Doctoral Thesis

Evolutionary Features of Serine Protease and Its Inhibitor as Mosaic Proteins

1991

Doctor of Philosophy

Kazuho Ikeo

Department of Genetics
School of Life Science
The Graduate University for Advanced Studies

Mishima 411, Japan

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my supervisor, Prof. Takashi Gojobori, for his continuous guidance and encouragement during all stages of this work.

I also would like to express my deep indebtedness to Dr. Kei Takahashi of Shimane Medical University, without whom this dissertation would never have reached its present form. I am grateful to Dr. Yoshio Tateno and Dr. Etsuko N. Moriyama of The National Institute of Genetics, for their help and valuable advise. I also thank Ms. Mary Shimoyama for her useful comments on the manuscript.

Warm thanks to all the people of Department of Evolutionary Genetics and DNA Research Center in The National Institute of Genetics for their help and cooperation during various stages of this work.

ABSTRACT

The serine protease is a typical example of the mosaic protein. It is suitable materials for studying molecular evolution of genes and proteins, because a gene (and thus a protein) is now considered to be composed of domains of independent evolutionary origins. The serine protease has a protease domain and a domain called the kringle structure. The kringle structure is a typical motif that has a pattern of disulfide bridges. The Kuniz-type protease inhibitor, a protein inhibitor of the serine protease, has also a smaller size of the kringle structure called the minikringle.

The purpose of this dissertation is to study the molecular evolution of the serine proteases and the Kunitz-type protease inhibitors, in order to elucidate the evolutionary process of mosaic proteins. To accomplish this purpose, I studied the evolutionary dynamics of these proteins from various aspects of molecular evolution in the following chapters.

In Chapter 2, I explained the kringle structures of the serine proteases and showed characteristic features of its molecular evolution. In Chapter 3, I discussed the difference in evolutionary pattern between the protease domains and kringle domains, in order to exemplify domain shuffling occurring in the evolutionary course of mosaic proteins. The molecular evolution of the Kunitz-type protease inhibitors was discussed in Chapter 4, which explored the evolutionary origin of the Kunitz-type domains inserted into

various proteins. In Chapter 5, I described the evolutionary relationships between the kringle and kringle-like structures, for the purpose of showing the possibility that these structures have undergone independent courses in molecular evolution of the mosaic proteins.

In conclusion, as mentioned in Chapter 6, I emphasized the importance of domain shuffling in evolution of mosaic proteins.

<TABLE OF CONTENTS>

		Page
ACKNOWLEDGMENTS		1
ABSTRACT		2
TABLE OF CONTENTS		4
LIST OF TABLES AND FIGURES		9
CHAPTER 1 INTRODUCTION		
1.1	Mosaic protein	14
1.2	Exon shuffling and gene duplications	15
1.3	Motifs	16
1.4	Classification of proteases	18
1.5	Protease inhibitor as a regulatory element	20

CHAPTER 2 MOLECULAR EVOLUTION OF KRINGLE STRUCURE

2.1 INTRODUCTION

2.1.1 Molecular mechanism of blood coagulation 36

2.1.2 Structure of proteases in blood coagulation 38

2.1.3 Apolipoprotein(a) 39

2.1.4 Hepatocyte growth factor 40

2.2 MATERIALS AND METHODS

2.2.1 Sequences used 41

2.2.2 Genetic distance 42

2.3 RESULTS

2.3.1 Evolutionary origin of kringle structures 43

2.3.2 Molecular evolution of apolipoprotein(a) 43

2.4 DISCUSSION

2.4.1 A model for the evolution of apolipoprotein(a) 45

2.4.2 Function of kringle structures 47

CHAPTER 3 EVOLUTION OF SERINE PROTEASE

3.1 INTRODUCTION

3.1.1 Serine protease as an enzyme 69

3.1.2 Molecular mechanism of catalytic reactions 71

3.1.3 Physiological function of some serine proteases 72

3.1.4 Evolutionary study of protease domains 74

3.2 MATERIALS AND METHODS

3.2.1 Sequences used 75

3.2.2 A methods of multiple alignment 75

3.3 RESULTS

3.3.1 Evolutionary origin of serine protease domain 76

3.3.2 Evolution of the blood coagulation and fibrinolysis systems 77

3.4 DISCUSSION

3.4.1 Different patterns in evolution between protease and kringle domains 79

CHAPTER 4 EVOLUTION OF SERINE PROTEASE INHIBITOR

4.1 INTRODUCTION

4.1.1 Minikringle structure 99

4.1.2 Amyloid β protein in Alzheimer's disease 100

4.1.3 α3 chain of collagen VI 101

4.1.4 Dendrotoxin as a serine protease inhibitor 103

4.2 MATERIALS AND METHODS

4.2.1 Sequences used 104

4.3 RESULTS

4.3.1 Evolution of serine protease inhibitor 105

4.3.2 Evolution of snake toxins 108

4.3.3 Evolution of the insertion sequence in amyloid β protein 108

4.4 DISCUSSION

4.4.1 Relationships between amino acid sequences and specificity of inhibitors 110

4.4.2 Biological function of protease inhibitor 112

CHAPTER 5 EVOLUTION OF MOSAIC PROTEINS

5.1 Introduction 143

5.2 Molecular evolution of kringle-like structures

5.2.1 Kringle-like structures in defensins 145

5.2.2 Kringle-like structures in an envelope protein of TBE virus 148

5.3 Evolution of serine proteases involved in the system of biological cascades

5.3.1 Biological function of regulatory domains in serine proteases 150

5.3.2 Structural similarity in catalytic triads between serine protease and lipase 153

5.4 Evolutionary roles of domain shuffling 156

CHAPTER 6 CONCLUSIONS 174

LIST OF THE AUTHOR'S PUBLICATIONS 175

LIST OF TABLES AND FIGURES

CHAPTER 1

TABLES

Table 1.1 Families of proteolytic enzymes.

Table 1.2 Proteases found in mammalian cells.

Table 1.3 Families of protein inhibitors that inhibit serine proteases.

FIGURES

Figure 1.1 Members of proteases involved in the cascade of blood coagulation.

Figure 1.2 Motifs in the proteases of the blood coagulation cascade.

Figure 1.3 Kringle structure.

Figure 1.4 Topological relationships between the disulfide bridges and the location of the reactive site of serine protease inhibitors.

CHAPTER 2

TABLES

Table 2.1 Binding Functions of Kringle Domains.

FIGURES

Figure 2.1 Schematic pathways of the blood coagulation fibrinolysis.

Figure 2.2 Mosaic structures of the serine proteases in blood coagulation.

Figure 2.3 Structure of human and simian apolipoprotein(a).

Figure 2.4 Structure of plasminogen and HGF.

Figure 2.5 Sequence alignment of kringle structures.

Figure 2.6 A phylogenetic tree of kringle domains.

Figure 2.7 Evolutionary changes of apolipoprotein(a).

CHAPTER 3

TABLES

Table 3.1 Physiological processes regulated by protease and its inhibitors.

Table 3.2 Serine proteases used in the present analysis.

FIGURES

Figure 3.1 The enzymatic mechanism of serine protease.

Figure 3.2 The phylogenetic tree of serine proteases.

Figure 3.3 A different in the evolutionary patterns between the kringle domains and the serine protease domains.

CHAPTER 4

TABLES

Table 4.1 Tissue Distribution of Type VI Collagen Using Immunofluorescence.

Table 4.2 Serine protease inhibitors used in the present analysis.

Table 4.3 Amino acid residues at the reactive site (P_1).

Table 4.4 The biological activity and amino acid sequences of the inhibitors in group I.

Table 4.5 Contact regions of the Kunitz-type inhibitors in group III.

FIGURES

Figure 4.1 The secondary structure of Kunitz-type protease inhibitors.

Figure 4.2 The Kunitz-type protease inhibitor domain inserted into βAPP.

Figure 4.3 The Kunitz-type protease inhibitor domain inserted into α3 chain of collagen type VI.

Figure 4.4 Sequence alignment of Kunitz-type protease inhibitors.

Figure 4.5 The phylogenetic tree constructed using amino acid sequences of Kunitz-type protease inhibitors.

Figure 4.6 Evolutionary changes of Kunitz-type protease inhibitors.

Figure 4.7 The phylogentic tree constructed using nucleotide sequences of Kunitz-type protease inhibitors.

Figure 4.8 The Schecter and Berger's nomencultures for binding of a peptide substrate to a peptidase.

Chapter 5

FIGURES

Figure 5.1 Amino acid sequences of defensins.
Figure 5.2 Phylogenetic tree of defensins.
Figure 5.3 The kringle structure of TBE virus.
Figure 5.4 Serine residue as active center.
Figure 5.5 Charge-relay system of serine proteases.
Figure 5.6 The catalytic triad of lipase, and serine protease.
Figure 5.7 The structure of aspartic acid and glutamic acid.

CHAPTER ONE

INTRODUCTION

1.1 Mosaic protein

In the primary world of life, an organism must have had, at least, a minimum set of "mini" genes, which were simpler in function, shorter in length, and plainer in structure than the currently existing genes. It is quite possible that the mini genes were ancestors of functional domains of the present genes. The mosaic protein, which is made up with different functional domains, is a good material to study evolution of such organismic macromolecules as DNA, RNA, and proteins. The mosaic proteins are thought to have evolved by the fusion of ancestral mini genes.

Serine proteases are typical mosaic proteins having domains of finger, epidermal growth factor (EGF), kringle, and serine protease. They exist and function in the blood coagulation and fibrinolysis cascades. The first serine protease as a mosaic protein was found in tissue-type plasminogen activator (Figure 1.1) (Pennica et al. 1983). Tissue-type plasminogen activator is composed of four domains. Of the four domains, the first three have a regulatory function and the last has a catalytic function (Patthy, 1985). The junctions between these domains precisely

coincide with the borders between the exons and introns of the corresponding gene. This structure supports that the gene was formed with ancestral mini genes of different evolutionary origins.

1.2 Exon shuffling and gene duplications

How has a mosaic protein evolved from ancestral mini genes and obtained some new functions? This problem is very important to understand molecular evolution of functional proteins. Exon shuffling and gene duplication must have played essential roles in generating mosaic proteins. Gilbert's (1978) proposal of exon shuffling was motivated by the fact that a gene is organized with a series of exon-intron arrangements. Gilbert asserts that genes must have evolved in the way that exons had been shuffled and recombined with each other during evolution. As a result, such genes acquired a new biological function. The exon-shuffling might have speeded up evolution, because exons each of which already played a useful role must have been quickly assembled by this mechanism. Thus, exon shuffling can be a major mechanism for generating a new genes. The fact that a protein often has a partial sequence homology with other evolutionally unrelated proteins endorses a significant evolutionary role of exon shuffling.

Another mechanism for producing a new gene is gene duplication. If two duplicate genes emerge from a gene, one of them may be able to mutate itself drastically and thereby acquires an entirely different function from the other. A gene with improved function may also arise by gene elongation. Gene elongation is usually caused by duplication of a gene itself or a part of the gene. Many proteins have internal repeats of amino acid sequences, and these repeats often correspond to separate functional or structural domains of the protein. This suggests that the genes coding for these proteins were formed by internal gene duplication. It seems that most genes were produced by duplication and elongation of mini genes that have existed in the early stage of the gene evolution (Darnel 1978; Doolittle 1978; Ohno 1981; Blake 1985).

1.3 Motifs

Protein families often have their characteristic nature in their amino acid sequence (Figure 1.2). For example, a family of DNA-binding proteins is known to have such a characteristic pattern that leucine residues are located every seven amino acids in a particular region. Such a pattern known as leucine zipper is called a motif. A motif is usually related with function of a protein. Leucine-zipper,

for example, is directly involved in protein-nucleotide recognition (Mcknight, 1991). Thus, a motif is an important key for inferring the function and structure of a protein from the amino acid sequences. Recently, many kinds of motifs have been found in various proteins.

The serine protease also has its own motifs. It consists of a motif of about 80 amino acids in which six cysteines are contained. The six cysteines make three particular sets of disulfide bridges which are called the kringle (Figure 1.3). The kringle is a typical motif. The serine protease also includes the epidermal growth factor domain. This domain has six cysteine residues which form a pattern of disulfide bridges different from the kringle. This is also a motif. Interestingly, Kunitz-type protease inhibitor also contains six cysteines by which a kringle is made. Because this inhibitor consists of about 50 amino acids, its kringle is smaller than that of serine protease. We call the smaller one a mini-kringle.

A kringle was also found in other molecules which have apparently no connection to serine protease or Kunitz-type protease inhibitor. For example, some virus envelope protein has a kringle different in length. Difensin, which is associated to an immune reaction, also has the same pattern of disulfide bridges as Kunitz-type protease inhibitor. These two structures are new members of the protein having the kringle structures. They each have their own length different from kringle or mini-kringle.

In this study, I focused the discussion on the molecular evolution of this pattern of the disulfide bridges.

1.4 Classification of proteases

The International Union of Biochemistry classified proteases into four classes according to the mechanism of enzymatic actions (Table 1.1). They are serine proteases, cysteine proteases, aspartic proteases, and metallo-proteases. The class of serine proteases is divided into two families, serine proteases I and II. The class of metallo-proteases is also separated into two families, metallo-protease I and II. Together with other two classes, there are six families of proteases (Table 1.1). Each family has a characteristic set of functional amino acid residues that are arranged in a particular configuration to form the active site (Table 1.1).

The serine protease has a serine residue at the center of catalytic sites. The serine protease occupies two of the six protease families; a family of serine protease I for the mammalian serine proteases and a family of serine protease II for the bacterial serine proteases such as subtilisin (EC 3.4.21.24). They differ from each other in amino acid sequences, despite that the locations of active sites in the three-dimensional structure of the protein molecule and the enzymatic mechanism are similar to each other.

The cysteine protease, which occupies one family, includes several mammalian lysosomal cathepsins, cytosolic calcium activated proteases, and plant proteases such as papain and actinidin. Papain is the best-studied member in this family (Neurath 1989). The major catalytic residue of the cysteine protease is cysteine, which is just like serine at the active site in the serine protease.

Another family is the asparatic protease, which includes bacterial penicillopepsin, mammalian pepsin, renin, chymosin, and certain fungal proteases. In this family of proteases, the characteristic residues at reactive sites are asparatic acids.

The metallo-protease includes the last two families; metallo-protease I for the mammalian pancreatic carboxypeptidase and metallo-protease II for the bacterial thermolysin. They differ in chemical structures, even though both are zinc metallo-enzymes and have similar configurations of active sites. They need a metallo-ion for their enzymatic activity.

Among these four families, the serine proteases are best-studied for the amino acid sequences of various organisms, the enzymatic mechanism, and the three-dimensional structures. For this reason, I will focus on the molecular evolution of the serine proteases in the present study.

1.5 Protease inhibitor as a regulatory element

The protease inhibitors can be divided into two classes; (1) active-site specific and low-molecular inhibitors that irreversibly modify an amino acid residue at the active site of the corresponding proteases, and (2) naturally occurring protein protease inhibitors, many of which act as pseudosubstrates (Laskowski et al. 1980).

Natural protein protease inhibitors are ubiquitous in the living world. The protein protease inhibitors have been isolated from animal, plant and bacteria. Several crystalline protease inhibitors isolated from pancreas of mammals, bird eggs, and certain legumes (beans) have served as materials to elucidate the biological mechanism of protease inhibition (Laskowski et al. 1974).

Many inhibitors of the proteases have been reported so far. Those having known active sites and target enzymes are listed in Table 1.2. In particular, well-characterized inhibitors of serine proteases are shown in Table 1.3. They can be classified into ten families or probably more. While the major criterion for establishing a family is extensive sequence homology among the members, the location of the reactive site and topological relationships between the disulfide bridges are used as notable characteristics (Figure 1.4). In particular, the positions of all intrachain disulfide bridges are completely conserved in each of the families.

The protein protease inhibitors are irreversibly bound to the active site of a protease in order to inhibit the protease activity, and they are converted into their modified forms which cannot be digested by the protease. Their physiological function is apparently to prevent unwanted proteolysis. However, a physiological target enzyme of the protease inhibitor is usually unknown because protease inhibitors for most organisms were discovered by the assay of inhibition with bovine trypsin and chymotrypsin. In other words, bovine trypsin and chymotrypsin are unlikely to be true target enzymes for most of the isolated inhibitors. Although the lack in knowledge of the true target enzymes for the majority of inhibitors is one of the major obstacles in understanding the evolutionary mechanism of protease inhibitors, the accumulation of amino acid and nucleotide sequence data will help us to study the molecular evolution of this protein family.

As explained in all sections of this chapter, the purpose of this dissertation is to study the molecular evolution of serine proteases and their Kunitz-type protease inhibitors, in order to elucidate the evolutionary process of mosaic proteins. To accomplish this purpose, I am going to discuss the evolutionary dynamics of these proteins from various viewpoints of molecular evolution in the following chapters. In Chapter 2, I am going to explain the kringle structures of the serine proteases and to show characteristic features of its molecular evolution. In Chapter 3, I will discuss the difference in the evolutionary pattern between the protease

domains and kringle domains, in order to exemplify the domain shuffling occurring in the evolutionary process of mosaic proteins. The molecular evolution of Kunitz-type protease inhibitors will be discussed in Chapter 4, in order to explore the evolutionary origin of the Kunitz-type domains inserted into various proteins. In Chapter 5, I am going to describe the evolutionary relationships between kringle and kringle-like structures, for the purpose of showing the possibility that these structures have independent histories in the molecular evolution of mosaic proteins. In conclusion, as will be mentioned in Chapter 6, I will emphasize the evolutionary importance of domain shuffling in mosaic proteins.

References (Chapter 1)

Blake C C F (1985) Exons and the evolution of proteins. *Intl. Rev. Cytol.* 93, 149-185.

Darnel J E (1978) Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science* 202, 1257-1260.

Doolittle W F (1978) Genes in pieces: Were they ever together? *Nature* 272, 581-582.

Gilbert W (1978) Why genes in pieces? *Nature* 271, 501.

Laskowski M, Kato I, Lery T R, Schrode J, and Sealock R W (1974) Evolution of Specificity of Protein Protease Inhibitors. Bayer-Symposium V "Protease Inhibitors" 597-611.

Laskowski M Jr and Kato I (1980) Protein inhibitors of proteases. *Ann. Rev. Biochem.* 49, 593-626.

McKnight S L (1991) Molecular Zippers in gene relation. *Scientific American* 264, 32-39.

Ohno S (1981) Original domain for the serum albumin family arose from repeated sequences. *Proc. Natl. Acad. Sci. USA* 78, 7657-7661.

Patthy L (1985) Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. *Cell* 41, 657-663.

Pennica D, Holmes W E, Kohr W J, Harkins R N, Vehar G A, Bennet W M, Yelventon E, Seeburg P H, Heyneker L and Goeddel D V (1983) Cloning and expression of human tissue-type plasminogen activator cDNA in *E. coli*. *Nature* 301, 214-221.

Table 1.1 Families of proteolytic enzymes.

Family[a]	Representative protease(s)	Characteristic residues[b] of active sites
Serine protease I	**Chymotorypsin**	Asp^{102}, Ser^{195}, His^{57}
	Trypsin	
	Elastase	
	Pancreatic kallikrein	
	Sinbdis virus core protein	
Serine protease II	**Subtilisin**	Asp^{32}, Ser^{221}, His^{64}
Cysteine protease	Actinidin	
	Rat liver cathepsins B and H	
Aspartic protease	**Penicillopepsin**	Cys^{25}, His^{159}, Asp^{158}
	Rhizopus chineses and Endothia parasitica, acid protease	
Metallo-protease I	**Bovine carboxypeptidase A**	Zn, Glu^{270}, Try^{248}
Metallo-protease II	**Thermolysin**	Zn, Glu^{143}, His^{231}

[a]This table includes only the enzymes whose amino acid sequences and three dimensional structures have been determined. For rat liver cathepsins B and H, the three dimensional structure has been surmised by analogy to that of papain.

[b]The number on the residues corresponds to the residue number in the amino acid sequences of the enzymes listed in **bold** at column 2.

Table 1.2 Proteases found in mammalian cells[a].

Proteases as integral cellular components	Secretory proteases
Serine proteases	
prolyl endopeptidase of cytosol, ATP-dependent protease of mitochondria, processing protease of mitochondria and ER, trypsin-like protease of plasma membrane, chromatin proteinases of nuclei, cathepsin R of ER	tissue kallikrein, submandibular proteinase A, nerve-growth-factor γ-subunit, acrosin of sperm, cathepsin G of neutrophils, elastase of neutrophils, tryptase of liver, chymase of mast cells, chemotactic proteinase of skin and lymphocytes, plasminogen activator
Cysteine proteases	
cathepsins B, L, H, N, S, M, T of lysosome, calpains of cytosol, ATP-dependent protease of cytosol, metal-dependent protease of cytosol	
Asparatic proteases	
cathepsin D, E of lysosome, protease of red blood cell membrane	renin, processing protease of pituitary
Metallo-proteases	
endopeptidase 24.11 of plasma membrane, meprin of plasma membrane, PABA-peptide hydrolase of plasma membrane, metallo-endopeptidase of cytosol, signal peptidases of mitochondria and ER	

Table 1.3 Families of protein inhibitors that inhibit serine proteases.

I.	Bovine pancreatic trypsin inhibitor "BPTI" (Kunitz) family
II.	Pancreatic secretory trypsin inhibitor (Kazal) family
III.	Streptomyces subtilisin inhibitor (SSI) family
IV.	Soybean trypsin inhibitor "STI" (Kunitz) family
V.	Soybean trypsin inhibitor (Bowman-Birk "BBI") family
VI.	Potato I inhibitor family
VII.	Potato II inhibitor family
VIII.	Ascaris trypsin inhibitor family
IX.	α1-proteinase inhibitor
X	Other families

Figure 1.1 Members of proteases involved in the cascade of blood coagulation.

Diagrammatic representation of the structural and functional relationships among protein components of blood coagulation and fibrinolysis. Protein families include the serine proteases, procoagulant blood clothing proteins, regulatory anticoagulant proteins, and fibrinolytic proteins. Abbreviations are as follows: tPA, tissue plasminogen activator; uPA, urokinase; PK prekallikrein; T, trypsin; CT, chymotrypsin; ATIII, antithrombin III; TM, thrombomodulin; PS, protein S; PC, protein C, PAI I and II, plasminogen activater inhibitor I and II.

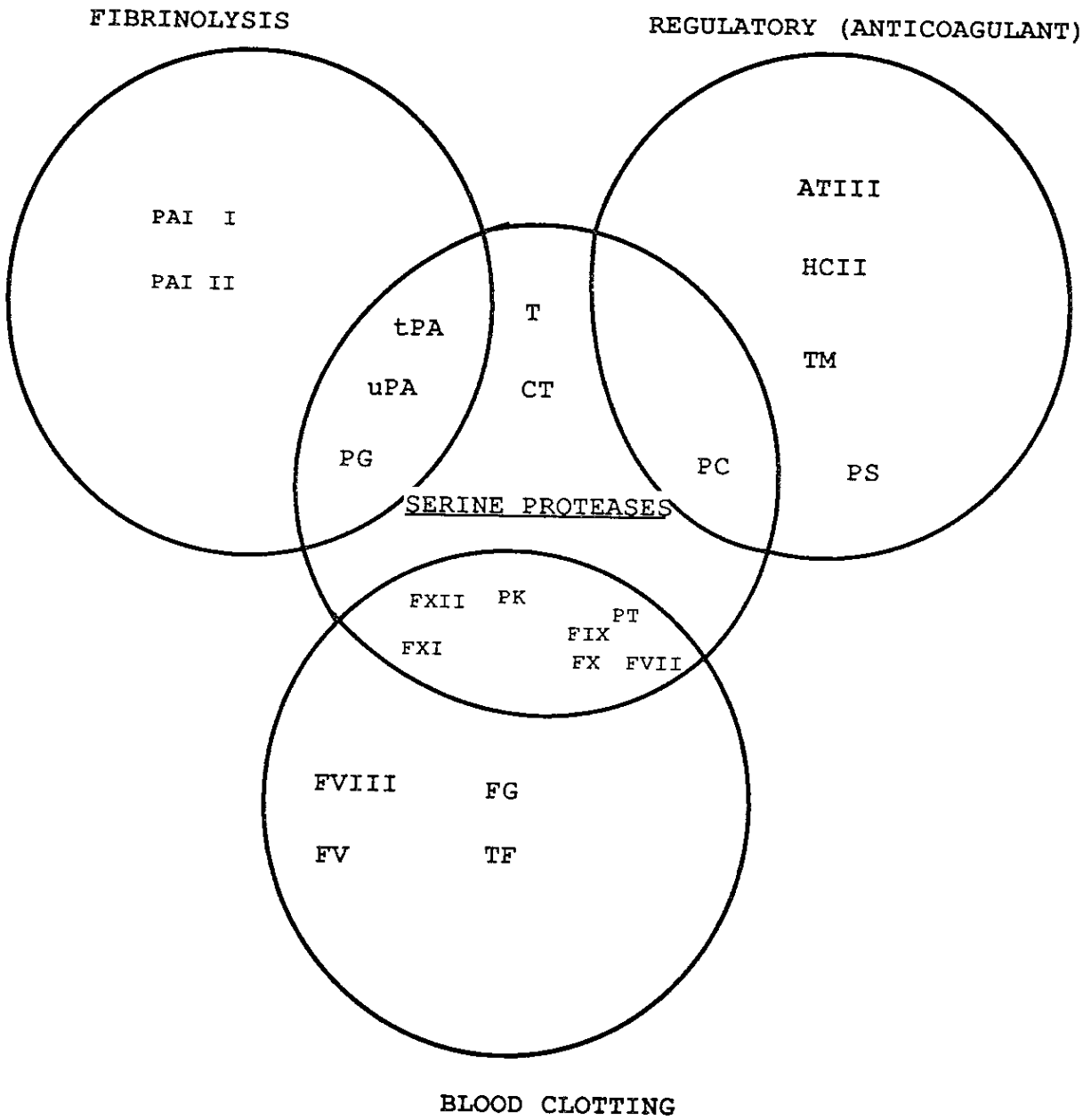

FIBRINOLYSIS
REGULATORY (ANTICOAGULANT)
PAI I
PAI II
ATIII
HCII
TM
PS
tPA
uPA
PG
T
CT
PC
SERINE PROTEASES
FXII
PK
PT
FIX
FXI
FX
FVII
FVIII
FG
FV
TF
BLOOD CLOTTING

Figure 1.2 **Motifs in the proteases of the blood coagulation cascade.**

This figure shows different motifs of the regulatory domains in proteases. C, vitamin K-dependent calcium-binding domain; K, kringle domain; G, growth factor domain; F, finger domain.

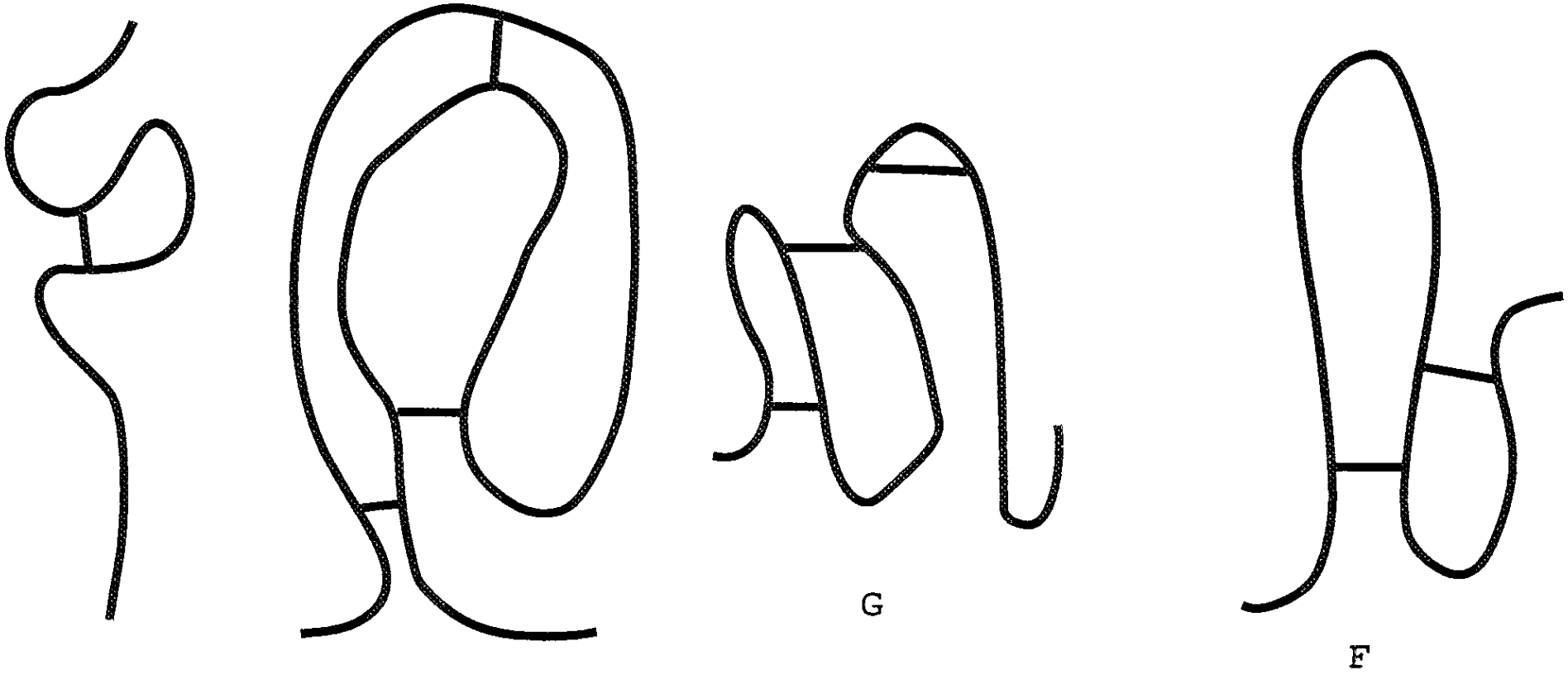

C
K
G
F

Figure 1.3 Kringle structure.

The solid lines show the disulfide bridges.

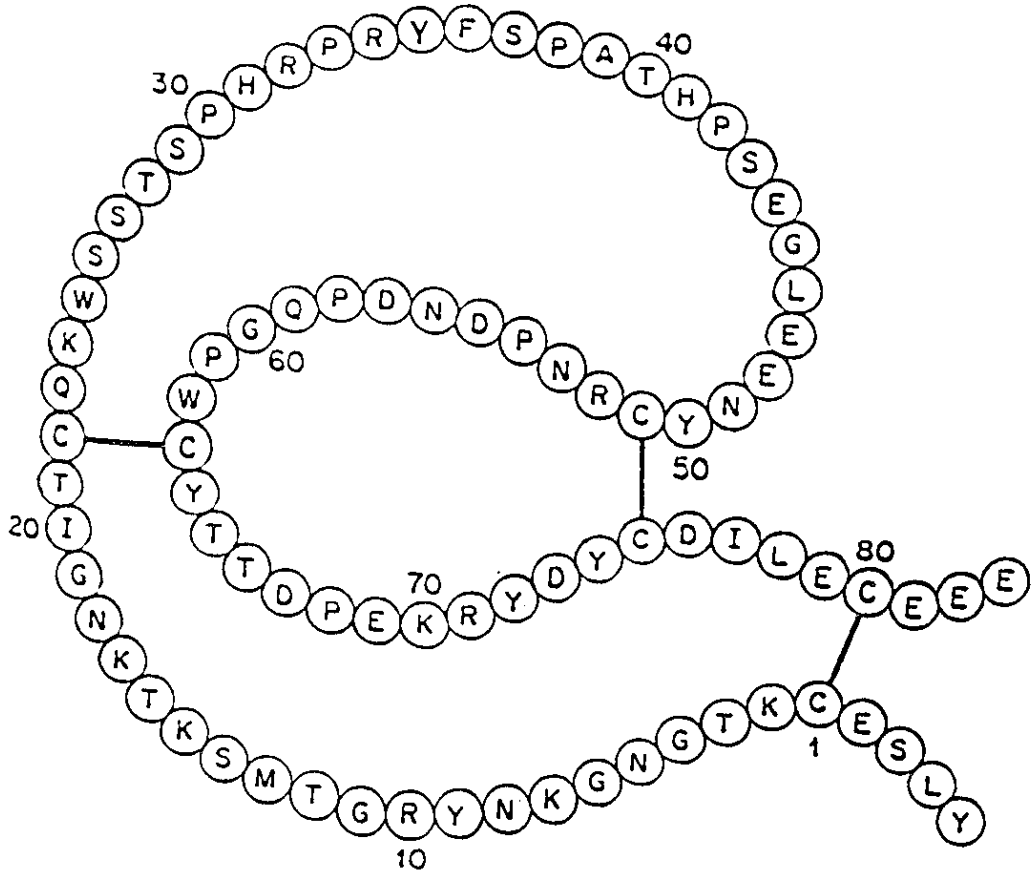
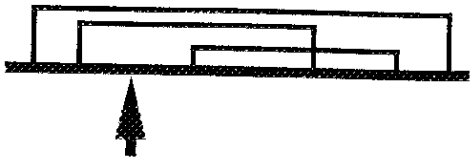

1
10
20
30
40
50
60
70
80

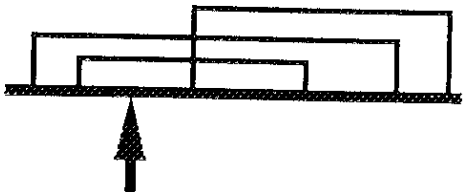
Figure 1.4 **Topological relationships between the disulfide bridges and the location of the reactive site in serine protease inhibitors.**

Each figure shows the pattern of the disulfide bridges. The arrow indicates the location of a reactive site.

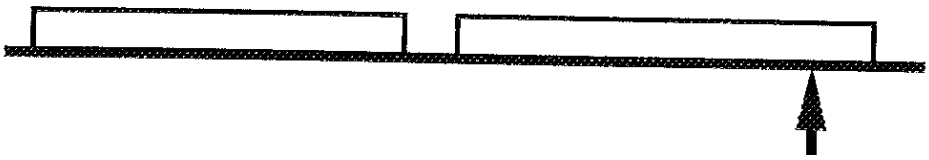
Kunitz-type



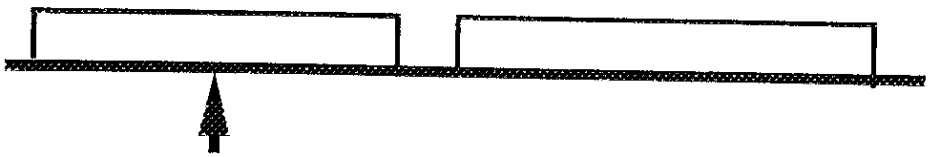
Kazal-type



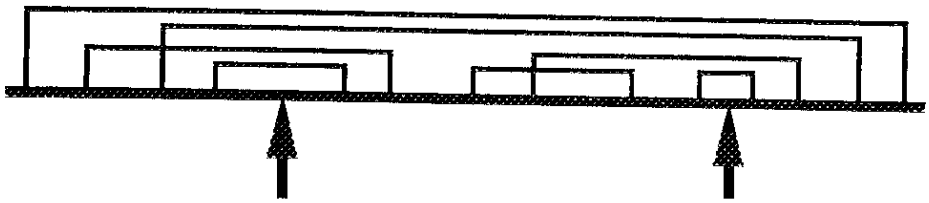
SSI type



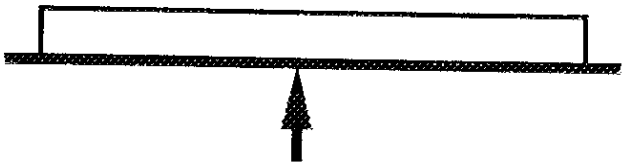
STI (Kunitz) type



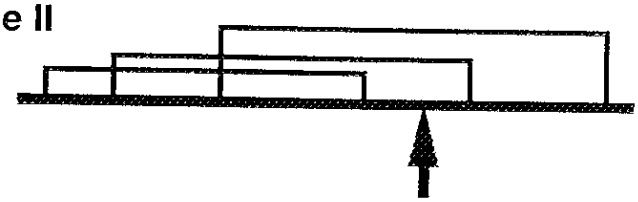
BBI type



Potato type I



Potato type II



CHAPTER TWO

MOLECULAR EVOLUTION

OF KRINGLE STRUCTURE

2.1 INTRODUCTION

2.1.1 Molecular mechanism of blood coagulation

From the view point of biological functions, proteases can be classified into two groups, digestive proteases and regulatory proteases. Although a digestive protease is the one that digests so-called junk proteins, a regulatory protease is the one that plays a regulatory role in a metabolic pathway or a biological cascade. The best-known regulatory proteases and their zymogens are those of the blood coagulation, fybrinolysis, and complement systems.

Blood coagulation is initiated by two separate pathways, the intrinsic and extrinsic pathways. In the extrinsic pathway, the tissue factor of cell-membrane proteins plays a critical role in the process (Furie et al. 1988). In the intrinsic pathway, on the other hand, all of the protein components are present in blood. Although

the extrinsic pathway consists of a single step of an enzymatic reaction, the intrinsic pathway includes seven consecutive steps of enzyme activation (Figure 2.1). Most of these consecutive steps are catalyzed by serine proteases. The reactions of these serine proteases are regulated by their specific plasma protein inhibitors. These two pathways eventually converge at a single pathway through which fibrinogen is converted into fibrin at the terminal stage. Fibrin is further degraded by plasmin that is produced in the process of fibrinolysis.

The precursor forms of the proteases in blood coagulation, as well as other regulatory proteases, are larger in molecular weights and more complex in molecular structures than those of the digestive proteases. In fact, the molecular weight of the tissue-plasminogen activator, a typical regulatory protease, is much larger than that of the pancreatic proteases which is one of the well-known digestive proteases. This is mainly because the tissue-type plasminogen activator has extra domains called "regulatory domains" in addition to a protease domain, although the pancreatic protease has only a protease domain. The molecular weights of the regulatory domains are often large, and in some cases it exceeds those of the protease domains. As will be explained in the next section, the regulatory domains play an essential role in interaction with other proteins and other non-protein cofactors. Thus, these domains certainly have regulatory functions.

2.1.2 Structure of proteases in blood coagulation

Regulatory plasma serine proteases are involved in the cascades of blood coagulation and fibrinolysis (Neurath 1984, 1989). Each of these active serine proteases consists of two polypeptide chains, A and B. The B-chain contains a catalytic domain that resembles trypsin, whereas the A-chain has no counterpart among the digestive serine proteases. The A-chain contains regulatory domains that interact with cofactors essential for physiological regulation. The regulatory domains are classified into four major types; kringle domains, vitamin K-dependent calcium-binding domains, finger domains, and growth factor domains (Figure 2.2).

The number of kringles varies with the type of proteases, but each of "kringle" domains consists of approximately 80 amino acids. A kringle is a characteristic secondary structure that is formed by three pairs of intrachain disulfide bonds (Figure 1.3). Such a structure was first found in human prothrombin (Waltz et al. 1977).

It is known that urinary urokinase (namely, urinary plasminogen activator: uPA) and coagulation factor XII contain one kringle each (McMullen and Fujikawa 1985). Tissue-type plasminogen activator (tPA) and prothrombin have two each (Penica et al. 1983, Sandra et al. 1983). The growth factor found in hepatocytes has four kringles (Nakamura et al. 1989), and plasminogen has five (Schaller et al. 1985). It is of particular

interest to know how such a mosaic protein emerged from its ancestor during evolution (Patthy 1985).

In this chapter, I would focus our attention on the evolution of kringle structures.

2.1.3 Apolipoprotein(a)

The lipoprotein(a) [Lp(a)] is a complex molecule of lipids and a protein that is present in human plasma. Its concentration in human plasma is significantly correlated with the risk of heart disease (Utermann 1989). Recent studies also demonstrate a correlation between high plasma concentrations of Lp(a) and atherosclerosis (McLean et al. 1987). Lp(a) is assembled from a low-density lipoprotein and a large hydrophobic glycoprotein. The low-density lipoprotein is composed of lipids and apolipoprotein B-100, and the hydrophobic glycoprotein is called apolipoprotein(a) [apo(a)] (McLean et al. 1987). In other words, Lp(a) is a form of apo(a) which is bound to apolipoprotein B-100 by a disulufide linkage.

Recently, 38 kringles have been found in human apolipoprotein(a) (Mclean et al. 1987). Although apo(a) has a great size heterogeneity, it usually consists of 38 kringle domains in the amino terminal and a serine protease domain in the carboxyl

terminal (Figure 2.3). Ten kringles, at least, are in the sequence of rhesus macaque (Tomlinson et al. 1989); each of kringles is very similar to its human counterpart (Figure 2.3). More kringles in rhesus macaques may be possibly found by a further analysis of DNA. A large number of kringles in apo(a) seem to be derived from the kringles of a plasminogen-type protein, since they are homologous to the fourth kringle of plasminogen (plgen 4).

2.1.4 Hepatocyte growth factor

Four kringles were also found in hepatocyte growth factor (HGF). HGF is a hepatotrophic factor that acts as a trigger for liver regeneration after partial hepatectomy and liver injury, and it is the most potent mitogen for mature parenchymal hepatocytes in the primary culture. In practice, the HGF stimulates DNA synthesis of adult rat hepatocytes in the primary culture. HGF is also known as a human lung fibroblast-derived mitogen that was isolated from conditioned medium of fibroblast in human embryonic lungs (Rubin et al. 1991, Okajima et al. 1990).

The HGF has a molecular mass of 82kDa and is found to be a heterodimer which is dissociated into a large subunit of 69 kDa and a small one of 34 kDa (Nakamura et al. 1989). Four kringles are found in the large subunit. Thus, HGF is structurally related to

plasminogen (Figure 2.4). The nucleotide sequence of HGF reveals that both subunits are encoded by a single open reading frame. These charcteristics show that HGF is not identical to any known growth factors such as a platelet-derived growth factor.

To elucidate the evolutionary origin of these kringle domains of serine proteases in blood coagulation, apo(a), and HGF, I constructed a phylogenetic tree for these kringles in various organisms.

2.2 MATERIALS AND METHODS

2.2.1 Sequences used

I constructed phylogenetic trees by use of all the available sequences of amino acids and nucleotides for the kringle domains. The 50 amino acid sequences and the 45 nucleotide sequences of various serine proteases were compiled from the PIR (version 25) protein database and from GenBank (version 64) of the DDBJ/EMBL/Genbank DNA database.

2.2.2 Genetic distance

Using the sequence data, I first conducted the homology alignment with maximum match (Figure 2.5). The substitution numbers were estimated from comparisons between each pair of sequences. For the correction of multiple substitutions at a site, Kimura's equation was used; $d_a = -\log_e(1 - p - 0.2p^2)$ (Kimura 1983), where d_a and p are the number of amino acid substitutions per site and the proportion of different amino acids, respectively. A phylogenetic tree for amino acid sequences was constructed by use of the d_a values by the unweighted pair-grouping (UPG) method, assuming that a substitution rate is constant over time (Nei 1975).

I also constructed a phylogenetic tree by the neighber-joining (NJ) method (Saitou and Nei 1987), which does not require the assumption of the rate constancy. The tree topology obtained by the NJ method was virtually the same as that obtained by the UPG method. In the case of a phylogenetic tree for nucleotide sequences, the total number of nucleotide substitutions for each pair of kringles was estimated by the six-parameter method (Kimura 1983, Gojobori et al. 1982).

2.3 RESULTS

2.3.1 Evolutionary origin of kringle structures

Figure 2.6 shows a phylogenetic tree of all kringles that was constructed by the UPG method, using amino acid sequences. The majority of kringles could be separated into three major groups indicated by I, II, and III. By this classification, all kringles of apo(a) except the 38th (apo 38) belong to group I. The second and third kringles (plgen 2 and plgen 3) of plasminogen and the second kringle (HGF 2) of hepatocyte growth factor are in group II. Apo 38, plgen 1, and plgen 5 in group III are differentially separated from the other 37 kringles of apo(a).

2.3.2 Molecular evolution of apolipoprotein(a)

The rate of amino acid substitution was calculated to be 0.956 x 10^{-9} per amino acid site per year, assuming that the divergence of mammals occurred 80 million years ago. Using this rate, I estimated the time of the evolutionary events that must have taken place in the past.

The evolutionary development of kringles and apo(a) in human and simian species is summarized in Figure 2.7. The followings are based on the estimation of the divergence times for various branching points in the phylogenetic tree. An ancestral gene must have been of the plasminogen type with one kringle and one serine protease domain. About 500 million years ago, the kringle was duplicated into two domains, one similar to an ancestral domain in groups I and II, and the other similar to an ancestor in group III. The former could have been similar to plgen 4 (and plgen 4') and the latter similar to plgen 5 (plgen 5'; corresponding to apo 38). Some 300 million years ago, plgen 4' was triplicated into three domains; plgen 1', plgen 2', and plgen 3'. Thus, the tree shows that all the five kringles have existed in the plasminogen since about 300 million years ago. Therefore, it is considered that the ancestral molecule of apo(a) was a plasminogen-type protein having five kringles.

2.4 DISCUSSION

2.4.1 A model for the evolution of apolipoprotein(a)

The phylogenetic tree obtained in the present analysis suggests that the evolution of apo(a) can be explained by the following model. About 80 million years ago, the plasminogen-like gene was duplicated into two genes; one formed apo(a) and the other became the present plasminogen. It is supported by a cytogenetical study that genes for human apo(a) and plasminogen are located very close to each other on chromosome 6 with loci at q2.6 - 2.7 and q2.7, respectively (Frank et al. 1988). Since apo38 is the most similar to plgen 5, apo 38 was probably derived from plgen 5 about 80 million years ago. Separation of the ancestral gene of the 37 other kringles from plgen 4' was at the almost same time as when apo 38 diverged from its ancestor. Then, a portion containing plgen 1' to plgen 3' was deleted in apo(a) of human and simian species. Thereafter, plgen 4' underwent multiple duplications about 2 to 3 million years ago, giving rise to the 37 kringles in humans. In particular, the 36th kringle acquired the ability to bind to apolipoprotein B-100 at an early stage of these duplications. In simian species, the process was similar but independent of the human homologue because the divergence

between human and rhesus macaque had occurred (about 45 million years ago) much earlier than the multiple duplication of plgen 4' in human. In particular, plgen 5' was deleted in rhesus macaque while plgen 4' (apo28 - apo37) was multiplied. A similar result was obtained by another tree that was constructed by the neighbor-joining method (Saitou and Nei 1987) (data not shown). Note that this tree-making method does not require the assumption of a constant rate of amino acid substitution.

To clarify the evolutionary processes of kringles in human and simian species further, I constructed a tree by use of the nucleotide sequences. The tree is much more informative at the nucleotide sequence level than at the amino acid sequence level, because the identical sequences of amino acids such as (apo2 - 29) were different sequences as the corresponding sequences of nucleotides. However, this tree topology was essentially the same as the one in Figure 2.6. The rate of nucleotide substitution was calculated to be 1.48×10^{-9} per site per year. Using this rate, the multiple duplications of plgen 4' in human apo(a) was estimated to have occurred 2 to 3 million years ago. This is consistent with the results that were obtained from the tree of amino acid sequences.

Tomlinson et al. (1989) speculated that apo(a) emerged from its ancestor about 40 million years ago. However, the present analysis shows that the ancestral molecule of apo(a) evolved from the plasminogen-like gene about 80 million years ago. It implies that apo(a) appeared at the almost same time that the mammalian

divergence took place. If this is the case, apo(a) may be found in a certain variety of mammalian species. It is also possible that the number of kringles in the proteins homologous to apo(a) would be found to vary largely with the species if the DNA of apo(a) in other mammals was sequenced. The regions that were multiply duplicated in apo(a) may be unstable because further duplications are likely to occur in this region. These duplications of plgen 4' can be the cause of the size heterogeneity of apo(a). In fact, it is known that apo(a) has several isoforms in human and simian species (Uterman 1989).

Kringles are structurally autonomous units (Trexler and Patthy 1983). However, their consensus function is not well understood. Accumulation of sequence data from various proteins would provide us with a better understanding of kringles and an answer to the question of where they came from (Uterman 1989).

2.4.2 Function of kringle structures

Although the function of the kringle structures has many of unclear points, Table 2.1 shows possible functions that have been known so far (Holland and Blake 1990). A known function of the kringle structures is basically a binding activity. For example, the fourth kringle of plasminogen binds to fibrin clots in the blood

vessels, so that plasminogen digests fibrin clots by its protease activity. It is thus thought that kringles have an regulatory role in the activation of plasminogen by locating their substrates in the blood vessels.

There may be two mechanisms of regulating the protease activity. One is, of course, the regulation by a protease inhibitor. The protease inhibitors act as an on-and-off switch for the activation of proteases. The other is the regulation by the kringle domain of protease itself, as discussed in this section. The binding activity of the kringles might help the protease to identify and locate its appropriate substrates in organismic tissues. These regulatory features of kringle structures are a key for understanding the evolutionary process of proteases.

References (Chapter 2)

Belin D, Vassalli J-D, Combepine C, Godeau F, Nagamine Y, Reich E, Kocher H P and Duvoisin R M (1985) Cloning, nucleotide sequencing and expression of cDNAs encoding mouse urokinase-type plasminogen activator *Eur. J. Biochem.*148, 225-232.

Forsgren M, Raden B, Israelsson M, Larsson K and Haden L-O (1987) Molecular cloning and characterization of a full-length cDNA clone for human plasminogen. *FEBS Lett.* 213, 254-260.

Frank S L, Klisak I, Sparkes R S, Mohandas T, Tomlinson J E, McLean J W, Lawn R M and Lusis A J (1988) The apolipoprotein(a) gene resides on human chromosome 6q26-27, in close proximity to the homologous gene for plasminogen *Hum. Genet.* 79, 352-356.

Furie B and Furie C B (1988) The molecular basis of blood coagulation. *Cell* 53, 505-518.

Gojobori T, Ishii K and Nei M (1982) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *J. molec. Evol.* 18, 414-423.

Holland S K and Blake C C F (1990) Proteins, Exons, and molecular evolution. in Intervening sequences in evolution and development. Edited by Stone E M and Schwartz R J Oxford university press, New York.

Kimura M. (1983) The Neutral Theory of Molecular Evolution (Cambridge University Press, Cambridge, 1983).

MacGillivray R T A and Davie E W (1984) Characterization of bovine prothrombin mRNA and its translation product. *Biochemistry* 23, 1626-1634.

McLean J W, Tomlison J E, Kuang W-J, Eaton D L, Ellson Y C, Fless G M, Scanu A M and Lawn R M (1987) cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature* 330, 132-137.

McMullen B A and Fujikawa K. (1985) Amino acid sequence of the heavy chain of human α-Factor XIIa (activated Hageman factor). *J. Biol. Chem.* 260, 5328-5341.

Miyazawa K, Tsubouchi H, Naka D, Takahashi K, Okigaki M, Arakaki N, Nakayama H, Hirono S, Sakiyama O, Takahashi K, Gohda E, Daikuhara Y, and Kitamura N (1989) Molecular cloning and sequence analysis of cDNA for human hepatocyte growth factor *Biochem. Biophys. Res. Commun.* 163, 967-973.

Nagamine Y, Pearson D, Altus M S and Reich E (1984) cDNA and gene nucleotide sequence of porcine plasminogen activator. *Nucli. Acid. Res.* 12, 9525-9541.

Nakamura T, Nishizawa T, Hagiya M, Seki T, Shimonishi M, Sugimura A, Tashiro K. and Shimizu S. (1989) Molecular cloning and expression of human hepatocyte growth factor. *Nature* 342, 440-443.

Nei M (1975) Molecular population Genetics and Evolution (North-Holland, Amsterdam, 1975).

Neurath H (1984) Evolution of proteolytic enzymes. *Science* 224, 350-357.

Neurath H (1989 Proteolytic enzymes edited by Beynon R., J. and Bond, J., S. (IRL PRESS, Oxford, 1989)

Okajima A, Miyazawa K, and Kitamura N. (1990) Primary structure of rat hepatocyte growth factor and induction of its mRNA during liver regeneration following hepatic injury. *Eur. J. Biochem* 193, 375-381.

Patthy, L. (1985) Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. *Cell* 41, 657-663.

Pennica D, Holmes W E, Kohr W J, Harkins R N, Vehar G A, Bennet W M, Yelventon E, Seeburg P H, Heyneker L and Goeddel D V (1983) Cloning and expression of human tissue-type plasminogen activator cDNA in *E. coli. Nature* 301, 214-221.

Rubin S J, Chan A M.-L, Bottaro D P, Burgess W H, Taylor G W, Cech C A, Hirshfield W D, Wong J, Miki T, Finch W P, and Aaronson S. (1991) A broad-spectrum human lung fibroblast-derived mitogen is a variant of hepatocyte growth factor. *Proc. Natl. Acad. Sci. USA* 88, 415-419.

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Estimation of

average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *Mol. Biol. Evol.* 4, 406-425.
Sandra J F D, Ross T, MacGillivary A and Davie E W (1983) Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin. *Biochemistry* 22, 2087-2097.

Schaller J, Moser P W, Dannegger-Muller G A K, Roselt S J, Kampfer V and Rikli E E (1985) Complete amino acid sequence of bovine plasminogen comparison with human plasminogen. *Eur. J. Biochem* 149, 267-278.

Tomlinson J E, McLean J W and Lawn R M (1989) Rhesus monkey apolipoprotein(a). *J. Biol. Chem.* 264, 5957-5965.

Trexler M and Patthy L (1983) Folding autonomy of the kringle 4 fragment of human plasminogen. *Proc. Natl. Acad. Sci. USA* 80, 2457-2461.

Utermann G (1989) The mystery of lipoprotein(a). *Science* 246, 904-910.

Verde P, Stoppelli M P, Galeffi P, Nocera P D and Blasi F (1984) Identification and primary sequence of an unspliced human urokinase poly(A)$^+$ RNA. *Proc. Natl. Acad. Sci. USA* 81, 4727-4731.

Walz D A, Hewett-Emmett D and Seegers W H (1977) Amino acid sequence of human prothrombin fragments 1 and 2. *Proc. Natl. Acad. Sci. USA* 74, 1969-1972.

Table 2.1. Binding Functions of Kringle Domains.

Kringle domains in proteins	Binding
prothrombin (A)	Factor V (?)
prothrombin (S)	Factor V
plasminogen 1	Arginine
plasminogen 2	Lysin
plasminogen 3	Fibrin
plasminogen 4	Arginin, tetranectin
plasminogen 5	Lysin
tissue plasminogen activater 2	Fibrin
urokinase	?
apolipoprotein(a) 36	apolipoprotein B

Figure 2.1 **Schematic pathways of the blood coagulation and fibrinolysis.**

The proteins in the intrinsic pathway include factors XII, XI, IX, VIII, X, and V, as well as prothrombin and fibrinogen. The proteins in the extrinsic pathway, which is initiated by the action of tissue factors located on the cell surface, include factors VII, X, and V, as well as prothrombin and fibrinogen. The cascade reactions culminate in the conversion of fibrinogen into fibrin and the formation of a fibrin clot. Certain reactions such as the activation of factor X and prothrombin take place on the membrane surface. The macromolecular complexes on the membrane surfaces are indicated by shaded rectangles. Numbered factors are abbreviated by FXII, FXI, and so forth. Other abbreviations are as follows: HMWK, high molecular weight kininogen; TF, tissue factor; PT, prothrombin; FG, fibrinogen; F, fibrin; tPA, tissue-type plasminogen activator; uPA, urokinase; PG, plasminogen; P, plasmin; Fdeg, degraded fibrin.

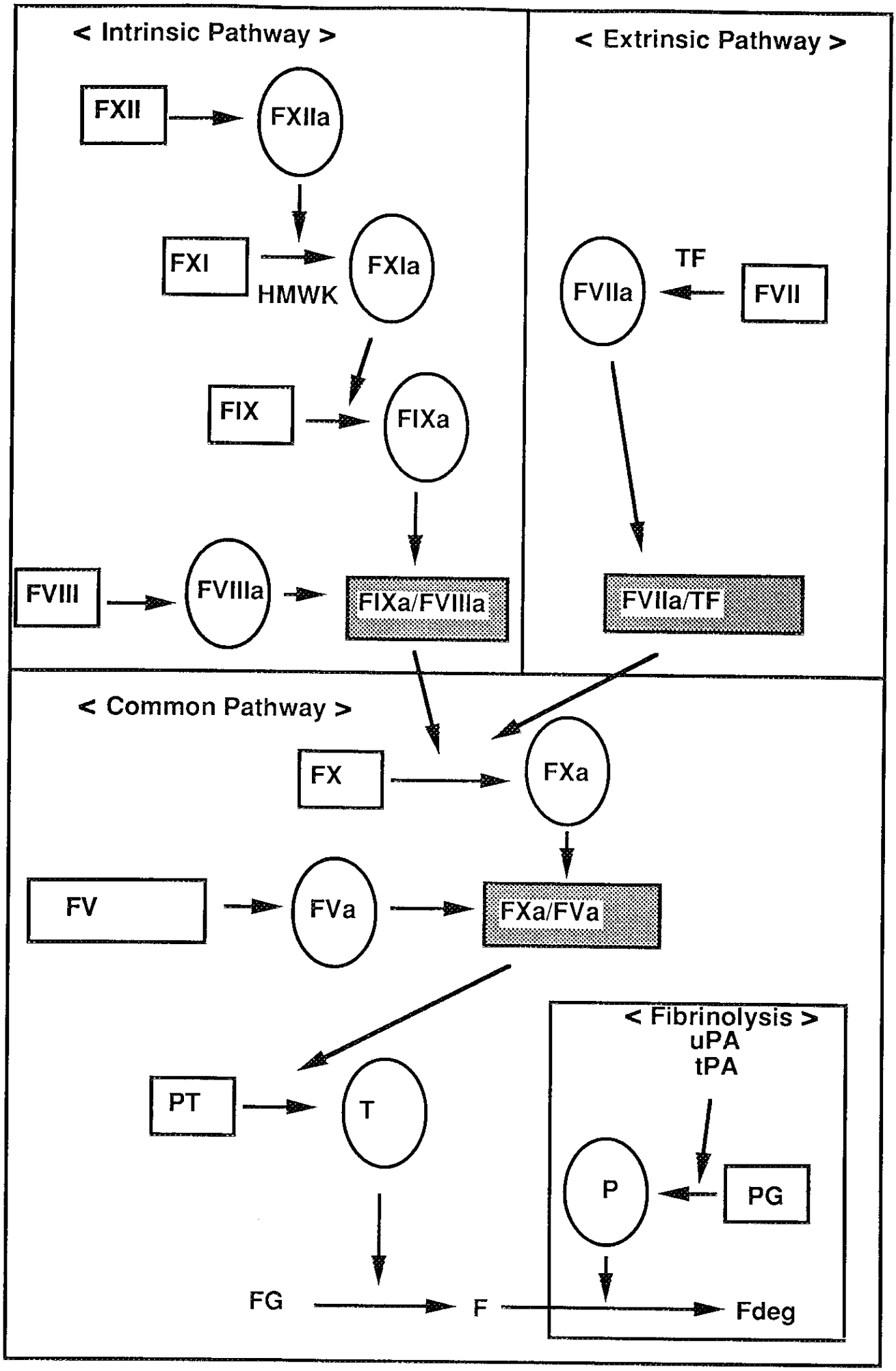
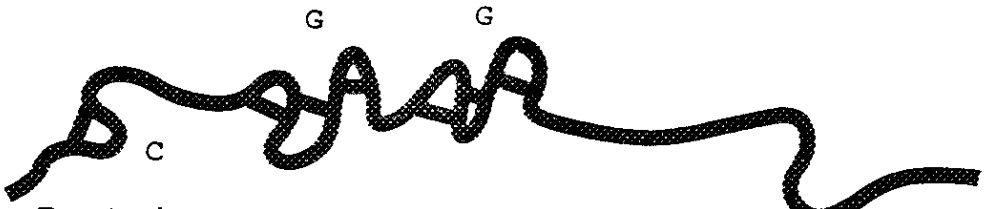
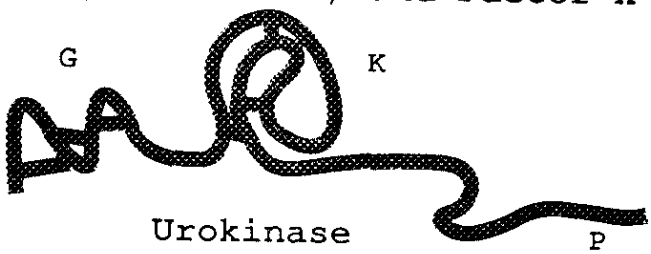

< Intrinsic Pathway >
FXII
FXIIa
FXI
HMWK
FXIa
FIX
FIXa
FVIII
FVIIIa
FIXa/FVIIIa
< Extrinsic Pathway >
TF
FVIIa
FVII
FVIIa/TF
< Common Pathway >
FX
FXa
FV
FVa
FXa/FVa
PT
T
FG
F
Fdeg
< Fibrinolysis >
uPA
tPA
P
PG

Figure 2.2 **Mosaic structures of the serine proteases in blood coagulation.**

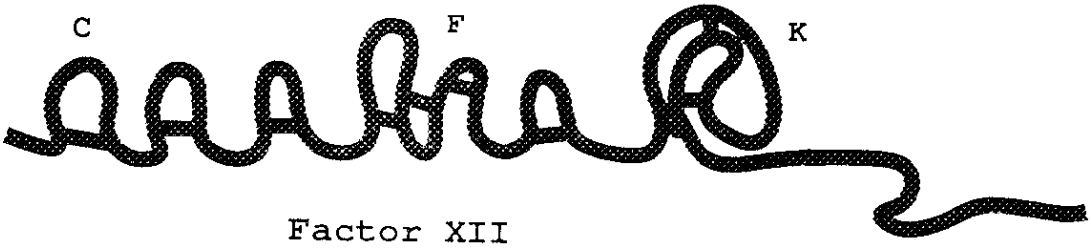
A serine protease consists of several regulatory domains. It shows a typical feature of the mosaic protein. Abbreviations: C, vitamin K-dependent calcium-binding domain; G, growth factor domain; K, kringle domain; F, finger domain; P, protease domain.



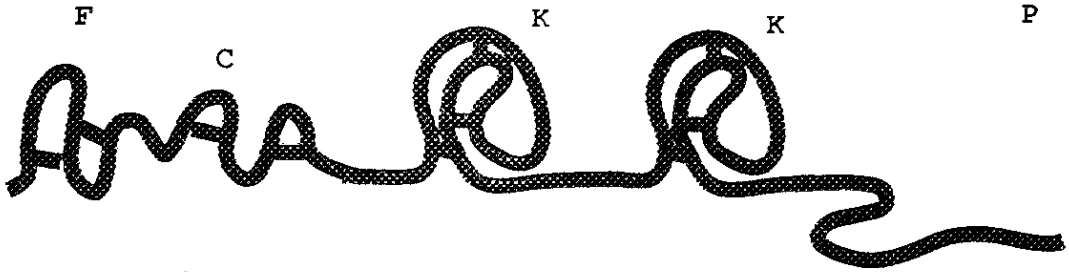
Protein C, Factor IX, and Factor X



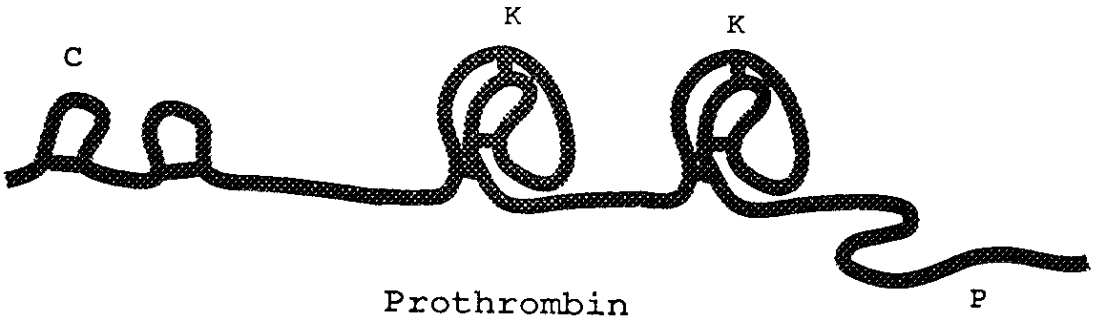
Urokinase



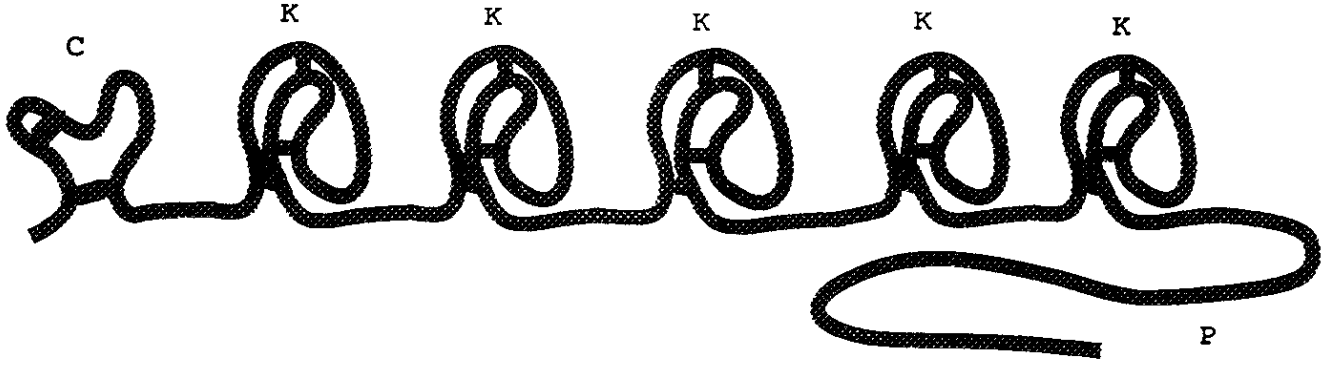
Factor XII



Tissue-type plasminogen activator



Prothrombin



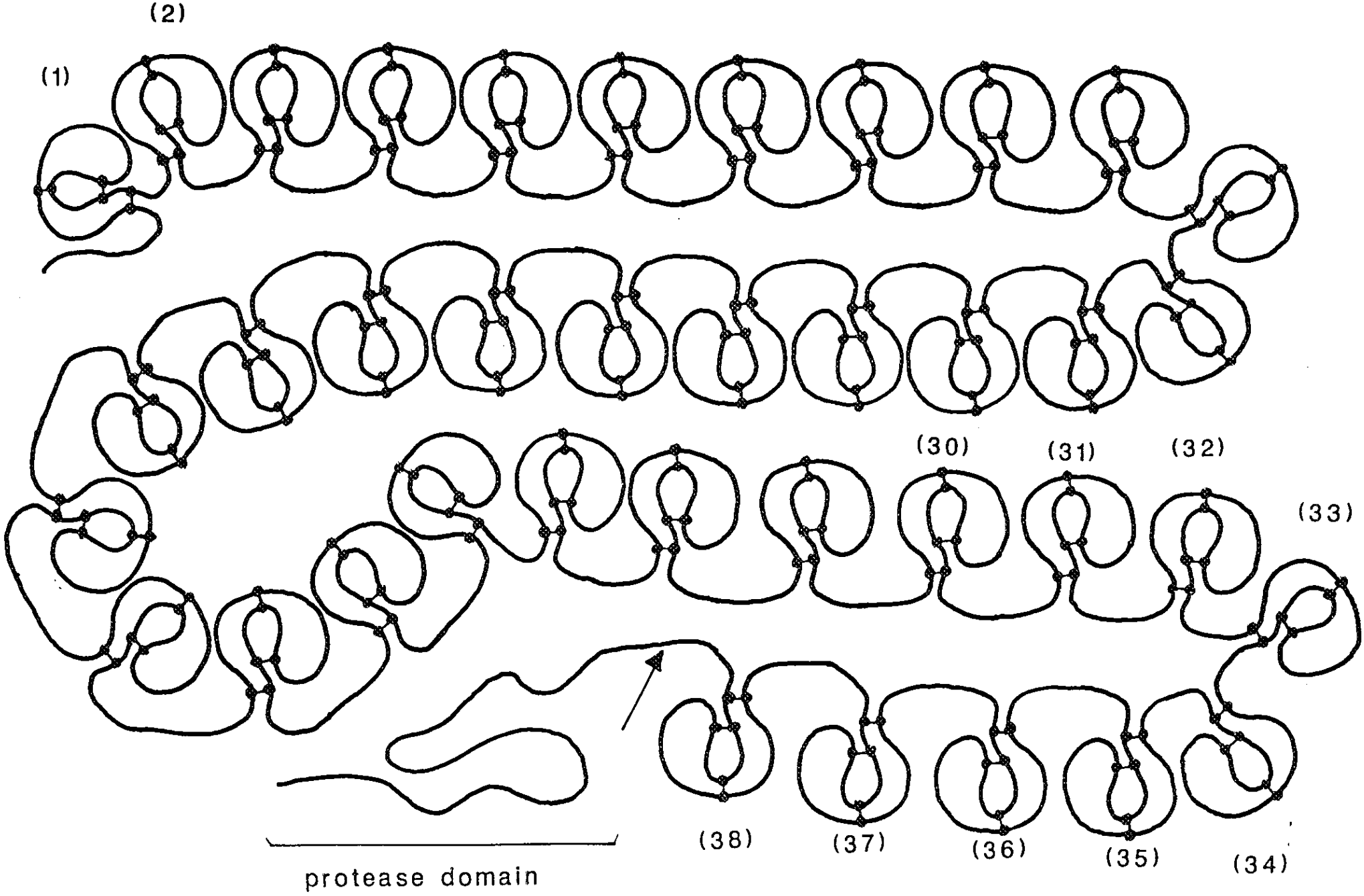
Plasminogen

Figure 2.3 **Structure of human and simian apolipoprotein(a).**

(a) The structure of human apolipoprotein(a).

(b) The structure of rhesus apolipoprotein(a).

(A) Structure of apolipoprotein (a)



(B) Structure of Rhesus apolipoprotein(a)

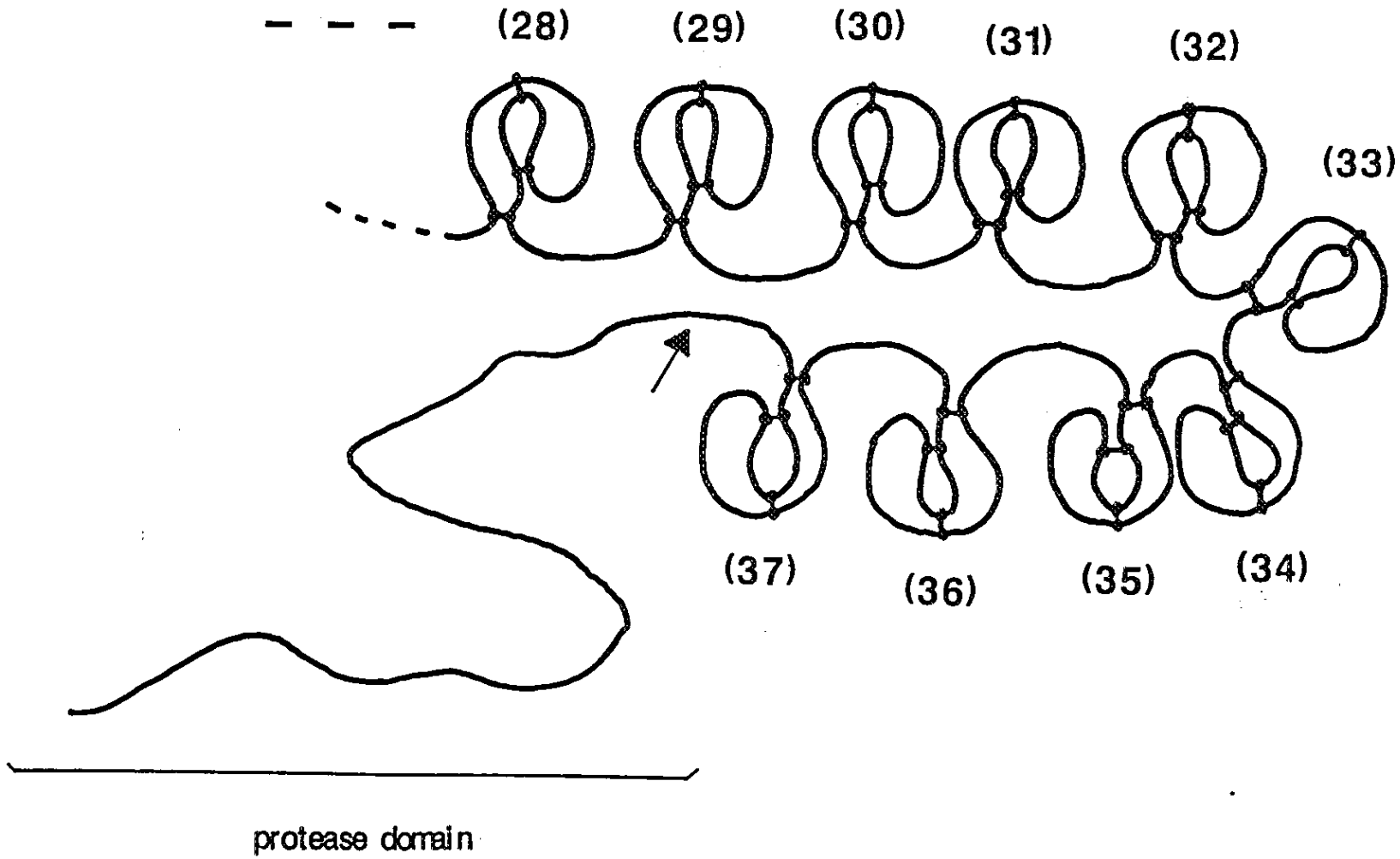
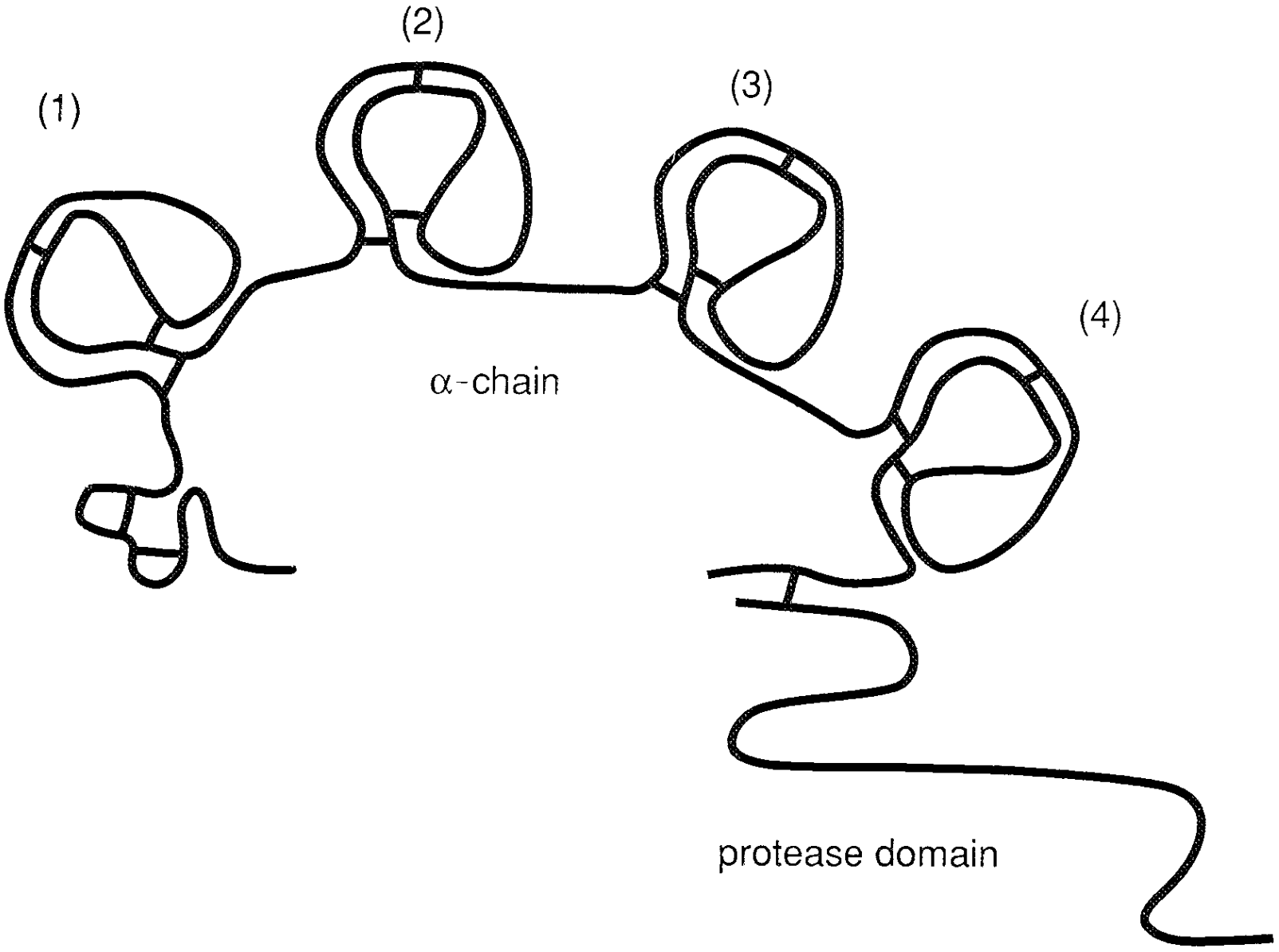


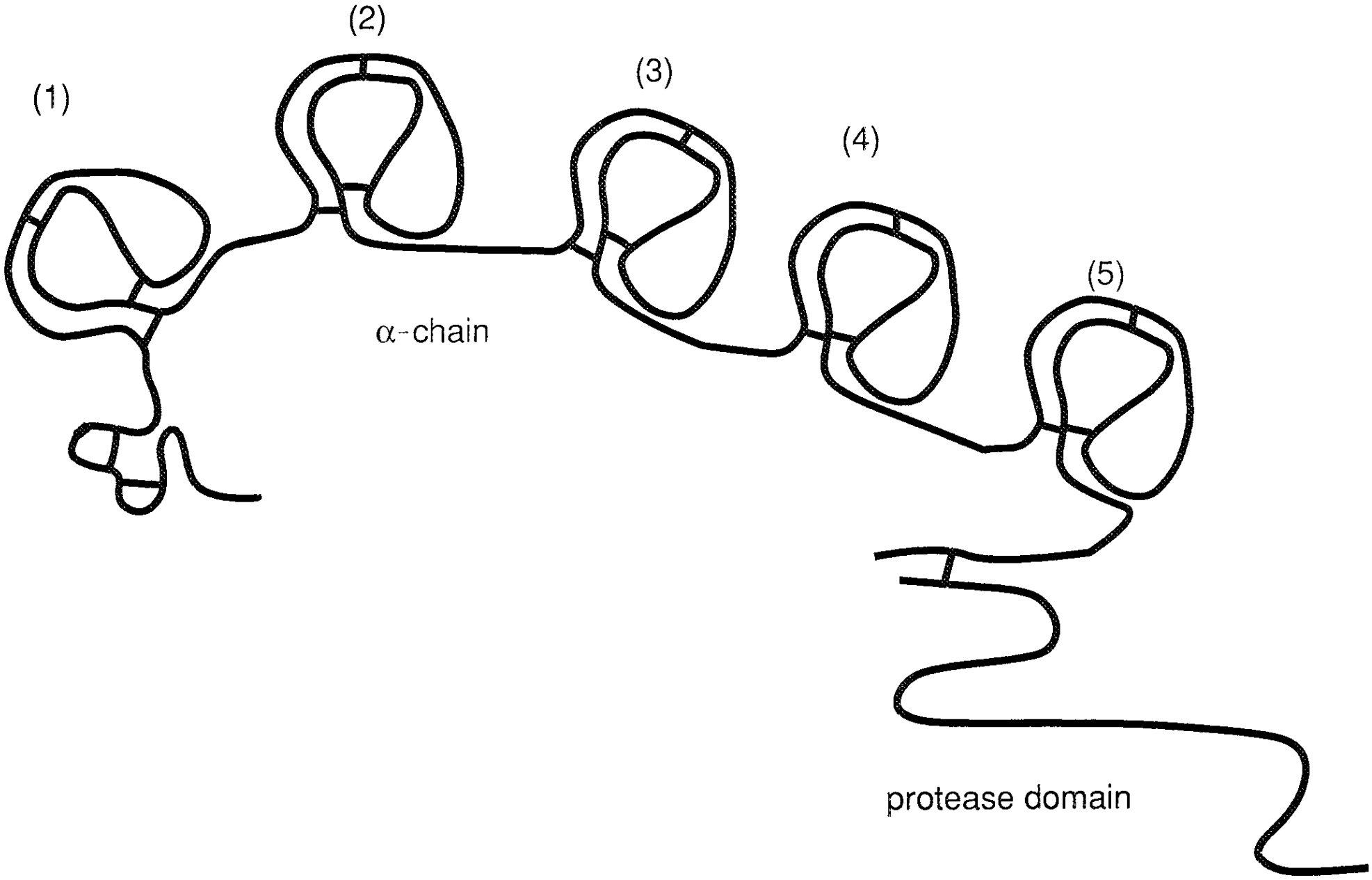
Figure 2.4 **Structures of plasminogen and HGF.**

(a) The structure of human plasminogen.

(b) The structure of human HGF.



(B) Structure of human HGF



(A) Structure of human plasminogen

Figure 2.5 Sequence alignment of kringle structures.

This figure shows a part of the alignment of sequences used in this study. The solid circles show the position of cysteine residues. Abbreviations: tpa, tissue-type plasminogen activator; uk, urokinase; proA, A chain of prothrombin; proS, S chain of prothrombin; Plg, plasminogen; apo, apolipoprotein(a); hu, human; mo, mouse; po, porcine; bo, bovine; mn, rhesus. The number in the name of a sequence represents the kringle number from the N-terminus of each protein.

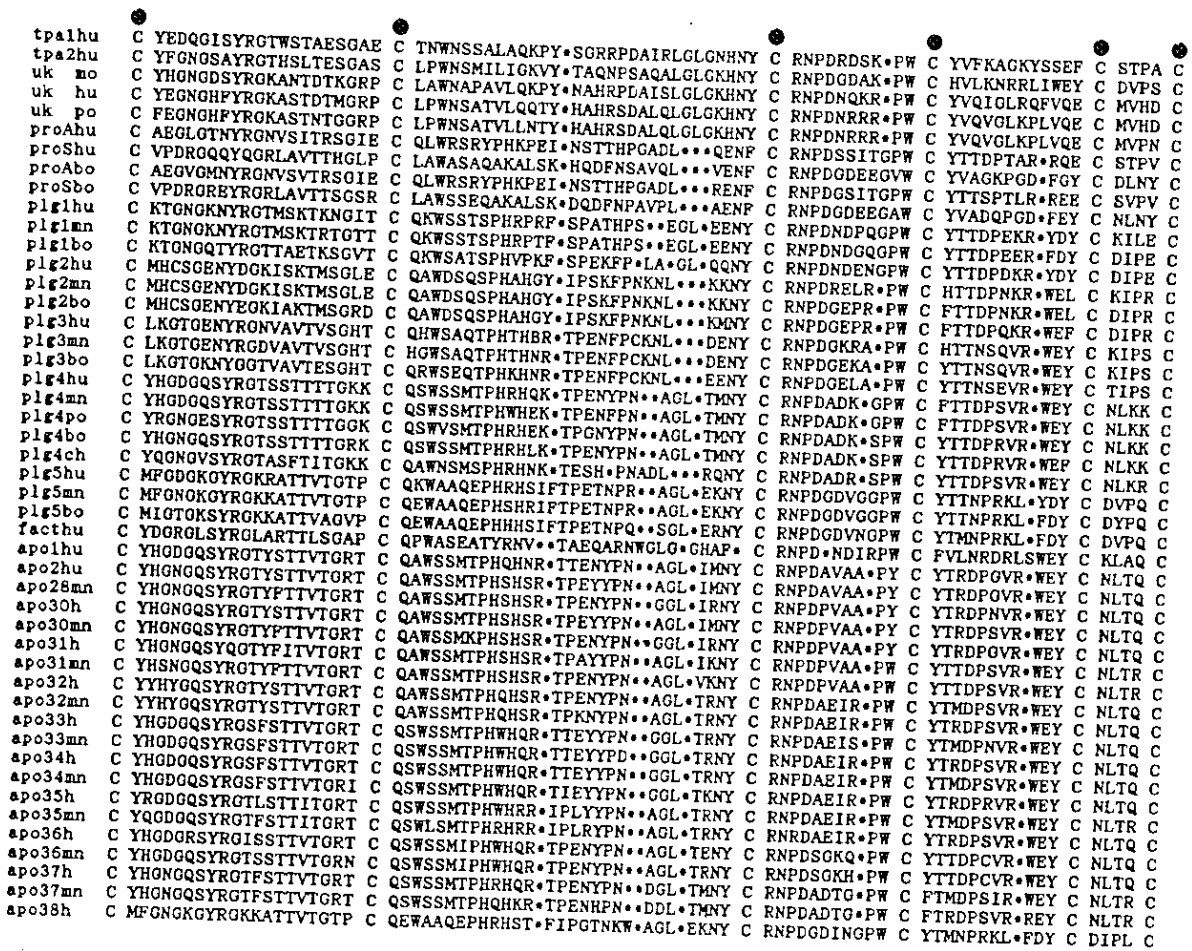


Figure 2.6 A phylogenetic tree of kringle domains.

The tree was constructed by analysis of amino acid sequences. Three major groups are indicated by I, II, and III. Abbreviations: tPA, tissue-type plasminogen activator; uk, urokinase; plgen, plasminogen; apo, apolipoprotein(a); HGF, human hepatocyte growth factor; hu, human; bo, bovine; po, porcine; mo, mouse. The number in the name of a sequence represents the kringle number from the N-terminus of each protein. Data sources: pro(hu) (Walz etal. 1977, Sandra et al. 1983), factorXII(hu) (McMullen et al. 1085), tPA(hu) (Pennica et al. 1983), HGF(hu) (Nakamura et al. 1989), plgen(bo) (Schaller et al. 1985), apo(hu) (Mclean et al. 1987), apo(rhesus) (Tomlinson et al. 1989), plgen(rhesus) (Tomlinson et al. 1989), pro(bo) (MacGillivray et al. 1984), uk(mo) (Belin et al. 1985), uk(hu) (Verde et al. 1984), uk(po) (Nagamine et al. 1984), and plgen(hu) (Forsgren et al 1987).

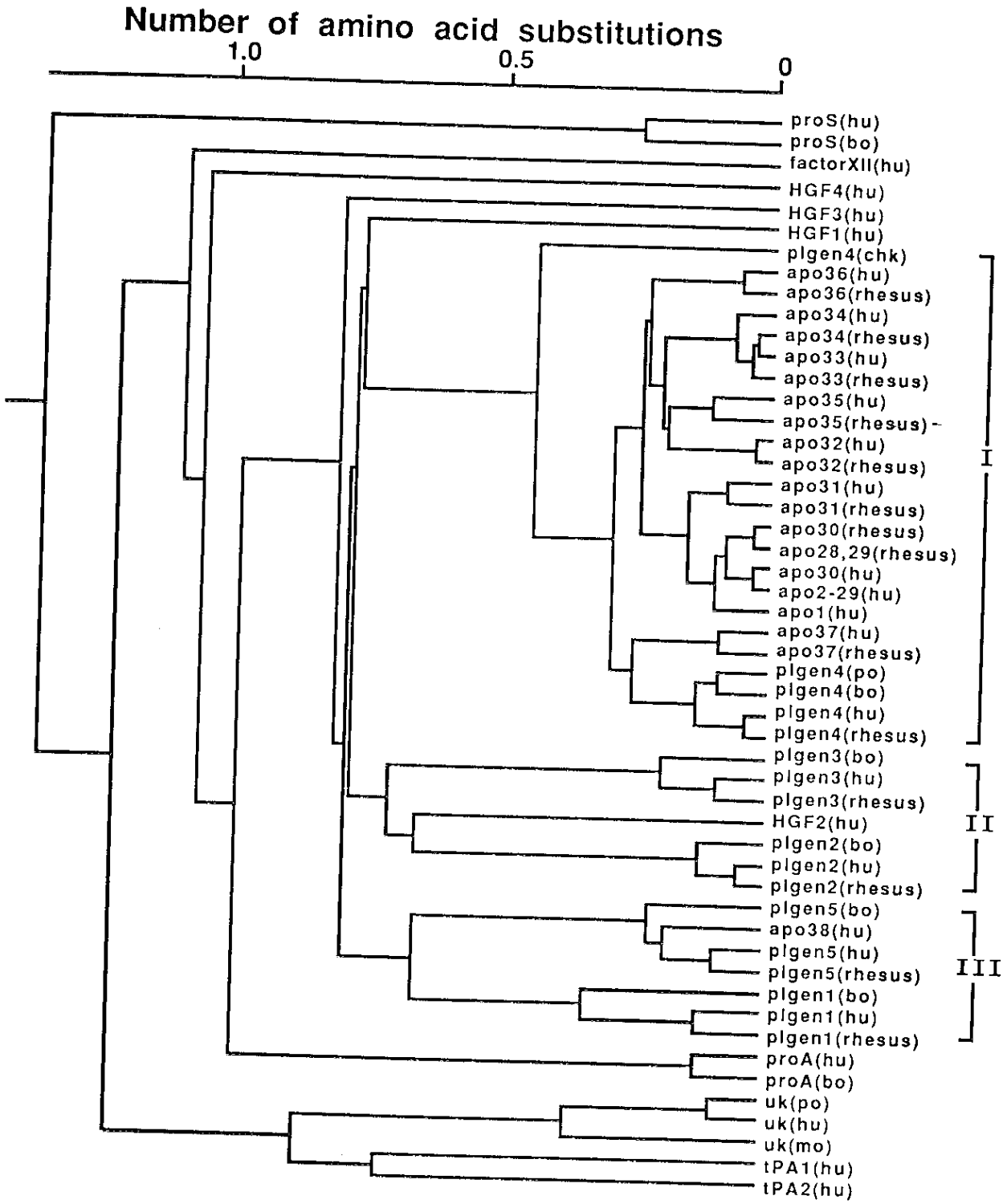
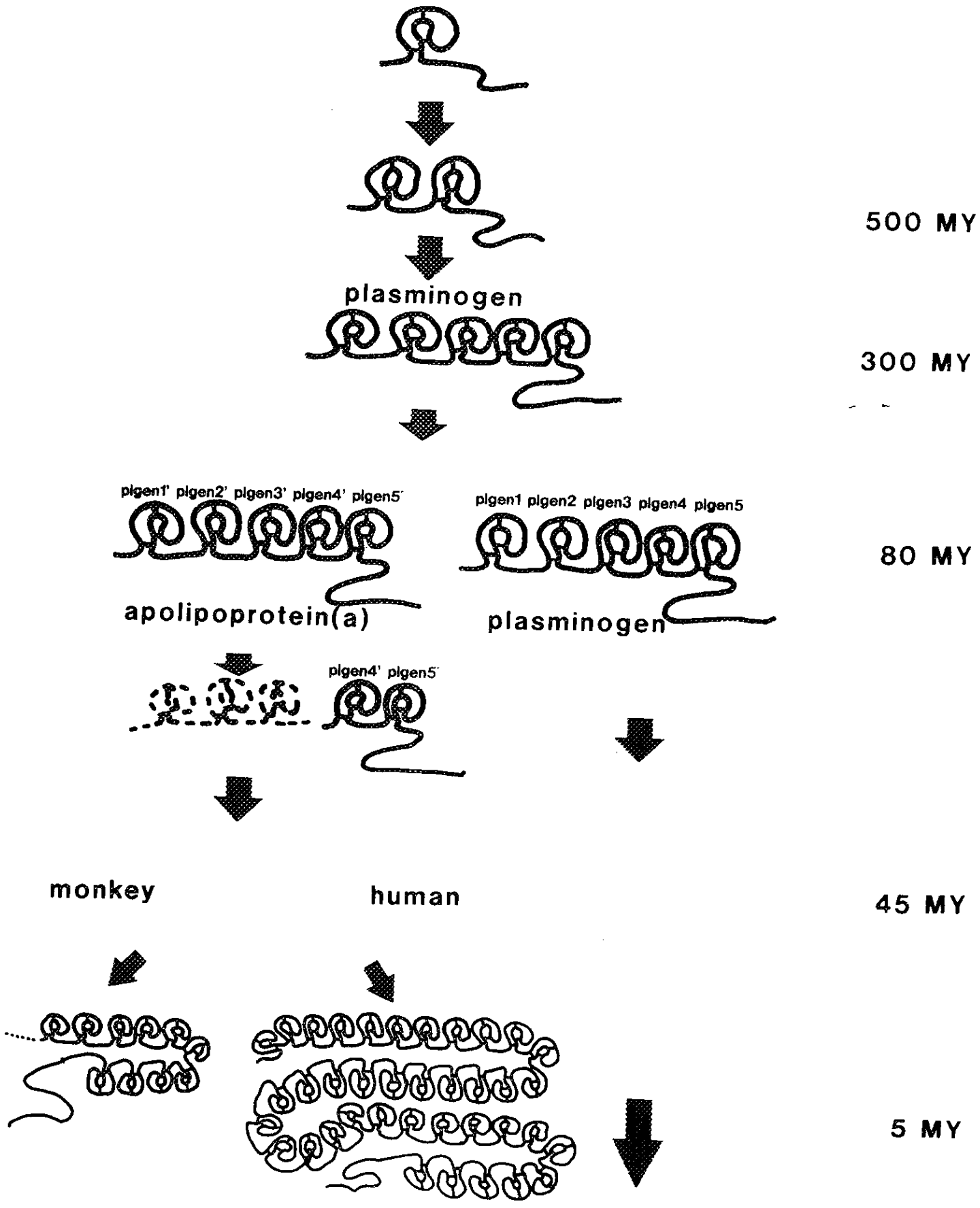

Number of amino acid substitutions
1.0
0.5
0
proS(hu)
proS(bo)
factorXII(hu)
HGF4(hu)
HGF3(hu)
HGF1(hu)
plgen4(chk)
apo36(hu)
apo36(rhesus)
apo34(hu)
apo34(rhesus)
apo33(hu)
apo33(rhesus)
apo35(hu)
apo35(rhesus)
apo32(hu)
apo32(rhesus)
apo31(hu)
apo31(rhesus)
apo30(rhesus)
apo28,29(rhesus)
apo30(hu)
apo2-29(hu)
apo1(hu)
apo37(hu)
apo37(rhesus)
plgen4(po)
plgen4(bo)
plgen4(hu)
plgen4(rhesus)
plgen3(bo)
plgen3(hu)
plgen3(rhesus)
HGF2(hu)
plgen2(bo)
plgen2(hu)
plgen2(rhesus)
plgen5(bo)
apo38(hu)
plgen5(hu)
plgen5(rhesus)
plgen1(bo)
plgen1(hu)
plgen1(rhesus)
proA(hu)
proA(bo)
uk(po)
uk(hu)
uk(mo)
tPA1(hu)
tPA2(hu)
I
II
III

Figure 2.7 Evolutionary changes of apolipoprotein(a).

See the text for detailed explanation.


plasminogen
500 MY
300 MY
plgen1' plgen2' plgen3' plgen4' plgen5'
apolipoprotein(a)
plgen1 plgen2 plgen3 plgen4 plgen5
plasminogen
80 MY
plgen4' plgen5'
monkey
human
45 MY
5 MY

CHAPTER THREE

EVOLUTION OF SERINE PROTEASE

3.1 INTRODUCTION

3.1.1 Serine protease as an enzyme

As mentioned in Chapter 1, the serine proteases that have been characterized in prokaryotes and eukaryotes fall into two families; families of serine protease I and II (Table 1.1). They are also called the chimotrypsin family and the substilysin family, respectively (Barret 1986). The substilysin family contains the bacterial serine proteases. Only the chimotrypsin family has been found in eukaryotes. This family includes many extracellular proteases such as trypsin, elastase, thrombin, plasma kallikrein, plasma coagulation proteases, and cellular proteases.

Proteolytic processing occurs in many different ways and is triggered by many different kinds of proteases. Moreover, it has become evident that cellular proteolysis is a highly controlled and complex process that takes place in virtually all compartments of cells. Although all serine proteases have essentially proteolytic function, the serine protease can be further divided into digestive

serine proteases and regulatory proteases, as explained in Chapter 2.

The digestive serine proteases have a role in the removal of defective or 'abnormal' polypeptides from cells. By removing these polypeptides, they control the concentrations of polypeptide and enzymes in cells, and contribute to the creation of amino acids from proteins and the constant renewal of cellular contents.

The regulatory serine proteases have also many physiologically important functions (Table 3.1). These functions include recognition of cytoskelton, myoblast fusion and differentiation (Rapaport et al. 1985), memory (Lynch et al. 1984), protein synthesis (Schwartz 1986), hormone action and inactivation (Duckworth 1975), fertilization (Bond et al. 1987), growth and aging (Bond et al. 1987), creation of immunologically recognizable molecules (Whitaker 1979), and degradation of endocytosised material and necrosis (deDube 1983). In addition, the regulatory serine proteases play an important role in the onset of diseases such as muscular dystrophy, diabetes, cachexia, cancer, and multiple sclerosis (Whitaker 1979). In this way, the regulatory serine proteases affect the possibly irreversible modifications of a protein molecule such as the cleavage of a peptide bond, the changes in protein conformation, and the determination of physiological process. Thus, the regulatory serine proteases have profound consequences in biological systems.

Although the regulatory domains such as kringle domains are deeply involved with the function of regulatory serine proteases, the protease domains are essential for proteolytic function of both regulatory and digestive proteases. With the aim of elucidating the evolutionary mechanism of the serine proteases, I studied the molecular evolution of the protease domains in a variety of serine proteases. In particular, the most sophisticated function of the serine proteases found in mammalian tissues is the protease activities of the fibrinolytic and blood coagulation systems (Neurath 1990). Thus, most of the protease domains in the serine proteases studied in this chapter are those that are involved in the fibrinolytic and blood coagulation systems. I will also discuss the evolutionary implication of physiological and catalytic functions of the serine proteases.

3.1.2 Molecular mechanism of catalytic reactions

Serine protease is the most thoroughly-studied class of proteases in catalytic reactions. The current understanding of the catalytic mechanism of serine protease is illustrated in Figure 3.1. In particular, mammalian serine proteases are a well-characterized family in terms of the molecular mechanism of catalytic function. The active sites of mammalian serine proteases are conformed by the so-called "catalytic triad" of Asp 102, His 57, and Ser 195 in chymotorypsin numbering. Note that the number

following an amino acid residue represents the residue number that starts from the N-terminus in the amino acid sequence. Catalysis proceeds through an intermediate of tetrahedral transition state during both acylation and deacylation steps. The same molecular mechanism underlies the catalytic reactions of all other serine proteases (Creighton 1984). In this case, a difference in substrate specificities can be explained by amino acid substitutions at the active sites where the primary or secondary substrates are bound to the serine protease. Thus, it is of interest to examine the relationship between the evolutionary changes of substrate specificities and amino acid substitutions during evolution.

3.1.3 Physiological function of some serine proteases

We dealt with various kinds of serine proteases in this study. Some of the serine proteases are unfamiliar to many readers. Thus, the physiological functions of these serine proteases used in this study are explained in the following.

A family of arginine-specific serine proteases are found in secretary or special granules in a number of cells (Ashley et al. 1985). This family of serine proteases includes tissue kallikrein (EC 3.4.21.35), the β- and γ–subunits of a nerve growth factor, endopeptidase, tonin, rat urinary esterase A, EGF-binding protein, and submandibular proteinase A (EC 3.4.21.40). These enzymes appear to have special processing functions (Bond et al. 1987) In particular, they are all involved in the processing of

precursors of polypeptide hormones, kinins, and growth factors, and have many physiochemical and immunological characteristics similar to each other.

Prolyl endopeptidase (EC 3.4.21.26) is an example of a cytosolic serine protease. It is found in many tissues including brain, liver, spleen, kidney, heart, lung, skeletal muscle, and pancreas. This protease is important in the neuropeptide metabolism, although it does not degrade large proteins.

Cathepsin G (EC 3.4.21.20), elastase (EC 3.4.21.37), and acrosin (EC 3.4.21.20) are the tissue-specific proteases that are found in special granules. Cathepsin G and elastase are found in the azurophil granules of human neutrophil leukocytes, and they act extracellularly in the inflammatory processes or in phagosomes of neutrophils (Bond et al. 1987). Acrosin is found in a modified lysozyme called the acrosome that covers the anterior part of sperm heads. This enzyme is a peripheral membrane protein that is involved in fertilization (Muller-Esterl, 1981).

Chymase (EC 3.4.21.31) and hepsin are thought to be involved with inflammatory reactions. They are present, in a high concentration, in normal skeletal muscle, lung, and skin (Bond et al. 1980).

3.1.4 Evolutionary study of protease domains

In the study of this chapter, I would focus the discussion on the protease domains of various serine proteases. As mentioned earlier, serine proteases are materials suitable to the evolutionary study of the mosaic proteins. In Chapter 2, I have already discussed the evolution of kringle domains as a typical example of regulatory domains in the serine proteases. The evolution of the protease domains must be studied in order to elucidate the whole features of the evolutionary process of serine proteases. In particular, the protease domains are contained by both regulatory and digestive serine proteases, and the domains are essential for proteolytic function of both serine proteases.

It is of particular interest to examine if a topology of the phylogenetic tree for the protease domains is identical to that for the kringle domains, when the phylogenetic tree for the same sets of serine proteases are compared with each other. If these topologies are essentially different, we may be able to conclude that the evolutionary history of the protease domains is different from that of the kringle domains. If this is case, it supports the idea that the domain shuffling played an important role in the evolution of mosaic proteins such as serine proteases. Thus, I would conduct the evolutionary analysis of the protease domains in the following sections.

3.2 MATERIALS AND METHODS

3.2.1 Sequences used

Seventy-five amino acid sequences of serine proteases were collected from the PIR protein database and Genbank of the DDBJ/EMBL/GenBank DNA databases (Table 3.2). The nucleotide sequences were translated into amino acid sequences. I then made a multiple alignment of these sequences for the protease domains with maximum match.

3.2.2 A method of multiple alignment

In conducting the multiple alignment, the sequences were tentatively aligned with each other by the pairwise comparison method. On the basis of this tentative alignment, the phylogenetic tree was then constructed by the unweighted pairwise grouping method. This was the first step of the multiple alignment. At the next step, the sequences were realigned by following the branching order of the phylogenetic tree obtained. Using the second alignment, I newly constructed the phylogenetic tree. I continue to repeat these steps many times until the sequence alignment becomes unchanged. In constructing the tree, I used

the estimated numbers of amino acid substitutions between a pair of the sequences compared. For estimating the numbers, d_a, of amino acid substitutions per site, I used the formula of $d_a = -\log_e(1 - p - 0.2p^2)$, where p is the proportion of different amino acids (Kimura, 1983). The phylogenetic trees were constructed by the unweighted pair grouping method (Nei, 1975) and the NJ method (Saitou and Nei 1987). As mentioned earlier, the former method requires a rate constancy of amino acid substitution over time whereas the latter method does not. Essential features of the phylogenetic tree constructed by the UPG method were the same as those constructed by the NJ method.

3.3 RESULTS

3.3.1 Evolutionary origin of serine protease domain

Figure 3.2 shows a phylogenetic tree for the protease domains of serine proteases. According to this tree, serine proteases could be separated into two major groups that were indicated by I and II. By this classification, the proteases having an activity of a growth factor belong to group I. However, the proteases which have the capability of binding a growth factor belong to group II.

The proteases that are related to the blood coagulation cascade are in group II. Elastases belong to group II. Acrosin which is found in sperm as well as hepsin which is found on the cell membrane of hepatocyte, belong to group II. Carboxyl peptidase, which is related to the protein degradation at lysosomes, are in group II. However, cathepsin G which is also related to the protein degradation at lysosomes belongs to group I. Collagenase is also contained by group I.

The serine proteases of the chymotrypsin type in group I appeared independently of those of group II in the evolutionary course. For example, the bovine chymotrypsins belong to group II and the other chymotrypsins to group I. Moreover, a trypsin of Drosophila belongs to group I, but trypsins of other species belong to group II. Furthermore, kallikreins belong to either groups I or II, depending on the species.

3.3.2 Evolution of the blood coagulation and fibrinolysis systems

In the evolution of the blood coagulation system, the phylogenetic tree in Figure 3.2 suggests that the proteases in the blood coagulation systems were divided into three groups. Let us call these groups A, B, and C. Group A contains uPA, tPA and factor XII. Group B contains protein C, prothorombin, and factors VII, IX, and X. Group C contains plasminogen and factor XI. The system of blood coagulation consists of two different pathways and a common pathway, as mentioned in Chapter

2. Kallikrain, factors IX and XII are involved in one of the two different pathways that is called the "intrinsic" pathway, and factor V is in the other pathway called the "extrinsic" pathway. The common pathway contains factor X. All the enzymes related to these three pathways are distributed over all of groups A, B, and C, although Group B contained only the enzymes of blood coagulation such as prothrombin, factor VII, factor IX, factor X, and protein C. Thus, it does not seem that the classifications of pathways in the blood coagulation system is directly related to evolutionary grouping of the protease domains of the serine proteases.

The enzymes related to the blood fibrinolysis (tPA, uPA, and plasminogen) belong to groups A and C. The enzymes in group A diverged first from the common ancestor of the three groups. Then, groups B and C were separated from their common ancestor. Therefore, the evolution of the blood coagulation cascade may have taken place together with the evolution of the fibrinolysis cascade.

3.4 DISCUSSION

3.4.1 Different patterns in evolution between protease and kringle domains

With the aim of examining whether the evolutionary pattern of the protease domains is the same as that of the kringle domains, as explained in section 3.1.4, I compared the tree topology for the kringle domains with that for the protease domains (Figure 3.3). The evolutionary pattern for the kringle domains is different from that for protease domains. In particular, the evolutionary position of HGF in the phylogenetic tree for the kringle domains is totally different from that for the protease domains. The phylogenetic tree for the kringle domains suggests that HGF is evolutionarily the closest to plasminogen and apo(a). It implies that the origin of kringle domain of HGF is plasminogen because apo(a) is known to have originated from plasminogen. However, the phylogenetic tree for the protease domains shows that HGF is evolutionarily the remotest among the serine proteases compared. It rather shows that the origin of the protease domain of HGF is related to acrosin. This fact suggests that the recombinational events must have occurred, in a certain stage of the evolution, between the protease domain of the ancestral gene of acrosin and the kringle domain of the plasminogen. For this reason, the protease domain of HGF may have a function differentiated from the proteases which have kringle structures.

The biological function of the kringle domains in HGF may be the same as that in plasminogen because the kringles in the former protein are evolutionarily close to those in the latter protein. Since the kringle structures of plasminogen are known to have an activity of binding other molecules in the cellular matrix, this activity of the kringle structures may be also important for the biological function of HGF. Because plasminogen does not seem to have the function of a growth factor, however, the function of a growth factor in HGF may be attributable to its protease domain that is remotely related to that of plasminogen. If the origin of the protease domains of HGF is much older than that of other proteases having kringle structures, it suggests that the original function of the serine proteases may have been like a growth factor.

Reference (Chapter 3)

Adham I M, Klemm U, Maier W M, Hoyer-Fender S, Tsaousidou S, and Engel W (1989) Eur. Molecular cloning of preproacrosin and analysis of its expression pattern in spermatogenesis. *J. Biochem.* 182, 563-568.

Arlaud G J and Gagnon J (1983) Complete amino acid sequence of the catalytic chain of human complement subcomponent C1r. *Biochemistry* 22, 1758-1764.

Ashley P L and Macdonald R J (1985) Tissue-specific expression of kallikrein-related genes in the rat. *Biochemistry* 24, 4520-27.

Baba T, Watanabe K, Kashiwabara S I, and Arai Y (1989) Primary structure of human proacrosin deduced from its cDNA sequence. *FEBS Lett.* 244, 296-300.

Belin D, Vassalli J-D, Combepine C, Godeau F, Nagamine Y, Reich E, Kocher H P and Duvoisin R M (1985) Cloning, nucleotide sequencing and expression of cDNAs encoding mouse urokinase-type plasminogen activator *Eur. J. Biochem.*148, 225-232.

Bentrey D R (1986) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G, and Bartunik H (1983) Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase. *J. Mol. Biol.* 164, 237-282.

Bond J S and Butler P E (1987) Intracellular proteases *Ann. Rev. Biochem.* 56, 333-64.

Bleackley R C, Duggan B, Ehrman N, and Lobe C G (1988) Isolation of two cDNA sequences which encode cytotoxic cell proteases. *FEBS Lett.* 234, 153-159.

Brown J R and Hartley B S (1966) Location of disulphide bridges by diagonal paper electrophoresis. *Biochem. J.* 101:214-228.

Brune J L, Yang F, Barnet D R, and Bowman B H (1984) Evolution of haptogrobin: composition of complementary DNA encoding $Hp\alpha^{1S}$ and $Hp\alpha^{2FS}$. *Nucleic Acids Res.* 12, 4531-4538.

Catterall C F, Lyons A, Sim R B, Day A J, Harris T J R (1987) Characterization of the primary amino acid sequence of human complement control protein factor I from an analysis of cDNA clones. *Biochem. J.* 242, 849-856.

Cook K S, Groves D L, Min H Y, and Spiegelman B M (1985) A developmentally regulated mRNA from 3T3 adipocytes encodes a novel serine protease homologue. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6480-6484.

Craik C S, Choo Q L, Swift G H, Quinto C, MacDonald R J, Rutter W J (1984) Structure of two related rat pancreatic trypsin genes. *J. Biol. Chem.* 259, 14255-14264.

Creigton T E (1984) PROTEINS; structures and molecular properties.- W H Freedman Co. New York.

Davis C A, Riddell D C, Higgins M J, Holden J J A, and White B N (1985) A gene family in Drosophila melanogaster coding for trypsin-like enzymes. *Nucleic Acids Res.* 13, 6605-6619.

deDuve C (1983) Lysosomes revisited. *Eur. J. Biochem.* 137, 391-397.

DeLotto R and Spierer P (1986) A gene required for the specification of dorsal-ventral pattern in Drosophila appears to encode a serine protease. *Nature* 323, 688-692.

Drinkwater C C, Evans B A, and Richards R I (1987) Mouse glandular kallikrein genes: identification and characterization of the genes encoding the epidermal growth factor binding proteins. *Biochemistry* 26, 6750-6756.

Drinkwater C C, Evans B A, and Richards R I (1988) Sequence and expression of mouse gamma-renin. *J. Biol. Chem.* 263, 8565-8568.

Ducworth W C, Heineman M, and Kitabchi A E (1975) Proteolytic degradation of insulin and glucagon. *Biochem, Biophys. Acta* 377, 421-430.

Evans B A and Richards R I (1985) Genes for the α and γ subunits of mouse nerve growth factor are contiguous. *EMBO J.* 4,133-138.

Forsgren M, Raden B, Israelsson M, Larsson K and Haden L-O (1987) Molecular cloning and characterization of a full-length cDNA clone for human plasminogen. *FEBS Lett.* 213, 254-260.

Foster D C, Yoshitake S, and Davie E W (1985) The nucleotide sequence of the gene for human protein C. Proc. *Nat. Acad. Sci. USA.* 82, 4673-4677.

Fujikawa K, Chung D W, Hendrickson L E, and Davie E W (1986) Amino acid sequence of human factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma prekallikrein. *Biochemistry* 25, 2417-2424.

Fukushima D, Kitamura N, and Nakanishi S (1985) Nucleotide sequence of cloned cDNA for human pancreatic kallikrein. *Biochemistry* 24, 8037-8043.

Fung M R, Cambell R M, and MacGillvray T A (1984) Blood coagulation factor X mRNA encodes a single polypeptide chain containing a prepro leader sequence. *Nucleic. Acid. Res.* 12, 4481-4492.

Fung M R, Hay C W, and MacGillvray R T A (1985) Characterization of an almost full-length cDNA coding for human blood coagulation factor X. P*roc. Nat. Acad. Sci. USA* 82, 3591-3595.

Gershenfeld H K, Hershberger R J, Shows T B, Weissman I L (1988) Cloning and chromosomal assignment of a human cDNA encoding a T cell- and natural killer cell-specific trypsin-like serine protease. *Proc. Nat. Acad. Sci. USA* 85, 1184-1188.

Goldberger G, Bruns G A P, Rits M, Edge M D, and Kwiatkowski D J (1987) Human complement factor I: analysis of cDNA-derived primary structure and assignment of its gene to chromosome 4. *J. Biol. Chem.* 262:10065-10071.

Goldstein L A, and Heath E C (1984) Nucleotide sequence of rat haptoglobin cDNA. *J. Biol. Chem.* 259, 9212-9217.

Grant G A, Henderson K O, Eisen A Z, and Bradshaw R A. (1980) Amino acid sequence of a collagenolytic protease from the hepatopancreas of the fiddler crab, *Uca pugilator. Biochemistry* 19, 4653-4659.

Hermodson M A, Ericsson L H, Neurath H, and Walsh K A (1973) Determination of the amino acid sequence of porcine trypsin by sequence analysis. *Biochemistry* 12, 3146-3153.

Holland S K and Blake C C F (1990) Proteins, exons, and molecular evolution in Intervening Sequences in Evolution and Development edited by Stone E M and Schwartz R J Oxford university press New York.

Itoh N, Tanaka N, Funakoshi I, Kawasaki T, Mihashi S, and Yamashina I (1988) The complete nucleotide sequence of the gene for batroxobin, a thrombin-like snake venom enzyme. *Nucleic Acids Res.* 16, 10377-10378.

Jany K D, Bekelar K, and Ishay (1981) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Kraut J Serine (1977) proteases: structure and mechanism of catalysis. *Ann. Rev. Biochem.* 46, 331-58.

Lazure C, Leduc R, Seidah N G, Thibault G, Genest J, and Chretien M (1984) Amino acid sequence of rat submaxillary tonin reveal similarities to serine proteases. *Nature* 307, 555-558.

Lecroisey A, Gilles A M, De Wolf A, and Keil B (1987) Complete amino acid sequence of the collagenase from the insect Hypoderma lineatum. *J. Biol. Chem.* 262, 7546-7551.

Le Trong H, Parmelee D C, Walsh K A, Neurath H, Woodbury R G (1987) Amino acid sequence of rat mast cell protease I (chymase). *Biochemistry* 26, 6988-6994.

Leytus S P, Loeb K R, Hagen F S, Kurachi K, Davie E W (1988) A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. *Biochemistry* 27, 1067-1074.

Lobe C G, Upton C, Duggan B, Ehrman N, Letellier M, Bell J, McFadden G, and Bleackley R C (1988) Organization of two genes encoding cytotoxic T lymphocyte-specific serine proteases CCPI and CCPII. *Biochemistry* 27, 6941-6946.

Long G L, Baragaje R M, Macgillivray (1984) Cloning and sequencing of liver cDNA coding for bovine protein C. *Proc. Natl. Acad. Sci. USA* 81,5653-5656.

Lundgren S, Ronne H, Rask L, and Peterson P A (1984) Sequence an epidermal growth factor-binding protein. *J. Biol. Chem.* 259, 7780-7784.

Lynch G, and Baudry M, (1984) The biochemistry of memory: a new and specific hypthesis. *Science* 163, 1057-63.

MacDonald R J, Swift G H, Quinto C, Swain W, Pictet R L, Nikovits W, and Rutter W J (1982) Primary structure of two distinct rat pancreatic preproelastases determined by sequence analysis of the complete cloned messenger ribonucleic acid sequences. *Biochemistry* 21, 1453-1463.

MacGillivray R T A and Davie E W (1984) Characterization of bovine prothrombin mRNA and its translation product. *Biochemistry* 23, 1626-1634.

Maeda N (1985) Nucleotide sequence of the haptoglobin and haptoglobin related gene pair. *J. Biol. Chem.* 260, 6698-6709.

Marti T, Schaller J, and Rickli E E (1985) Determination of the complete amino-acid sequence of porcine miniplasminogen. *Eur. J. Biochem.* 149, 279-285.

Mason A J, Evans B A, Cox D R, Shine J, and Richards R I (1983) Structure of mouse kallikrein gene family suggests role in specific processing of biologically active peptides. *Nature* 303, 300-307.

McLean J W, Tomlison J E, Kuang W-J, Eaton D L, Ellson Y C, Fless G M, Scanu A M and Lawn R M (1987) cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature* 330, 132-137.

McMullen B A and Fujikawa K. (1985) Amino acid sequence of the heavy chain of human α-Factor XIIa (activated Hageman factor). *J. Biol. Chem.* 260, 5328-5341.

Miyazawa K, Tsubouchi H, Naka D, Takahashi K, Okigaki M, Arakaki N, Nakayama H, Hirono S, Sakiyama O, Takahashi K, Gohda E, Daikuhara Y, and Kitamura N (1989) Molecular cloning and

sequence analysis of cDNA for human hepatocyte growth factor *Biochem. Biophys. Res. Commun.* 163, 967-973.

Mucmullen B A, Fujikawa K, and Kisiel W (1983) The occurrence of β-hydroxyaspartic acid in the vitamin K-dependent blood coagulation zymogens. *Biochem. Biophys. Res. Commun.* 115, 8-14.

Nagamine Y, Pearson D, Altus M S and Reich E (1984) cDNA and gene nucleotide sequence of porcine plasminogen activator. *Nucli. Acid. Res.* 12, 9525-9541.

Nakamura T, Nishizawa T, Hagiya M, Seki T, Shimonishi M, Sugimura A, Tashiro K. and Shimizu S. (1989) Molecular cloning and expression of human hepatocyte growth factor. *Nature* 342, 440-443.

Neurath H (1984) Evolution of proteolytic enzymes *Science* 224, 350-357.

Nieman M A, Bhown A S, Bennett J C, and Volanakis J E. (1984) Amino acid sequences of human D of the alternative complement pathway. *Biochemistry* 23, 2482-2486.

Ny T, Leonardsson G, Hsueh A J W (1988) Cloning and characterization of a cDNA for rat tissue-type plasminogen activator. *DNA* 7, 671-677.

O Hara P J, Grant F J, Haldeman B A, Gray C L, Insley M Y, Hagen F S, and Murray M.J. (1984) Nucleotide sequence of the gene coding for human factor VII, a vitamin K-dependent protein participating in blood coagulation. *Proc. Nat. Acad. Sci.* USA 84, 5158-5162.

Okajima A, Miyazawa K, and Kitamura N. (1990) Primary structure of rat hepatocyte growth factor and induction of its mRNA during liver regeneration following hepatic injury. *Eur. J. Biochem* 193, 375-381.

Okano K, Aoki Y, Sakurai T, Kajitani M, Kanai S, Shimazu T, Shimizu H, and Naruto M (1987) Molecular cloning of complementary DNA for human medullasin: an inflammatory serine protease in bone marrow cells. *J. Biochem.* 102, 13-16.

Olafson R W, Jurasek L, Carpenter M R, Smillie L B (1975) Amino acid sequence of *Streptomyces griseus* trypsin. Cyanogen bromide fragment and complete sequence. *Biochemistry* 14, 1168-1177.

Pennica D, Holmes W E, Kohr W J, Harkins R N, Vehar G A, Bennet W M, Yelventon E, Seeburg P H, Heyneker L and Goeddel D V (1983) Cloning and expression of human tissue-type plasminogen activator cDNA in *E. coli. Nature* 301, 214-221.

Pinsky S D, LaForge K S, and Scheele G (1985) Different regulation of trypsinogen mRNA translation: full-length mRNA sequences encoding two oppositely charged trypsinogen isozymes in the dog pancreas. *Mol. Cell. Biol.* 5, 2669-2676.

Rapaport S, Daniel W, and Muller M (1985) Proteolysis of mitochondria in reticulocytes during maturation is ubiquitin-dependent and is accompanied by a high rate of ATP hydrolysis. *FEBS Lett.* 180, 249-252.

Salvesen G, Farley D, Shuman J, Przybyla A, Reilly C, and Travis J (1987) Molecular cloning of human cathepsin G: structural similarity to mast cell and cytotoxic T lymphocyte proteinases. *Biochemistry* 26, 2289-2293.

Sandra J F D, Ross T, MacGillivary A and Davie E W (1983) Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin. *Biochemistry* 22, 2087-2097.
Sappino A-P, Hurate J, Belin D, and Vassalli J-D (1989) Plasminogen activators in tissue remodeling and invasion: mRNA localization in mouse ovaries and implanting embryos. *J. Biol. Chem.* 109, 2471-2479.

Schaller J, Moser P W, Dannegger-Muller G A K, Roselt S J, Kampfer V and Rikli E E Complete amino acid sequence of bovine plasminogen comparison with human plasminogen. (1985) *Eur. J. Biochem* 149, 267-278.

Schaller J, Akiyama K, Tsuda R, Hara M, Marti T, and Rickli E E (1987) Isolation, characterization and amino-acid sequence of gamma-seminoprotein, a glycoprotein from human seminal plasma. *Eur. J. Biochem.* 170, 111-120.

Schwartz T W (1986) The processing of peptide precursors. 'Proline-directed arginyl cleavage' and other monobasic processing mechanisms *FEBS Lett.* 200, 1-10.

Shirasu Y, Yoshida H, Mikayama T, Matsuki S, Tanaka J I, and Ikenaga H (1986) Isolation and expression in Escherichia coli of a cDNA clone encoding porcine pancreatic elastase. *J. Biochem.* 99, 1707-1712.

Smillie L B, Furka A, Nagabushan N, Stevenson K J, and Parkes C O (1968) Structure of chymotrypsinogen B compared with chymotrypsinogen A and trypsinogen. *Nature* 218, 343-346

Titani K, Ericsson L H, Neurath H, Walsh K A (1975) Amino acid sequence of dogfish trypsin. *Biochemistry* 14, 1358-1366.

Titani K, Sasagawa T, Woodbury R G, Ericsson L H, Dorsam H, Kraemer M, Neurath H, and Zwilling R (1983) Amino acid sequence of crayfish (*Astacus fluviatils*) trypsin If. *Biochemistry* 22, 1459-1465.

Tomlinson J E, McLean J W and Lawn R M (1989) Rhesus monkey apolipoprotein(a). *J. Biol. Chem.* 264, 5957-5965.

Verde P, Stoppelli M P, Galeffi P, Nocera P D and Blasi F (1984) Identification and primary sequence of an unspliced human urokinase poly(A)$^+$ RNA. *Proc. Natl. Acad. Sci. USA* 81, 4727-4731.

Venot N, Sciaky M, Puigserver A, Desnuelle P, and Laurent G (1986) Amino acid sequence and disulfide bridges of subunit III, a defective endopeptidase present in the bovine pancreatic 6 S procarboxypeptidase A complex. *Eur. J. Biochem.* 157, 91-99.

Walz D A, Hewett-Emmett D and Seegers W H (1977) Amino acid sequence of human prothrombin fragments 1 and 2. *Proc. Natl. Acad. Sci. USA* 74, 1969-1972.

Woodbury R G, Katunuma N, Kobayashi K, Titani K, and Neurath H. (1978) Covalent structure of a group-specific protease from rat small intestine. *Biochemistry* 17,811-819.

Yamasaki N, Sugimura K, Hiida M, Naito T, and Watanabe T (1987) Sequence analysis of a cDNA clone of a gene encoding a component of a putative phosphorylcholine-specific T suppressor factor and functional property of its gene product. *Eur. J. Immunol.* 17, 247-253.

Yoshitake S, Schach B G, Foster D C, Davie E W, and Kurachi K. (1985) Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). *Biochemystry* 24, 3736-3750.

Table 3.1 **Physiological Process Regulated by Proteases and its Inhibitors**

Processing of proteins
Activating of enzymes
Blood coagulation
Complement system
Kinin system
Renin-angiotensin system
Maturation of phage
Digestion of microorganism
Sporulation and germination
Fertilization and development
Inflamentation
Immune system
Oncogenesis
Degradation of proteins

Table 3.2 **Serine proteases used in the present analysis.**

	Abbreviation	Species	Data source
7S NGF α-chain	NGMSA	Mouse	Evans et al. 1985
7S NGF γ-chain	NGMSG	Mouse	Evans et al. 1985
Epidermal growth factor (EGF) binding protein A	EGFBA	Mouse	Drinkwater 1987
EGF-binding protein type B	EGFBB	Mouse	Lundgren et al. 1984
	EGFMSB	Mouse	Drinkwater et al. 1987
Protein C	Protein C	Bovine	Long et al. 1984
		Human	Foster et al. 1984
Apolipoprotein(a)	apo	Rhesus	Tomlinson et al. 1989
		Human	Mclean et al. 1987
Plasminogen	plgen	Bovine	Schaller et al. 1985
		Porcine	Marti et al. 1985
		Rhesus	Tomlinson et al. 1989
		Human	Forsgren et al. 1987
Hepatocyte growth factor	HGF	Human	Nakamura et al. 1989
Tissue-type plasminogen activator	tPA	Rat	Ny et al. 1988
		Human	Pennica et al. 1983
Urokinase	uPA	Mouse	Belin et al. 1984
		Porcine	Nagamine et al. 1984
		Human	Verde et al. 1984
Prothrombin	pro	Bovine	Macgillivrey et al. 1984
		Human	Waltz et al. 1977
Factor XII	KFHU12	Human	McMullen et al. 1985

Factor VII	favtorVII	Human	O Hara et al. 1987
Factor IX (Christmas factor)	factor IX	Bovine	Mcmillen et al. 1983
		Human	Yoshitake et al. 1985
Factor X	factor X	Bovine	Fung et al. 1984
		Human	Fung et al. 1985
Factor XI	factor XI	Human	Fujikawa et al. 1986
Tissue kallikrein	KQPG	Porcine	Bode etal. 1983
Tissue prokallikrein	KQHU	Human	Fukushima et al. 1985
Tonin	KQRTTN	Rat	Lazure et al. 1984
Collagenolytic protease	KCUF	Atrantic sand fiddler crab	Grant et al. 1980
Mast cell proteinase II	PRRTG	Rat	Woodbury et al. 1978
Complement factor D	DBHU	Human	Niemann et al. 1984
Complement subcompornent C1r b-chain	C1HURB	Human	Arlaud et al. 1983
Complement C2	C2HU	Human	Bentrey 1986
Complement favtor I	C1HUM	Human	Goldberg et al. 1987
Complement controll protein factor I	C1FC1	Human	Catterall et al. 1987
T-cell supressor factor	MSTSUP	Mouse	Yamasaki et al. 1987
Haptoglobin 1	HPHU1	Human	Brune et al. 1984
Haptoglobin-related protein	HPHUR	Human	Maeda et al. 1985
Haptoglobin	HPRT	Rat	Goldstein et al. 1984
28K protein	WMMS28	Mouse	Cook et al. 1985
Trypsinogen, anionic	TRDG	Dog	Pinsky et al. 1985
Trypsinogen, cationic	TRDGC	Dog	Pinsky et al. 1985
Trypsinogen	TRPGTR	Porcine	Hermodson et al. 1973
Trypsinogen1	TRRT1	Rat	Craik et al. 1984
Trpsinogen	TRDFS	Spiny dogfish	Titani et al. 1975
Trypsin I	TRCY1	Cray fish	Titani et al. 1983

Trypsin	TRSMG	Streptomyces griseus	Olafson et al. 1975
Trypsinogen-like proenzyme	TRFF	Fruit fly	Davis et al. 1985
Chymotrypsinogen A	KYBOA	Bovine	Brown et al.
Chymotrypsinogen B	KYBOB	Bovine	Smille et al. 1968
Chymotrypsin II	KYVH20	Oriental hornet	Jany et al. 1981
Serine proteinase	COLLAG	Fruit fly	Delotto et al. 1986
Collagenase	RTMATI	Early cattle grub	Lecroisey et al. 1987
Chymase	DROPRO	Rat	Le Trong et al. 1987
Cathepsin G	CATHEPG	Human	Salvesen et al. 1987
Cytotoxic T-lymphocyte proteinase II	MSCTL2	Mouse	Lobe et al. 1988
	CTCMS	Mouse	Bleackley et al. 1988
Trypsin-like serine protease (Hanukah factor)	TRPHUM	Human	Gershenfeld et al. 1988
Acrosin	ACROSIN	Human	Baba et al. 1989
	POACRO	Porcine	Adham et al. 1989
Hepsin	HEPSIN	Human	Leytus et al. 1988
Bothrops atrox serine protease	VESER	Bothrops atrox	Itoh et al. 1988
Gamma-semino protein human	SEMNOG	Human	Schaller et al. 1987
Gamma-Renin	MSGREN	Mouse	Drinkwater et al. 1988
Tissue prokallikrein mGK-1 DROPRO	KQMS1	Mouse	Mason et al. 1983
Medullasin	MEDUHU	Human	Okano et al. 1987
Proelastase precursor I	ELRT1	Rat	MacDonald et al. 1982
Elastase	ELPG	Porcine	Shirus et al. 1986
Procarboxypeptidase A complex component II	CPBOA3	Bovine	Venot et al. 1986

Figure 3.1 **The enzymatic mechanism of serine protease.**

This figure shows the hydrolysis of protein by the protease. In this figure, **E** means the protease. **R_1CO-NHR_2** shows the peptide bond of the protein as the substrate.

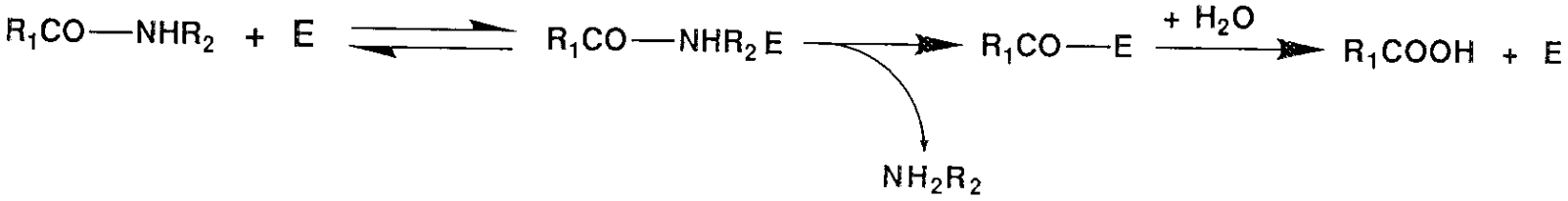

R1CO—NHR2 + E
R1CO—NHR2 E
R1CO—E
+ H2O
R1COOH + E
NH2R2

Figure 3.2 The phylogenetic tree of serine proteases.

The tree was constructed by the analysis of amino acid sequences. Two major groups are indicated by I and II. Group I can be further divided into three subgroup A, B, and C. The abbreviated names of the amino acid sequences were explained in Table 3.2.

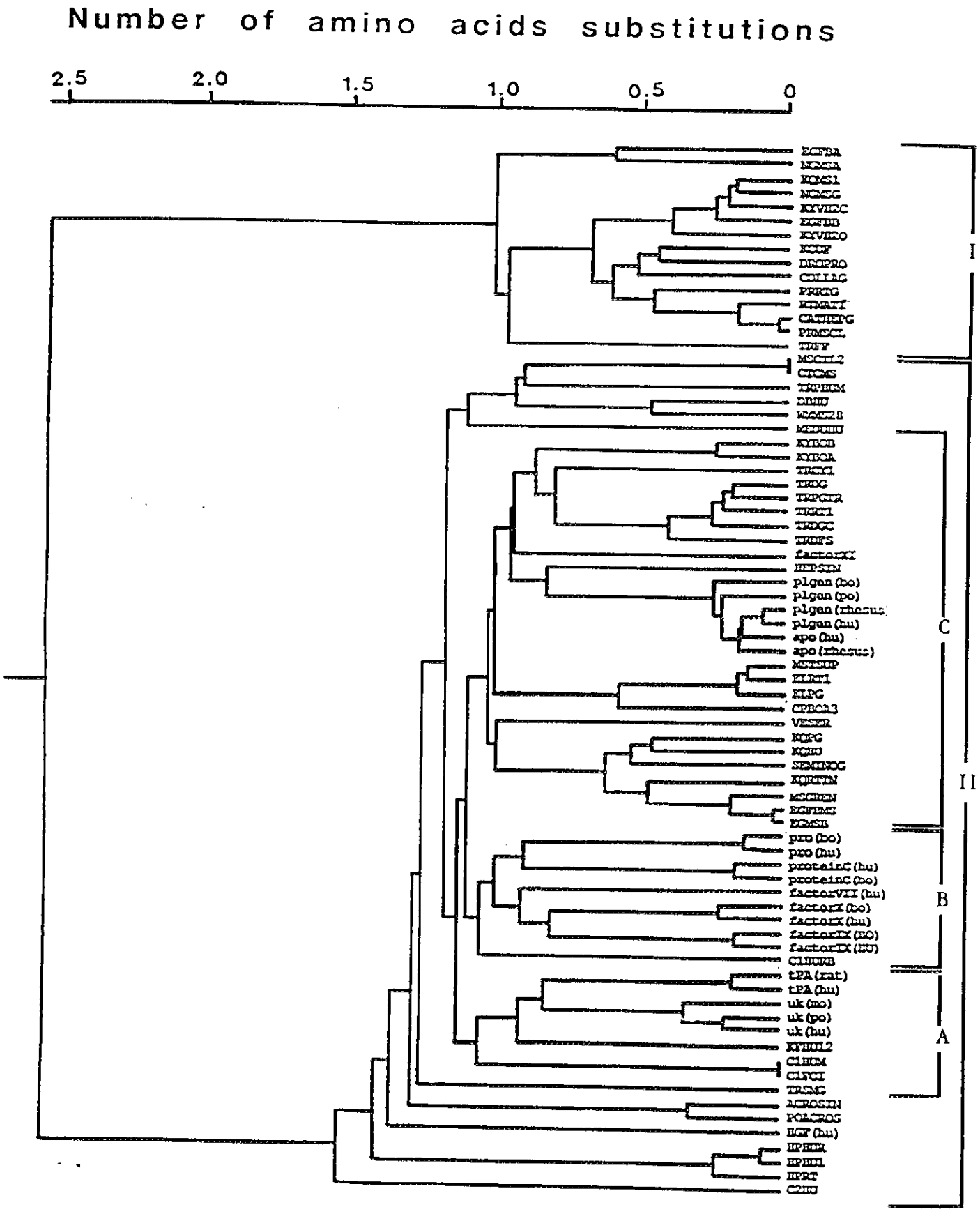
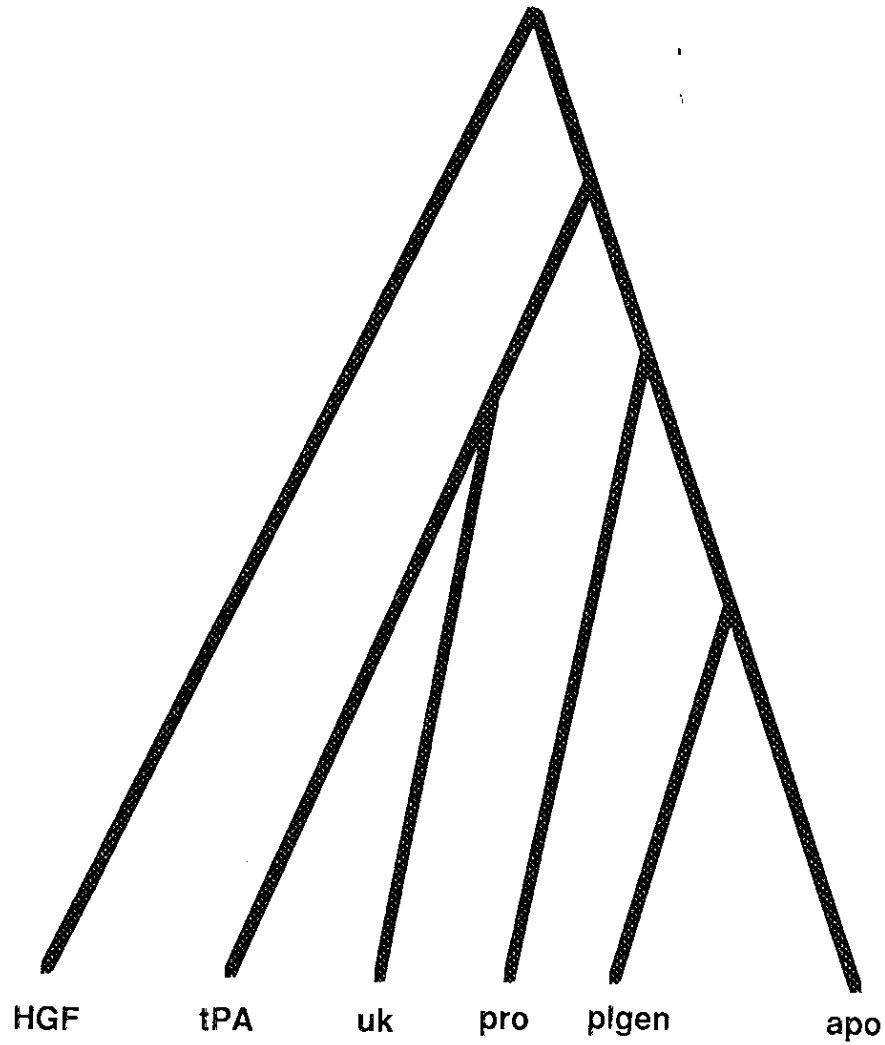
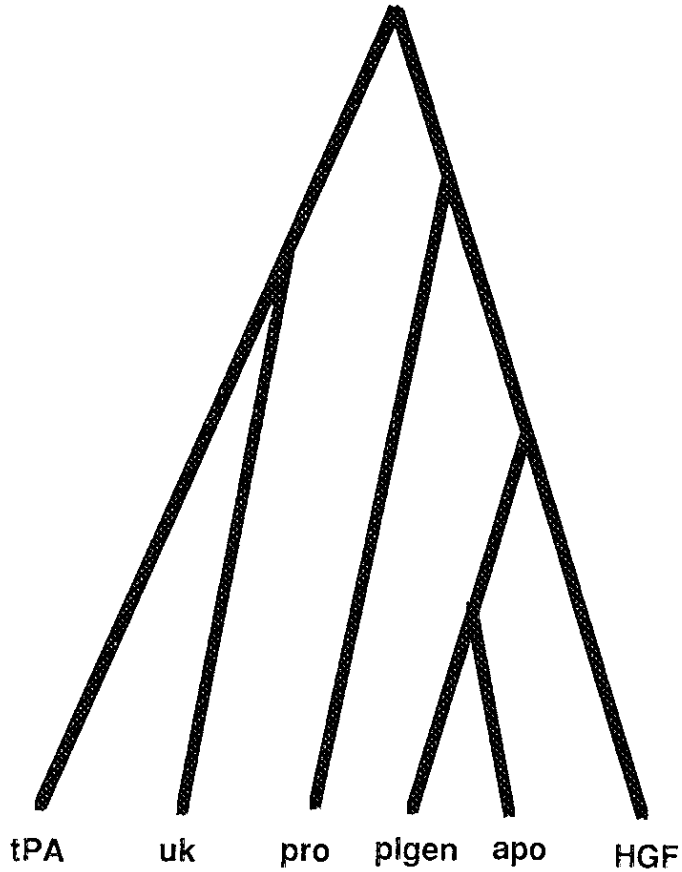

Number of amino acids substitutions
2.5
2.0
1.5
1.0
0.5
0
I
II
C
B
A
plgen (bo)
plgen (po)
plgen (rhesus)
plgen (hu)
apo (hu)
apo (rhesus)
pro (bo)
pro (hu)
proteinC (hu)
proteinC (bo)
factorVII (hu)
factorX (bo)
factorX (hu)
tPA (rat)
tPA (hu)
uk (mo)
uk (po)
uk (hu)
ACROSIN
HGF (hu)

Figure 3.3 A difference in the evolutionary patterns between the kringle domains and the serine protease domains.



protease domain



kringle domain

CHAPTER FOUR

EVOLUTION OF SERINE PROTEASE INHIBITOR

4.1 INTRODUCTION

4.1.1 Minikringle structure

The Kunitz-type inhibitor is one of the serine protease inhibitors. It usually has a low relative molecular mass, a basic isoelectric point, and one or several inhibitory domains (Salvesen and Nagase 1989). This inhibitor was first isolated as a trypsin inhibitor from bovine pancreas (Laskowski et al. 1974). It has an amino acid sequence of 58 residues with three disulfide bridges (Figure 4.1). The pattern of its disulfide bridges is the same as that in the kringle domains of serine proteases which are involved with the blood coagulation cascade (Ikeo et al. 1991). Because the kringle domain of the inhibitor is much smaller than that of the protease, the former kringle is called the "minikringle' structure.

The enzymatic reaction between a protease and its inhibitor is characterized by the formation of a pseudo-irreversible inhibitor-protease complex (Lascowski and Kato 1980). The Kunitz-type

inhibitor is found in virtually all tissues and has a wide range of specificity (Fioretti et al. 1983). This inhibitor has also been found in snake venom and in some insects. Very little, however, is known about the biosynthesis and physiological function of the Kunitz-type inhibitor.

4.1.2 Amyloid β protein in Alzheimer's disease

In the last few years, it has become apparent that various proteins contain Kunitz-type inhibitor domains. It is very likely that these proteins were assembled as a consequence of exon shuffling which sometimes results in incorporation of genetic elements or domains from different evolutionary precursors. The Kunitz-type inhibitor sequence was found in a precursor protein of the amyloid β protein which is a component of cerebral depositions in the brains of patients with Alzheimer's disease (Figure 4.2).

Alzheimer's disease is the most common neurodegenerative disorder, affecting over 2 million people in the United States alone. This disease is characterized by progressive dementia with neuropathological features such as extracellular deposits of amyloid plaques and vascular amyloid as well as intracellular deposits of neurofibrillary tangles (Muller-Hill and Beyreuther 1989). Amyloid plaques have also been found in Down's syndrome patients and, in

smaller numbers, in normally aged human brains and in the brains of aged mammals. The amyloid forming β-peptide of Alzheimer's disease is synthesized as part of a larger membrane precursor protein (βAPP). This precursor protein has three alternatively spliced versions of APP695, APP751, and APP770. APP751 and APP770 both have an insertion sequence of the Kunitz-type trypsin inhibitor although APP695 does not (Kang et al. 1987; Ponte et al. 1988; Tanzi et al. 1988; Kitaguchi et al. 1990). Thus, it is of particular interest to study the evolutionary feature of the Kunitz-type protease inhibitor domain in the amyloid β protein.

4.1.3 α3 chain of collagen VI

A sequence similar to the Kunitz-type inhibitor was also found in the chicken α3 chain of collagen VI (Bonaldo and Colombatti 1990). The type VI collagen is a unique component of the family of collagenous proteins found in soft connective tissues and cartilage. The type VI collagen is the major component of 100nm long periodic filaments that form a major class of tissue microfibrils (von der Mark et al. 1984). These fibrils are present in both embryo and adult tissues (Koller et al. 1989). The type VI collagen also forms a flexible network that anchors large interstitial

structures such as nerves, blood vessels, and other collagen fibers (Table 4.1).

The protein molecule of the type VI collagen is composed of three different polypeptide chains that form a heterotrimeric molecule with a short triple helix and large globular domains at both ends. Type VI collagen is synthesized and secreted by mesenchymal cells.

The amino acid sequence of the α3 chain of chicken type VI collagen was reported by Bonaldo et al. (1989). Although the amino acid sequence of the α3 chain is homologous to those of the α1 and α2 chains of chicken type VI collagen, the carboxyl-terminal region of the α3 chain is not present in the α1 and α2 chains and is a mosaic region with extensive similarities to several other proteins (Figure 4.3). This region is preceded by sequences analogous to the platelet glycoprotein Ib. This region is also followed by one segment that closely resembles the type III domains of fibronectin. At the C-terminus of the sequence, there is a domain which is very similar to the Kunitz type protease inhibitors. This collagen might play an important role in growing and remodeling of connective tissues (Bonaldo et al. 1989). It is interesting to know the evolution of this protein for the understanding of the evolution of mosaic proteins.

4.1.4 Dendrotoxin as a serine protease inhibitor

The dendrotoxin is a typical example of Kuniz-type inhibitors (Figure 4.1). The dendrotoxins are pharmacologically active proteins that have been identified in the venom of the eastern green mamba (*Dendroaspis angusticeps*) and the black mamba (*Dendroaspis polylepis polylepis*). They are also known as 'pre-synaptic facilitatory toxins' because their action is to faciliate acetylcholine release from the presynaptic membranes of cholinergic synapses. Their most notable effect *in vivo* is to produce exessive muscular activity via neuromuscular synapses, but synapses also affects in the sympathetic and parasympathetic branches of the autonomic and central nervous systems are also affected.

To elucidate the evolutionary origin and biological function of these insertion sequences, particularly in βAPP of Alzheimer's disease, I constructed a phylogenetic tree for the Kunitz-type protease inhibitors.

4.2 MATERIALS AND METHODS

4.2.1 Sequences used

Forty-nine amino acid sequences and 13 nucleotide sequences of the Kunitz-type protease inhibitors were collected from the PIR (version 28) protein data base and Genbank (version 68) of the DDBJ/EMBL/GenBank DNA database (Table 4.2), respectively. I made a homology alignment of these sequences with maximum match after translating the nucleotide sequences into amino acid sequences. I then computed the number of amino acid substitutions per site from comparisons between each pair of amino acid sequences (Figure 4.4). For the correction of multiple substitutions at a site, I used Kimura's formula of $d_a = -\log_e(1 - p - 0.2\ p^2)$, where d_a and p are the number of amino acid substitutions per site and the proportion of different amino acids, respectively (Kimura 1983).

Phylogenetic trees of the Kunitz-type protease inhibitor were constructed by the unweighted pair-grouping (UPG) method (Nei 1975) and the neighbor-joining (NJ) method (Saitou and Nei 1987), using the d_a values. Note that the UPG method requires a rate constancy of amino acid substitution over time whereas the NJ method does not. I also constructed a phylogenetic tree using nucleotide sequences. The number of nucleotide substitutions was

calculated by the four-parameter method (Takahata and Kimura 1981), because the six-parameter method (Gojobori et al. 1982) was inapplicable in a few cases.

4.3 RESULTS

4.3.1 Evolution of serine protease inhibitor

Figure 4.5 shows a phylogenetic tree for all Kunitz-type inhibitors available. This tree was constructed by the NJ method, using amino acid sequences. A majority of the inhibitors were separated into three groups that were indicated by I, II, and III. Group I contains only the Kunitz-type inhibitors known to be neurotoxins in snakes. The insertion sequence in collagen IV of chicken, all the larval protease inhibitors of insects, and a few mammalian inhibitors belong to group II. The βAPP precursor, the c-terminal region of the inter-α-trypsin inhibitor (IαTI), and trypstatin are found in group III.

I obtained a reliable estimate of the rate of amino acid substitution from the comparisons of IαTI2 and LACI (lipoprotein associated coagulation inhibitor). Although the sequence of rabbit LACI is recently determined, it is not included in the phylogenetic

trees. I estimated the rate of amino acid substitution for IαTI2 to be, on the average, 1.2 x 10^{-9} per site per year from comparisons between human and bovine, porcine, sheep, and horse. In this case, I assumed that the divergence of these species occurred 80 million years ago. In the same way, I estimated the rates of amino acid substitution for LACI1, LACI2, and LACI3 to be respectively 1.6 x 10^{-9}, 1.6 x 10^{-9}, and 0.52 x 10^{-9} per site per year from comparisons between human and rabbit. Thus, I estimated the average rate of amino acid substitutions for IαTI2, LACI1, LACI2, and LACI3 to be 1.2 x 10^{-9} per site per year. It is known that IαTI1 may have lost the activity as a protease inhibitor (Laskowski and Kato 1980), because it does not inhibit any types of the enzymes. For this reason, I excluded IαTI1 from this calculation.

Using the average rate of amino acid substitution, I estimated the time of evolutionary events that must have taken place in the past. The following results are based on estimations of divergence times for various branching points in the phylogenetic tree (Figure 4.6). The ancestral gene of the Kunitz-type inhibitor must have appeared about 500 million years ago. Then, about 50 million years later, the Kunitz-domain was duplicated into two domains. One became the ancestral domain of group I, and the other of groups II and III. The inhibitor sequences in group I were further separated into two subgroups. One subgroup evolved into the lineage of snake toxins which exhibit protease inhibitor activity. The other became the snake toxins with neuromuscular activities.

The two subgroups were estimated to have diverged from the ancestral gene about 400 million years ago. A little more than 400 million years ago, the ancestor of groups II and III was duplicated into two domains to become the respective members of groups II and III. Group II contains IαTI1, basic protease inhibitor (BPI), and LACI1 and LACI2, although LACI3 belongs to group III. Group III includes LACI3, IαTI2, trypstatin, and βAPP, each of which diverged from their ancestor about 270 million years ago. Thus, it was found that the Kunitz-type inhibitor domains of βAPP are most closely related to that of IαTI2. Moreover, the ancestor of the insertion sequence in chicken α collagen IV was shown to have diverged from the ancestral gene of LACI1 and LACI2 about 360 million years ago.

Figure 4.7 shows the phylogenetic tree that was constructed by the NJ method, using the nucleotide sequence data. The tree topology in Figure 4.7 is somewhat different from that in Figure 4.5 in the relationship between groups II and III. I believe that the tree in Figure 4.5 is more reliable than the other, because the former was obtained using more data than the latter. In fact, the tree in Figure 4.5 is consistent with functional classifications of the proteins studied. In any event, these results suggest that the gene duplication occurred at a relatively early stage in the evolution of Kunitz-type inhibitors.

4.3.2 Evolution of snake toxins

The evolution of snake toxins can be seen in group I of the phylogenetic tree (Figure 4.5), because all the members of group I are snake toxins. In group I, the two subgroups were separated soon after the duplication about 320 million years ago. One of the two subgroups contains dendrotoxins. They were further separated into two clusters, dendrotoxins with the activity of protease inhibitor and those without it. The other subgroup includes the members of other snake protease inhibitors. These snake toxins are the venom basic protease inhibitors of viper russelli, a western sand viper, a sand viper, and a kind of krait (*Bungars fasciatus*).

4.3.3 Evolution of the insertion sequence in amyloid β protein

It has been argued that the functionally important sites of serine protease inhibitors evolved faster and exhibited more interspecific variability than other sites (Hill and Hastie 1987). This speculation was based upon the assumption that all amino acid changes at the reactive sites of a protease inhibitor are due to nucleotide substitutions which might have occurred in the same inhibitor gene. In the case of the Kunitz-type inhibitor, however,

this assumption may not always be warranted because the genetic variation at the reactive sites is sometimes created by gene conversion of the gene segment from duplicated genes (Li and Graur 1988).

As described above, the rate of amino acid substitution of the Kunitz-type inhibitor is estimated to be 1.2×10^{-9} per site per year. This value is nearly equal to that of hemoglobins, and it is not particularly high among the proteins studied. The rate of amino acid substitution for βAPP is 1.2×10^{-10} per site per year which is even lower. This rate was obtained from the sequence comparison between mouse and human βAPPs. These βAPPs differ at only one of the 51 amino acid residues compared, none of which are among the 12 residues of the enzyme-inhibitor contact regions. Because the amyloid β precursor itself is a very conservative protein, the inserted domain may have also been under strong functional constraints.

The rate of nucleotide substitution at all codon positions for βAPP was calculated to be 0.33×10^{-9} per site per year. This value was computed from a comparison between mouse and human βAPP sequences, assuming that the divergence time of these two species is 80 million years ago. The rate of synonymous substitution for βAPP is estimated to be 1.7×10^{-9} per site per year and the rate of nonsynonymous substitution is 5.2×10^{-11} per site per year. These values show that the evolution of these inserted domains is extremely slow at the amino acid level.

4.4 DISCUSSION

4.4.1 Relationships between amino acid sequences and specificity of inhibitors

Detailed information on the physiological functions of Kunitz-type protease inhibitors and their target-enzymes has not been compiled yet. It is known, however, that inhibitors having Lys or Arg at the reactive site (called the P_1 residue) tend to inhibit trypsin and trypsin-like enzymes. The amino acid residues on the amino-terminal side of the scissile bond, which is cleaved on the substrate, are numbered P_1 where P stands for peptide, P_2, P_3, and so forth (Figure 4.8) (Benyon and Bond, 1989). Those having Tyr, Phe, Trp (for artificial sequences only), Leu, and Met are likely to inhibit chymotrypsin and chymotrypsin-like enzymes. Those having Ala and Ser most inhibit elastase-like enzymes (Laskowski and Kato 1980). Table 4.3 showed the relationships between the P_1 residues and the target proteases.

In the case of snake toxins, those having neuromuscular activity have lost the protease inhibitor activity, even though they have characteristic residues such as Lys or Tyr at reactive sites within the contact regions with the protease (Table 4.4). Thus, the region responsible for the neuromuscular activity of snake toxins may be the contact regions other than the reactive sites. This does

not mean that the contact regions of the snake toxins with a protease are unnecessary for activity as a toxin. The contact of the toxin with a serine protease may be important for playing a regulatory role in enzymatic activities on the presynaptic membrane (Hawkins and Seeds 1986). It is indeed known that snake toxins are able to contact the protease even though they have lost the protease inhibitor activity (Dufton, 1985).

The phylogenetic tree in Figure 4.5 shows that an ancestral gene of mammalian Kunitz-type inhibitors was duplicated into groups II and III. Judging from the P_1 residue of the inhibitors in group II, I predict that the target enzymes for IαTI1 are chymotrypsin or chymotrypsin-like enzymes whereas the target enzymes for the other members in group II are trypsin or trypsin-like enzymes. Since all the inhibitors in group III have the identical P_1 residue, it is possible that the target enzymes are also identical. From the results of a comparison between the reactive sites of IαTI2 and βAPP, I predict that the target enzymes of the Kunitz-type inhibitor of βAPP are trypsin or trypsin-like enzymes. In the case of the snake toxins in group I, the changes at the P_1 residue might have occurred independently of the inhibitors in groups I and II. The results show that the gene duplications of group II took place after the duplication of IαTI1 and IαTI2. After this event, the inhibitors of group II changed their target enzymes.

As already mentioned, the inhibitors in group III have identical P1 residues. In particular, IαTI2, βAPP, trypstatin, and

fetal cell inhibitor have identical amino acid residues at the contact regions except for the residues which are variable in mammalian IαTI2. This observation suggests that these four inhibitors may have the same or at least similar target enzymes. This hypothesis is supported by the following facts. It is known that the V3 domain of HIV-1 envelope glycoprotein (gp120) is able to bind specifically tryptase (TL_2), which is a membrane-bound serine esterase in cultured human $T4^+$ lymphocytes. This binding activity was selectively blocked by trypstatin which is an inhibitor of tryptase. Trypstatin has a Gly-Pro-Cys-Arg (GPCR) sequence in its reactive site and the V3 domain of gp120 also has a similar sequence (Table 4.5) (Kido et al. 1991; Javaherian et al. 1989). This shows a possibility that other members of group III, which also have a GPCR sequence in the reactive sites, have a trypstatin-type protease as their target enzyme. It is also possible that βAPP works on the interaction between cells and the extra cellular matrix in the same way as those for gp120 and tryptase.

4.4.2 Biological function of protease inhibitor

The possibility cannot be ruled out that substitutions at sites other than the reactive sites change the target enzyme of the inhibitor, indicating that each inhibitor in group III has a different type of enzymes for its target. Interestingly, other serine proteases

such as urokinase are known to bind the cell membrane and to be thus related to the cell movement (Takahashi et al. 1991, Hawkins and Seeds 1986). The insertion site of the Kunitz-type inhibitor domain in the amino acid sequence of βAPP is located between Arg and Val. This site is known as a target site of the urokinase-type protease. Therefore, urokinase-type enzyme may be the target enzyme of the Kunitz-type inhibitor of βAPP.

There are many unclear points with respect to the physiological functions of the Kunitz-type inhibitor. The inserted Kunitz-type inhibitor may have a tertiary structure different from the uninserted Kunitz-domain. However, it is known that the inserted domain of the Kunitz-type inhibitor has the protease inhibitor activity when a protease cuts off the insertion sequence from amyloid β protein (Kitaguchi et al. 1990). This suggests that the insertion sequences of the Kunitz-domain are able to bind serine protease in the same way as other Kunitz-type inhibitors.

The phylogenetic tree in Figure 4.5 shows that the insertion occurred independently of βAPP and collagen type IV. The Kunitz-type domain of collagen IV has an evolutionarily close relationship with the Kunitz-type inhibitor of a fruit fly. This inhibitor is produced in the male accessory glands of this insect and strongly inhibits acrosin in sperm cells (Schmidt et al. 1989). Acrosin has an important role in the acrosomal reaction in the fertilization process. In particular, this enzyme seems to be directly involved in the interaction between sperm and egg. It is thought that the other

Kunitz-type inhibitors also have variable physiological functions as protease inhibitors, because they are found in virtually all tissues and have a wide range of specificities of target enzymes. The Kunitz-type inhibitor may also have an important role, through the contact with a protease on the membrane, in cell-to-cell recognition or interaction. These functions of proteases and the protease inhibitors on the cell membrane have yet to be elucidated.

References (Chpter 4)

Becker J, Schuppan D, Benzian H, Bals T, Hahn E G, Cantulppi C, and Reichart P (1986) Immunohistochemical distribution of collagen type IV, V, and VI and pro-collagen types I and III in human alveolar bone and dentine. *J. Histochem. Cytochem.* 343, 1417-1429.

Benyon, R J, and Bond J. S (1989) The Schecter and Berger nomenclature for protease subsites. in: Beynon R J, and Bond J S (eds) Proteolytic enzymes IRL PRESS New York pp231.

Bonaldo P, Russo V, Bucciotti F, Doliana R, Colombatti A (1990) Structural and functional features of the alpha-3 chain indicate a bridging role for chicken collagen VI in connective tissues. *Biochemistry* 29, 1245-1254.

Cechova D, Jonakova V, Sorm F (1971) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Creighton T E, Charles I G (1987) Sequences of the genes and polypeptide precursors for two bovine protease inhibitors. *J Mol Biol* 194, 11-22.

Dufton M J (1985) Proteinase inhibitors and dendrotoxins. sequence classification, structural prediction and structure/activity *Eur J Biochem* 153, 647-654.

Fioretti E, Binotti I, Barra D, Citro G, Ascoli F, Anonini E (1983) Heterogeneity of the basic pancreatic inhibitor (Kunitz) in various bovine organs. *Eur J Biochem* 130, 13-18.

Fioretti E, Iacopino G, Angreletti M, Barra D, Bossa F, Ascoli F (1985) Primary structure and antiproteolytic activity of a Kunitz-type inhibitor from bovine spleen. *J Biol Chem* 260, 11451-11455.

Fukuchi K, Martin G M, Deeb SS (1989) Sequence of the protease inhibitor domain of the A4 amyloid protein precursor of Mus domesticus. *Nucleic Acids Res* 17, 5396-5396.

Gbson M A, and Cleary E G (1983) Distribution of CL glycoprotein in tissues: an immunohistochemical study. *Collagen. Relat. Res.* 3, 469-488.

Gojobori T, Ishii K, Nei M (1982) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *J Mol Evol* 18, 414-423.

Graur D, Li W H (1988) Evolution of protein inhibitors of serine Proteinases: Positive Darwinian Selection or Compositional Effects? *J Mol Evol* 28, 131-135.

Hawkins R L, Seeds N W (1986) Effect of protease and their inhibitors on newrite outgrowth from neonatal mouse sensory ganglia in culture. *Brain Research* 398, 63-70.

Hill R E. and Hastie N D (1987) Accelerated evolution in the reactive center regions of serine protease inhibitors. *Nature* 326, 96-99

Hochstrasser K, Wachter E, Albrecht G J, Reisinger P (1985) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Hokama Y, Iwanaga S, Tatsuki T, Suzuki T (1976) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC.

Ikeo K, Takahashi K, Gojobori T (1991) Evolutionary origin of numerous kringles in human and simian apolipoprotein(a). *FEBS Lett* 287, 146-148.

Javaherian K, Langlois A J, McDanal C, Ross K L, Eckler L I, Jellis C L, Profy A T, Rusche J R, Bolognesi D P, Putney S D, Matthews T J (1989) Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. *Proc Natl Acad Sci USA* 86, 6768-6772.

Joubert F J, Strydom D J (1978) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Kang J, Mueller-Hill B (1989) The sequence of the two extra exons in rat preA4. *Nucleic Acids Res* 17, 2130-2130.

Kassell B, Laskowski M (1965) The basic inhibitor of bovine pancreas V. the disulfide linkages. *Biochem Biophys Res Commun* 20, 463-468

Kato I, Tominaga N (1979) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Kaumeyer J F, Polazzi J O, Kotick M P (1986) The mRNA for a proteinase inhibitor related to the HI-30 domain of inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin (protein HC). *Nucleic Acids Res* 14, 7839-7850.

Kido H, Yokogoshi Y, Katunuma N (1988) Kunitz-type protease inhibitor found in rat mast cells. Purification, properties, and amino acid sequence. *J Biol Chem* 263, 18104-18107.

Kido H, Fukutomi A, Katsunuma N (1991) Tryptase TL_2 in the membrane of human $T4^+$lymphocytes is a novel binding protein of the V3 domain of HIV-1 envelope glycoprotein gp120. *FEBS Lett* 287, 233-236.

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK

Kitaguchi N, Takahashi Y, Oishi K, Shiojiri S, Tokushima Y, Utsunomiya T, Ito H (1990) Enzyme specificity of proteinase inhibitor region in amyloid precursor protein of Alzheimer's disease: different properties compared with protease nexin I. *Biochim Biophys Acta* 1038, 105-113.

Kondo K, Toda H, Narita K, Lee C Y (1982) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Koo E H, Sisodia S S Price D L (1989) The sequence of the monkey amyloid precursor protein protease inhibitor domain. Genetic Sequence Data Bank (Genbank), California.

Laskowski M, Kato I (1980) Protein Inhibitors of Proteinases. *Annu Rev Biochem* 49, 593-626.

Laskowski M, Kato I, Lery T R, Schrode J, Sealock R W (1974) Evolution of Specificity of Protein Proteinase Inhibitors. Bayer-Symposium V "Proteinase Inhibitors", pp597-611.

Linseinmayer T F, Bruns R R, Mentzer A, and Mayne R (1986) Type VI collagen: immunohistochemical identification as a filament of the extracellular matrix of the developing avian cornel stroma. *Dev. Biol.* 118, 425-431.

McComb R D, Moul J M, and Bigner D D (1987) Distribution of type VI collagen in human gliomas: comparison with fibronectin and gliomamescenchymal matrix glycoprotein. *J. Newropathol. & Exp. Neurol.* 46, 623-633.

Muller-Hill B, Beyreuther K (1989) Molecular Biology of Alzheimer's Disease. *Annu Rev Biochem* 58, 287-307.

Nakamura T, Hirai T, Tokunaga F, Kawabata S, Iwanaga S (1987) Purification and amino acid sequence of Kunitz-type protease inhibitor found in the hemocytes of horseshoe crab (*Tachypleus tridentatus*). *J Biochem* 101, 1297-1306.

Nei M (1975) Molecular Population genetics and Evolution. North-Holland, Amsterdam.

Ponte P, Gonzalez-DeWhitt P, Schilling J, Miller J, Hsu D, Greenberg B, Davis K, Wallace W, Lieberburg I, Fuller F, Cordell B (1988) A new A4 amyloid mRNA contains a domain homologous to serine inhibitors. *Nature* 333, 525-527.

Ramesh N, Sugumaran M, Mole J.E (1988) Purification and characterization of two trypsin inhibitors from the hemolymph of Manduca sexta larvae. J *Biol Chem* 263, 11523-11527.

Ritonja A, Meloun B, Gubensek F (1983) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Sailer J P (1990) Inter-α-trypsin inhibitor: emergence of a family within the Kunitz-type protease inhibitor superfamily. *TIBS* 15, 435-439.

Saitou N, Nei M (1987) The neighber-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425.

Salvessen G, Nagase H (1989) Inhibitor of proteolytic enzymes Proteolytic enzymes. In: Beynon RJ, Bond JS (eds) Proteolytic enzymes IRL PRESS, UK, pp83-104.

Sakaki L Y, Keene D R, and Engvall E (1986) Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J. Cell Biol.* 103, 2499-2509.

Sasaki T (1984) Amino acid sequence of a novel Kunitz-type chymotrypsin inhibitor from hemolymph of silkworm larvae, Bombyx mori. *FEBS Lett* 168, 227-230.

Sasaki T. (1988) Amino-acid sequences of two basic chymotrypsin inhibitors from silkworm larval hemolymph. *Biol. Chem. Hoppe-Seyler* 369, 1235-1241.

Schmidt T, Stumm-Zollinger E Chen P S, Bohlen P, Stone S R (1989) A male gland peptide with protease inhibitory activity in Drosophila funebris. *J. Biol. Chem.* 264, 9745-9749.

Siekmann J, Wenzel H R, Schroeder W, Tschesche H (1988) Characterization and sequence determination of six aprotinin homologues from bovine lungs. *Biol. Chem. Hoppe-Seyler* 369, 157-163.

Strydom D J (1973) Protease Inhibitors as Snake Venom Toxins. *Nature new biol* 243, 88-89.

Strydom D J, Joubert F J (1981) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Takahashi K, Gojobori T, Tanifuji M (1991) Two-color Cytofluorometry and Cellular Properties of the Urokinase Receptor Associated with a Human Metastatic Carcinomatous Cell Line. *Exp Cell Res* 192, 405-413.

Takahashi H, Iwanaga S, Kitagawa T, Hokama Y, Suzuki T (1974) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Takahata N, Kimura M (1981) A model of evolutionary base substitutions and its aplication with special reference to rapid change of pseudogenes. *Genetics* 98, 641-657.

Tanzi R E, McClatchey A I, Lamperti E D, Villa-Komaroff L, Gusella J F, Neve R L (1988) Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease. *Nature* 33, 1528-530.

Tschesche H, Dieti T (1975) The amino-acid sequence of isoinhibitor K from snails (Helix pomata). *Eur J Biochem* 58, 439-451.

von der Mark H, Aumailley M, Wick G, Fleishmajer R, and Timpl R (1984) Immunochemistry, genuine size and tissue localization of collagen VI. *Eur. J. Biochem.* 142, 493-502.

Yamada T, Sasaki H, Dohura K, Goto I, Sakaki Y (1989) Structure and expression of alternatively-spliced forms of mRNA for the mouse homolog of Alzheimer's disease amyloid beta precursor. *Biochem Biophys Res Commun* 15, 8906-912.

Wachter E, Deppner K, Hochstrasser K, Lempart K, Geiger R (1980) Protein Identification Resource (PIR). National Biomedical Reserch Foundation, Washington DC

Wesselschmidt R L, Gifard T J, Broze Jr G J (1991) cDNA sequence of rabbit lipoprotein-associated coagulation inhibitor Genetic Sequence Data Bank (Genbank), California.

Wun T C, Kretzmer K K, Girard T J, Miletich J P Broze G J (1988) Cloning and characterization of a cDNA coding for the lipoprotein associated coagulation inhibitor shows that it consists of three tandem Kunitz-type inhibitory domains. *J Biol Chem* 263, 6001-6004.

Wunderer G, Machleidt W, Fritz H (1981) The broad-specificity proteinase inhibitor 5 II from the sea anemone Anemonia sulcata. *Meth Enzymol* 80, 816-820.

Table 4.1 **Tissue Distribution of Type VI Collagen Using Immunofluorescence.**

Tissue	Reference
Skin	McComb et al. (1987)
Kidney	McComb et al. (1987)
Muscle	Limsenmayer et al. (1986)
Brain, liver	McComb et al. (1987)
Blood vessels	von der Mark et al. (1984)
Amnion/chorion	Hessel and Engvall (1986)
Cornea, sclera, perichondrium	Linsenmayer et al. (1986)
Albeolar bone, dentine	Becker et al. (1986)
Spleen	McComb et al. (1987)
Thyroid, pancreas, heart	McComb et al. (1987)
Elastic cartilage	Sakai et al. (1986)
Nuchal ligament, lung, tendon	Gibson and Cleary (1983)

Table 4.2 Serine protease inhibitors used in the present analysis.

Protein	Abbreviation	Species	Data sorce
β-2 bungarotoxin B chain	β-BGT1	many-banded krait	Kondo et al. (1982)
β-2 bungarotoxin B chain	β-BGT2	many-banded krait	Kondo et al. (1982)
venom basic protease inhibitor	HHV II	Ringhals	Hokama et al. (1976)
venom basic protease inhibitor II	NNV II	cape cobra	Hokama et al. (1976)
venom basic protease inhibitor II	RRV II	viper russelli	Takahashi et al. (1974)
venom basic protease inhibitor B	dendrotoxin B	Black mamba	Strydom et al (1981)
venom basic protease inhibitor E	dendrotoxin E	Black mamba	Joubert et al (1978)
venom basic protease inhibitor I	dendrotoxin I	Eastern green mamba	Strydom, D. J. (1973)
venom basic protease inhibitor K	dendrotoxin K	Black mamba	Strydom, D. J. (1973)
dendrotoxin	dendrotoxin 1	Dendroaspis angusticeps	Dufton, M. J. (1985)
dendrotoxin	dendrotoxin 2	Dendroaspis angusticeps	Dufton, M. J. (1985)
venom basic protease inhibitor I	CTI	western sand viper	Ritonja et al. (1983)
venom chimotrypsin inhibitor	TI	sand viper	Ritonja et al. (1983)
venom basic protease inhibitor	VIII B	Bungars fasciatus	Dufton, M. J. (1985)
basic protease inhibitor	redturtle 1	red sea turtle	Kato et al. (1979)
basic protease inhibitor	redturtle 2	red sea turtle	Kato et al. (1979)
colostrum trypsin inhibitor	colostrum BPI	bovine	Cechova et al. (1971)
basic protease inhibitor	BPI	bovine	Kassell et al. (1965)
isoinhibitor I	BPI I	bovine	Siekman et al. (1988)
protease inhibitor II	BPI II	bovine	Fioretti et al. (1985)
spleen inhibitor precursor	SIG2	bovine	Creighton et al. (1987)
serum basic protease inhibitor	serum BPI	bovine	Wachter et al (1980)
chymotrypsin inhibitor I	silkworm I	silkworm	Sasaki, T. (1988)
	silkworm II	silkworm	Sasaki, T. (1988)

	silkworm III	silkworm	Sasaki, T. (1984)
hemolymph trypsin inhibitor B	inhibitor B	Tobacco hawk moth	Romesh et al. (1988)
proteinase inhibitor (BPI-type)	Horse crab	horseshe crab	Nakamura et al. (1987)
isoinhibitor K	snaile K	Roman snale	Tschesche et al (1975)
proteinase inhibitor 5.II	SA5 II	sea anemone	Wunderer et al. (1981)
hypothetical acrosin inhibitor	acrosin inhibitor	fruit fly	Wun et al. (1989)
lipoprotein-associated			
coagulation inhibitor	LACI 1	human	Girard et al. (1989)
	LACI 2	human	Girard et al. (1989)
	LACI 3	human	Girard et al. (1989)
Alzheimer's disease			
amyloid β protein precursor	βAPP	human	Tanzi etal. (1988)
		rhesus macaque	Koo et al. (1989)
		mouse	Fukuchi et al. (1989)
		rat	Kang et al. (1989)
inter α trypsin inhibitor	IαTI 1	human	Kaumeyer et al. (1986)
		bovine	Hochstrasser et al. (1985)
		horse	Hochstrasser et al. (1985)
		sheep	Rasp etal (1987)
		porcine	Hochstrasser et al. (1985)
inter α trypsin inhibitor	IαTI 2	human	Kaumeyer et al. (1986)
		bovine	Hochstrasser et al. (1985)
		horse	Hochstrasser et al. (1985)
		sheep	Rasp etal (1987)
		porcine	Hochstrasser et al. (1985)
	TRYPSTATIN	rat	Kido et al. (1988)
collagen α 3(VI) chain	collagen	chicken	Bonald et al. (1989)

Table 4.3 Amino acid residues at the reactive site (P_1).

Amino acid residue at reactive site (P_1)	Target protease
Lys, Arg	Trypsin Trypsin-like enzymes
Tyr, Phe, Trp, Leu, Met	Chymotrypsin Chymotrypsin-like enzymes
Ala, Ser	elastase-like enzymes

Table 4.4 **The biological activity and amino acid sequences of the inhibitors in group I.**

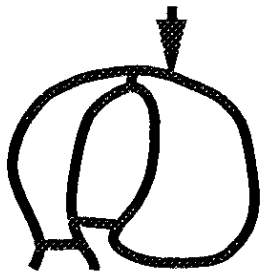
Source	Reactive sites	Activity	
		Neuromuscle	Antiprotease
dendrotoxin I	GRCY	+	-
dendrotoxin 1	GRCY	+	-
dendrotoxin 2	GPCK	+	-
dendrotoxin K	GPCKR	+	-
HHV II	GLCK	-	+
NNV II	GLCK	-	+
dendrotoxin E	GRCY	-	+
CTI	GRCL	-	+
redturtle2	GPCK	-	+
TI	GRCK	-	+
VIII B	GRCN	-	+
RRV II	GRCR	-	+

Table 4.5 **Contact regions of the Kunitz-type inhibitors in group III.**

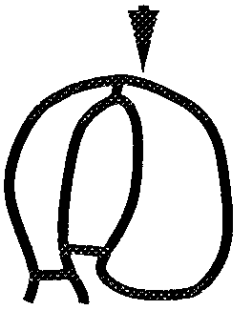
Inhibitor	Contact regions	
	7 15	30 35
IaTI2 (bovine)	VQGPCRAFI	SYGGCK
IaTI2 (human)	R--------	D----Q
IaTI2 (horse)	---------	T----R
IaTI2 (porcine)	---------	N----Q
IαTI2 (sheep)	R------G-	I----N
trypstatin (rat)	-------A-	I-----
APP (human)	ET-----M-	F----G
APP (rhesus)	ET-----M-	F----G
APP (mouse)	ET-----M-	F----G
APP (rat)	ET-----M-	F----G
tobaccow honeworm	EV-----GF	T----Q
HIV 1	I Q R--G---V	

Figure 4.1 **The secondary structure of Kunitz-type protease inhibitors.**

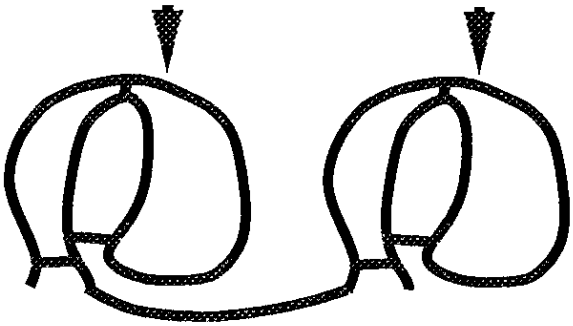
Arrows indicate the reactive center. Abbreviation; BPTI, bovine pancreatic trypsin inhibitor; snake, venom protease inhibitor; IaTI, inter α trypsin inhibitor; LACI, lipid associated coagulation inhibitor.



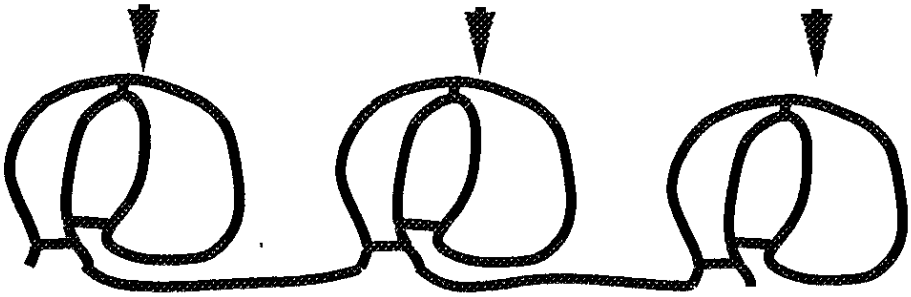
BPTI



snake



IαTI



LACI

Figure 4.2 **The Kunitz-type protease inhibitor domain inserted into βAPP.**

This figure shows the schematic model of βAPP (Kang et al. 1987, Kitaguchi et al. 1987).

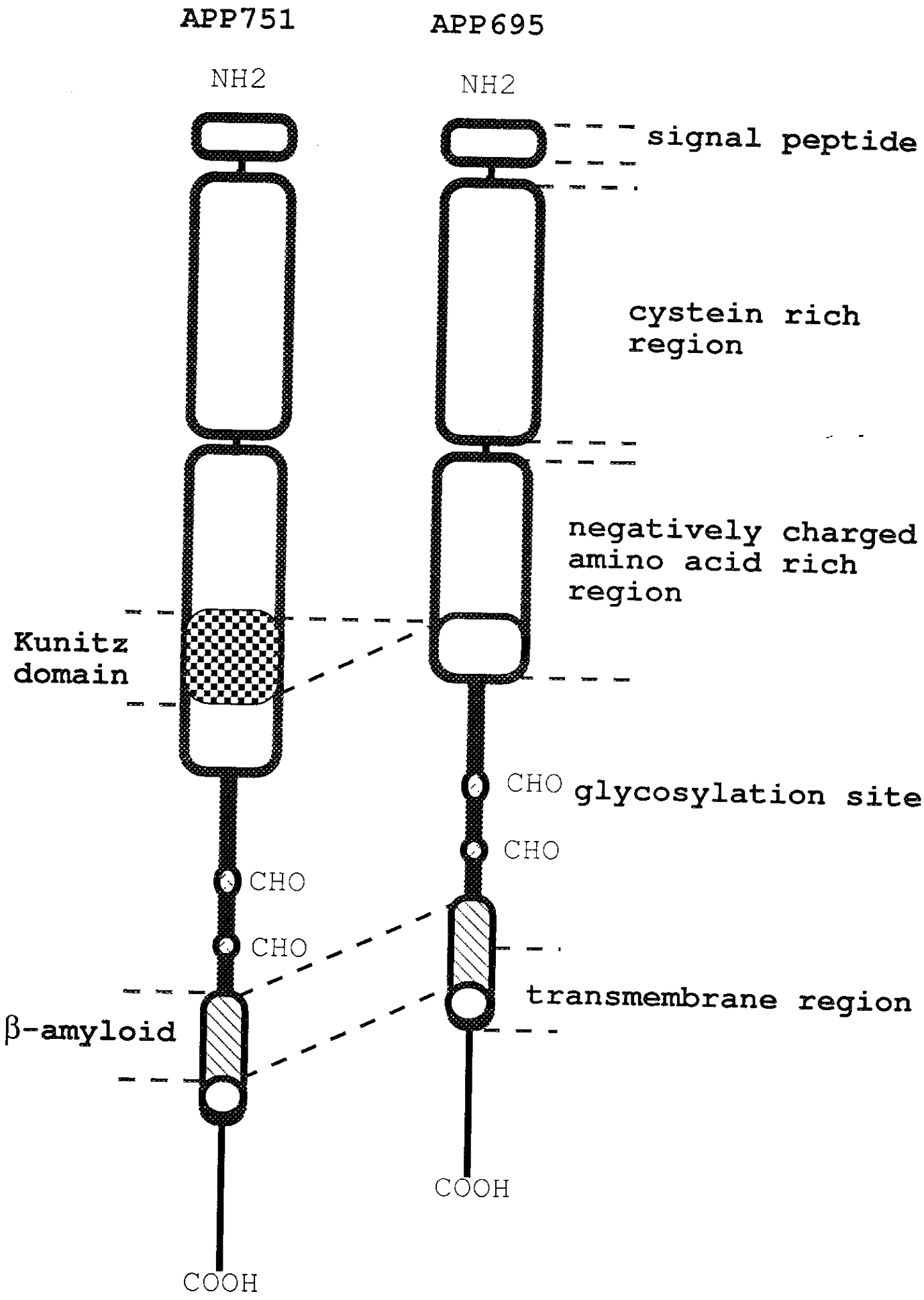

APP751
APP695
NH2
NH2
signal peptide
cystein rich region
negatively charged amino acid rich region
Kunitz domain
CHO
glycosylation site
CHO
CHO
CHO
transmembrane region
β-amyloid
COOH
COOH

Figure 4.3 **The Kunitz-type protease inhibitor domain inserted into α3 chain of collagen type VI.**

This figure shows the schematic model of the Kunitz-type protease inhibitor of collagen type VI Bonald et al. 1989). COL indicates the position of the triple helix.

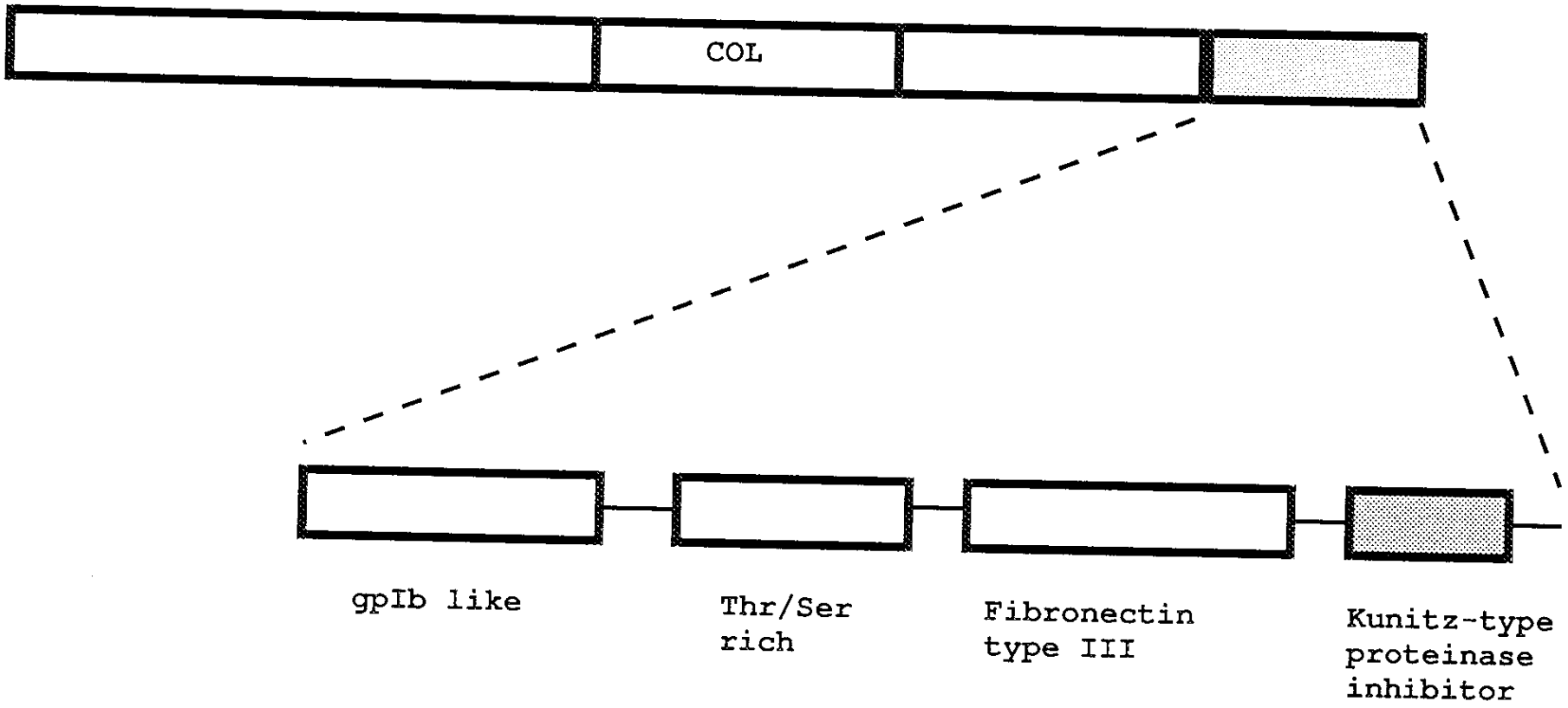

COL
gpIb like
Thr/Ser
rich
Fibronectin
type III
Kunitz-type
proteinase
inhibitor

Figure 4.4 **Sequence alignment of Kunitz-type protease inhibitors.**

This figure shows a part of sequences used in this study.

```
                C———————— C———————— ————————C- ——YGGC—— ———————— —C——C
I
  BPI (bov)     CLEPPYTGP‡ CKARIIRYFY NAKAGL‡‡CQ TFVYGGCRAK RNNFKSAED‡ ‡‡CMRTC
  SERBPI (bov) CLEPPYTGP‡ CKAAMIRYFY NAKAGF‡‡CE TFYYGGCRAK SNNFKSAED‡ ‡‡CMRTC
  INTBPI (hu)   CQLGYSAGP‡ CMGMTSRYFY NGTSMA‡‡CE TFQYGGCMGN GNNFVTEKE‡ ‡‡CLQTC
  BPIREDTAT     CRLPPEQGP‡ CKGRIPRYFY NPASRM‡‡CE SFIYGGCKGN KNNFKTKAE‡ ‡‡CVRAC
  SNBPI (lcob) CELPAETGL‡ CKAYIRSFHY NLAAQQ‡‡CL QFIYGGCGGN ANRFKTIDE‡ ‡‡CRRTC
  SNBPI (kaco) CELPAETGL‡ CKARIRSFHY NRAAQQ‡‡CL EFIYGGCGGN ANRFKTIDE‡ ‡‡CHRTC
  COLBPI (bov) CQLPQARGP‡ CKAALLRYFY BSTSNA‡‡CE PFTYGGCQGN NBNFETTEM‡ ‡‡CLRIC
II
  SNBPIE (blm) CKLPAEPGP‡ CKASIPAFYY NWAAKK‡‡CQ LFHYGGCKGN ANRFSTIEK‡ ‡‡CRHAC
  SNBPI. III    CYLPADPGR‡ CLAYMPRFYY NPASNK‡‡CE KFIYGGCRGN ANNFKTWDE‡ ‡‡CRHTC
  SNBPI. I L    CYLPADPGR‡ CKAHIPRFYY DSASNK‡‡CK FFIYGGCPGN ANNFKTWDE‡ ‡‡CRQTC
  SNBPI. II     CNLAPESGR‡ CRGHLRRIYY NLESNK‡‡CK VFFYGGCGGN ANNFETRDE‡ ‡‡CRETC
  SNBPIK (blm) CKLPLRIGP‡ CKRKIPSFYY KWKAKQ‡‡CL PFDYSGCGGN ANRFKTIEE‡ ‡‡CRRTC
  SNBPIB (blm) CELIVAAGP‡ CMFFISAFYY SKGANK‡‡CY PFTYSGCRGN ANRFKTIEE‡ ‡‡CRRTC
  MILK (bov)    CQLPQARGP‡ CKAALLRYFY NSTSNA‡‡CE PFTYGGCQGN NNNFKSAED‡ ‡‡CMRTC
  SN. HEM       CNLAPESGR‡ CRGHLRRIYY NLESNK‡‡CK VFFYGGCGGN ANRFKTIDE‡ ‡‡CRRTC
  SN. DPOL I    CILHRNPGR‡ CYQKIPAFYY NQKKKQ‡‡CE GFTWSGCGGN SNRFKTIEE‡ ‡‡CRRTC
  INBP (horse) CQLDHAQGP‡ CLGMISRYFY NGTSMA‡‡CE TFQYGGCLGN GNNFASQKE‡ ‡‡CLQTC
  INBPI (poc)   CQLGYSQGP‡ CLGMIKRYFY NGSSMA‡‡CE TFHYGGCMGN GNNPVSQKE‡ ‡‡CLQTC
  INB (sheep)   CQLGYSQGP‡ CLGMFKRYFY NGTSMA‡‡CE TFYYGGCMGN GNNFPSEKE‡ ‡‡CLQTC
  BPI2 (redta   CYLPADDGR‡ CKAHIPRFYY DSASNK‡‡CN KFIYGGCPGN ANNFKTWDE‡ ‡‡CRQTC
  SA5. II       CELPKVVGP‡ CRARFPRYYY NSSSKR‡‡CE KFIYGGCGGN ANNFHTLEE‡ ‡‡CEKVC
  snale. K      CNLPAETGP‡ CKASFRQYYY NSKSGG‡‡CQ QFIYGGCRGN QNRFDTTQQ‡ ‡‡CQGVC
  horsecrab     CTSPPVTGP‡ CRAGFKRYNY NTRTKQ‡‡CE PFKYGGCKGN ‡‡GNRYKSEQ ‡DCLDAC
  LACI1         CAFKADDGP‡ CKAIMKRFFP NIFTRQ‡‡CE EFIYGGCEGN QNRFESLEE‡ ‡‡CKKMC
  LACI2         CFTEEDPGI‡ CRGYITRYFY NNQTKQ‡‡CE RFKYGGCLGN MNNFETLEE‡ ‡‡CKNIC
  LACI3         CLTPADRGL‡ CRANENRFYY NSVIGK‡‡CR PFKYSGCGGN ENNFTSKQE‡ ‡‡CLRAC
III
  INTBPI2 (hu) CNLPVYRGP‡ CRAFIQLWAF DAVKGK‡‡CV LFDYGGCQGN GNKFYSEKE‡ ‡‡CREYC
  INTBP2 (bov) CNLPIVQGP‡ CRAFIQLWAF DAVKGK‡‡CV RFSYGGCKGN GNKFYSQKE‡ ‡‡CKEYC
  INB2 (horse) CNLPIVQGP‡ CRAFIRLWAF DAAQGK‡‡CV LFTYGGCRGN GNKFYSQKE‡ ‡‡CKEYC
  INBP2 (poc)   CSLPIVQGP‡ CRAFIRLWAF DAAQGK‡‡CV LFNYGGCQGN QNKFYSQKE‡ ‡‡CKEYC
  INB2 (sheep) CNLPIVRGP‡ CRAGIELWAF DAVKGK‡‡CV RFIYGGCNGN GNQFYSQKE‡ ‡‡CKEYC
  ALZAPP        CSEQAETGP‡ CRAMISRWYF DVTECK‡‡CA PPFYGGCGGN RNNFDTEEY‡ ‡‡CMAVC
  TRYPSTA       CNLPIVQGP‡ CRAFAELLAF DAAQGK‡‡CI QFIYGGCKGN NNKFYSEPK‡ ‡‡CKWYC
  FCELLBPTI     CNLPIVQGP‡ CRAFAELLAF DAAQGK‡‡CI QFIYGGCKGN NNKFYSEPK‡ ‡‡CKWYC
  VIchain       CLLQKEEGT‡ CRDFVLKWHY DLKTK‡‡SCA RFWYGGCNEN ‡‡RFNTQKE‡ ‡‡CEKAC
  TOBACCOW      CSLPPEVGP‡ CRAGFLKFAY YSELNK‡‡CK LFTYGGCQGN ‡‡‡ENNFETL QACXQA‡
  1BNGBCOB      CDKPPDKGN‡ CGP‡VRAFYY DTRLKT‡‡CV QFRYGGCDGD HGNFKSDHL‡ ‡‡CRCEC
  2BNGBCOB      CDKPPDTKI‡ CQT‡VRAFYY KPSAKR‡‡CV QFRYGGCDGD HGNFKSDHL‡ ‡‡CRCEC
  TAP           CIKPRDWIDE CDSNEGGERA YFRNGKGGCD SFWI‡‡CPED HTGADYYSSY RDCFNAC
```

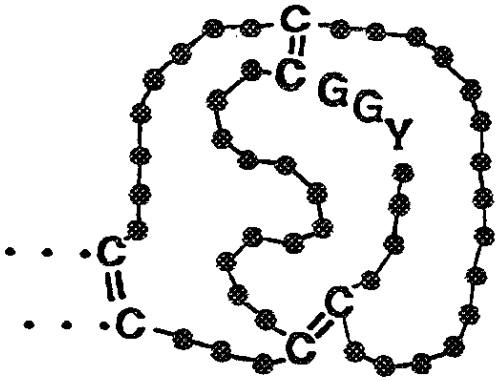


Figure 4.5 **The phylogenetic tree constructed using amino acid sequences of Kunitz-type protease inhibitors.**

The tree was constructed by the analysis of amino acid sequences. Three major groups are indicated by I, II, and III.

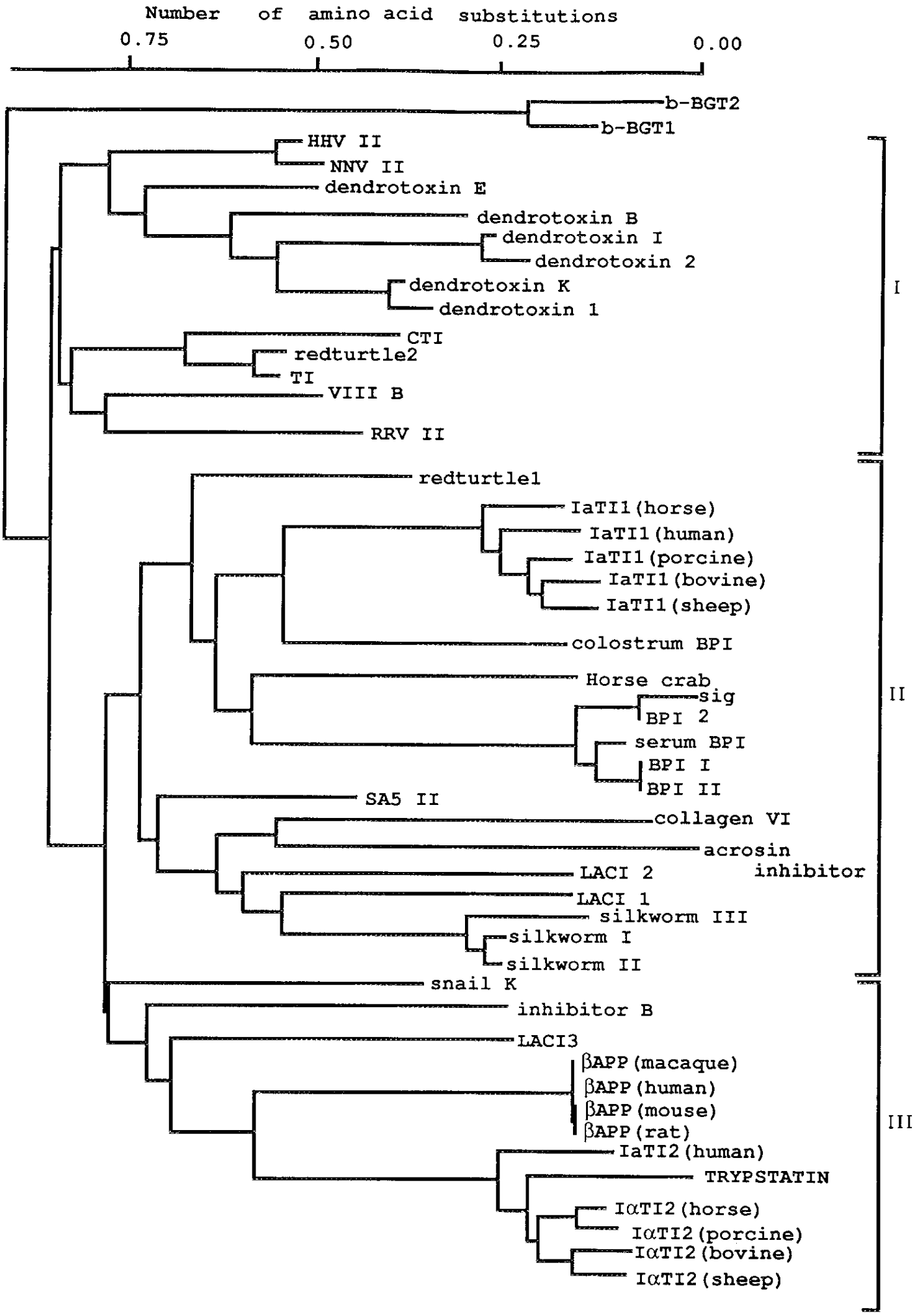

Number of amino acid substitutions
0.75
0.50
0.25
0.00
b-BGT2
b-BGT1
HHV II
NNV II
dendrotoxin E
dendrotoxin B
dendrotoxin I
dendrotoxin 2
dendrotoxin K
dendrotoxin 1
CTI
redturtle2
TI
VIII B
RRV II
redturtle1
IaTI1(horse)
IaTI1(human)
IaTI1(porcine)
IaTI1(bovine)
IaTI1(sheep)
colostrum BPI
Horse crab
sig
BPI 2
serum BPI
BPI I
BPI II
SA5 II
collagen VI
acrosin inhibitor
LACI 2
LACI 1
silkworm III
silkworm I
silkworm II
snail K
inhibitor B
LACI3
βAPP(macaque)
βAPP(human)
βAPP(mouse)
βAPP(rat)
IaTI2(human)
TRYPSTATIN
IαTI2(horse)
IαTI2(porcine)
IαTI2(bovine)
IαTI2(sheep)
I
II
III

Figure 4.6 Evolutionary changes of Kunitz-type protease inhibitors.

The ancestral molecule of the Kunitz-type domain appeared at least 500 million years ago. The Kunitz-type domain was duplicated to form two types, groups I and II, about 450 million years ago. An ancestral IαTI protein emerged by duplication about 400 million years ago. Each domain of IαTI formed groups II and III. Group III includes βAPP which diverged from its ancestor about 270 million years ago.

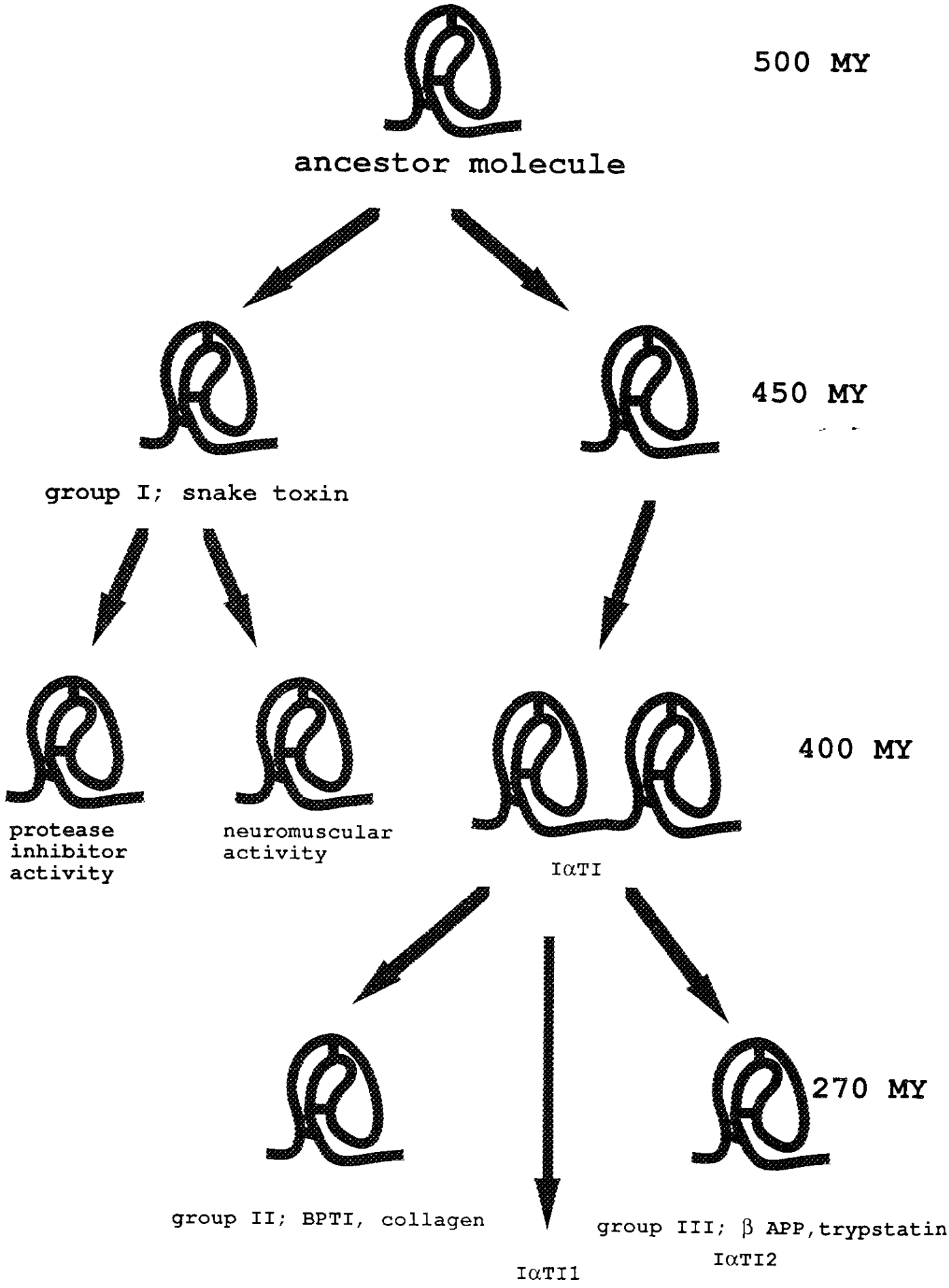

500 MY
ancestor molecule
450 MY
group I; snake toxin
protease inhibitor activity
neuromuscular activity
400 MY
IαTI
270 MY
group II; BPTI, collagen
IαTI1
group III; β APP,trypstatin
IαTI2

Figure 4.7 **The phylogentic tree constructed using nucleotide sequences of Kunitz-type protease inhibitors.**

The total number of nucleotide substitutions for each pair of Kunitz-type sequences was first estimated by the four-parameter method. Using this number, the phylogenetic tree was constructed by the neighbor-joining method.

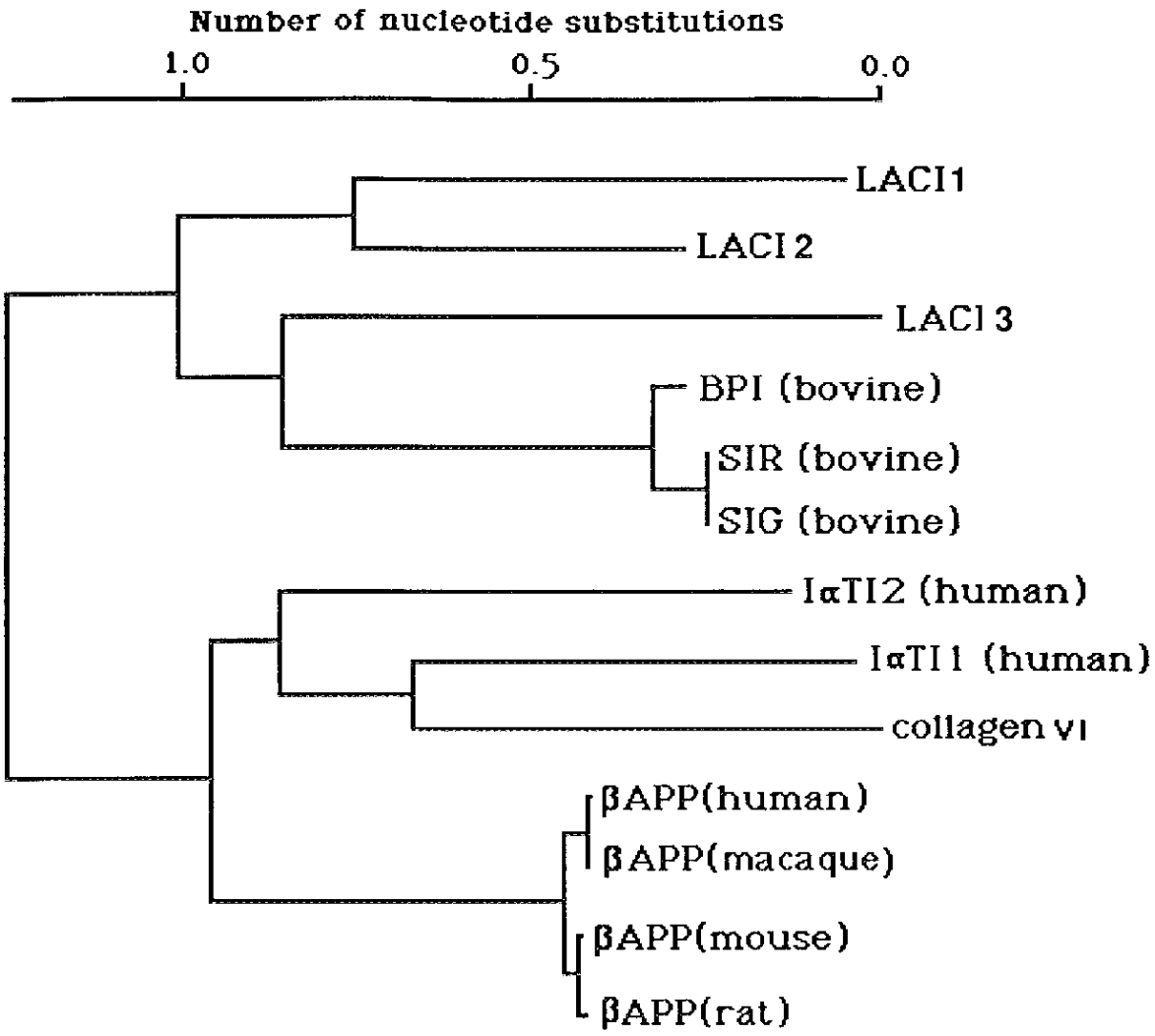
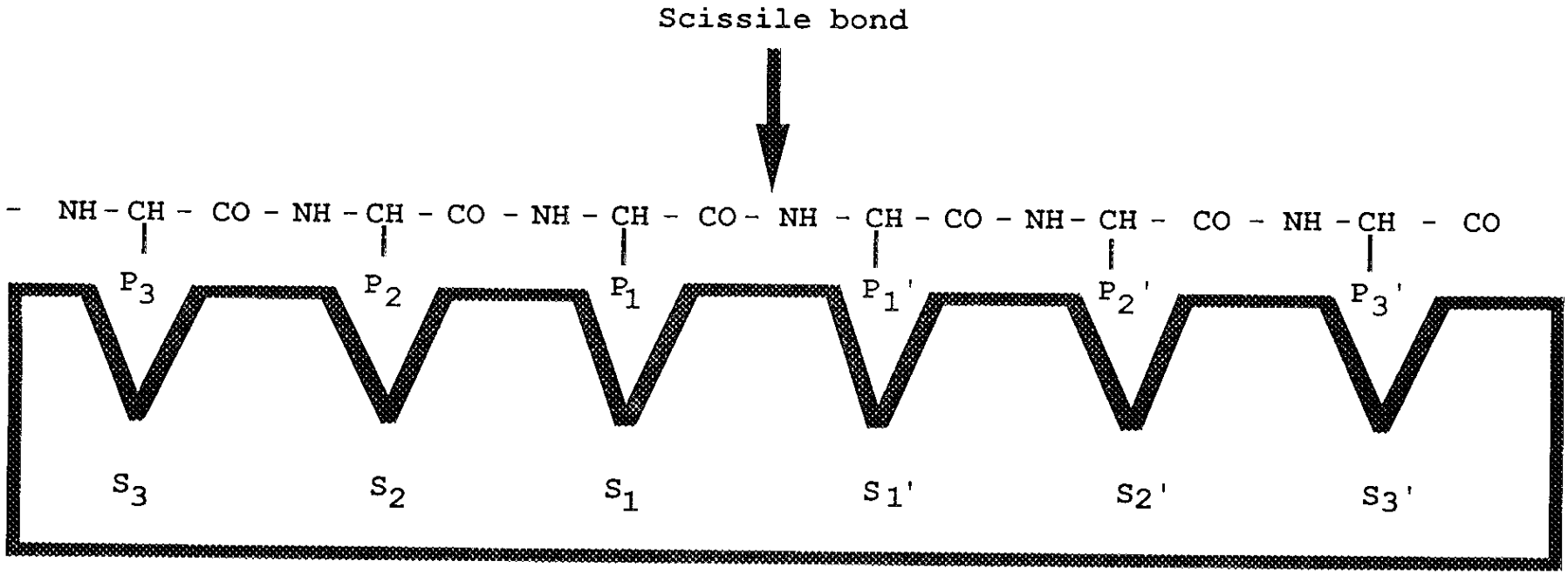

Number of nucleotide substitutions
1.0
0.5
0.0
LACI1
LACI2
LACI3
BPI (bovine)
SIR (bovine)
SIG (bovine)
IαTI2 (human)
IαTI1 (human)
collagen VI
βAPP(human)
βAPP(macaque)
βAPP(mouse)
βAPP(rat)

Figure 4.8 The Schecter and Berger's nomenclatures for binding of a peptide substrate to a peptidase.

The schecter and Berger's nomenclatures for binding of a peptide substrate to a peptidase are shown. The protease is represented as the solid line. P_1,.P_2, P_3, P_1', P_2', and P_3' represent the side chains of six amino acids, and S_1,.S_2, S_3, S_1', S_2', and S_3' are the corresponding subsites on the enzyme. The "scissile bond" indicate the peptide bond that is digested by enzyme.


Scissile bond
- NH - CH - CO - NH - CH - CO - NH - CH - CO - NH - CH - CO - NH - CH - CO - NH - CH - CO
P3
P2
P1
P1'
P2'
P3'
S3
S2
S1
S1'
S2'
S3'

CHAPTER FIVE

EVOLUTION OF MOSAIC PROTEINS

5.1 Introduction

Molecular evolution of mosaic proteins has been concretely exemplified by serine proteases and its protein inhibitors, as discussed in the preceding chapters. In particular, the evolutionary process of the kringle domains in serine proteases is different from that of the protease domains, which was clearly shown in Chapter 3. This observation strongly suggests that for the serine proteases, the unit of evolution is a functional domain rather than the whole protein. It also implies that mosaic proteins must have been created by a fusion of different functional domains, which usually result from genetic recombinations between their encoding gene segments at the DNA level.

As shown in Chapter 2, the evolutionary study of kringle domains in the serine proteases exquisitely revealed an interesting feature that a functional domain was duplicated many times, in the

evolutionary process, into more than two domains. These duplicates were evolutionarily shuffled so that they were used as parts of a certain mosaic protein. Although amino acid substitutions and other mutational events must have disturbed the amino acid sequence of a functional domain during evolution, a certain pattern of the amino acid sequence, which is called a motif, has remained unchanged. The kringle structure was also shown as a typical example of such evolutionary motifs, in Chapter 2.

The minikringle of a Kunitz-type protease inhibitor is structurally similar to the kringle structure of a protease. However, it was shown, in Chapter 4, that the minikringles have their own features of molecular evolution, and particularly that their evolutionary process is independent of that of the kringles.

Recently, it is found that there are kringle-like structures which are different from the kringles in serine proteases and the minikringles of the Kunitz-type protease inhibitors. In fact, a much smaller size of the kringle structure can be found in defensin that is an anti-bacterial protein in the white blood cells called neutrophils (Hill et al. 1991). The tick-borne encephalititis virus is also known to have a kringle-like structure.

In the present chapter, I would discuss the evolutionary relationship between the kringle-like structures and the kringle structures in serine protease. In particular, I would examine whether the kringle-like structures had shared a common ancestor with the kringle structures. Moreover, I would discuss the

evolutionary significance of various serine proteases as mosaic proteins in the system of biological cascades.

5.2 Molecular evolution of kringle-like structures

5.2.1 Kringle-like structures in defensins

Defensin is a protein that is present in neutrophils. Neutrophils are the cells which constitute fifty to seventy percent of the total white blood cells in humans. The neutrophils are involved in the immune response by ingesting invading microorganisms, which are then destroyed by either one of the two general mechanisms, the "oxygen-dependent" and "oxygen-independent" mechanisms. The former mechanism results from the production of superoxides which are converted to potent oxidants termed "reactive oxygen intermediates" (Klebanoff 1988). The latter defense mechanism is activated when the microbicidial-cytotoxic proteins of cytoplasmic granules are discharged into the phagocytic vacuole (White et al. 1986). In this "oxygen-independent" mechanism, the defensin plays a vital role in attacking cell membranes of the invading microorganisms.

Defensins account for about thirty percent of the total proteins in azurophil granules in human neutrophils. A protein molecule of the defensin is small in molecular weight of only 3500 to 4000. It is a cationic and disulfide cross-linked protein. It has an *in vitro* activity of destroying Gram-negative and Gram-positive bacteria, fungi, mammalian cells, and enveloped viruses. For example, defensins permiabilize both inner and outer membranes of *Escherichia coli*, so that the cell death takes place by the membrane permeabilization (Hill et al. 1991).

The three-dimensional structure of a human defensin called HNP-3 was determined by the method of x-ray diffraction (Hill et al. 1991). The pattern of conserved residues in the amino acid sequences of the defensin family (Figure 5.1) suggests that all defensins have structural conformation similar to HNP-3. These protein molecules of the defensin family show the same topology with the kringles and minikringles. Although the amino acid sequences of a kringle and a minikringle are, respectively, about eighty amino acids and fifty-one amino acids long, the amino acid sequence of defensin is only twenty-nine amino acids long. Thus, the kringle-like structure of a defensin is much smaller than the minikringles. Moreover, there is no sequence homology among the kringle-like structures, the minikringles, and kringles. It suggests that these three similar structures do not share the common ancestor.

In order to know the evolutionary history of defensins, I constructed the phylogenetic tree for defensins of various organisms (Figure 5.2). The defensins could be separated into two groups which are indicated by groups A and B. Group A contains defensins of human and guinea pig, and rabbit defensins NP-1, -2, -3, and -3b. Group B contains rat defensins and rabbit defensins NP-4 and -5. Thus, the phylogenetic tree obtained shows that the rabbit defensins were separated into the two groups before the divergence of humans and rabbits. It implies that the evolutionary origin of defensins are much older than the mammalian divergence.

For both groups A and B, the evolutionary divergence of the defensins within a group appear to have occurred recently. In particular, the divergence time between human defensins HNP-2 and HNP-3 may be only a few million years ago. This divergence time was obtained from the rate of amino acid substitution that was estimated to be 6.24×10^{-9} per site per year. The substitution rate was estimated by comparing the amino acid sequence of human defensin HNP-1 with that of rabbit NP-1 and making the assumption that these two sequences diverged at the time of speciation between humans and rabbits (80×10^6 years ago). Thus, many of defensins must have emerged from their recent ancestors possibly by repeated gene duplications. This evolutionary feature of the defensins is similar to that of the kringle structures in the serine proteases, although the evolutionary origins of these two structures are different from each other.

The kringle-like structures of the defensins may have their own biological significance. However, it is possible that the kringle, minikringle, and kringle-like structures share a certain type of functional constraints, because the particular conformations of the amino acid sequences by three pairs of cysteine-cysteine bonds are evolutionarily conserved among the three different structures.

5.2.2 Kringle-like structures in an envelope protein of TBE virus

Tick-borne encephalitis virus (TBEV) represents one of the major human pathogens of the family Flaviviridae. The disease caused by this virus is endemic in many European countries, Russia, and China. This disease can be effectively prevented by a highly purified inactivated whole-virus vaccine (Kunz et al. 1980). The envelope protein, E, plays a central role in the biological function of flaviviruses. It mediates hemaglutination, binding to cell receptors and probably fusion with endosomal membranes. It also induces a protective immune response after active immunization or natural infection.

It was revealed, by competitive binding analyses, that the epitope clusters in the E protein make three major antigenic domains (designated by I, II, and III) (Figure 5.3). All cysteins are highly conserved in the amino acid sequences of all flavivirus E

proteins. The I domain has five characteristic cysteine bridges. Three of them are composed of a characteristic topology that is the same as kringles, but its amino acid sequence of the kringle-like structure is 62 amino acids long. Although the size of this kringle-like structure is closer to that of the minikringle rather than the kringle, there are no homology in the amino acid sequences between these structures. Moreover, it does not show any sequence homology with the defensin, either. This observation thus suggests that there are no evolutionarily common ancestor among these structures. Since the biological function of the kringle-like structure in the E protein has not been known, we cannot rule out the possibility that the kringle-like structure has been evolutionarily exposed by functional constraints similar to that of the kringles.

5.3 Evolution of serine proteases in the system of biological cascades

5.3.1 Biological functions of regulatory domains involved in serine proteases

In the preceding sections, I showed that the kringle-like structures of the defensin and the E protein of TBEV did not appear to share a common ancestor with the kringle structures in the serine proteases. In Chapter 2 and section 5.1 in this chapter, I have emphasized the evolutionary importance of the kringle structure in exhibiting its regulatory function of the serine proteases in the biological cascades such as the blood coagulation system and the system of fibrinolysis. As I have already mentioned before, the regulatory function of the serine proteases can be attributed to not only the kringle domains but also other regulatory domains. Although it is commonly understood that the essential function of a serine protease is the proteolysis of a target enzymes or zymogens, I would now point out the possibility that serine proteases may play more important role in regulating the biological cascades by a function of their regulatory domains.

Plasminogen activators catalyze the conversion of the plasminogen into the plasmin, which can then activate or degrade extracellular proteins such as zymogens, fibrinogen, and fibronectin

(Robbins et al. 1967). Through this process, plasminogen activators regulate extracellular proteolysis, fibrin clotlysis, tissue remodeling, developmental cell migration, and pathological processes as inflammation and metastasis (Reich et al. 1978). These processes appear to be associated with a variety of biological cascades. In particular, it is recently found that urinary plasminogen activator, uPA, has a regulatory role in other forms of extracellular proteolysis (Appella et al. 1987). Usually, the activity of extracellular proteolysis is simply directed by the production of a protease such as uPA without any other control mechanisms. However, it was newly discovered that extracellular proteolysis is regulated by the expression of the uPA gene which is controlled by stimulation of a growth factor in normal quiescent cells (Grimaldi et al. 1986, Appella et al. 1987). It is also shown that normal cells possess a uPA-specific receptor. The existence of this receptor might ensure that the proteolytic activity of uPA can take place at a particular location on the cell membrane. In fact, it is known that the receptor-bound uPA is not internalized in the cytoplasm (Appella et al. 1987). Thus, the newly discovered form of extracellular proteolysis seems to constitute a certain cascade, and uPA plays a regulatory role in the cascade. Because the receptor-binding domain of uPA is located in the amino-terminal fragment (ATF) of uPA which is different from the catalytic domain (Stoppelli et al. 1985), the regulatory role appeared to be played by a regulatory domain, but not the protease domain.

Tissue-type plasminogen activator, tPA, may also play a regulatory role in a biological cascade which might be different from the blood coagulation system. This is because tPA-specific receptors are found to be widely distributed on peripheral blood cells and vascular endotherial cells (Scott 1989). Although the receptor-binding domain of tPA has not been identified at present, it is possible that a domain different from the catalytic domain of tPA is responsible for this receptor binding activity.

It is quite possible that uPA and tPA are involved in a cascade system of signal transduction, because both uPA and tPA are phosphorylated when they bind their own receptors (Takahashi et al. 1989). It is thought that uPA and tPA may be involved in the signal transduction through the phosphorelation of possibly their regulatory domains, since it binds the receptors on the cell membrane. Thus, the regulatory domains may be more important than the catalytic domain, if the serine proteases are, indeed, related to signal transduction. In practice, the kringle and EGF domains seems to play an important role in binding to the receptor and in making specific modification of the enzymes, by which the enzymatic activity is controlled in the system of signal transduction.

It has been already known that the protease activities are directly regulated by the protease inhibitors and the substrate specificity. Only the protease domain is primarily responsible for this type of direct regulation. Since the direct regulation by the protease domain is common to all proteases, a variety of more

exquisite regulations such as differentiated activations of the protease and the selective timing of the protease production cannot be carried out by the protease domain. Thus, the regulatory domains of proteases may be more important for the regulation of a biological cascade. It is also possible that proteases having a wide range of specificities for substrates may be able to delicately control their enzymatic activity with the aid of the regulatory domain of the protease.

5.3.2 Structural similarity in catalytic triads between protease and lipase

In the above subsection, I pointed out the evolutionary significance of the regulatory domains of serine proteases in biological cascades. On the other hand, the protease domain is responsible for catalytic reactions of proteolysis that is an essential function of serine proteases. The protease domain contains a region that is critically important for a proteolytic function. This region is called the active site or the catalytic site. In this subsection, I would discuss the evolutionary significance of the active site in the protease domain with special reference to a structural similarity between serine protease and lipase.

At the active site of all serine proteases, three residues of serine (Ser), histidine (His), and aspartic acid (Asp) are strongly

conserved in the evolution of the protease domains. The three-dimensional structure of chymotrypsin revealed that Ser is located within a hydrogen-bonding distance of a ring nitrogen ($N\varepsilon 2$) of His, and that the other nitrogen of His is hydrogen-bonded to the carboxylate group of Asp (Figure 5.4). This system composing of three residues at the catalytic site is called the catalytic triad. As shown in Figure 5.5, the catalytic triad is considered to ensure the enzymatic activity that is mediated by the electric charge-relay system within the catalytic triad (Dunn 1989). The same system was also discovered in the structure of subtilisin (Creighton 1984). Because there is no similarity in the structures of subtilisin and the chymotrypsin family, the catalytic triad of the chymotrypsin family may have evolved independently of that of subtilisin (Kraut 1977, Neurath 1989).

Interestingly, the lipase is found to have a catalytic triad at its active site (Brady 1990). The enzymatic mechanism of the lipase is the same as that of the serine proteases. For example, triacylglycerol lipase contains a catalytic triad that forms the catalytic center of a triacylglycerol. Moreover, pancreatic lipase has also a catalytic triad. Although the serine at the 152th residue of pancreatic lipase is the nucleophilic residue essential for catalysis, it is a part of an Asp-His-Ser triad which is chemically analogous to serine proteases (Figure 5.6). It is known that the molecular structures of the whole proteins for triacylglycerol and pancreatic lipase are different from those of any serine proteases

(Winkler et al. 1990). It implies that although lipases and serine proteases do not share the common origin of catalytic triads, the system of catalytic triads has been evolutionarily conserved under similar functional constraints. This evolutionary situation of the protease domains in the serine proteases is just like that of kringle-like and kringle domains.

In the case of the lipase of *Geotrichum candidum* (GCL), the catalytic triad of GCL has a configuration of Ser-His-Glu where an aspartic acid of the serine proteases is replaced by a glutamic acid (see Figure 5.7, Schrag et al. 1991). Although this lipase has a sequence similarity with the other lipases in only the region near Ser at the active site, there is some similarity in their tertiary structures. Sequence comparisons with proteins from the cholinesterase family suggest that they also contain the Ser-His-Glu triad (Blow 1990). This observation suggests that if the motifs in the amino acid sequence are the same between different proteins, the molecular structures and possibly biological functions are similar to each other. However, it also suggests that the same motifs at the amino acid sequences do not necessarily reflect the evolutionarily common origin in some cases. Thus, we need some cautions when the evolutionary features are discussed only on the basis of the sequence motifs.

5.4 Evolutionary roles of domain shuffling

As explained in the previous sections, serine proteases and lipases have the same enzymatic mechanisms that are realized by similar catalytic triads. According to the sequence comparisons of the whole proteins between serine proteases and lipases, their evolutionary origins appear to be different. Thus, the sequence motifs of the catalytic triads may represent the so-called "convergent evolution." However, the concept of "convergent evolution" is somewhat vague because it does not imply any clear picture of the ancestral form.

Taking into account the observation that a single sequence motif can represent the same enzymatic mechanisms that had undergone the different evolutionary process, I would propose the concept of an "evolutionary motif." The evolutionary motif can be defined as a sequence motif the amino acid sequences of which share a common ancestor in the evolution. Therefore, the sequence motifs of the catalytic triad for lipases and serine proteases constitute different evolutionary motifs, even though their sequence motifs are the same. Thus, in some case, a single sequence motif can be separated into more than one evolutionary motifs.

The kringle domain in serine proteases constitute an evolutionary motif. Of course, the minikringle is another

"evolutionary motif." It is now clear that the kringle-like structures of the defensin and the E protein in TBEV separately constitute their own evolutionary motif. The evolutionary motif may reflect a fundamental function that has been conserved in the evolutionary process. Thus, it is understandable that a functional domain can be characterized by its evolutionary motif.

In the primary world of life, exon shuffling may have been effective to create new functional genes. In fact, a fusion of different exons by exon shuffling must have been also effective to increase the variety of proteins. Under the situation that there existed a large variety of proteins, the world of life might have started establishing the complicated and elaborative system. In establishing the elaborative system, the primordial life form had to acquire the molecular mechanisms of regulatory and maintaining the system. For this purpose, domain shuffling may have been more advantageous than exon shuffling. Because a single domain is often encoded by more than one exon at the gene level, exon shuffling must have taken a risk at destroying an once well-made system. On the other hand, the evolutionary changes by domain shuffling may have caused additions or deletions of the already established function or exchanges between these different functions. These evolutionary changes are considered to be less deleterious than those by exon shuffling. Then, the amino acid substitutions may play a role as a fine tuner of the roughly established system.

References (Chapter 5)

Appella E, Robinson E A, Ullrich S J, Stoppelli M S, Corti A, Cassani G, and Blasi F (1987) The receptor-binding sequence of urokinase. *J. Biol. Chem.* 262, 4437-4440.

Blow D (1990) More of the catalytic triad. *Nature* 343, 694-695.

Brady L, Brzozowski A M, Derewenda Z S, Dodson E, Dodson G, Tolley S, Turkenburg J P, Christiansen L, Huge-Jenses B, Norskov L, Thim L, and Menge U (1990) A serine protease triad forms catalytic centre of a triacylglycerol lipase. *Nature* 343, 767-770.

Creighton T E (1984) Proteins; structure and molecular properties. (1984) W H Freeman and Company New York.

Grimaldi G, DiFiore P, Kajtaniak E, Falco J, and Blasi F (1986) Modulation of urokinase plasminogen activator gene expression during the transition from quiescent to proliferative state in normal cells *EMBO J.* 5, 855-861.

Heinz F X, Mandel M, Guirakhoo F, and Kunz C (1989) A structural and antigenic model of the tick-bone encephalitis virus envelope protein E. *Vaccines 89* 341-345 Cold Spring Harbor Laboratory.

Hill C P, Yee J, Selsted M E, and Eisenberg D (1991) Cristal structure of defensin HNP-3, an amphiphilic dimer: membrane permiabilization. *Science* 251, 1481-1485.

Kelly K, Cochran B H, Stiles C D, and Leder P (1983) Cell-specific regulation of the c-myc gene by lymphocyte mitogens and platelet-derived growth factor *Cell* 35, 603-610.

Klebanoff S J, (1988) in Inflammation: Basic Principles and Clinical Correlates. Gallin, J. I., Goldstein, I. M., and Snyderman, R. Eds. (Raven, New York) pp.391-444.

Kraut J, (1977) Serine proteases: Structure and mechanism of catalysis. *Ann. Rev. Biochem.* 46, 331-358.

Reich E (1978) In Molecular Basis of Biological Degradative Processes. Berlin R D, Herrmann H, Lepow I H, and Tanzer J M eds (Academic Press, New York) pp155-169.

Robbins K C, Summaria L, Hsieh B, and Shah R J (1967) The peptide chains of human plasminogen *J. Biol. Chem.* 242, 2333-2342.

Schrag J D, Li Y, Wu S, and Cygler M (1991) Ser-His-Glu triad forms the catalytic site of the lipase from *Geotrichum candidum Nature* 351, 761-764.

Scott J (1989) Thrombogenesis linked to atherogenesis at last. *Nature* 341, 22-23.

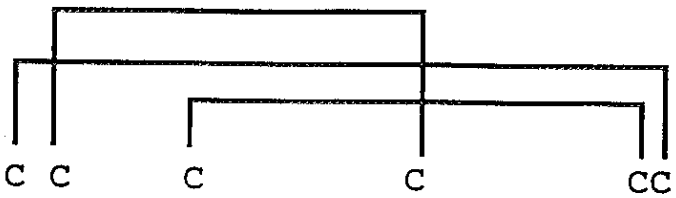
Stoppelli M P, Corti A, Soffientini A, Cassani G, Blasi F, and Assoian R K (1985) Differentiation-enhanced binding of the amino-terminal fragment of human urokinase plasminogen activator to a specific receptor on U937 monocytes. *Proc. Natl. Acad. Sci. USA* 82, 4939-4943.

Takahashi K, Gojobori T, and Tanifuji M (1991) Two-color cytofluorometry and cellular properties of the urokinase receptor associated with a human metastatic carcinomatous cell line. *Exp. Cell Res.* 192, 405-413.

Winkler F K, D'Arcy A, and Hunziker W (1990) Structure of human pancreatic lipase. *Nature* 343, 771-774.

Figure 5.1 Amino acid sequences of defensins.

The disulfide bridge is indicated by solid line. Abbreviations, HNP-1, 2, 3, and 4, defensins of human; GPNP, defensins of guinea pig; NP-1, 2, 3_a, 3_b, 4, and 5, defensins of rabbit; RNP-1, 2, 3, and 4, defensins of rat.



HNP-1	*ACYCRIP*ACIAGERRYGTCIYQGRLWAFCC***
HNP-2	**CYCRIP*ACIAGERRYGTCIYQGRLWAFCC***
HNP-3	*DCYCRI*PACIAGERRYGTCIYQGRLWAFCC***
HNP-4	*VCSCRL*VFCRRTELRVGNCLIGGVSFTYCCTRV
GPNP	RRCICTT*RTCRFPYRRLGTCIFQNRVYTFCC***
NP-1	VVCACRR*ALCLPRERRAGFCRIRGRIHPLCCRR*
NP-2	VVCACRR*ALCLPLERRAGFCRIRGRIHPLCCRR*
NP-3	GICACRR*RFCPNSERFSGYCRVNGARYVRCCSRR
NP-3b	GRCVCRKQLLCSYRERRIGDCKIRGVRFPFCCPR*
NP-4	VSCTCRGFLCGSGERASGSCTINGVRHTLCCRR**
NP-5	VFCTCRGFLCGSGERASGSCTINGVRHTLCCRR**
RNP-1	VTCYCRRTRCGFRERLSGACGYRGRIYRLCCR***
RNP-2	VTCYCRSTRCGFRERLSGACGYRGRIYRLCCR***
RNP-3	**CSCRTSSCRFGERLSGACRLNGRIYRLCC****
RNP-4	*ACYCRIGACVSGERLTGACGLNGRIYRLCCR***

Figure 5.2 Phylogenetic tree of defensins.

The tree was constructed through analysis of amino acid sequences. Two major groups are indicated by A, B.

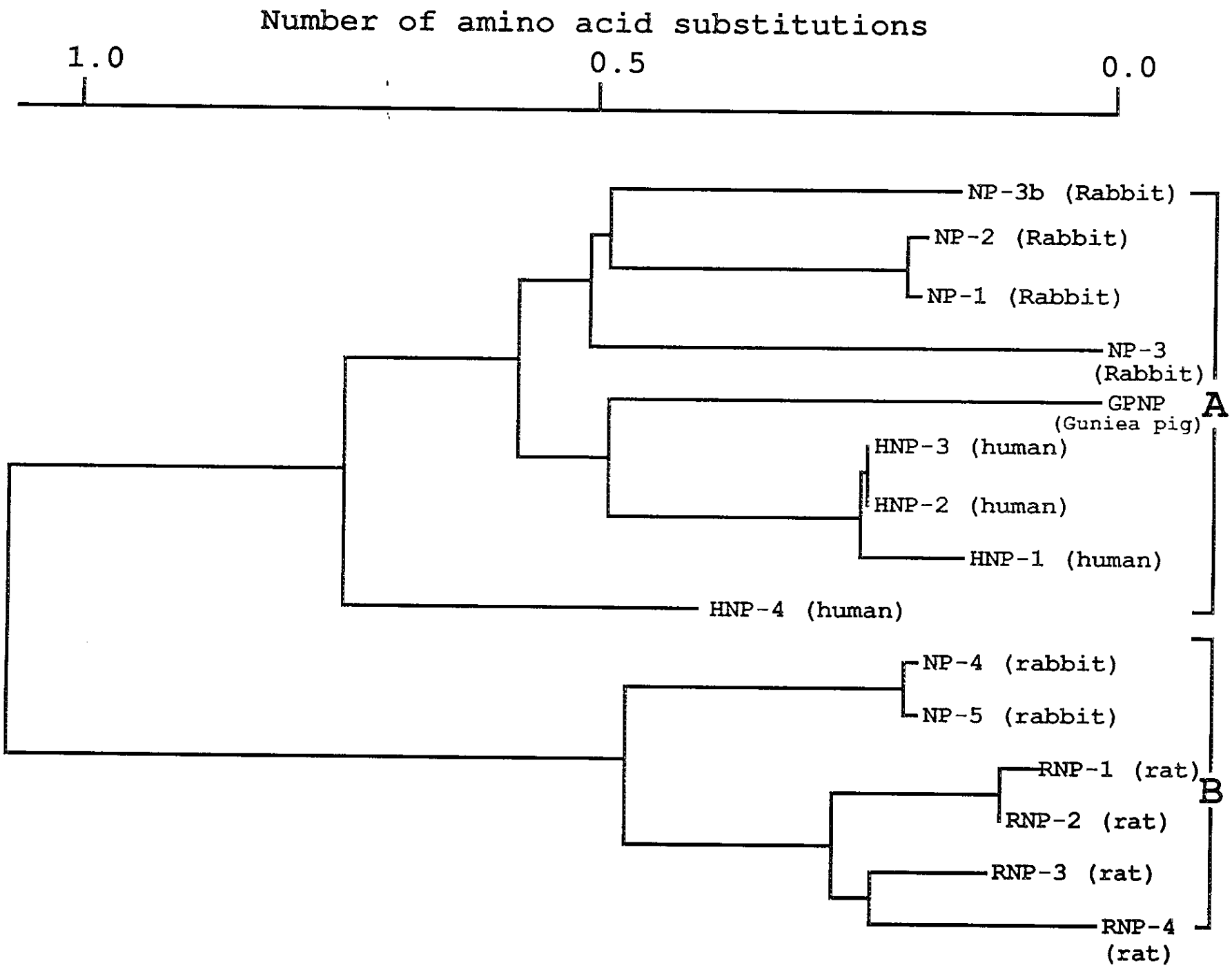

Number of amino acid substitutions
1.0
0.5
0.0
NP-3b (Rabbit)
NP-2 (Rabbit)
NP-1 (Rabbit)
NP-3 (Rabbit)
GPNP (Guniea pig)
A
HNP-3 (human)
HNP-2 (human)
HNP-1 (human)
HNP-4 (human)
NP-4 (rabbit)
NP-5 (rabbit)
RNP-1 (rat)
RNP-2 (rat)
RNP-3 (rat)
RNP-4 (rat)
B

Figure 5.3 The kringle structure of TBE virus.

Cisteine residues forming disulfide bridges are connected by solid lines. Two parallel solid lines indicate the lipid membrane that is spanned by two transmembrane regions of protein E (Heinz et al. 1989).

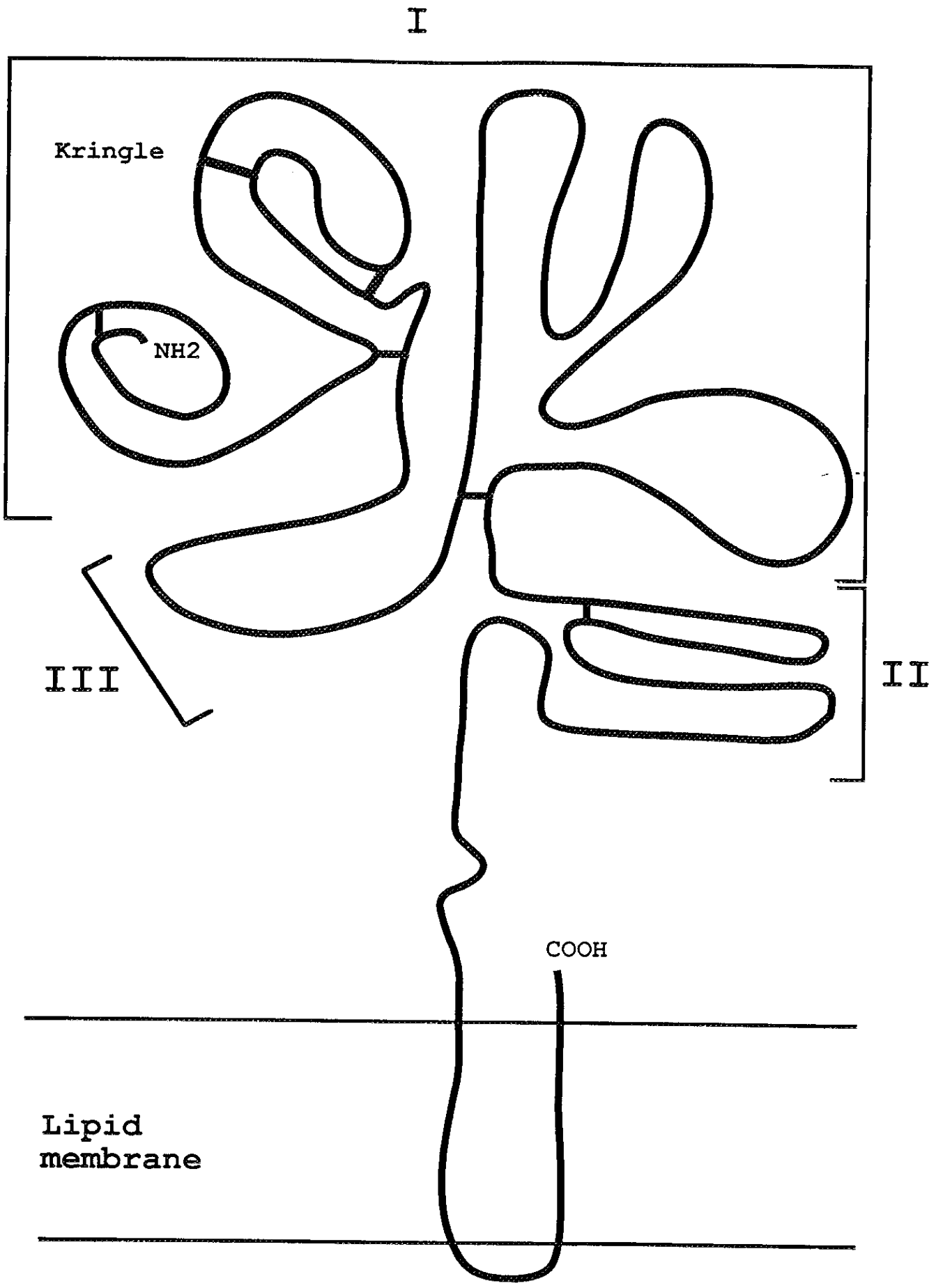

I
Kringle
NH2
III
II
COOH
Lipid
membrane

Figure 5.4 **Serine residue as active center.**

This figure shows the catalytic triad of a-chymotrypsin (Creighton, 1984). Dashed lines show the salt bridges.

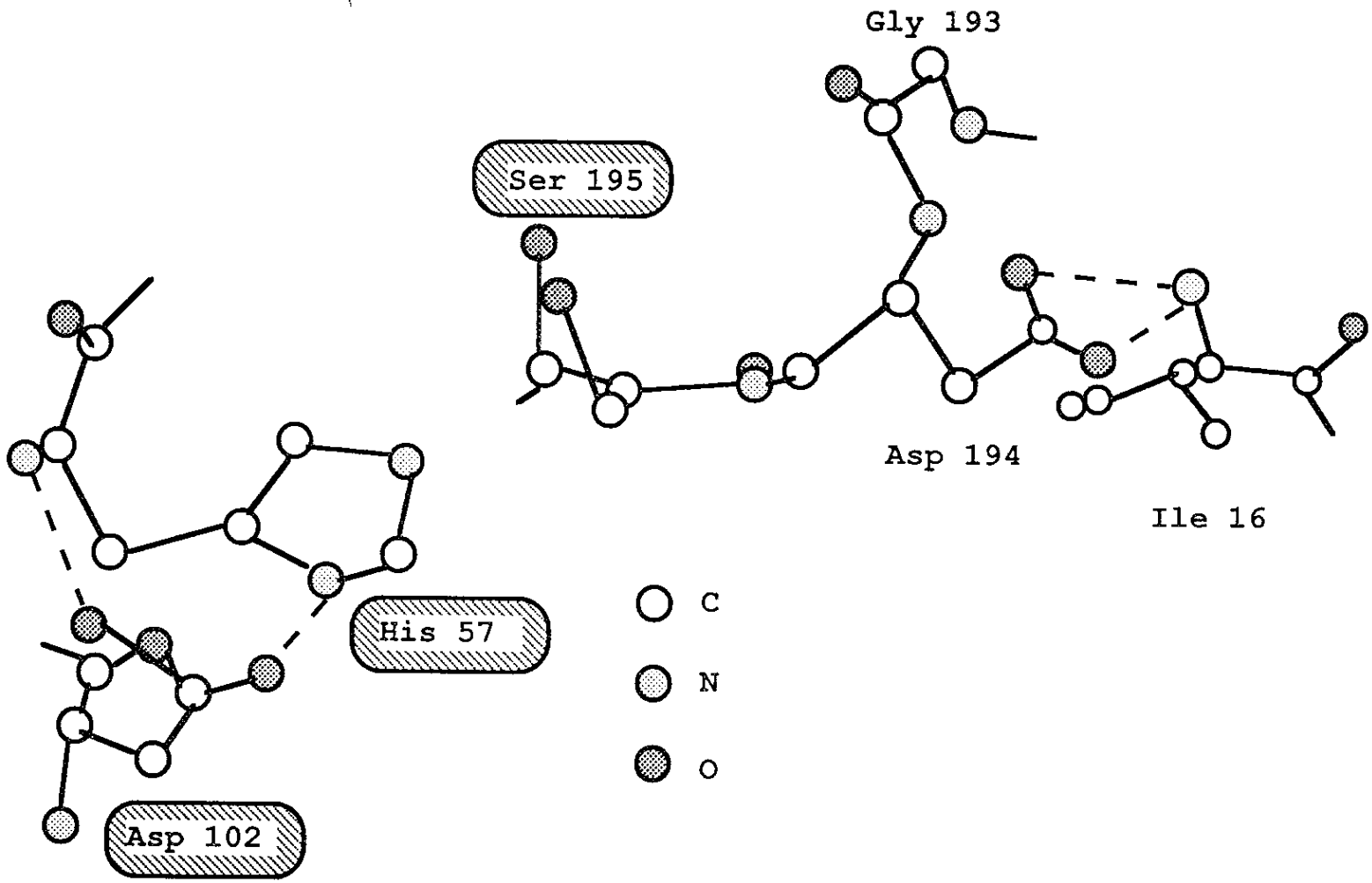

Gly 193
Ser 195
Asp 194
Ile 16
His 57
Asp 102
C
N
O

Figure 5.5 Charge-relay system of serine proteases.

The reaction proceeds through formation of a tetrahedral intermediate followed by loss of the right-hand half of the substrate to give an acyl enzyme intermediate. Breakdown of that intermediate occurs by enzyme-catalysed attack of water to generate the product acid.

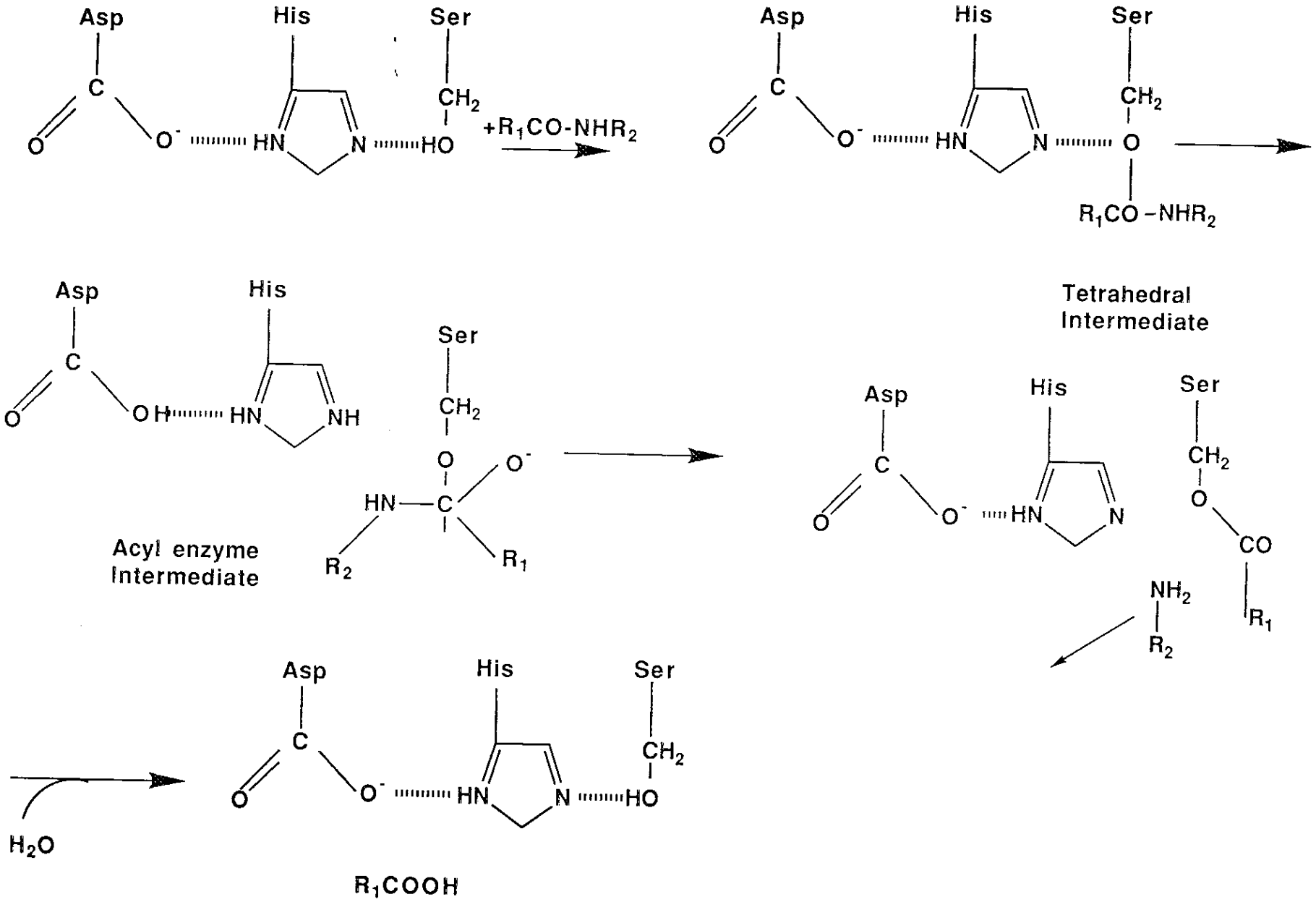

Asp
His
Ser
+R1CO-NHR2
R1CO-NHR2
Tetrahedral Intermediate
Acyl enzyme Intermediate
H2O
R1COOH

Figure 5.6 **The catalytic triad of lipase, and serine protease.**

The open bonds show the catalytic triad of trypsin and the full bonds show the catalytic triad of lipase.

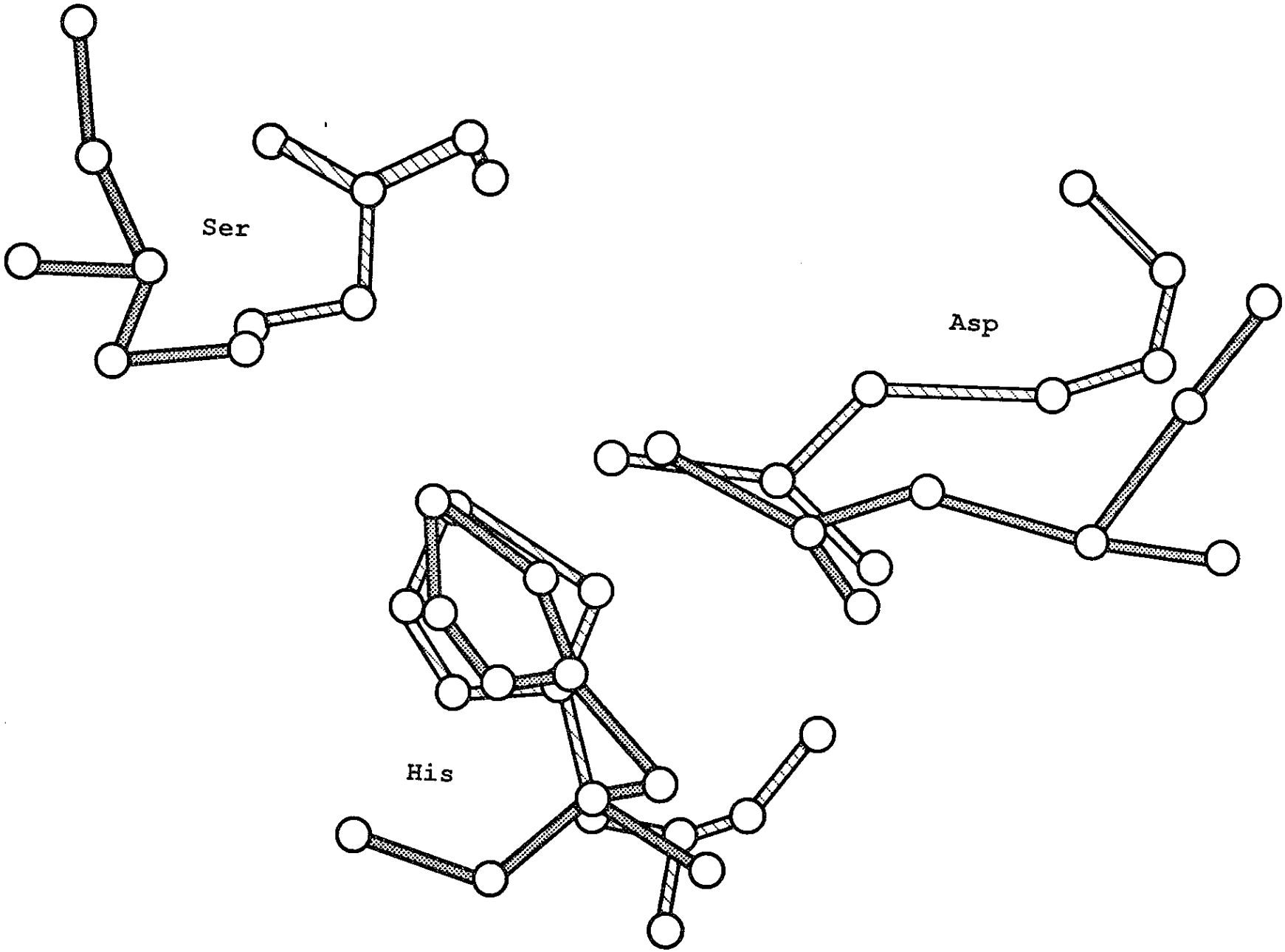
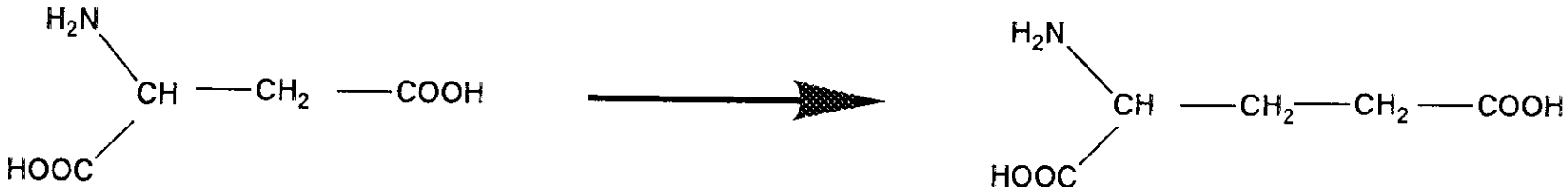

Ser
Asp
His

Figure 5.7 The structure of aspartic acid and glutamic acid.



Aspartic acid

Glutamic acid

CHAPTER SIX

CONCLUSIONS

In this study, I have clearly shown that the serine proteases are a good example of the mosaic protein and materials suitable for studying molecular evolution. The serine proteases have a protease domain and a domain called the kringle structure, which is a typical motif that has a pattern of disulfide bridges. Because this pattern is also found in other proteins which have no evolutionary relationships with each other, the kringle domains must have been evolutionarily shuffled and inserted into other proteins. I have indicated that the domain shuffling is attributable to genetic recombinations or gene conversion between gene segments at the DNA level. By studying the molecular evolution of the Kunitz-type protease inhibitors, I have shown that the minikringle structures of these inhibitors have evolutionary characteristics similar to the kringles in the serine proteases. I have also suggested that these structures are important as a regulatory domain of the enzymes involved in the biological cascades.

In conclusion, the origins and the evolutionary processes of serine proteases and its inhibitors as mosaic proteins were clarified, which revealed new evolutionary features, by molecular evolutionary studies of DNA and amino acid sequence data of the kringle and minikringle structures.

Publications by K. Ikeo
(As of January 10, 1992)

1. Ikeo, K., K. Takahashi, and T. Gojobori (1991) Molecular evolution of Kunitz-type protease inhibitor. J. Mol. Evol. (in press)

2. Takahashi, K., H. C. Kwaan, K. Ikeo, and E. Koh Phosphrylation of surface receptor bound urokinase-type plasminogen activator in a human metastatic carcinomatous cell line. Biochem. Biophys. Res. Commun. (in press).

3. Ikeo, K., K. Takahashi, and T. Gojobori (1991) Evolutionary origin of numerous kringles in human and simian apolipoprotein(a). FEBS Letters 287,146-148.

4. Moriyama, E. N., Y. Ina, K. Ikeo, N. Shimizu, and T. Gojobori (1991) Mutation pattern of human immunodeficiency virus genes. J. Mol. Evol., 32:360-363.

5. Gojobori, T., E. N. Moriyama, Y. Ina, K. Ikeo, T. Miura, H. Tsujimoto, M. Hayami, and S. Yokoyama (1990) Evolutionary origin of human and simian immunodeficiency viruses. Proc. Natl. Acad. Sci. USA, 87:4108-4111.

6. Ikeo, K., K. Takahashi, and T. Gojobori (1989) Molecular evolution of kringle domains in apolipoprotein(a) and other serine protease. Thrombosis and Haemostasis, 62:345.

7. Takahashi, K., K. Ikeo, T. Gojobori, and M. Tanifuji (1989) Local function of urokinase receptor at the adhesion contact sites of a metastatic tumor cell. Thromb. Res. Suppl., 10:55-61.

8. 五條堀　孝、池尾　一穂(1990)　「分子進化時計」　数理科学　No.324, pp67-71.

9. 内藤　公敏、河合　正人、岸野　敦子、森山　悦子、池尾　一穂、伊奈　康夫、池坂　守夫、佐藤　弘幸、五條堀　孝(1990)　「高並列計算機の適用による大量遺伝情報解析技術の実現」　並列処理シンポジウムJSPP'90 pp329-336.

10. 五條堀　孝、池尾　一穂　分子進化工学の将来　「分子生物学からバイオテクノロジーへ」　三浦　謹一郎編　共立出版　(印刷中)

発表講演

池尾　一穂(1991)　「セリン・プロテアーゼのモザイク蛋白としての進化」日本遺伝学会第６３会大会：ワークショップ「分子進化学の最前線ＩＩ」、九州大学、10月

池尾　一穂、高橋　敬、五條堀　孝(1991)　「巨大遺伝子に挿入されたプロテアーゼ・ドメインの進化的起源」日本遺伝学会第６３会大会、九州大学、10月

池尾　一穂、高橋　敬、五條堀　孝(1991)　「クリングル構造を持つ肝再生因子および肺繊維芽分裂促進因子にみられるモザイク構造の分子進化」第１４回日本分子生物学会年会、福岡国際センター、11月

Shiobara, T., T.Tokumitsu, E. N. Moriyama, Y. Ina, K. Ikeo, T. Imanishi, Y. Tateno, A. Kishino, M. Kawai, K. Naito, Y. Matsuura, and T. Gojobori(1991) Molecular evolutionary analysis system: SINCAIDEN. Human Genome III New York.

Gojobori, T., A. Kishino, T. Shiobara, T.Tokumitsu, K. Ikeo, Y. Ina, T. Imanishi, Y. Tateno, M. Kawai, K. Naito, Y. Matsuura, and E. N. Moriyama(1991) Classification of nucleotide sequence data in the DNA database by computers and its application to molecular evolutionary studies. Human Genome III New York

塩原　立也、徳光　孝之、五條堀　孝、森山　悦子、伊奈　康夫、池尾　一穂、今西　規、舘野　義男、内藤　公敏、河合　正人、　岸野　敦子(1991)　「スーパーコンピューターを用いた相同性分類アラインメントデータベースの作成」第１４回日本分子生物学会年会、福岡国際センター、11月

池尾　一穂、高橋　敬、五條堀　孝(1990)　「クリングル構造の分子進化学的解析」日本遺伝学会第６２会大会、お茶の水女子大学、10月

池尾　一穂、高橋　敬、五條堀　孝(1990)　「プロテアーゼ・インヒビターの分子進化学的解析」第１３回日本分子生物学会年会、国立京都国際会館、11月

Moriyama, E. N., Y. Ina, K. Ikeo, T. Miura, H. Tsujimoto, S. Yokoyama, M. Hayami, and T. Gojobori(1990) The origion of human and simian immunodeficy viruses. Sixth International Conference on AIDS. San Francisco. 6月

五條堀　孝、森山悦子、池尾　一穂、伊奈　康夫、今西　規、舘野　義男、岸野敦子、河合　正人、内藤　公敏(1990)　「大量ＤＮＡ配列データの分子進化学的解析によるモチーフの抽出」第１３回日本分子生物学会年会、国立京都国際会館、11月

池尾　一穂、高橋　敬、五條堀　孝(1989)「セリンプロテアーゼインヒビターの分子進化学的解析」日本遺伝学会、北海道大学学術交流会館、10月

池尾　一穂、五條堀　孝、高橋　敬(1989)「アポリポプロテイン(a)に見られるクリングル構造の分子進化学的解析」第12回日本分子生物学会年会、東北大学斉藤報恩記念会場、11月

高橋　敬、池尾　一穂、五條堀　孝(1989)「ウロキナーゼ受容体と癌細胞浸潤移動」第４１会日本細胞生物学会、名古屋

Takahashi, K., K. Ikeo, T. Gojobori, and M. Tanifuji (1989) Local function of urokinase receptor at the adhesion contact sites of a metastatic tumor cell. The 2nd Osaka Symposium on Fibrinolysis, Osaka.

Gojobori, T., E. N. Moriyama, Y. Ina, K. Ikeo, K. Oguchi, K. Hirai, A. Kishino, K. Naito, and T. Ikemura (1989) An integrated menu-driven computer system for molecular evolutionary analysis. International Symposium and Workshop "MacroMolecule, Genes and Computers: Chapter two," White Mountain Conference Center at Waterville Valley, New Hampshire. 8月

Gojobori, T., K. Ikeo, and K. Takahashi (1989) Molecular evolution of kringle domains in apolipoprotein (a) and other serine proteases. XIIth Congress of the International Society of Thrombasis and Haemostasis. Keio Plaza Hotel, Tokyo. 8月

五條堀　孝、森山　悦子、伊奈　康夫、池尾　一穂、大口　恵子、池村　淑道、内藤　公敏、平井　佳奈子、岸野　敦子(1989)「ヒトゲノムプロジェクトに呼応した大量ＤＮＡデータの総合的遺伝情報解析メニューシステムの開発研究」生物物理学会、東京大学教養学部、10月

森山　悦子、五條堀　孝、伊奈　康夫、池尾　一穂、大口　恵子、内藤　公敏、平井　佳奈子、岸野　敦子(1989)「ワークステーション用遺伝情報解析システムの開発」第12回日本分子生物学会年会、東北大学斉藤報恩記念会場、11月

森山　悦子、伊奈　康夫、池尾　一穂、大口　恵子、池村　淑道、五條堀　孝、内藤　公敏、河合　正人、平井　佳奈子、岸野　敦子(1989)「大量ＤＮＡデータを対象とする総合的遺伝情報解析システムの開発」公開ワークショップ：ヒト・ゲノム研究の現状と展望(第１回)、東京ガーデンパレス、12月

五條堀　孝、森山　悦子、池尾　一穂、伊奈　康夫、内藤　公敏、河合　正人、岸野　敦子(1989)「超高速並列計算機によるモチーフ探索と機能ドメインの分類－ヒト・ゲノムプロジェクトにおける分子進化学的方法論の確立－」公開ワークショップ：ヒト・ゲノム研究の現状と展望(第１回)、東京ガーデンパレス、12月

池尾　一穂、高橋　敬、五條堀　孝(1988)「アポリポプロテイン(ａ)に存在す38個のクリングル構造とその分子進化」日本遺伝学会第60回大会、京都大学農学部、10月

五條堀　孝、伊奈　康夫、池尾　一穂、森山　悦子(1988)「ヒトエイズウイルスとサルエイズウイルスの分子進化的関係」日本遺伝学会第60回大会、京都大学農学部、10月

高橋　敬、池尾　一穂、五條堀　孝(1988)「ウロキナーゼ受容体と癌細胞浸潤移動」第41回細胞生物学会大会、愛知県産業貿易館、11月

森山　悦子、伊奈　康夫、池尾　一穂、五條堀　孝(1988)「SIVとHIVの分子系統学的解析」エイズ研究会第2回学術集会、農協ビル(東京)、12月

五條堀　孝、伊奈　康夫、池尾　一穂、森山　悦子(1988)「AIDSウイルスの分子進化学(1)SIVとHIVの分子系統学的解析」第11回日本分子生物学会年会、日本大学農獣医学部、12月

森山　悦子、清水　宣明、伊奈　康夫、池尾　一穂、五條堀　孝(1988)「AIDSウイルスの分子進化学(2)SIVとHIVの塩基アミノ酸置換パターン」第11回日本分子生物学会年会、日本大学農獣医学部、12月

池尾　一穂、石川　勝利(1987)「ＨＣＧにより誘起されるアフリカツメガエル卵成熟のＰＭＳＦによる阻害」第５８会日本動物学会、富山大学理学部、10月