGTGTGGCCTGTGACAAGCAGATAGTGACCAGCTTGCTCCAGGCTGGACACAAGGGGCGGGTGACAGGCTGGGGCAAC

LKEMWTNVMNVEQPSVMLQMVNPILVERPIC
CTGAAGGAGATGGGACCCTGAACATGAGAAGGGTGAGCGCCACGGTGTCACTAGATGGAATCTGCCTCTTGGAGCCGGCCCATGTC

KASTGIRVTDNMFCCAGYKKEEGKRGDACGE
AAGGGCGTCCACCGGATCCCGATACCAGAACATGGTTTGGCGGGTTAACAACCGCTGAAGAGGGAAACCGAGGGAATGCGTCTGAGGTT

DSGGPFVMKNPYNWRWYQMGIVSWGEgcdR
GACAGTGGGGACCCTTGTTGTATGAGAACCCTCAACAAACGACTTGATCACGACTGGCCATTGGCTCCCTGGGTTGAAGGCCTTGAGC

DGKYGFYTHVFVRLLKKWIRKMVDRF
GATGGGAGATGGCCTTACACACCGACCTTCGCGCTCAAGAAGTGGATACCGCAAGATGGTGTAGGCTTAGGCTAAAGGGGCCCCACACAC

ATTCTGGGCTTCACACTGGCATGTTACAAAGAACCAACCGACTCAAGAAGATGGTTTGGGTTTGCTAAGAGGTGTTCCCAATAAA

AGTGACTCTGTCACTG

145
D Partial cDNA sequence and restriction map of gekko prothrombin

Q DLLCGASLISDRWILTAAHCIFYPPWDDKN
CAAGACTTGCTGTGGTGCCAGCCTCATCAGTGATCGCTGGATCCTGACTGCTGCTCACTGTATCTTCTACCCACCCTGGGATAAGAAC 90

FTADDLVVRIGKHNRRIHEKTREKIALLDK
TTCACGGCAGATCACCTTTGTGGAAATGGAAACATACAAAGAGAAGAAGACCAGGGAAAAATTGCCTTGCTGGACAAA 180

I I I H P K Y N W K E N L D R D I A L L R R L R K P V P F S D
AICATCATCCATCCAAAATATACATGGAAGAGAATCTAGATCGGGATATTGCACTGTTGCGCTTGAGGAAACCTGTTCCTTTCAGTGAC 270

Y IQPVCLPLPTKETYQVLQLLLTGYKGRVGTGWGN
TACATCCAGCCAGTCTGTGGCCACAAAGAAACTGTGAAACCGCTGTGGACAGGGTAGCAAAAGAGCTGTGACTGGCTGGGGAAAC 360

LFETWGSTSPALPTYQLQLVNLPIVDRDTC
CTTTTTGAAACCTGGGGTTCTAGCTTCCTGCCCTACATATTTGCGCTGTGAACCTCTCCCTACAGAGACTGGACACCTGTAAG 450

ASTKIKITDNMFCAGYSPEDSKRGDACEDG
GGATCAACACAAAAATTAAACTACAGAGAAATATGTCTCTGTGCAAAGGTGACAGTCAACATCAAGAGGGGATGCTTGTGAAGGTAAC 540

SGGFVFWMKNPDNRWYQVTVSVWEGCGCRR
AGCGGGGGCCCATTTGTCATGAGAACCCCACAACAGATTCTCAGTGCCTGACATCCCTACAGAAGCTCACTGTTGAGCGATCTCAGGGGAAAC 630

GKYGFYTHFLRKWKLLKTVKHEGN
GGTAAATATGGATTCTATACCCATGTGTTGGTCCGTGGATTGAAGAACTGTGAAACACATGGGATTTAAGGATTGTATC 720

TCAACCTACAGCCAGATTTGAAAGCTATGTCTACAAAAGGCAACATTACAGTACCCAAAAGCTAGAAGACTTCCTTGAATTTAAACGCAACCTCCA 810

ATAAAACTCTCTGTACCCAGAAAATCACTCTGTACCCAGAAAATCACCGAAGACCCGAAATCACTCTGTACCAGAAGGAACCTTCACTGAATCCAGG 900

GTTCTGTTCCAGTAACATGTTGCTGC
E  Partial cDNA sequence and restriction map of newt prothrombin

Q E L I C G A S I I S D R W V L T A A R C I F Y P P W D K N
CAGGAGCTCATATGTGGAGCAAGCATCATCAGTGACCGCTGGGTTCTCACTGCAGCGCACTGCATTTTCTACCCACCCTGGGACAAAAAC 90
Y T T E D I L V R I G K H Y R T K Y E R Q Q E K I R M L E R
TACACCACAGAAGACATCCTGTGGAATTGGAAAAACACTACAGGACCAAGTACGAGAGACAACAGGAGAAGATTCGAATGCTGGAGGG 180
I I I H P K Y N W R E N L D R D I A L I Q L K R P I G F T N
ATCATCATTACCCCCAAGTAAGACCGGACCGTACGCCCCGATCCAGTGAAACCGGACCCATTGGCCTCCACCAAC 270
Y I H P V C L P T K E I V Q T L M L N R H K G R V S G W G N
TACATCCATCCCGCTCTGCTGGCACCAGAGATCGTCCAGTGTAAGACAGAGACAACAGGAGAAGATTCGAATGCTGGAGGGG 360
L H E T W T S G G Q A L P Q V L Q Q V N L P I V D Q E T C K
CTGCATGAGACCTGGAACCTGAGGGGGCCAGGCCCCTTCCAGGCTGCTGCAAGTCCAAACCTGTGCAAAGAAACCGGAGGGGGGAC 450
GCCTCAAACAAAATCAAAGTCATCGACATGTTTCTGTGCAAGTTAIAACACAGATGCAAAACAGAGGGGAGCGCTGGCAGGGGAC 540
AGTGTCGTCATCGATCGACATCGACATCGACACTACCGTCTCCTCCTCTGGGGCCAGGTTTGATGGATCGGGAT 630
G K Y G F Y T H L H R M R Q W M M K I I E K C G S *
GGCAAGTATGGGATTCTATAGCACCCTGCAGCAGGGATCGGAGATGGAAGATCATCGAGAAATGTGGGAGGTAGGATGGATGCA 720
CCAGCCTCCTCATGCCAGAAAAGAACAGCCAAACATATCCATAGAAATGCTGGAAAATACAGCGCAAATAAAAGCCTCATTCTCGGGGAC 810

AGC
Partial cDNA sequence and restriction map of rainbow trout prothrombin

QELLCGASLISDEWILTAABCLYLPPWKN
CAGGAGCTGCTGTGTGGGGCCAGTCTAATCTAGTGATGATGGATCCTCACTGCAGCCCACTGCATCCTCTATCCACCATGGAACAAAAAC

FTINDILVRGLKHKNRRAKFEKGTEKIVAI
TTCACCATTAGACATCCTGGCTGGGACCTGCAAACATTAACAGAGTGTAAGTTTGAGAAGGGCACAGAGAAGATTGTGGCTATTGATGAG

IIIVHPKYNWKENNLNRDIALLHLMRMRPPITFTTD
ATAATTTGTCACCCCAAGTACAACTGGAAGACCTGACCCCCGGAATTCGGCTCTGCACATGAGGAGGCCCATTACTTTCACAGAT

EIHHPVCLPTKQVAKTLMFAGYKGRVTGWGN
GAGATTCATCTCTGTCTCTACCAACCAACGTTTGGACTAAGACACTGATGTTTGCTGGCTTAAACCAGAGGACGGCTGGGGGAAC

LYETWSSSPSFKSLPTVLOQIHLPIVEQDICR
CTATATGAGACATGAGGATTCTCCTCCCCAAGTCTCTACCCACAGTTCTCCAGGAGCTACATCCATATCCTTAACAGAGAAGGAAC

DSTSIRITDNFMCAFCKPKPEEQKTGDACEGD
GACTCTACCTCTACCCCACTGACAACGTTGATGCTCCATCGTGGGAACAGAAGGAGATTGGGACAGGCTGGGGAAC

SGGPFVMSKSPDDNRWYQIGIVSVWEGGC
AGCGGTGGTCCCTTTAGCATGAGAACCGAGCAACCGTTGTACCGATCGGCTATTGTCTCTGGGGAAGGAGATGACAGGATGAG

GKYGFYTHTFLFRMRWWMKKVIDKTK
GGGAATATGATACCCATCTTTTTCGTTAGGACCGGTGGAAGAAAGATTATCTGCAAAAACAGGGCCGAGGACAGGATGACATG

TTGTTATTTCTCCTACTTCTCTCAGCATCAGGAAAACCTGATGATTAATTTGGAATATACACCATCGGTCTCCTGATCATG
**Partial cDNA sequence and restriction map of sturgeon prothrombin**

**Hhal**  
**Kpnl**  
**PvuI**

```
QELLCGASLISDQWILTAAHCILYPPWNKN
CAGGAGCTGCTTTTGAGCCAGTCCTCAATAGTGACCAGTGGATACTGACGGCTGCCACTGCATTCTCTACCCACCCTGGGAACAAAAAC 90
FTANDILVRVGKYRAKFEKQTEKIVALDE
TTCCACAGGAAATGACATCTCTTAGTCAGAGTGCCAAAAACACTACCAGCAGGAACTTTTGACAGATATTGCTCTTCACCTGAGAAACAATTGGTCTGATGAA
180
IILHPKYNWKENLDRLIALLHLRKLPLTFTE
ATTATCCTCCCTCCCAATATAACTGGAAAGAGAATTGGACAGAGATATTGCTCTTTGGCAGACCTGAGAAACTGGAGATGGGGAAAC 270
NIVPICLPTKKVKALTLMFAGFKGRVTGWN
AACATCGTGCCCATCTGTTTGCCCACTAAGAAAGTTGCTAAGACGCTGATGTTTGCCGGGTTCAAGGGGCGTGTGACGGGATGGGGAAAC 360
LYETWTSSSPQLPQLQQIHLLPIVQQETCR
CTGTATGAGACGTGGATAAGCTCTCCCCAATCTCTCTCCAGGCTCCATCTGGACAGATTGCCCTGGAAGATCTATAAGTGGAGATGGGGAAAC 450
DSTKIRVTVDMFCAAGFSPEDSISGDSCEGD
GACTCCACCAAGATAGTGACTGACAACATCTCTGTTGGACAGATTTGACCCAGGAGAAGATTCTCTATTGAAGGTGAGATTCTCTGGTGACAGGACGAGGGTTGAC
540
SGGPFVSMKNPPEDDRWYQIGIVSWSGWEGCDRS
AGTGGGGGTCATTTGACTGAAGAAACCCAGGAGAGTCGCTTGACGATGTCTTGGGAGAAGGCTTGCCACTGGATCGTACCCAGGGGAGACTCTATTTGAGGACAACTGGGAACAAAAAC 630
GKYGFYTHLFRMRKWMKLKTIVDTE*
GGCAATAACGGATTTTACACACCTTTTCCGTATGGCAAGAATACCTAGTGGACACCGAGTAGTGAAAGTTTAAAGAA
720
TACACCATTGTTGCTTTTTTCTCTGGACAGTCTTTTTTTGTTAGTATATATAAAAATATAGT
785
```
V. REFERENCES


Sudhoff, T.C., Russell, D.W., Goldstein, J.L., Brown, M.S., Sanchez-Pescador, R.,

815-822.


van de Ven., W.J.M, Voorberg, J., Fontijn, R., Pannekoek, H., van den Ouweland,

van den Ouweland, A.M.W., van Duijnhoven, H.L.P., Keizer, G.D., Dorssers, C.J., and


1068.

Biophys. 164, 527-535.


63, 24-26.

3922-3926.


