Prostasin: A Novel Human Serine Proteinase
Purification, Characterization and Cloning of Its cDNA and Gene

by

Xuezheng (Jack) Yu

A dissertation submitted to
the faculty of the Medical University of South Carolina
in partial fulfillment of the requirement for the degree of Doctor of Philosophy in the
College of Graduate Studies

Department of Biochemistry and Molecular Biology
1995

Approved by:

Julie Chao
Chairperson, Advisory Committee

[Signatures of other committee members]
ABSTRACT

XUEZHENG (JACK) YU. Prostasin: A Novel Human Serine Proteinase; Purification, Characterization and Cloning of Its cDNA and Gene. (Under the direction of JULIE CHAO and LEE CHAO).

A novel serine proteinase, designated as prostasin, has been purified from human seminal fluid to apparent homogeneity by DEAE-Sepharose CL-6B and aprotinin-affinity chromatography. The purified protein has an apparent molecular mass of 40 kDa on SDS-PAGE. It can be labeled with $[^{14}C]$diisopropylfluorophosphate and has pIs ranging from 4.5 to 4.8. Prostasin has trypsin-like activity with a pH optimum of 9.0. Its enzymatic activity on synthetic substrates is inhibited by trypsin inhibitors: aprotinin, antipain, leupeptin, and benzamidine. Prostasin is present at a high level in the human prostate gland, moderate levels in colon, lung, kidney, pancreas, salivary gland, liver and bronchi, but it is not detected in the brain, muscle, testis, ventricle, atrium or aorta.

A full-length cDNA sequence encoding prostasin was elucidated by PCR and amplification of the 5' and 3' ends of the cDNA. It contains a 1,032-base coding region, a 574-base 3' noncoding region and a 138-base 5' noncoding sequence. Prostasin mRNA encodes a protein of 343 amino acids, which consists of a 32-amino acid signal peptide and a 311-amino acid proprostasin. Proprostasin is converted into an active two-chain form by a cleavage between Arg$^{12}$ and Ile$^{13}$. The generated light chain has 12 amino acids and the heavy chain consists of 299 amino acids. The two chains are associated through a disulfide bond. The deduced amino acid sequence of the heavy chain has 34-42% identity to human acrosin, plasma kallikrein and hepsin. A potential N-glycosylation site of Asn$^{127}$ and the catalytic triad of His$^{53}$, Asp$^{102}$ and Ser$^{206}$ have been identified. The deduced prostasin has
a unique 19-amino acid hydrophobic portion at the COOH-terminus. Carboxyl-terminal sequencing of purified prostasin indicates that the hydrophobic portion is removed by a cleavage between Arg$^{290}$ and Pro$^{291}$ during secretion, indicating that the COOH-terminal hydrophobic portion is likely to be a transmembrane domain. Southern blot analysis, following a reverse transcription polymerase chain reaction, indicates that prostasin mRNA is expressed in prostate, liver, salivary gland, kidney, lung, pancreas, colon, bronchus, renal proximal tubular cells and prostate carcinoma LNCaP cells. Immunohistochemical localization reveals that prostasin is present in epithelial cells and ducts of the prostate gland. Cellular localization of prostasin mRNA was identified within epithelial cells of the human prostate gland by \textit{in situ} hybridization histochemistry. These results indicate that prostasin in the prostate gland is synthesized in epithelial cells and then secreted into the ducts.

A full-length prostasin gene has been isolated and sequenced. The gene consists of six exons and five introns. A number of potential regulatory elements have been revealed in the 5'-flanking region, including an AP2 site, two erythroid-specific promoter elements and a sterol regulatory element. In addition, there is a variant GC box and a variant AP1 site in the promoter region. The transcriptional initiation site of prostasin has been defined at the G residue and its adjacent A residue in a sequence CTCATGACT, which is similar to an initiator element CTCANTCT. These potential regulatory elements may be involved in differential expression of prostasin in tissues. Prostasin is a single-copy gene and has been localized on human chromosome 16p11.2 by \textit{in situ} hybridization.
ACKNOWLEDGMENT

I would like to express my deep appreciation to Drs. Julie Chao and Lee Chao for their strong support, encouragement and guidance during my Ph. D. study. I would like to thank the members of my advisory committee for their valuable suggestions and guidance. I would also like to thank all my colleagues for their support and assistance during my study.
# Table of Contents

I. General Introduction .................................................. 1

II. Prostasin: A Novel Human Serine Proteinase from Seminal Fluid; Purification, Tissue Distribution and Localization in Prostate Gland

   Summary ................................................................. 8
   Introduction ............................................................. 10
   Materials and Methods ............................................... 12
   Results ................................................................. 18
   Discussion .............................................................. 21

III. Molecular Cloning, Tissue-Specific Expression and Cellular Localization of Human Prostasin mRNA

   Summary ................................................................. 35
   Introduction ............................................................. 37
   Materials and Methods ............................................... 39
   Results ................................................................. 44
   Discussion .............................................................. 48

IV. Structure and Chromosomai Localization of Prostasin Gene

   Summary ................................................................. 60
   Introduction ............................................................. 61
   Materials and Methods ............................................... 63
   Results ................................................................. 66
   Discussion .............................................................. 69
V. General Discussion ................................................. 78

VI. List of References .................................................. 85

VII. Appendix: Experimental Procedures

DFP Labeling .............................................................. 99
Preparation of Protein for NH₂-terminal Sequencing .................. 100
ELISA for Prostasin ..................................................... 101
Immunohistochemistry .................................................. 103
*In Situ* Hybridization .................................................. 105
Protein Digestion by Trypsin and Separation by RP-HPLC .......... 107
Southern Blot ............................................................. 109
Primer Extension Assay ................................................. 112
Isolation of Lambda DNA ............................................... 114
Colony Hybridization ................................................... 116
Direct DNA Sequencing from Colonies ................................. 118
RT-PCR Amplification ................................................. 119
**List of Figures**

Fig. 1. SDS-PAGE of prostasin under reducing and non-reducing conditions

Fig. 2. Active site labeling of prostasin by DFP

Fig. 3. Isoelectric focusing of prostasin on a polyacrylamide slab gel

Fig. 4. NH₂-terminal amino acid sequence comparison of prostasin with human serine proteinases

Fig. 5. Optimal pH determination of prostasin enzyme activity

Fig. 6. Log-logit transformation of a specific radioimmunoassay

Fig. 7. Immunohistochemical analysis of prostasin in the prostate gland

Fig. 8. Nucleic acid sequence of prostasin cDNA and the deduced amino acid sequence

Fig. 9. Comparison of the deduced prostasin sequence with other serine proteinases

Fig. 10. Hydrophathy plot of the deduced prostasin

Fig. 11. Diagram of prostasin structure.

Fig. 12. Tissue-specific expression of prostasin mRNA
Fig. 13. Localization of prostasin mRNA by *in situ* hybridization in the human prostate gland

Fig. 14. Structure of human prostasin gene

Fig. 15. Nucleic acid sequence of human prostasin gene

Fig. 16. Determination of the transcription initiation site

Fig. 17. Genomic Southern blot analysis

Fig. 18. Chromosomal localization
I. GENERAL INTRODUCTION

Specific and limited proteolyses are involved in many biological processes, including removal of signal peptides from secretory protein, blood coagulation, fibrinolysis, complement activation, processing of precursors of protein hormones, growth factors and neuropeptides, processing of polyprotein in viruses, and activation of precursors of proteinases (Neurath, 1986; Hook et al., 1994). A great number of proteinases have been found to be involved in these processes. They vary in substrate specificity, mechanism of cleavage, three-dimensional structure, tissue distribution and cellular localization.

It was first observed by William Beaumont that substances other than hydrochloric acid were involved in gastric digestion (Beaumont, 1833). This finding led to the identification of digestive enzymes: pepsin and rennin (also called chymosin) and elucidation of their chemical structures and biological functions. The term "enzyme" was introduced by Willy Kühne to describe the substances he extracted from pancreatic tissues (Kühne, 1876). Before then, such substances had been called as ferment and diastase. It was first recognized by A. Schmidt in 1872 that trypsin and thrombin were secreted as inactive zymogens which required specific processing to become active (Magnusson et al., 1976). Zymogen activation is achieved by a specific cleavage of a peptide bond. This cleavage results in one of the following changes to convert the inactive precursor to the active enzyme: improving the substrate binding site, inducing the formation of an effective catalytic site, or exposing the active site to substrates (Neurath & Walsh, 1976). Proteolytic enzymes were among the first enzymes which were highly purified and crystallized. Crystallization of pepsin by Northrop (1930) provided direct evidence that pepsin which participates in gastric digestion is a protein. Proteolytic enzymes had long
been associated primarily with digestive processes and protein degradation until increasing evidence demonstrated that many proteinases displayed regulatory functions by selective and restrictive peptide bond cleavage. Linderstrom-Lang and Ottesen (1949) first introduced the term "limited proteolysis" to describe these regulatory processes, in which only one or a limited number of peptide bonds of a target protein are split. It is believed that limited proteolysis is the last step in the production of many biologically active proteins and probably the beginning of their degradation. Since then the functions of proteinases in biological and physiological regulation have been explored extensively and many of them have been elucidated (Neurath, 1989).

So far, more than 600 peptidases have been identified and sequenced either at the protein level or at the DNA level (Rawlings & Barrett, 1993). These peptidases are currently classified in three different ways. (1) Based on the reactions they catalyze, proteolytic enzymes are classified into two groups: exopeptidases and endopeptidases. (2) According to the active sites, the mechanism of action and the three dimensional structure, proteolytic enzymes are grouped into four families, namely, the serine proteinases, cysteine proteinases, aspartic proteinases and metalloproteinases (McDonald, 1985). (3) Recently, proteolytic enzymes were classified into different families based on their similarities in primary structures (Rawlings & Barrett, 1993), such as the chymotrypsin family. Among these three classifications, the second one is most often used because it is based on the nature of the catalytic groups and proteinases from different families can easily be distinguished from each other by simple inhibition assays. Serine proteinases, cysteine proteinases, aspartic proteinases and metalloproteinases are sensitive to diisopropyl fluorophosphate (DFP), peptide epoxide E-64, acetylated pentapeptide pepstatin and metal chelator ethylenediaminetetraacetic acid (EDTA), respectively. Proteinases in each of the four families are thought to have originated from a common ancestor by duplication and mutation during divergent evolution. In addition to the regulation at the transcriptional level by cis-acting elements and trans-acting factors, expression of the activities of proteinases is
regulated by posttranslational proteolytic processes and the interaction with specific protein inhibitors, oligopeptides, or chemicals. Most proteinases are secreted as inactive zymogens, which require a specific process called activation to convert the zymogens into active proteinases. This process is accomplished by a specific cleavage of a proteinase which has unique substrate specificity. An activation of one proteolytic enzyme may be capable of activating the precursor of another proteinase, leading to a cascade of proteolytic reactions. These proteolytic cascades are observed in blood coagulation, fibrinolysis and complement activation. Expression of proteinase activity can also be regulated by inhibition of specific protein inhibitors. Serine proteinases are inhibited by a family of proteins called serpins, cysteine proteinases by cystatins and metalloproteinases by a group of tissue inhibitors of metalloproteinases (TIMP). Most of the proteolytic enzymes in these four families can be inhibited by a protein called α2-macroglobulin. Alpha2-macroglobulin inhibits the activities of proteinases by irreversibly entrapping them regardless of the specific active sites for each proteinase family. It is the only protein inhibitor which does not show class-specific inhibition. Proteolytic processing is an effective way to regulate physiological processes at the posttranslational level. Activities of proteinases are strictly regulated inside the body and any abnormality may cause severe problems.

Proteinases are rich in bodily fluids and are present in most cellular organelles, such as the lysosome, endoplasmic reticulum, Golgi apparatus, secretory vesicle, mitochondrion and plasma membrane (Barrett, 1992). Proteolytic enzymes in plasma have important physiological functions in blood coagulation, fibrinolysis and complement activation. Point mutations in factor VII or factor X may cause severe clinical phenotypes (Chaing et al., 1994; Watzke et al., 1991). Proteinases in seminal fluid have crucial roles in semen coagulation, liquefaction and the sperm-egg interaction. The endoplasmic reticula, Golgi apparatus and secretory vesicles are where proteolytic enzymes are synthesized, posttranslationally processed and/or secreted out of cells. Proteinases in mitochondria are involved in posttranslational processing to remove NH2-terminal targeting peptides of
proteins imported from the cytosol. In the lysosome, proteolytic enzymes are in charge of the degradation of fast turnover proteins as well as damaged and abnormal protein molecules. Some proteinases are associated with the plasma membrane, either directly or indirectly. The proteinases which are directly associated with the membrane, either have a transmembrane domain, like serine proteinase hepsin (Leytus et al., 1988) or interact with the membrane lipid like acrosin (Baba et al., 1989). For the proteinases which are indirectly associated with the membrane, they are linked through specific receptors, such as urokinase-type plasminogen activator (Pyke et al., 1991). These proteinases anchored in the membrane have the ability to focus their activities on the surface of cells and therefore have important roles in the processing of protein prohormones, growth factors and receptors and have functions in tissue remodeling during development.

Proteolytic enzymes have been found to be involved in many important physiological processes. Kallikreins are a subfamily of serine proteinases that take part in the release of kinins from kininogens. Kinins are a group of biological active peptides, which interact with specific receptors to mediate some important biological processes, such as the regulation of blood pressure and reproduction (Burch & Kyle, 1992; Regoli & Barabe, 1980). Recent studies on transgenic mice of the human tissue kallikrein gene provide direct evidence that tissue kallikrein is a strong modulator of blood pressure (Wang et al., 1994) and delivery of the human kallikrein gene has gene therapy potential for treatment of human hypertensive diseases (Wang et al., 1995; Xiong et al., 1995). Furthermore, delivery of the rat kallikrein inhibitor gene can reverse hypertension in transgenic mice expressing human tissue kallikrein (Ma et al., 1995). Alzheimer’s disease is characteristic of the aggregation and deposition of β-protein in the brain. This β-protein is generated from its precursor, the β-amyloid precursor protein. So far, no alternative splicing events that directly produce β-protein can be identified. Recently, a chymotrypsin-like mast cell proteinase was found to be capable of generating the NH2-terminus of β-protein (Nelson et al., 1993), indicating that proteolytic enzymes may play a critical role in
the development of Alzheimer's disease. Angiotensin converting enzyme is a membrane-anchored metalloproteinase on vascular endothelial cells. It converts angiotensin I into vasopressor and aldosterone-stimulating peptide angiotensin II by removal of a single carboxyterminal dipeptide. Furthermore, evidence indicates that angiotensin converting enzyme is a kininase that inactivates bradykinin (Wiemer et al., 1991). Recent studies (Yuan et al., 1993; Miura et al., 1993; Tewari et al., 1995) revealed that IL-1β-converting enzyme and Yama/CPP32β, mammalian homologs of the C. elegans cell death gene ced-3, may play a critical role in programmed cell death. The second member of human kallikrein family, prostate-specific antigen (PSA), has become the most important and sensitive marker in serum for the detection of prostate cancer. Recent studies revealed that some viruses, such as RSV and HIV-1, also encode proteinases. These proteinases have important roles in the viral replication cycle or in virus maturation (Miller et al., 1989; Navia et al., 1989). A point mutation at the active site Asp25 of HIV-1 proteinase results in the production of non-infectious virions which contain unprocessed polyprotein (Kohl et al., 1988). Thus, HIV-1 proteinase has become the target for developing specific inhibitors to block viral replication (Humber et al., 1993; Caflisch et al., 1993; Wlodawer, 1994).

The roles of proteinases in tumor invasion, metastasis and angiogenesis are receiving more and more attention. Some of them are membrane-anchored proteinases; some of them are associated with the membrane through specific receptors; some of them are secreted into the extracellular matrix. The activities of proteinases are regulated by proteolytic processes such as activation of precursors and by specific inhibitors. In many tumors, proteolytic activities are increased or the levels of specific inhibitors are decreased (Mignatti & Rifkin, 1993). Elastases, a group in the serine proteinase family, can degrade elastin, a main component of elastin fibers, which are resistant to most other proteinases. Collagenases can degrade collagen, gelatin and fibronectin, which are main components of the extracellular matrix. In addition, stromelysin and some cysteine proteinases have also
been found to be involved in the degradation of the extracellular matrix in the progression of tumor development. Plasmin/plasminogen activator plays important roles in triggering the proteolytic cascade in the extracellular matrix during tumor cell invasion, migration and metastasis. A recent study revealed a new group of metalloproteinases, the membrane-type matrix metalloproteinases (MT-MMP; Sato et al., 1994). The MT-MMPs have a transmembrane domain through which they are anchored in the membrane and can activate pro-gelatinases and degrade extracellular matrix macroproteins, such as type IV collagen, a main component of the basement membrane. This study further strengthens the importance of proteinases in the development of tumor.

Serine proteinases are the well-known and best studied family of proteolytic enzymes. Based on the linear order of the components of the catalytic triad in the primary structure, serine proteinases are further divided into two subclasses: (1) the mammalian serine proteinases, which have chymotrypsin-like active sites and three-dimensional structure and (2) the bacterial serine proteinases, which have subtilisin-like active sites and three-dimensional structure. Serine proteinases have an active serine residue involved in catalytic process and it is this active serine residue that can specifically react with is inhibited by disisopropyl fluorophosphate (DFP), which distinguishes serine proteinases from other kinds of proteinases. The reaction of serine proteinase cleavage of peptide bonds consists of two steps. The first step is the production of an acyl-enzyme intermediate through a tetrahedral transition state. The second step is the hydrolysis of this acyl-enzyme intermediate by a water molecule to release the second product peptide. Serine proteinases include exopeptidases and endopeptidases and are present in viruses, bacteria and eukaryotes. It is believed there are still many serine proteinases remaining to be identified and their biological and physiological functions have yet to be established.

Human seminal plasma is an important bodily fluid and a rich source of proteinases. These proteinases are the products of the sex glands and the accessory sex glands. Semen coagulation occurs immediately after ejaculation and is subsequently
liquefied within 30 min (Tauber et al., 1980). These processes are thought to be mediated by various proteinases, including the aminopeptidases, dipeptidyl-peptidases, dipeptidyl carboxypeptidases, peptidyl dipeptidases, plasminogen activators, prostate-specific antigen, elastase, and kallikrein (Shivaji et al., 1989). However, the detailed processes are far from clear. Studies showed that many distinct proteolytic activities present in human seminal plasma remain to be identified and characterized (Yin et al., 1990; Kobayashi et al., 1992). They differed from known proteinases in specific enzymatic activities or affinity and sensitivity to trypsin inhibitors. In this dissertation, I describe methods for the purification and characterization of a new serine proteinase, designated as prostatin, from human seminal fluid, molecular cloning of its cDNA and gene. In addition, I established its posttranslational proteolytic processes and chromosomal localization, and determined its tissue-specific expression and cellular localization.
II. Prostasin: A Novel Human Serine Proteinase from Seminal Fluid
Purification, Tissue Distribution and Localization in Prostate Gland

SUMMARY

A novel serine proteinase, designated as prostasin, has been purified from human seminal fluid to apparent homogeneity by DEAE-Sepharose CL-6B and aprotinin-affinity chromatography. The purified protein migrates as two close bands with an apparent molecular mass of 40 kDa on SDS-PAGE under reducing conditions. It can be labeled with \[^{14}\text{C}]\text{diisopropylfluorophosphate}\) and has a pI ranging from 4.5 to 4.8. Sequence analysis reveals that the two protein bands have an identical NH2-terminal amino acid sequence which is different from any known protein sequence in SwissProt or GenBank database. The NH2-terminal 20-amino acid sequence shares 50-55\% identity with human \(\alpha\)-tryptase, elastase 2A and 2B, chymotrypsin, acrosin and the catalytic chains of hepsin, plasma kallikrein and coagulation factor XI. Prostasin has trypsin-like activity with a pH optimum of 9.0, hydrolyzing peptidyl fluorogenic substrates: D-Pro-Phe-Arg-MCA, D-Phe-Phe-Arg-MCA, D-Val-Leu-Arg-MCA and Z-Gly-Pro-Arg-AFC. It is inhibited by aprotinin, antipain, leupeptin, and benzamidine. Tissue distribution of prostasin was determined by a newly developed radioimmunoassay. Linear displacement curves for immunoreactive prostasin in body fluids and tissues were parallel with the standard curve of purified prostasin, indicating their immunological identity. Immunoreactive prostasin levels were 8.61 ± 0.42 \(\mu\text{g/ml}\) in the seminal fluid and 0.201 ± 0.029 \(\mu\text{g/ml}\) in urine. Prostasin is present at high levels in the prostate gland (143.7 ± 15.9 ng/mg protein), moderate levels (2-6 ng/mg protein) in colon, lung, kidney, pancreas, salivary gland, liver
and bronchi, but is not detected in the brain, muscle, testis, ventricle, atrium and aorta. Immunohistochemical localization reveals that prostatic is present in epithelial cells and ducts of the prostate gland. These studies indicate that prostatic purified from seminal fluid is a novel serine proteinase and originates from the prostate gland.
INTRODUCTION

Many biological processes require specific and limited proteolysis. These include: removal of signal peptides from secretory proteins; processing precursors of protein hormones, growth factors and proteinases; blood coagulation; fibrinolysis; and complement activation (Neurath, 1986). Human seminal fluid is a rich source of various proteolytic activities. Many of the proteinases in seminal fluid are involved in postejaculatory hydrolysis of proteins and semen coagulation and liquefaction. These activities can influence sperm mobility and male fertility. In addition, acrosin which was purified from human sperm extract (Siegel et al., 1986) is a membrane-bound serine proteinase with trypsin-like substrate specificity. Acrosin may be responsible for the limited proteolysis of the oocyte zona pellucida glycoprotein matrix (Dunbar et al., 1985) and it may be involved in sperm-egg interaction (Topfer-Petersen et al., 1987; Jones et al., 1988).

A number of proteinases have been purified and characterized from human seminal fluid, such as prostate-specific antigen (PSA) (Wang et al., 1981), kininase II (also called angiotensin converting enzyme; Miska et al., 1989), and a basic carboxypeptidase (Skidgel et al., 1988). PSA is produced in epithelial cells of the prostate gland and is secreted into the seminal fluid (Wang et al., 1981). The physiological substrate of PSA is semenogelin, a major seminal vesicle protein (Lilja, 1985), indicating that PSA may play an important role in semen liquefaction (Lilja et al., 1987). Recently, serum PSA has been established as the most important tumor marker for patients with prostatic carcinoma (Guinan et al., 1987; Stamey et al., 1989). Kininase II is a zinc-metalloproteinase which has been found in several mammalian tissues in addition to human seminal fluid. This particular proteinase releases a dipeptide His-Leu from the carboxyl terminus of angiotensin I to produce potent vasoactive angiotensin II. The basic carboxypeptidase, that has been purified from human seminal fluid, cleaves the COOH-terminal arginine or lysine from peptides. It is thought to be implicated in the control of fertility by activating or inactivating peptide hormones in
human seminal fluid and in the degradation of basic proteins during semen liquefaction (Skidgel et al., 1988).

Over the last several years, a gene and a cDNA for human glandular kallikrein-1 (hGK-1), highly similar to prostate specific antigen, have been cloned and characterized (Schedlich et al., 1987; Young et al., 1992). Its mRNA was found to be exclusively expressed in the prostate gland (Chapdelaine et al., 1988; Morris, 1989), however, the protein has not been isolated. A number of basic and acidic arginine amidase activities were also detected in human seminal fluid (Kobayashi et al., 1992) and showed different affinities to aprotinin, lima bean trypsin inhibitor and lysine. However, none of them have been purified and characterized. These studies demonstrate that there are many proteolytic enzymes yet to be identified in human seminal fluid and this prompted us to search for them. In this study, we purified and characterized a novel serine proteinase from human seminal fluid. We identified its tissue distribution by a newly developed radioimmunoassay and its cellular localization in the prostate gland by immunohistochemical techniques.
MATERIALS AND METHODS

Materials: DEAE-Sepharose CL-6B was purchased from Pharmacia Fine Chemicals (Piscataway, NJ); low molecular weight protein standards from Bio-Rad Laboratories (Richmond, CA); [14C]diisopropylfluorophosphate (DFP) and EN3HANCE from New England Nuclear (Boston, MA); D-Pro-Phe-Arg-MCA, D-Val-Leu-Arg-MCA, D-Phe-Phe-Arg-MCA, Z-Gly-Pro-Arg-AFC from Enzyme System Products (Livermore, CA); 3,3'-diaminobenzidine tetrahydrochloride, aprotinin-agarose, Carrageenan Lambda, Coomassie Brilliant Blue R, soybean trypsin inhibitor (SBTI), complete Freund's adjuvant, polyethylene glycol (PEG) compound with molecular mass of 15-20 kDa and PEG with molecular mass of 8 kDa from Sigma Chemical Co. (St. Louis, MO); Immobilon-P transfer membranes from Millipore Corp. (Bedford, MA); Excellulose GF-5 column and Iodogen from Pierce Chemical Co. (Rockford, IL); goat IgG fraction against rabbit IgG F(ab')2 and peroxidase-antiperoxidase complex (rabbit) from Organon Teknika Co. (Durham, NC and West Chester, PA).

Purification of Prostasin: Human seminal fluid (kindly provided by Dr. Subbi Mathur, Department of Obstetrics and Gynecology, Medical University of South Carolina and Dr. Fletcher C. Derrick, Jr., Urology Clinic, Charleston, SC) was dialyzed against 4 liters of 25 mM Tris-HCl, pH 7.6 (buffer A) overnight at 4 °C, transferred to the same buffer and dialyzed for an additional 24 hr. The sample was centrifuged at 12,000 x g for 30 min at 4 °C and the supernatant was applied to a DEAE-Sepharose CL-6B column (2.2 x 20 cm) which had been preequilibrated with buffer A at room temperature. After being washed, the column was eluted with a linear gradient of 0-0.15 M NaCl in buffer A, followed by a gradient of 0.15-0.40 M NaCl. The eluted fractions from 0.15-0.25 M NaCl, which bound covalently with [14C]-DFP and exhibited trypsin-like activity toward D-Pro-Phe-Arg-MCA, were pooled and concentrated by a PEG compound with a
molecular mass of 15-20 kDa. The sample was then loaded on an aprotinin-agarose column (1.5 x 20 cm) that had been preequilibrated with buffer A. The column was washed with buffer A until the absorbance at 280 nm dropped to baseline and was then eluted with 0.1 N glycine-HCl, pH 3.0. The eluate was neutralized immediately with 1 M Tris base, concentrated and dialyzed with 20 mM sodium phosphate buffer, pH 7.0, through a Centricon-10 concentrator (Amicon Inc., Beverly, MA).

**Determination of Protein Concentration:** Protein concentration was measured by the method of Lowry et al. (1951) using bovine serum albumin (BSA) as a standard. The concentration of purified prostasin was determined by measuring its extinction coefficient (E\(^{0.1\%280}\)) spectrophotometrically at 205 nm and 280 nm (Van Lersel et al., 1985).

**SDS-polyacrylamide Gel Electrophoresis:** SDS-PAGE was performed as previously described by Xiong et al. (1990). Under reducing conditions, dithiothreitol was added to the extraction buffer to a final concentration of 20 mM.

**Isoelectric Focusing:** Isoelectric focusing was performed on a 0.5 mm thin layer polyacrylamide slab gel with a pH gradient of 3.5-10.0 in an LKB-2117 multiphor electrophoresis system. The gel was run for 15 min at 15 °C, and for 2.5 hr at 4 °C. After completion of focusing, the edge portion of the gel along the pH gradient was cut into a 1 cm wide strip, which was then cut transversely at 5 μm intervals. Each section was put into a tube containing 1 ml of distilled water and allowed to stand overnight at room temperature. Tubes were vortexed and the pH value of each gel section was measured. A pH gradient was obtained by plotting the pH value of each section versus its original position on the gel. The remaining gel slab was fixed in 25% trichloroacetic acid overnight, stained in 0.2% Coomassie Brilliant Blue solution for 15 min, destained in methanol/acetic acid/water (30/12/110) for several days and air dried.

**Analysis of NH\(_2\)-Terminal Amino Acid Sequence:** NH\(_2\)-terminal amino acid sequence was analyzed in two ways. First, after electrophoresis on a SDS-PAGE
under reducing conditions, the protein was transferred from the gel to an Immobilon-P membrane, and stained with 0.2% (w/v) Ponceau S in 3% trichloroacetic acid for 1 min, then destained with 1 M acetic acid for 2 min. The two protein bands around 40 kDa were cut off separately and subjected to NH2-terminal sequence analysis using a gas phase protein sequencer equipped with an on-line narrow-bore PTH-amino acid analyzer (ABI model 470 A, Applied Biosystems Inc.). Secondly, purified prostasin was directly applied to the sequencer for sequence analysis.

**Determination of Optimal pH:** Hydrolytic reactions were carried out in 20 mM Tris-HCl buffer (pH 7.0 to 9.0) and 20 mM glycine-NaOH buffer (pH 9.0 to 10.5) at 0.5 pH intervals. The initial velocities of the reactions were measured at λex 380 nm and λem 460 nm at 25 °C toward substrate D-Pro-Phe-Arg-MCA (a final concentration of 30 μM) using a Perkin Elmer LS-5 spectrofluorometer. The prostasin concentration in each reaction mixture was 5 nM.

**Kinetic studies:** To determine the kinetic constants, an appropriate amount of prostasin was mixed with various concentrations of a peptidyl fluorogenic substrate in 20 mM Tris-HCl, pH 9.0, at 25 °C. The enzyme (1-4 nM) and four substrates: D-Phe-Phe-Arg-MCA, D-Pro-Phe-Arg-MCA, D-Val-Leu-Arg-MCA, and Z-Gly-Pro-Arg-AFC, were used for kinetic analyses. The initial hydrolytic velocities were determined by measuring the change of the concentration of the released products, 7-amino-4-methyl-coumarin (AMC) at λex 380 nm and λem 460 nm or 7-amino-4-trifluoromethyl coumarin (AFC) at λex 400 nm and λem 505 nm on a Perkin Elmer LS-5 spectrofluorometer.

**Active Site Labeling:** Aliquots (40 μl) of fractions from a DEAE-Sepharose CL-6B column were incubated with 0.4 μCi of [14C]-DFP in 10 mM sodium phosphate, pH 7.0, in a total volume of 60 μl, and purified prostasin (5 μg) was incubated with 0.2 μCi of [14C]-DFP in 10 mM sodium phosphate, pH 7.0 in a total volume of 30 μl with gentle shaking for 2 hr at 37 °C, then placed at 4 °C overnight. The samples were loaded on SDS-PAGE under reducing conditions. After staining and destaining, the gel was
treated with EN3HANCE for 1 hr and was washed several times with water for 30 min. The gel was dried under vacuum at 80 °C and exposed to an X-ray film at -70 °C for several days.

**Inhibition Studies:** The effects of five serine proteinase inhibitors on prostasin activity were measured using D-Pro-Phe-Arg-MCA as a substrate. The enzyme (12 pmole in 20 μl of 0.2 M Tris-HCl buffer, pH 9.0) was mixed with an equal volume of various concentrations of each inhibitor and incubated at 37 °C with gentle shaking for 30 min. Each mixture was then added to 1.96 ml of 20 mM of Tris-HCl buffer, pH 9.0, containing 20 μM of substrate D-Pro-Phe-Arg-MCA. Initial velocities of hydrolytic reactions were measured on a Perkin Elmer LS-5 spectrofluorometer at λex 380 nm and λem 460 nm.

**Preparation of Antiserum:** Purified prostasin (100 μg) was mixed and emulsified with an equal volume of complete Freund's adjuvant. Subcutaneous and intramuscular injections of purified prostasin were administered to a New Zealand female rabbit (~2 kg body weight). Booster injections of 50 μg of prostasin were performed every three weeks until a high titer of antiserum was detected.

**Development of a Prostasin-Specific Radioimmunoassay (RIA):** Prostasin (5 μg) was labeled with 125I by using Iodogen according to the manufacturer's instructions. One mg of Iodogen was dissolved in 25 ml of methylene chloride and a 50-μl aliquot was dried in an Eppendorf tube. To the tube containing Iodogen, 50 μl of 250 mM sodium phosphate, pH 7.0, 5 μg of prostasin and 0.5-0.9 mCi of Na125I were added, mixed well and incubated at room temperature for 10 min. The reaction mixture was applied to an Excellulose GF-S column preequilibrated with 50 mM sodium phosphate (pH 7.4) containing 0.1% BSA. The column was then eluted with the same buffer and 0.5 ml fractions were collected. The radioactivity of each fraction was measured with a gamma counter. A specific radioimmunoassay for prostasin was developed as described by Chao et al. (1989). In the antibody titration, the antiserum was diluted from 1:100 to 1:1,000,000 in phosphate-buffered saline (PBS, 0.14 M sodium chloride in 0.01 M
sodium phosphate, pH 7.0) containing 1% BSA. A mixture of 100 μl of PBS containing 1% BSA, 100 μl of diluted antibody solution and 200 μl of the 125I-labeled prostasin in PBS containing 1% BSA was incubated overnight at room temperature, and the reaction was stopped with 400 μl of 1% bovine γ-globulin and 800 μl of 25% PEG-8,000. The antibody titration curve of prostasin showed the titer of the antiserum to be 1 : 80,000 indicating that 35% of 125I-labeled prostasin bound to antibody in the absence of competitive antigen.

**Tissue Distribution and Quantitation:** Human tissues from autopsy (kindly provided by Dr. Sandra Conradi, Department of Pathology, Medical University of South Carolina) were minced and homogenized as previously described by Chao et al. (1990). Protein concentrations were determined by the method of Lowry et al. (1951) with BSA as a standard. The distribution and quantitation of prostasin was measured by the prostasin-specific RIA as described above. Levels of prostasin in human seminal fluid and urine were also measured.

**Immunohistochemical Localization in Prostate Gland:** Sections (5 μm) of a formalin-fixed, paraffin-embedded human prostate gland were deparaffinized and rehydrated by standard procedures. Sections were then incubated in 1% hydrogen peroxide in distilled water for 20 min to remove endogenous peroxidase activity. After washing three times with 0.05 M Tris-HCl, pH 7.6, sections were incubated in the following solutions: 1) rabbit antiserum against prostasin (1:400) overnight; 2) goat IgG fraction against rabbit IgG F(ab')2 (1:50, Organon Teknika Co., Durham, NC) for 1 hr; 3) peroxidase-antiperoxidase complex (1:100) for 1 hr; 4) 0.1% 3,3'-diaminobenzidine tetrahydrochloride plus 0.01% hydrogen peroxide. The antiserum, antibody and peroxidase-antiperoxidase complex was diluted in 0.05 M Tris-HCl (pH 7.6) containing 0.5% Triton X-100 and 0.7% Carrageenan Lambda. Between incubations, sections were washed three times with 0.05 M Tris-HCl (pH 7.6) for 10 min each. After staining, which was monitored under a light microscope, sections were washed with distilled water,
dehydrated and mounted with Permount. A negative control consisted of a human prostate
gland section incubated with a pre-immune serum (1:200 dilution) instead of rabbit
antiserum against prostasin.
RESULTS

Purification and Characterization: The purification of prostasin from human seminal fluid is summarized in Table I. Prostasin was eluted in gradient fractions of 0.15 to 0.25 M NaCl in 25 mM Tris-HCl, pH 7.6, from a DEAE-Sepharose CL-6B column and monitored by $[^{14}\text{C}]-\text{DFP}$ labeling, enzyme assay, and a prostasin-specific RIA after the antiserum was obtained. Fractions containing prostasin were pooled, concentrated and applied to an aprotinin-agarose column. The eluate from the aprotinin-affinity column showed high purity on SDS-PAGE in the form of two close bands with molecular mass of 40 kDa under reducing conditions (Fig. 1, lane 3). Although it migrated slightly faster under non-reducing conditions (Fig. 1, lane 4), its protein pattern remained the same. Human urinary (tissue) kallikrein is shown in lane 2. The extinction coefficient ($E_{\text{0.1%280}}$) for prostasin was determined to be 1.63 based on the absorbances at 205 nm and 280 nm (Van Lersel et al., 1985).

Active Site Labeling: Gradient fractions of 0.15-0.25 M NaCl from a DEAE-Sepharose CL-6B column contained one protein with a molecular mass of 40 kDa, which reacted to $[^{14}\text{C}]-\text{DFP}$. Prostasin, purified from above fractions by an aprotinin-affinity column, has a molecular mass of 40 kDa and can also be labeled by $[^{14}\text{C}]-\text{DFP}$. Both of the close bands can react with $[^{14}\text{C}]-\text{DFP}$ (Fig. 2, lane 1). This result indicates that prostasin is a serine proteinase.

Isoelectric Focusing: Isoelectric focusing of prostasin showed that purified prostasin contains five major bands with a pI ranging from 4.5 to 4.8 (Fig. 3, lane 1). Human urinary (tissue) kallikrein has a pI ranging from 3.7 to 4.3 (Fig. 3, lane 2) and human prostate-specific antigen (PSA) has a pI of 6.3 (Fig. 3, lane 3).

Analysis of NH$_2$-Terminal Amino Acid Sequence: The two close protein bands in the eluate from the aprotinin-agarose column (Fig. 1, lane 3) were cut off an Immobilon-P membrane and sequenced separately. The result showed that they had an
identical NH2-terminal sequence. Purified prostasin was also sequenced directly in solution. Both methods showed the NH2-terminal 20-amino acid sequence to be ITGGSSAVAGQWPWQVSITY. This sequence is different from any known protein or gene product sequence in SwissProt and GenBank databases. Therefore, we named this new serine proteinase "prostasin". Figure 4 shows the comparison of the NH2-terminal 20-amino acid sequence of prostasin with several human serine proteinases: α-tryptase, elastase 2A and 2B, chymotrypsin, acrosin and the catalytic chains of hepsin, plasma kallikrein and coagulation factor XI. Prostasin shares 50-55% identity with these serine proteinases. Interestingly, prostasin contains a six-amino acid sequence, WPWQVS, which is highly conserved in the above mentioned serine proteinases with the exception of one amino acid difference in coagulation factor XI and acrosin (Fig. 4). The results also indicate that prostasin may be a member of the serine proteinase superfamily.

**Kinetic Studies and Inhibition Analysis:** The pH-activity profiles shows that prostasin has a pH optimum of 9.0 using tripeptidyl fluorogenic substrate D-Pro-Phe-Arg-MCA (Fig. 5). Four substrates, D-Pro-Phe-Arg-MCA, D-Phe-Phe-Arg-MCA, D-Val-Leu-Arg-MCA and Z-Gly-Pro-Arg-AFC were used to measure its kinetic constants, catalytic rate and substrate affinity (kcat and Km). The values of kcat/Km are 0.632, 0.351, 0.412 and 0.400 mM⁻¹s⁻¹, respectively (Table II). Inhibition studies revealed that prostasin was inhibited by aprotinin, benzamidine, antipain and leupeptin, but not by soybean trypsin inhibitor. IC₅₀ values (inhibitor concentration at which 50% of enzyme activity is inhibited) are shown in Table III. The result indicated that aprotinin is a potent inhibitor of prostasin with an IC₅₀ of 1.8 x 10⁻⁹ M.

**Tissue Distribution and Quantitation:** A specific RIA was developed for prostasin so that its distribution and levels could be analyzed in human tissues and body fluids. Table IV shows that prostasin is present at high levels in the prostate gland (143.7 ± 15.9 ng/mg protein, n=3) and seminal fluid (8.61 ± 0.42 µg/ml, n=3); moderate levels (2-6 ng/mg protein) in the colon, lung, kidney, pancreas, salivary gland, liver and bronchi.
as well as in the urine (0.201 ± 0.029 µg/ml, n=3), but was not detected in the brain, ventricle, atrium, muscle, testis and aorta (data not shown). Linear displacement curves for immunoreactive prostasin in body fluids and tissue extracts were parallel with the standard curve for purified prostasin, indicating their immunological identity. Fig. 6 shows a representative log-logit transformation of a prostasin standard curve and dilution curves for seminal fluid and prostate gland extracts.

**Immunohistochemical Localization in Prostate Gland:** Immunoreactive staining for prostasin was observed in the epithelia and ducts of prostate glands (Fig. 7A). In some areas, the ducts were stained more intensely than the epithelial cells. In the control sections, no positive staining could be identified (Fig. 7B). These results show that prostasin is synthesized in epithelial cells of the prostate gland and then secreted into ducts.
DISCUSSION

In this report, we have described the purification, characterization, tissue distribution and localization of a novel human serine proteinase, designated as prostasin. Prostasin is an acidic protein with a molecular mass of 40 kDa and a pI ranging from 4.5 to 4.8. It displays trypsin-like activity and is inhibited by several trypsin inhibitors, in particular, with a high affinity to aprotinin. Based on the following criteria: similarity of the NH2-terminal amino acid sequence with other serine proteinases, substrate specificity, susceptibility to trypsin inhibitors and binding to DFP, we conclude that prostasin is a new serine proteinase. Prostasin in seminal fluid is considered to originate from the prostate gland where it has been found to be present at high levels and to be localized in the epithelial cells and ducts.

Prostasin migrates as two close bands with an apparent molecular mass of 40 kDa on SDS-PAGE under both reducing and non-reducing conditions (Fig. 1). These two protein bands have an identical NH2-terminal amino acid sequence, suggesting that they may have the same primary structure. Heterogeneity in the molecular weight is probably the result of their various carbohydrate content. This phenomenon was also observed in human plasma kallikrein (Heimark and Davie, 1981) and in human urinary kallikrein (Shimamoto et al. 1980; Fig. 2, lane 2) in which the difference in their isoelectric focusing (Fig. 2) was attributed to different numbers of sialic acids in the carbohydrate content of each protein molecule. Prostasin moved faster under non-reducing conditions than under reducing conditions on the SDS-PAGE (Fig. 1), indicating that there are intrachain disulfide bonds in the molecule. A similar result was observed in the precursor of murine IL-1β (Gunther et al., 1991). It is believed that intrachain disulfide bonds hold molecules compact whereas, under reducing conditions, disulfide bonds are cleaved and the conformation of the molecule becomes more extended. Therefore, proteins meet more resistance and move slower on SDS-PAGE under reducing conditions.
Prostasin is a serine proteinase displaying arginine amidase activity on peptidyl fluorogenic substrates. It has relatively high activity toward tripeptidyl-MCA substrates D-Pro-Phe-Arg-MCA, D-Phe-Phe-Arg-MCA, D-Val-Leu-Arg-MCA and D-Gly-Pro-Arg-AFC (Table II) and minimal activity to Z-Phe-Arg-AFC (data not shown), suggesting that the P3 residue in the substrates is important in determining the catalytic efficiency of prostasin. We also observed high activity of prostasin toward Z-Val-Lys-Lys-Arg-AFC (data not shown). Although prostasin shows a high degree of similarity (~55%) to chymotrypsin in the NH2-terminal amino acid sequence (Fig. 4), it does not hydrolyze chymotrypsin substrates such as Suc-Ala-Ala-Pro-Phe-MCA (data not shown).

Human tissue kallikrein, another serine proteinase, was also reported to be present in seminal fluid and was thought to be secreted from the prostate gland (Fink et al., 1985; Kumamoto et al., 1989). We detected human tissue kallikrein by a specific RIA as described by Shimamoto et al. (1980) and found that it was present in the fractions obtained at the gradient of 0.30-0.35 M NaCl from a DEAE-Sepharose CL-6B column after the elution of prostasin at 0.15-0.25 M NaCl. No immunological cross-reactivity was observed between purified prostasin and human tissue kallikrein by prostasin-specific RIA and Western blot with the antiserum either against prostasin or against human tissue kallikrein (data not shown). Therefore, the possibility that prostasin was contaminated by human tissue kallikrein has been eliminated.

Although the biological and physiological roles of prostasin are not known at present, some speculations can be made based on available data. High levels of prostasin in the prostate gland and in seminal fluid (over 20-fold higher than any other tissues examined) suggests that prostasin may play important physiological roles in the prostate gland and in the process of fertilization. We have identified the presence of prostasin in the prostate by immunohistochemical assay in which both the prostatic epithelium and the duct have been stained by a specific antiserum against prostasin (Fig. 7). In particular, the duct is stained more intensely, indicating that prostasin is secreted from the epithelium and then
probably accumulated in the duct. Human semen coagulates a few minutes after ejaculation and then liquifies within about 30 min. During liquefaction, proteolysis occurs. It is well known that a variety of proteolytic activities are present in seminal fluid, many of them may become involved in this process. Prostasin could play a role in this process through its arginine amidase activity. It could function to activate or inactivate specific proteinases to regulate the process of liquefaction. In addition, prostasin may have potential pathological functions in prostatic diseases. Studies revealed that the protein patterns of prostatic secretions from patients with prostatitis are different from those of healthy people based on SDS-PAGE and two-dimensional gel electrophoresis (Balerna et al., 1982; Tsai et al., 1984), suggesting that abnormal proteolysis might occur in those patients. Since prostasin appears to be present at high levels in the prostatic ducts (Fig. 7), it is a good candidate to be considered for involvement in the prostatic disease.

In contrast to the major seminal fluid enzyme, PSA, which is exclusively expressed in the prostate gland (Wang et al., 1979; Morris, 1989), prostasin has been detected in other tissues such as kidney, pancreas, liver, lung, salivary gland and bronchi, suggesting that it may play a role in other biological processes. Moreover, we detected prostasin in both male and female urine. The fact that prostasin is present in female urine and in kidney extracts suggests that it is expressed in the kidney and then secreted into the urine. Although the level of prostasin in urine samples is low (0.201 ± 0.029 μg/ml) compared to that in seminal fluid (8.61 ± 0.42 μg/ml), it seems that the expression of prostasin is very active in the kidney considering the large volume of urine excretion per day. The site of prostasin synthesis in the kidney and its physiological role will be a subject of further studies.
Fig. 1. SDS-PAGE of prostasin under reducing and non-reducing conditions. Five μg of each enzyme was electrophoresed on a 7.5-15% gradient SDS-polyacrylamide slab gel and proteins were stained with Coomassie Brilliant Blue R. Lane 1, standard markers: rabbit muscle phosphorylase b, 97,400; bovine serum albumin, 66,200; hen egg white ovalbumin, 42,699; bovine carbonic anhydrase, 31,000; soybean trypsin inhibitor, 21,500; and hen egg white lysozyme, 14,400; lane 2, human tissue kallikrein; lane 3, prostasin under reducing conditions; lane 4, prostasin under non-reducing conditions.
Fig. 2. Active site labeling of prostatin. [14C]DFP-labeled samples were applied to a 7.5 - 15% SDS-polyacrylamide slab gel. Lane 1, prostatin (5 μg); lane 2, rat tissue kallikrein (3 μg), purified from rat urine as described by Chao & Chao (1988).
Fig. 3. Isoelectric focusing of prostatin on a polyacrylamide slab gel. Lane 1, prostatin (5 μg); lane 2, human tissue kallikrein (6 μg), purified from pooled urine as described by Shimamoto et al. (1980); lane 3, human prostate-specific antigen (5 μg), purified from human seminal fluid according to the procedure reported by Sensabaugh & Blake (1990). Corresponding pH values are shown in the vertical axis.
Prostasin  ITGGSASSAVAGQWPQVSLITY
Chymotrypsin  IVNGEDAVPGSWPWQVSLQD
Serine protease hepsin  IVGGRDTSLSGRPWQVSLRY
Coagulation factor XI  IVGGTASVREGPWQVTLHT
Plasma kallikrein  IVGGTNSSWGEWPQVSLQV
Elastase 2A  VVGGEEARPNNSWPQVSLQY
Elastase 2B  MLGGGEEARPNNSWPQVSLQY
alpha-Trypsin  IVGGQEAPRSKWPQVSLRV
Acrosin  IVGGKAAQHGAWPWMSLQI

Fig. 4. NH2-terminal amino acid sequence comparisons of prostasin with human serine proteinases. The amino acid sequences of these serine proteinase correspond to the mature forms of chymotrypsin (Tomita et al., 1989), elastase 2A, elastase 2B (Kawashima et al., 1987), acrosin (Adham et al., 1990) and α-tryptase (Vanderslice et al., 1990) or the catalytic chains of serine protease hepsin (Leytus et al., 1988), coagulation factor XI (Fujikawa et al., 1986) and plasma kallikrein (Chung et al., 1986). Amino acid residues which are highly conserved are boxed.
Fig. 5. Optimal pH determination of prostasin enzyme activity. Hydrolytic activities are described as the percentage of that at pH 9.0 toward substrate D-Pro-Phe-Arg-MCA.
Fig. 6. Log-logit transformation of a specific radioimmunoassay standard curves of prostasin (●), serial dilution of human seminal fluid (▲) and human prostate gland extract (■). Prostasin standard curve ranged from 0.16 ng to 20 ng. The parallelism indicates the immunological identity among the protein detected in human seminal fluid, human prostate gland extract and prostasin. B/Bo is the value of bound radioactivity in the presence, divided by that in the absence, of the unlabeled prostasin using rabbit antiserum against prostasin.
Fig. 7. Immunohistochemical analysis of prostatin in the prostate gland. Formalin-fixed and paraffin-embedded human prostate gland was sectioned and incubated with A: rabbit antiserum against prostatin or B: preimmune rabbit serum, followed by goat IgG fraction against rabbit IgG F(ab')2, peroxidase-antiperoxidase complex and 3,3'-diaminobenzidine tetrahydrochloride plus hydrogen peroxide. (magnification: x160)
Table I. Purification of prostasin from human seminal plasma

<table>
<thead>
<tr>
<th>Step</th>
<th>Volume (ml)</th>
<th>Total protein (mg)(^a)</th>
<th>prostasin (µg)(^c)</th>
<th>prostasin/protein (mg/mg)</th>
<th>Yield (%)</th>
<th>Purification fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seminal plasma</td>
<td>75</td>
<td>1,428.3</td>
<td>729.1</td>
<td>0.00051</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>DEAE-Sepharose</td>
<td>240</td>
<td>348.0</td>
<td>483.2</td>
<td>0.0014</td>
<td>66.3</td>
<td>2.75</td>
</tr>
<tr>
<td>Aprotinin-agarose</td>
<td>16</td>
<td>0.1064 (^b)</td>
<td>106.4 (^b)</td>
<td>1.0</td>
<td>14.6</td>
<td>1,960</td>
</tr>
</tbody>
</table>

\(^a\) Determined by the method of Lowry, \textit{et al.} (1951).

\(^b\) Determined by spectrophotometric absorbance at 205 nm and 280 nm, using 

\[ E_{0.1\% 280} = 1.63. \]

\(^c\) Measured by a prostasin-specific RIA.
Table II. Kinetic constants for hydrolysis of tripeptidyl fluorogenic substrates by prostasin

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Km (µM)</th>
<th>kcat (s⁻¹)</th>
<th>kcat/Km (mM⁻¹s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Val-Leu-Arg-MCA</td>
<td>255</td>
<td>0.105</td>
<td>0.412</td>
</tr>
<tr>
<td>D-Phe-Phe-Arg-MCA</td>
<td>827</td>
<td>0.290</td>
<td>0.351</td>
</tr>
<tr>
<td>D-Pro-Phe-Arg-MCA</td>
<td>108</td>
<td>0.068</td>
<td>0.632</td>
</tr>
<tr>
<td>Z-Gly-Pro-Arg-AFC</td>
<td>717</td>
<td>0.287</td>
<td>0.400</td>
</tr>
</tbody>
</table>
Table III. Effects of trypsin inhibitors on the activity of prostasin

IC50 is the inhibitor concentration at which 50% of the activity of prostasin toward substrate D-Pro-Phe-Arg-MCA is inhibited.

<table>
<thead>
<tr>
<th>Inhibitors</th>
<th>IC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aprotinin</td>
<td>$1.8 \times 10^{-9} \text{ M}$</td>
</tr>
<tr>
<td>Benzamidine</td>
<td>$8.6 \times 10^{-4} \text{ M}$</td>
</tr>
<tr>
<td>Antipain</td>
<td>$6.4 \times 10^{-6} \text{ M}$</td>
</tr>
<tr>
<td>Leupeptin</td>
<td>$1.0 \times 10^{-4} \text{ M}$</td>
</tr>
<tr>
<td>Soybean trypsin inhibitor</td>
<td>No Inhibition</td>
</tr>
</tbody>
</table>
Table IV. Tissue distribution of prostatin

Prostatin levels in human tissues and body fluids were measured by a specific radioimmunoassay. Values represent a mean ± S.D. (n = 3).

<table>
<thead>
<tr>
<th>Tissue or Fluid</th>
<th>Prostatin (ng/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>143.7 ± 15.9</td>
</tr>
<tr>
<td>Bronchi</td>
<td>5.07 ± 0.17</td>
</tr>
<tr>
<td>Lung</td>
<td>5.17 ± 0.55</td>
</tr>
<tr>
<td>Colon</td>
<td>4.57 ± 1.23</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>4.24 ± 0.94</td>
</tr>
<tr>
<td>Pancreas</td>
<td>3.89 ± 1.79</td>
</tr>
<tr>
<td>Kidney</td>
<td>2.64 ± 0.50</td>
</tr>
<tr>
<td>Liver</td>
<td>2.59 ± 0.08</td>
</tr>
<tr>
<td>Seminal fluid</td>
<td>8.61 ± 0.42 µg/ml</td>
</tr>
<tr>
<td>Urine</td>
<td>0.201 ± 0.029 µg/ml</td>
</tr>
</tbody>
</table>
III. Molecular Cloning, Tissue-Specific Expression and Cellular Localization of Human Prostasin mRNA

SUMMARY

We have purified a novel human serine proteinase, designated as prostasin, from seminal fluid (Yu et al., 1994). In the present study, we have cloned and characterized the full-length cDNA encoding prostasin and identified its tissue-specific expression and cellular localization. A cDNA fragment was obtained by polymerase chain reaction using degenerate oligonucleotide primers derived from the NH2-terminal and internal amino acid sequences. A full-length cDNA sequence encoding prostasin was obtained by amplification of the 5’ and 3’ ends of the cDNA. It contains a 1,032-base coding region, a 574-base 3’ noncoding region and a 138-base 5’ noncoding sequence. Prostasin mRNA encodes a protein of 343 amino acids, which consists of a 32-amino acid signal peptide and a 311-amino acid proprostasin. Proprostasin is then cleaved between Arg12 and Ile13 to generate a 12-amino acid light chain and a 299-amino acid heavy chain, which are associated through a disulfide bond. The deduced amino acid sequence of the heavy chain has 34-42% identity to human acrosin, plasma kallikrein and hepsin. A potential N-glycosylation site at Asn127 and the catalytic triad of His53, Asp102 and Ser206 have been identified. The deduced prostasin has a unique 19-amino acid hydrophobic portion at the COOH-terminus, which makes it suitable to anchor in the cell membrane. Carboxy-terminal sequencing of purified prostasin indicates that the hydrophobic portion is removed and that there is a cleavage between Arg290 and Pro291 during secretion. Southern blot analysis, following a reverse transcription polymerase chain reaction, indicates that
prostasin mRNA is expressed in prostate, liver, salivary gland, kidney, lung, pancreas, colon, bronchus, renal proximal tubular cells and prostate carcinoma LNCaP cells. Cellular localization of prostasin mRNA was identified within epithelial cells of the human prostate gland by in situ hybridization histochemistry.
INTRODUCTION

Human seminal fluid is a rich source of proteolytic enzymes, many of which are involved in the postejaculatory hydrolysis of proteins and in semen coagulation and liquefaction (Shivaji et al., 1990). Prostate-specific antigen (PSA) and acrosin are two of the most important proteolytic enzymes found in human semen. PSA may play an important role in semen liquefaction through hydrolyzing semenogelin, a predominant seminal vesicle protein (Lilia, 1985). PSA levels in blood have been recognized recently as the most important marker for prostate cancer. Acrosin is a serine proteinase present in acrosomes where it covers the anterior part of the sperm head (Klemm et al., 1991). It is believed to be involved in recognition, binding and penetration of the zona pellucida of the ovum during interaction of the sperm and egg (Jones et al., 1988; Topfer-Petersen and Henschen 1988). Recently, we have identified and purified a new serine proteinase, designated as prostasin, from human seminal fluid (Yu et al., 1994). At the present time, the physiological functions of prostasin are unknown and its physiological substrate remains to be identified.

Prostasin has an apparent molecular mass of 40 kDa on SDS-PAGE and displays arginine amidolytic activity. The NH2-terminal 20-amino acid sequence of prostasin shares 50-55% identity with human α-tryptase, elastase 2A and 2B, chymotrypsin, acrosin and the catalytic chains of hepsin, plasma kallikrein and coagulation factor XI (Yu et al., 1994). It is present in many tissues and has the highest level in the prostate gland. In the prostate gland, prostasin has been localized in epithelial cells and ducts by immunohistochemistry. It is believed that prostasin is synthesized in prostatic epithelial cells, secreted into the ducts and excreted into the seminal fluid where it may serve a role in fertilization. The wide distribution of prostasin outside the prostate gland indicates that it may also play important roles in other biological processes.
In order to understand the structure, regulation and function of prostasin, it is essential to isolate and characterize its cDNA. In this study, we have cloned the full-length cDNA encoding prostasin through polymerase chain reaction (PCR) and 5' and 3' rapid amplification of cDNA ends (RACE) based on its NH2-terminal and internal amino acid sequences. We have elucidated its primary structure and defined the posttranslational processes which convert preproprostasin into proprostasin and prostasin. In addition, we have identified the cleavage site where prostasin is released from membranes. Tissue-specific expression of prostasin has been identified by reverse transcription PCR followed by Southern blot analysis. Cellular localization of prostasin mRNA in the human prostate gland has been determined by in situ hybridization.
**EXPERIMENTAL PROCEDURES**

*Internal Amino Acid Sequence Analysis:* Prostasin was purified from human seminal plasma as described previously (Yu *et al.*, 1994). Purified prostasin (40 µg) was digested with TPCK-trypsin (Sigma) at a ratio of 1:100 (w/w) at 37 °C for 16 h after reduction by dithiothreitol (DTT) and S-carboxymethylation by iodoacetic acid according to the procedure described by Stone *et al.* (1989). Generated peptide fragments were separated by a reverse phase HPLC (Model 5000 Liquid Chromatograph, Varian Associates, Inc.) with a µBONDAPAK C18 column (3.9 mm x 30 cm, Water, Inc.) and eluted by an acetonitrile gradient. The collected fractions were concentrated by Speedvac to a desired volume and subjected to amino acid sequencing using a gas phase protein sequenator equipped with an on-line narrow-bore PTH-amino acid analyzer (ABI model 470 A, Applied Biosystems Inc.).

*Amplification of a Partial cDNA Fragment:* A human prostate cDNA library in λ gt 11 (Clontech Lab, Inc.) was amplified with forward and backward primers corresponding to the two arms of the phage DNA. The reaction mixture contained 1 x PCR buffer (Perkin-Elmer Cetus Corp.), 0.5 mM dNTP, 100 pmol of primers, 0.1% Triton X-100, 2 x 10^7 phage of the cDNA library and 2.5 units of Ampli-Taq DNA polymerase (Perkin-Elmer Cetus Corp.) in a total volume of 50 µl. The reaction was conducted in a GeneMachine II (USA/Scientific Plastics, Ocala, FL) using the following program: 94 °C/1 min - 42 °C/2 min - 72 °C/3 min for 30 cycles followed by 5 min at 72 °C. According to the NH2-terminal amino acid sequence of purified prostasin, ITGGSSAVAGQWPWQVSITY (Yu *et al.*, 1994), and the internal amino acid sequences obtained above, two degenerate primers were designed. The sense primer is JY-1: 5’-GT(ACGT)GC(ACGT)GG(ACGT)CA(AG)TGGCC-3’, which corresponds to VAGQWP in the NH2-terminal sequence; the antisense primer is JY-T84-2: 5’-TT(ACGT)GC(AG)TC(AGT)AT(AG)TT(AG)TA-3’, corresponding to YNIDAK in
peptide T84 (Table V). To amplify a partial prostasin cDNA fragment, two µl of the amplified human prostate cDNA library was used as template in the following PCR reaction: 1 x PCR buffer, 0.25 mM dNTP, 0.05 mM of tetramethylammonium chloride (TMAC), 100 pmol of JY-1 and JY-T84-2 and 2.5 units of AmpliTaq DNA polymerase. The cycling program was the same as mentioned above. A second round of PCR was carried out with 10 µl of the first round product as template and the same primers under the same condition.

Sequence Analysis of PCR Product: After purification, the amplified products were subjected to Southern blot analysis using a nested degenerate oligonucleotide JY-4, which is located just downstream of the sense-primer JY-1. The sequence of JY-4 is: 5' -CA(AG)TGGCC(ACGT)TGGCA(AG)GT-3', which corresponds to QWPWQV at the NH2-terminus of the purified prostasin. A 300 bp DNA fragment which hybridized to JY-4 oligonucleotide was sequenced using JY-4 as a primer with the BRL dsDNA Cycle Sequencing System (Bethesda Research Laboratories, Bethesda, MD).

Elucidation of the Full-length cDNA Sequence of Prostasin: A partial prostasin cDNA sequence with a size of approximately 300 bases was obtained from the above experiment. Based on this sequence, a sense primer JY-F1, 5'-GTCCATGTTGTGGTGTTG-3', was used with total RNA from human renal proximal tubular cells in the 3' RACE reaction (Gibco BRL). A nested primer JY-AD, 5'-CTGTCAGCTGCTCACTGC-3', was used to sequence the product of the 3' RACE reaction to reveal the 3' end of prostasin mRNA. The antisense primer JY-B1, 5'-TGGGTCTGCTGAGTTGG-3', was exploited for reverse transcription in the 5' RACE reaction. After adding an oligo-(dC) anchor sequence to the 3' end with terminal deoxynucleotidyl transferase, the generated cDNA was amplified with a nested antisense primer JY-UA, 5'-CTCCTGGAGGTAGCTGG-3', and an anchor primer provided by the manufacturer in the 5' RACE system (Gibco BRL). The amplified product was sequenced.
with primer JY-AU: 5'-GCAGTGAGCAGCTGACAG-3' to reveal the 5' end of prostatin mRNA.

**Analysis of NH2-terminal Amino Acid Sequence of Purified Prostasin:** Prostasin was purified from human seminal fluid as previously described by Yu *et al.* (1994). In order to explore whether there were two chains linked by a disulfide bond, the purified prostasin was resolved on SDS-PAGE under nonreducing conditions and transferred to an Immobilon-P membrane (Millipore Corp.). The protein band was cut out for NH2-terminal amino acid sequencing.

**Analysis of Prostasin Amino Acid Composition:** To prove that the presence of a hydrophobic portion at the COOH-terminus of the deduced prostasin (Fig. 10) is responsible for anchoring it in the membrane and that the hydrophobic segment is lost during secretion, like angiotensin-converting enzyme (Wei *et al.*, 1991), the amino acid composition of the heavy chain of purified prostasin was analyzed. Purified prostasin was resolved on SDS-PAGE under reducing conditions and then transferred to an Immobilon-P membrane. The protein band was cut out and subjected to amino acid composition analysis as described by Xiong *et al.* (1990). The amino acid composition of the deduced heavy chain of prostasin was obtained through computer analysis according to its cDNA sequence.

**Carboxypeptidase Digestion:** Since the deduced prostasin has a putative transmembrane domain at the COOH-terminus (Fig. 8), this hydrophobic portion is likely to be removed in the secreted prostasin. To explore this matter, 60 μg of purified prostasin from human seminal fluid was dissolved in 0.2 M N-ethylmorpholine acetate (pH 8.3) and a mixture of carboxypeptidase A and carboxypeptidase B was added to prostasin at a molar ratio of 1:50. The reaction was carried out at 37 °C and an aliquot of the incubation mixture was removed for amino acid composition analysis at 1, 5 and 19 h. After 19 h incubation, the pH of the test protein mixture was lowered to 6.0 with acetic acid and carboxypeptidase
Y was added to the remaining prostasin at a 1:50 ratio. An aliquot of the incubation mixture was removed at 1 and 19 h for amino acid composition analysis.

**Tissue-specific Expression of Prostasin:** Human tissues were obtained from autopsy and kindly provided by Dr. Sandra Conradi at the Medical University of South Carolina. Renal proximal tubular cells were obtained from Dr. Debra Hazen-Martin at the Medical University of South Carolina (Detrisac et al., 1984) and prostate carcinoma LNCaP cells from the American Type Culture Collection. Tissues were handled as reported by Chai et al. (1993) and total RNA was isolated according to the procedure described by Davis et al. (1986). Total RNAs extracted from various tissues were subjected to reverse transcription PCR using the sense primer JY-F1 and antisense primer JY-B1. The PCR product was confirmed by Southern blot analysis with probe JY-AD (Maniatis et al., 1982).

**In Situ Hybridization:** Prostasin cDNA was amplified with two primers, JY-F1 and JY-B1, after reverse transcription using total RNA from LNCaP cells. A 255 bp fragment was gel-purified from an agarose gel and ligated into pSP73 at the HindIII site. The recombinant plasmid was used to transform E. coli JM101. Positive colonies were selected by colony-hybridization with oligonucleotide JY-AD and the orientation of the cDNA fragment in the vector was determined by DNA sequencing. The sense and the antisense RNAs, corresponding to the partial prostasin cDNA, were synthesized by using SP6 and T7 RNA polymerases, respectively. During the synthesis of the RNAs, digoxigenin-labeled uridine-triphosphate (DIG-UTP) was incorporated according to the protocol of Boehringer Mannheim Biochemica. The DIG-labeled antisense RNA (riboprobe) was used to detect prostasin mRNA in 5 μm sections of human prostate which was formalin-fixed and paraffin-embedded. DIG-labeled sense RNA was used as a control. An antibody-conjugate (anti-digoxigenin alkaline phosphate conjugate) was used to recognize the DIG-labeled RNA. A subsequent enzyme-catalyzed color reaction was conducted by the addition of 5-bromo-4-chloro-3-indolyl phosphate (X-phosphate) and
nitroblue tetrazolium salt (NBT) and incubated overnight at room temperature in the dark as described by the manufacture. After washing and dehydration, sections were mounted with Permount.
RESULTS

*Internal Amino Acid Sequence Analysis:* Purified prostasin was subjected to TPCK-trypsin cleavage and as many as 15 peptide fragments were isolated by reverse phase HPLC with a C18 column. Five of the peptide fragments were subjected to NH2-terminal amino acid sequencing (Table V). The peptides are T45: LGAHQLDSYSE, T59: ASSYASWIQSK, T75: NRPGVYTLASSYAS, T77: YIRPIXLPAAXASFP and T84: ETXNXLYNIDAKPEEPHFVQ (X stands for undetermined amino acids). A degenerate oligonucleotide JY-T84-2 with the least degeneracy (96) was synthesized as an antisense primer corresponding to the amino acid sequence YNIDAK in peptide T84. Two degenerate sense primers JY-1 and JY-4 with degeneracies of 128 and 16, respectively, were designed according to the previously obtained NH2-terminal amino acid sequence ITGGSSAVAGQWPQVSITY (Yu et al., 1994).

*Cloning and Analysis of Full-length cDNA Encoding Human Prostasin with Degenerate Oligonucleotide Primers:* Figure 8 shows the nucleic acid sequence of prostasin cDNA and its deduced amino acid sequence. Prostasin cDNA consists of a coding region of 1,032 bases, a 3' noncoding region of 574 bases and a 5' noncoding region. The coding region starts with an ATG codon which is present in a sequence of GTCCTGGCCA TGG, similar to the consensus sequence GCCGCC(AlG)CCATGG for the eukaryotic translational initiation site (Kozak, M., 1987). The cDNA encodes a 343-amino acid polypeptide, including a 32-amino acid signal peptide, a 12-amino acid light chain and a 299-amino acid heavy chain beginning with Ile-Thr-Gly-Gly, the NH2-terminal amino acid sequence of purified prostasin reported by Yu et al. (1994). A potential N-linked glycosylation site has been identified as Asn127 in a sequence of Asn-Ala-Ser and a variant polyadenylation signal, ATTAAA, was found twelve bases upstream of the poly (A). Figure 9 compares the deduced prostasin amino acid sequence with other serine proteinases. Prostasin shares 34-42% identity to human
plasma kallikrein, coagulation factor XI, β-tryptase, hepsin, plasminogen and acrosin. The catalytic triad of the deduced prostasin has been identified as His$^{53}$, Asp$^{102}$ and Ser$^{206}$ according to the multiple sequence alignment. In alignment with other serine proteinases (Fig. 9), prostasin contains an Asp residue at position 200, indicating that it has trypsin-like activity. A hydropathy plot of the deduced prostasin identifies two hydrophobic regions, one located at the NH2-terminus and the other at the COOH-terminus (Fig. 10). The one located at the NH2-terminus is likely to be a signal peptide, which could direct newly synthesized prostasin to enter the endoplasmic reticulum. The other region located at the COOH-terminus is indicated as a putative transmembrane domain and is double-underlined in figure 8.

**Analysis of Prostasin NH2-terminal Amino Acid Sequence and Defining the Cleavage Sites of the Preproprostasin:** Under nonreducing conditions, two distinct signals were observed on chromatograms in each Edman degradation cycle of the purified prostasin. However, under reducing conditions only one signal was obtained (Yu et al., 1994). The two amino acid sequences obtained are ITGGSSAVAGQW, which is the same as the previously published NH2-terminal amino acid sequence under reducing conditions (Yu et al., 1994), and AEAPXGVAPQ (X stands for an undetermined amino acid residue). Examination of the deduced prostasin amino acid sequence from its cDNA indicates that AEAPXGVAPQ is located just upstream from ITGGSSAVAGQW and the X is a cysteine residue. This result shows that the nascent prostasin is a preproenzyme, which is converted to proenzyme by removing a 32-amino acid signal peptide. The generated proprostasin is then activated by the cleavage of a peptide bond between Arg$^{12}$ and Ile$^{13}$ to give rise to active prostasin which contains a light chain of 12 amino acids and a heavy chain of 299 amino acids. The two chains are held together by a disulfide bond.

**Comparison of the Amino Acid Composition Between the Heavy Chains of the Purified and Deduced Prostasin:** The amino acid composition for
the heavy chain of purified prostasin was obtained through acidic hydrolysis and is shown in Table VI. Analysis of the amino acid composition indicates that most amino acids in purified prostasin are in slightly lower quantity than in the deduced one from cDNA. Several amino acids, such as Leu, Ser and Pro, are notably more abundant in the amino acid composition deduced from prostasin cDNA. Close examination of the deduced amino acid sequence indicates that the COOH-terminal portion, beginning with Ser (281), are rich in these amino acids (Fig. 8). The results suggest that the COOH-terminal hydrophobic portion is cleaved in the mature or secreted form of prostasin.

**Identification of COOH-terminal Residues of Purified Prostasin:** The result of carboxypeptidase digestion indicates that the first released amino acid residue is Arg, followed by either Leu or Ala. No release of His^{311}, the last amino acid residue of prostasin deduced from its cDNA, can be observed. The total amount of released Arg is substantially higher than that of Leu or Ala, indicating that there are two carboxy-terminal Arg residues. Considering that a light chain and a heavy chain exist in the purified prostasin, the signals obtained from carboxypeptidase digestion represent two COOH-terminal sequences. Since the last two amino acids of the light chain are Ala-Arg, the COOH-terminus of the heavy chain must be Leu-Arg. Compared with the deduced prostasin sequence, there is a Leu-Arg sequence just upstream from the putative transmembrane domain at the COOH-terminus. Therefore, it is believed that a cleavage occurs between Arg^{290} and Pro^{291}.

**Tissue-specific Expression of Human Prostasin:** Prostasin mRNA expression was detected in several human tissues with RT-PCR followed by Southern blot analysis using a specific oligonucleotide probe. The result shows that prostasin is expressed in the human prostate gland, liver, salivary gland, kidney, lung, pancreas, colon, bronchus, renal proximal tubular cells and LNCaP cells, but not in the testis, ovary, spleen, uterus, cortex, muscle, atrium, ventricle, aorta, vein, artery, umbilical vein endothelial cells, lymphocytes and polymorphonuclear cells (Fig. 12).
In Situ Hybridization of Prostasin mRNA in the Prostate Gland:

Cellular localization of prostasin mRNA has been identified within the epithelial cells of the human prostate gland using the antisense riboprobe by *in situ* hybridization histochemistry (Fig. 13a). No staining was observed in the control section using the sense riboprobe (Fig. 13b).
DISCUSSION

In this study, we have cloned and sequenced a full-length cDNA encoding prostasin, a human serine proteinase. This cDNA codes for a protein of 343 amino acids with a 32-amino acid signal peptide. The catalytic triad essential for enzymatic activity of prostasin is His^{53}, Asp^{102} and Ser^{206}. The sequences around these active sites are highly conserved in the serine proteinases as shown in figure 9 and the heavy chain of prostasin shares 34-42% identity with them. A potential glycosylation site, Asn^{127}, has been identified. A putative transmembrane domain is present at the COOH-terminus, suggesting that prostasin is likely to be a membrane-anchored serine proteinase. Prostasin is expressed in a variety of human tissues and its mRNA is localized within epithelial cells of the prostate gland.

Based on the amino acid sequence deduced from its cDNA, the posttranslational processing sites of prostasin have been defined. The posttranslational process of prostasin is very similar to that of acrosin, which produces a 23-amino acid light chain associated with the heavy chain by two disulfide bonds (Klemm et al., 1991). Prostasin is synthesized as a preproenzyme of 343 amino acids. During translocation into the endoplasmic reticulum, a cleavage occurs between Gly^{1} and Ala^{1} to remove the 32-amino acid signal peptide generating proprostasin. This cleavage site agrees with the motif that requires small neutral amino acid residues at positions -3 and -1 (von Heijine, 1990). The generated proenzyme with 311 amino acids is then activated by a specific cleavage between Arg^{12} and Ile^{13} to produce an active, two-chain form. The two chains are held together by an interchain disulfide bond between Cys^{5} in the light chain and a Cys residue in the heavy chain.

Like acrosin, prostasin has a unique protruding COOH-terminus (Fig. 9). Acrosin is a serine proteinase membrane-anchored through its COOH-terminus by an unknown mechanism and is then released after cleavage (Baba et al., 1989). At the COOH-terminus
of prostasin, there is a highly hydrophobic portion of 19 amino acids, which are rich in leucine and flanked by a positively or a negatively charged amino acid at either side (Arg and Glu). This indicates that prostasin is likely to be a membrane-bound serine proteinase. Similarly, angiotensin-converting enzyme has a 17-amino acid hydrophobic portion near its COOH-terminus, which has been identified as a transmembrane domain (Wei et al., 1991). After cleavage, angiotensin-converting enzyme is released and its COOH-terminal portion remains on the membrane (Beldent et al., 1993). Carboxypeptidase digestion shows that the COOH-terminal amino acids of the purified prostasin’s heavy chain are Leu-Arg, indicating that there is a cleavage between Arg\(^{290}\) and Pro\(^{291}\) in the deduced prostasin.

There are two Arg residues in the COOH-terminus of the deduced prostasin, Arg\(^{259}\) and Arg\(^{290}\). Obviously, the peptide bond between Arg\(^{290}\) and Pro\(^{291}\) is preferred over that between Arg\(^{259}\) and Val\(^{260}\) since Pro\(^{258}\) cannot be detected after carboxypeptidase digestion. This result was verified by immunoblot analysis using a specific antiserum against an 11-amino acid peptide, 262Pro-Gln-Thr-Gln-Glu-Gln-Pro-Asp-Ser-Asn\(^{272}\), which recognized purified prostasin (data not shown). Analysis of the amino acid composition of purified prostasin’s heavy chain also suggests that the Leu-rich hydrophobic portion is missing since the Leu content is much lower than that in the deduced one (Table VI). From these results, we conclude that prostasin loses its COOH-terminal hydrophobic portion during secretion.

Compared with other serine proteinases, the number and positions of 9 out of the 11 cysteine residues in the catalytic chain of the translated prostasin are highly conserved (Fig. 9). On the basis of the known disulfide bridge arrangement in serine proteinases (Young et al., 1978), four intrachain disulfide bonds are expected at cysteine pairs 38/54, 136/212, 169/191 and 202/230 in prostasin (Fig. 8). Except for the single-chained b-tryptase, Cys (122) is conserved in all of the serine proteinases listed in figure 9. This cysteine residue has been found to be involved in the formation of an interchain disulfide bond with the noncatalytic chain in plasma kallikrein, coagulation factor XI and acrosin.
Considering that purified prostasin consists of two chains held by a disulfide bond, we conclude that an interchain disulfide bond exists between Cys\textsuperscript{5} and Cys\textsuperscript{122}. In addition, prostasin has two unique cysteine residues at positions 171 and 274 (Fig. 8). Whether these cysteine residues form a disulfide bond remains to be explored.

A potential N-linked glycosylation site, Asn-Ala-Ser, has been identified at Asn\textsuperscript{127}. The presence of this site explains why purified prostasin under reducing conditions displayed a larger size (40 kDa) on SDS-PAGE than the deduced heavy chain of 32 kDa. Asn\textsuperscript{127}, which was included in tryptic peptide T77 and described as X, could not be identified in amino acid sequencing although the signals before and after the Asn residue were very strong (data not shown). Furthermore, Asn residues in tryptic peptide T84 were easily identified. This phenomenon indicates that Asn\textsuperscript{127} has been modified by carbohydrates which make it nondetectable. In addition to N-linked glycosylation, it is likely that O-linked glycosylation is present in prostasin, since there are many serine and threonine residues in the deduced amino acid sequence of prostasin. This modification might also contribute to the discrepancy between the molecular masses of the deduced and the purified protein.

On synthetic substrates, prostasin shows trypsin-like activities, such as arginine amidolytic activities on D-Pro-Phe-Arg-MCA and D-Phe-Phe-Arg-MCA (Yu et al., 1994), and lysine amidolytic activities on Suc-Ala-Phe-Lys-MCA and Boc-Val-Leu-Lys-MCA (data not shown). It has no enzymatic activity on chymotrypsin substrates such as Suc-Ala-Ala-Pro-Phe-MCA (Yu et al., 1994). These results are consistent with the fact that in the deduced prostasin, an Asp residue is present at position 200, which is located six residues before the active site Ser\textsuperscript{206}. This Asp residue, which is located at the bottom of the substrate-binding pocket in trypsin, is involved in an interaction with the Arg or Lys residue of a substrate (Ruhlmann et al., 1973). In addition, Gly\textsuperscript{227} and Gly\textsuperscript{237} are conserved in prostasin. The counterparts of these two Gly residues in trypsin are present at...
the entrance of the substrate-binding pocket and permit entry of large amino acid side chains. Thus, these features in prostasin's primary structure determine its trypsin-like cleavage preference.

Broad existence of prostasin mRNA in human tissues suggests that it may have important biological functions. Localization of prostasin mRNA in the epithelial cells of the prostate gland indicates that prostasin is synthesized in the cells and then secreted into the ducts. The presence of prostasin in prostatic epithelial cells and ducts was identified by immunohistochemistry in our previous studies (Yu et al., 1994). Since it is likely to be a membrane-bound serine proteinase, prostasin may be involved in some important processes on the surface of cell membranes, such as removal of propeptides from hormones and growth factors and the activation of proenzymes associated with membranes. In order to understand prostasin's physiological functions, further experimentation is needed.
Fig. 8. Nucleic acid sequence of prostasin cDNA and the deduced amino acid sequence. A variant polyadenylation signal, ATTAAA, is underlined and poly (A) is designated as (A)n. A solid triangle indicates a potential N-glycosylation site, open triangles indicate active sites of the catalytic triad and stars represent a stop codon. Amino acid numbering starts with the first amino acid of proprostasin.
Fig. 9. Comparison of prostasin sequence with other serine proteinases. The amino acid sequences of these serine proteinase correspond to the mature forms of α-tryptase (Vanderslice et al., 1990) or the catalytic chains of acrosin (Adham et al., 1990), plasma kallikrein (Chung et al., 1986), coagulation factor XI (Fujikawa et al., 1986), serine protease hepsin (Leytus et al., 1988) and plasminogen (Forsgren et al., 1987). Amino acid residues which are highly conserved are batched and the catalytic triad of histidine, aspartic acid and serine of the catalytic triad are indicated by trangles. Dots represent gaps to bring the sequences to better alignment.
Prostasin: ITGSSAVAG CAPOQVSQTV ...EGVHV CGGSSVSCAQ VLSSAhCFPS EHHKEAYEVK
Plasma kallikrein: I TVGGSGSSWG KNPQAVQYLV ...KLTAQQLH CGGGHCSQG LPLQWWRVRY
Coagulation factor XI: I TVGGSGSGRV GKPQVQTVT ...TSPTRQRL GCGSSINCO GILTAVCQG VPSKLRVY
β-Tryptase: I TVGGQAPRPS KNPQAVQYLV ...HGFMNNIF CGGGHSCQG LPLQAVGVP D.VKDLAAVL
Hepsin: I TVGKDSLG KNPQAVQYLV ...D.GAILL CGGGHSCQG LPLQAVCPE RNNVLSWRV
Plasminogen: I TVGCVANHP HSPQAVQLAT ...RPGMNH CGGSSPEVQ VLTAYAVL.E KSRRPSSYKV
Acrosin: I TVGKSAQHQ ABPMVLQI PTYNHRY HGT CGGSSNLVQ LPLQAVCVP GRSNHDRLV

LGAV....Q LDSYSEDAKP STLKKIXHP SYLQ...... EGSSDIALL OLSPFIPFSR YIRPILBAA NAFPPG.LH
SGIL....N LSDDTKDPPF SQIKEITIHQ NYKV...... SEGDKDIALL KLPAPLNYTE FQKPIEILPSK GDTSIIY.TN
SGIL....N QSEIKEDTSF FQGVIYIIDIH QYKM...... AESGDYIALL KLETTVNYTD SQRPILPSK GDKNVYI.TD
VQIL......E QHLTYQYAGL P.VSRIIIHVF QFTY...... AQIGADDIALL ELAIPEVYVSS HVTNTLPFA SETFPFG.MP
FA......GA VAQAPSHGILQ LGVQAVVYHG GYLPFRDPNS EENSNADAV HLSSPLPLTE YIQFVCLRAA QALV.DGKI
IL......GA HQFV......NLH PHQIEIPVSR LFL...... EPRKDIALL KLSSPFLTD KVIFEPILSP NYVVA.DRTE
FGAKEYTYGN NKPVKAPVQE RYVEKIIIHE KY......NS ATGENDIALV EITPPISGQR FIGFPVLHP LAGLPRGQS

CTOSNVNHVA PVSSELITPVP LQOQEVVLIS RETCNCLMYI DA.KPSSPFV QOEDMVCAQH VEGDKDASQ LSGDLSHCAV
CTOSNVSHFA EKGEQ...QT LQXVNIPLVT NEEQKRYQ...... DYK ITQRMVCAQG KSGDLDASH LGDSQVCKH
CTOSNVYKIR REDQ...QT LRQAKPLVT NEEQKRYA...... GHK ITQRMVCAQG RSGDLDASH LGDSQVCKH
CTOSNVGVOD NDERLPPPPF LQKVQPVME NHIIDAKYHL GAYTGDQAR VREMQMCAG ...NTPESDCO LGDSQVCKV
CTOSNVGVNTQ .YYG.QAGV LQEARPVII NDCVQGADFQ GN......Q IKPQFMCAG FQGPDHGOS LGDSQVCKED
CFIPOQGKTO GETG...AGL LKEAQLPVIE NKVANCYRFEL NG......R VQSTLCAAG LAGSDPCFQ LGDSQVCFPE
CMVAGAVYIE EKAP.RPSSI LMKVADILID LDCNSTDWY NG......R VQPTNVLAGE FGKNDTQGQ LGDSQVCFK

...EGLWYL TGHVSGDAC GARNRGVT Q LASSYASIQ SKVTLQPVR VFPQQESQPD SNLCGSHLAF SSAPAQGLLR
...NLMWRL VGTSVGEGC ATOIQGVT Q KVAEYMILL EKTOSSDGKA QMOSPA
...NEWWNL VGTSVGEGC AQERFGVT Q NNVEYVIIL EKTAQV
...NOVQL AGUSWGEC AQPNRGIT Q KVYIILIIH HYVPPK
SFSNPRWRL CGQTVGOGC ALOPQGVT Q KVSDPRLIP QAIKHSSEAS CMVTLQ
...KDVYIL CGVTSGGEC AIPRKQPVQ KSRPVYVIE GMVRN
...SKEEAYVGV QGTVSVGOC ARIKPGIT ATWYPYHIA SKIQSNALRM IQSAPPPPAC TRPPPQPPF SHPISAHLPW

PILFPLPLGLA LGLLSPWLSN H

YFPQPPRRFL PRPAAQCPF PPSSPPPPPFP PASPLLPPPPF PDDPPSSST KLQPQLSFAK KLQQLEVLK GKTYSQGHRH
Fig. 10. Hydropathy plot of the deduced prostasin. The hydropathy of prostasin’s amino acid sequence translated from its cDNA was predicted with MacProMass program and plotted with Kyte & Doolittle Hydropathic index using a window size of 10 residues. Amino acid numbering begins with the start codon Met.
Fig. 11. Diagram of posttranslational proteolytic processes of prostasin. The bridges of lines represent the predicted disulfide bonds based on comparison with known serine proteinases and stars stand for positions of the catalytic triad.
Fig. 12. Tissue-specific expression of prostasin mRNA. A specific RT-PCR was conducted using total RNA from 24 human tissues or cells. The top panel: Prostate, liver, testis, salivary gland, kidney, lung, pancreas, ovary, spleen, uterus, colon and cortex. The bottom panel: Muscle, atrium, ventricle, bronchus, aorta, vein, renal proximal tubular cells, human umbilical vein endothelial cells, LNCaP cells, lymphocytes and polymorphonuclear cells. The RT-PCR products were detected by a nested oligonucleotide probe specific for prostasin.
<table>
<thead>
<tr>
<th>Tissue/Cell Type</th>
<th>Prostate</th>
<th>Liver</th>
<th>Testis</th>
<th>Salivary gl.</th>
<th>Kidney</th>
<th>Lung</th>
<th>Pancreas</th>
<th>Ovary</th>
<th>Spleen</th>
<th>Uterus</th>
<th>Colon</th>
<th>Cortex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ventricle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronchus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aorta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prox. tubule</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huvec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNCaP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphocytes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 13. Localization of prostasin mRNA by \textit{in situ} hybridization in the human prostate gland. a) A digoxigenin-labeled antisense RNA of prostasin was used as a probe. b) A digoxigenin-labeled sense RNA of prostasin was used for hybridization as a control. The labeled RNA probes were detected with an enzyme-linked immunoassay. (Magnification: 80 x)
Table V. Amino acid sequences for TPCK-trypsin fragments of prostasin.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence*</th>
<th>Location in sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T45</td>
<td>LGAHQLDSYSE</td>
<td>68 to 78</td>
</tr>
<tr>
<td>T59</td>
<td>ASSYASWIQSK</td>
<td>242 to 252</td>
</tr>
<tr>
<td>T75</td>
<td>NRPGVYTLASSYAS</td>
<td>234 to 247</td>
</tr>
<tr>
<td>T77</td>
<td>YIRPIXLPAAAX’ASFP</td>
<td>117 to 131</td>
</tr>
<tr>
<td>T84</td>
<td>ETXNXLYNIDAKPEEPHVQ</td>
<td>167 to 186</td>
</tr>
</tbody>
</table>

† The number in each peptide’s designation stands for the elution time in minute from a mBONDAPAK C18 column.
* X and X' represent Cys and Asn, respectively, which were determined by comparison with the deduced amino acid sequence from the cDNA.
Table VI. Comparison of amino acid composition of the heavy chain between the purified and the deduced prostasins.

<table>
<thead>
<tr>
<th></th>
<th>Purified prostasin</th>
<th>Deduced prostasin</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>ASN</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>THR</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>SER</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td>GLU</td>
<td>30</td>
<td>17</td>
</tr>
<tr>
<td>GLN</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>PRO</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>GLY</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>ALA</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>VAL</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>MET</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ILE</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>LEU</td>
<td>21</td>
<td>34</td>
</tr>
<tr>
<td>TYR</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>PHE</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>HIS</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>LYS</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>ARG</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>1/2CYS</td>
<td>ND*</td>
<td>11</td>
</tr>
<tr>
<td>TRP</td>
<td>ND</td>
<td>8</td>
</tr>
</tbody>
</table>

* not determined.
SUMMARY

Prostasin is a newly identified human serine proteinase which shares high sequence similarity with acrosin, plasma kallikrein and hepsin (Yu et al., 1994; Yu et al., 1995). In the present study, a full-length prostasin gene has been isolated and characterized. A 7 kb prostasin gene fragment has been sequenced, including a 1.4 kb 5'-flanking region, the 4.4 kb prostasin gene and a 1.2 kb 3'-flanking region. The gene has six exons and five introns based on comparison with its cDNA sequence. The sizes of these exons are 417, 18, 163, 272, 167 and 899 bp while those of the introns are 243, 1763, 271, 85 and 92 bp. A number of potential regulatory elements have been revealed in the 5'-flanking region, including an AP2 site, two erythroid-specific promoter elements and a sterol regulatory element. In addition, there is a variant GC box and a variant AP1 site in the promoter region. The transcription initiation site of prostasin has been defined at the G residue and its adjacent A residue in a sequence CTCATGACT, which is similar to an initiator element CTCANTCT. Between the transcription initiation site and these putative regulatory elements, there is a AC-rich repetitive sequence which spans over 300 bp. A genomic Southern blot shows that human prostasin is a single-copy gene. The gene has been localized at human chromosome 16p11.2 by in situ hybridization.
INTRODUCTION

Proteinases are involved in many biological and physiological processes, including blood coagulation, fibrinolysis, activation of proteinases and processing of precursors of protein hormone and growth factors (Neurath, 1986). Recent studies showed that proteinases may play a critical role in programmed cell death (Yuan et al., 1993; Miura et al., 1993) and in triggering the cascade of proteolytic processes during tumor cell invasion and metastasis (Vassalli et al., 1991; Kleiner & Stetler-Stevenson, 1993; Sato et al., 1994). In addition, prostate-specific antigen, a member of the human kallikrein family, has become the most important marker for detecting prostate cancer (Guinan et al., 1987; Stamey et al., 1989).

Prostasin is a newly purified human serine proteinase which has trypsin-like activity (Yu et al., 1994). Recently, its cDNA was isolated and sequenced and posttranslational proteolytic processes of prostasin were established (Yu et al., 1995). The deduced sequence of prostasin indicates a preproenzyme consisting of 343 amino acids, which shows 34-42% identity to human acrosin, plasma kallikrein and hepsin. The preproprostasin is converted to proprostasin by removing a 32-amino acid signal peptide and the generated proprostasin is activated by a cleavage of a trypsin-like proteinase between Arg^{12} and Ile^{13} to give rise to a two-chain form. These two chains are linked together through a disulfide bond. Compared to other serine proteinases, the deduced prostasin has a unique protruding COOH terminus which contains a 19-amino acid hydrophobic portion. Carboxy-terminal microsequencing indicates that the hydrophobic portion has been removed from the secreted prostasin by a cleavage between Arg^{290}.Pro^{291} (Yu et al., 1995). These studies strongly suggest that prostasin is a membrane-anchored serine proteinase.

Prostasin is present in many tissues, including the prostate gland, kidney, liver, colon, pancreas, salivary gland and lung (Yu et al., 1994). Among them, the prostate
gland has the highest level, more than 20-fold higher than any other tissues. It can not be detected in testis, brain, ventricle, atrium, aorta or muscle. Similar result was obtained by reverse transcription-PCR (Yu et al., 1995). These results suggest that tissue-specific transcription factors are involved in regulation of prostasin gene's transcription. As a first step toward understanding how the prostasin gene expression is regulated, we have isolated full-length prostasin gene, characterized its exon-intron structure and revealed potential regulatory elements in the promoter region. We have also defined the transcriptional initiation site and determined its chromosomal localization by *in situ* hybridization.
MATERIALS AND METHODS

Screening of Human Genomic Library: A 255-bp prostasin cDNA fragment was amplified by PCR with primers JY-F1 and JY-B1 (Yu et al., 1995) and inserted into pUC18 plasmid at a Hind II site. The fragment was released from the recombinant plasmid by digestion with Pst I and Xba I and used as a probe to screen a human genomic library after labeling with 32P by nick-translation. The human genomic library in Lambda Fix II was purchased from Stratagene Cloning Systems (La Jolla, CA). Screening of the library was conducted as described by Ma et al. (1994). Two positive plaques were obtained from 24 plates which contained 720,000 plaques and these positives were confirmed in the secondary screening.

Direct Sequencing of Phage DNA: The two positive λ clones were amplified and their DNA was isolated as described by Chai et al. (1991). DNA from one of the two phages was used as template for DNA sequencing using a primer corresponding to prostasin cDNA sequence (Yu et al., 1995) and the BRL Double-Stranded DNA Cycle Sequencing System (Bethesda Research Lab., Bethesda, MD) according to the manufacturer's protocol.

Subcloning of Prostasin Gene: One of the two λ phage DNA was digested with a number of restriction enzymes and Southern blot analysis was performed with two probes which corresponded to the most upstream and the most downstream sequences available at the time. A 9 kb fragment from the Hind II digestion hybridized with the upstream probe whereas a 7 kb fragment was recognized by the downstream probe. The two fragments were then purified from an agarose gel and inserted into pUC18/Hind II. The two constructs were transferred into E. coli JM101 separately and transformants were identified by colony hybridization and confirmed by DNA sequencing.

Sequencing of Prostasin Gene: The two recombinant pUC18 plasmids which contained either the 7 kb or the 9 kb Hind II fragment were isolated as described by
Sambrook et al. (1989) and subjected to the Double-Stranded DNA Cycle Sequencing System. The 5' portion of the prostasin gene was characterized by sequencing the 9 kb fragment whereas the 3' portion was elucidated by sequencing the 7 kb fragment.

**Determination of Transcription Initiation Site:** The total RNA of LNCaP cells, a human prostate carcinoma cell line, was isolated by guanidine-cesium chloride gradient centrifugation (Sambrook et al., 1989). A primer extension assay was conducted as described by Docherty & Clark (1993) with a primer corresponding to the 5' end of the known prostasin cDNA sequence at the time. Ten µg of the total RNA was used as template for a reverse transcription reaction and the product was resolved on an 8% polyacrylamide gel together with a sequencing reaction using genomic DNA as template.

**Genomic Southern Blot Analysis:** Human genomic DNA was isolated from blood cells as described by Sambrook et al. (1989). Five µg of the DNA was digested for 4 h with 10 units of each restriction enzyme at 37 °C except for Bcl I which digested at 50 °C. The digestion products were resolved on a 0.8% agarose gel. After denaturing and neutralizing, the DNA was transferred onto an Immobilon-N membrane (Millipore Corp., Bedford, MA). Prehybridization and hybridization reactions were done as described by Ma et al. (1994). Prostasin cDNA was labeled with 32P by nick-translation and used as a probe for the hybridization.

**Chromosomal Localization:** The 9 kb fragment containing the prostasin gene was released by Hind II digestion from recombinant pUC18 and gel-purified. The fragment was then labeled with biotin-11-dUTP by nick translation and used as a probe for fluorescence in situ hybridization. Slides with metaphase chromosome spreads from lymphocyte cultures were heated at 80 °C for 2 min in 70% formamide and 2 x SSC to denature chromosomal DNA and were immediately dehydrated. The probe was denatured and added to the slides for hybridization at 37 °C overnight. Hybridization conditions and post hybridization washes were done as described by Lichter et al. (1990). The probe was detected with avidin-FITC (fluorescein isothiocyanate). Chromosomes were
counterstained with 4, 6-diamidino-2-phenylindole at 200 ng/ml for 5 min and the slides were washed in 2 x SSC and 0.05% Tween-20 for 10 min prior to mounting in antifade solution. The chromosome spreads were imaged with a Zeiss Axioskop fluorescence microscope equipped with a cooled CCD camera (Photometries) controlled by a Macintosh IIX computer.
RESULTS

Cloning and Sequencing of Human Prostasin Gene: After screening 720,000 plaques from a human genomic library with a human prostasin cDNA probe, two positive λ phages were isolated. The full-length prostasin gene was elucidated by combining the sequences obtained from the 7 kb and the 9 kb fragments produced by Hind II digestion from one of the two λ phage DNA. Figure 14 shows the structure and organization of the human prostasin gene. A stretch of about 7 kb was sequenced, which includes a 1.4 kb 5' flanking region, the 4.4 kb prostasin gene and a 1.2 kb 3' flanking region. Compared with its cDNA, the prostasin gene consists of six exons of 417, 18, 163, 272, 167 and 899 bp, which are represented by boxes, and five introns of 243, 1763, 271, 85 and 92 bp, which are represented by lines (Fig. 14). The first exon contains a 332 bp noncoding region and a 85 bp coding region, which are represented by a shaded box and a solid box, respectively. Exon VI consists a 324 bp coding region and a 575 bp noncoding region. Figure 15 shows the DNA sequence and the deduced amino acid sequences of the prostasin gene. The coding sequence matches the previously reported prostasin amino acid and cDNA sequences (Yu et al., 1994; Yu et al., 1995). The catalytic triad, His53, Asp102 and Ser206, has been located in exons III, IV and VI, as indicated by triangles. The cleavage site for removal of the signal peptide is present in exon II and the activation site through which proprostasin is converted into an active two-chain form is located in exon III as indicated by arrows. The only N-glycosylation site is Asn127, which is indicated by a solid circle, and is located in exon IV.

Analysis of the Promoter Region of Prostasin Gene: A sequence of 1.4 kb 5'-flanking region has been elucidated. A number of potential regulatory elements were revealed in this region by comparing with established consensus sequences, which are double underlined in figure 15. They are two erythroid-specific promoter elements (ESPE) from -387 to -382 and from -341 to -336, an AP2 site from -370 to -363 and a sterol
regulatory element (SRE) from -890 to -897. In addition, a variant GC box, GAGGCGGAGG, is present from -323 to -314 and a variant API binding site, TGTGTCA, is located from -644 to -638. No TATA box could be identified in the promoter region. However, a sequence, CTCATGACT, which is very similar to the initiator element, CTCANTCT, is present from -5 to +4. The initiator element is considered as an alternate TFIID binding site and has similar function to a TATA box. This initiator element is often present in genes that lack a TATA box (Smale et al., 1990). A highly AC-rich sequence spanning more than 300 bp is located between the transcription initiator element and those potential regulatory elements.

**Determination of Transcription Initiation Site:** The transcription initiation site of the prostasin gene has been determined by a primer extension assay. Figure 16 shows that there are multiple bands present in the product of the primer extension assay. The most upstream and strongest stop signals correspond to a C residue and its adjacent T residue which represent G and A residues in the prostasin gene, respectively. These two residues are present in the initiator element, CTCATGACT. This result indicates that transcription of prostasin gene starts from two major sites indicated by arrowheads in figure 16. The fact that the transcription initiation sites of prostasin gene are present within the initiator element is consistent with the study by Smale et al. (1990) that transcription of a gene directed by the initiator element starts from a nucleotide within the initiator element. Thus, based on the result of the primer extension assay, the size of the first exon is 417 bp and the 5' noncoding region of the prostasin gene has 332 nucleotides assuming that transcription starts from the G residue.

**Genomic Southern Blot Analysis:** Analysis of a human genomic Southern blot displays simple restriction patterns after digestion with each of the ten restriction enzymes and probed with the labeled prostasin cDNA (Fig. 17). This result indicates that human prostasin is a single-copy gene.
Chromosomal Localization: The 9 kb fragment of the prostasin gene was labeled with biotin-11-dUTP and used as a probe to hybridize with human metaphase chromosome spreads. Highly specific labeling was observed on both homologs of chromosome 16 in 22 of the 25 spreads examined. As shown in Fig. 18, the prostasin gene is localized on human chromosome 16 in band p11.2, close to the centromere.
DISCUSSION

In this study, we have isolated and characterized the human prostasin gene and determined its chromosomal localization. A sequence of 7 kb has been elucidated, which contains a 1.4 kb 5'-flanking region, the 4.4 kb prostasin gene and a 1.2 kb 3'-flanking region. The coding sequence of human prostasin gene is interrupted by five introns and all exon-intron junctions follow the GT/AG rule (Breathnach & Chambon, 1981). The sequence of exons is consistent with the published cDNA (Yu et al., 1995) and encodes a protein of 343 amino acids. The transcription initiation site has been defined and a number of potential regulatory elements have been revealed in the promoter region. A repetitive sequence, rich in the AC dinucleotide, is present between the transcription initiation site and these putative regulatory elements. Prostasin is a single-copy gene and this gene has been localized at chromosome 16p11.2 by in situ hybridization.

The prostasin gene consists of six exons and the coding sequence is spread over all exons. Consistent with other serine proteinases, each of the three components of the catalytic triad is encoded by sequences in different exons, namely, His$^{53}$ on exon III, Asp$^{102}$ on exon IV and Ser$^{206}$ on exon VI. Comparison of the prostasin gene with the trypsin gene indicates that they have very similar gene structure except that prostasin has one more exon (Craik et al., 1984). Close examination of these two genes indicates that the extra exon of the prostasin gene, namely exon II, seems to originate from an interruption of the second exon of trypsin-like serine proteinases by insertion of an intron during evolution. The consequence of this insertion is the production of an additional small exon which codes for a cleavage site for removal of the signal peptide from preproprostasin. In addition, prostasin has a longer exon VI than its counterpart in trypsin because this exon includes a longer 3'-noncoding sequence and a sequence coding for the unique protruding COOH-terminus of prostasin, which contains a 19-amino acid hydrophobic portion. This hydrophobic portion has been found to be removed from the
secreted prostasin by a cleavage between Arg^{290}-Pro^{291}, which provides strong evidence that prostasin is very likely to be a membrane-bound serine proteinase (Yu et al., 1995).

Mammalian serine proteinases are believed to have derived from a common ancestor through divergent evolution. They are widely dispersed throughout the genome. The chromosomal locus of the prostasin gene has been determined to be 16p11.2 by \textit{in situ} hybridization, close to the centromere (Fig. 18). This finding makes prostasin the first trypsin-like serine proteinase which has been localized on chromosome 16.

Prostasin is present in tissues at various levels and at the highest level in the prostate gland. This suggests that tissue-specific transcription factors may be involved in the regulation of prostasin expression. Evaluation of the 5'-flanking region of the prostasin gene reveals a number of potential regulatory elements. Compared with other serine proteinases, the prostasin gene has a unique promoter structure. No TATA box can be identified in the upstream sequence of the transcription initiation site. Instead, a sequence CTCATGACT, which has high homology with the initiator element, CTCANTCT, in the murine terminal deoxynucleotidyltransferase gene (Smale & Baltimore, 1989), is present from -5 to +4, and overlaps with the transcription initiation site. The initiator element is believed to be present in genes which do not have a TATA box and has similar functions to the TATA box. Namely, it has the ability to independently direct RNA polymerase II to initiate basal transcription from a specific position which is located within this element and the ability to be activated in the absence of a TATA box by an upstream activator element, such as a GC box (Smale et al., 1990). In the prostasin gene, the transcription initiation site starts with G residue or its adjacent A residue within the initiator element (Fig. 16). Furthermore, a variant GC box, which is 9 of 10 nucleotides homologous to the consensus sequence, KRGCGKRRY (Briggs et al., 1986), is found about 300 bp upstream from the initiator element. It is known that the GC box is a Sp1 binding site and Sp1 can activate transcription by interacting with TAF110, a component of the TFIID complex (Hoey et al., 1993). These results indicate that the highly AC-rich repetitive region may form a loop in
order to make Sp1 interact with the TFIID complex during the activation of prostasin gene transcription.

In the 5'-flanking region, there is an AP2 binding site, GCCAGGCC, and a variant AP1 binding site, TGTGTCA (Fig. 15). AP1 and AP2 are ubiquitous factors which respond to phorbol esters and kinase C (Angel et al., 1987; Yen et al., 1991; Mitchell et al., 1987; Imagawa et al., 1987; Leask et al., 1991), indicating that prostasin gene expression may be regulated by phorbol esters and kinase C. It is well known that phorbol esters are potent tumor promoters, which have the ability to potentiate the effect of a subcarcinogenic dose of an initiating carcinogen. An AP1 binding site is present in the promoters of collagenase gene and stromelysin gene (Whitham et al., 1986; Matrisian et al., 1986), and both proteinases play important roles in tumor invasiveness, metastasis and angiogenesis. Whether prostasin has a role in tumor development and tumor cell invasion has yet to be explored. There are also two erythroid-specific promoter elements just upstream the variant GC box. This element can act either as an activator (Martin & Orkin, 1990) or as a repressor (Kudla et al., 1990; Cunningham & Cooper, 1991). The role of these two elements in the regulation of prostasin transcription remains to be established. A negative element, sterol regulatory element GTGGCGTG, is present in the prostasin promoter region (Osborne et al., 1988; Rajavashisth et al., 1989). This element has been found to be involved in repression of gene transcription, suggesting that prostasin gene transcription be regulated by sterol hormones. Taken together, a number of potential regulatory elements have been revealed in the promoter of the prostasin gene. However, whether these elements have effects on the tissue-specific expression of prostasin gene remains to be determined.

In our previous studies, prostasin was purified and characterized and its cDNA was cloned and sequenced. In addition, its posttranslational proteolytic processes have been established. In the present study, the prostasin gene was isolated, sequenced and localized on human chromosome 16p11.2. However, its physiological substrate remains to be
identified and its biological function has yet to be established. Since prostasin has trypsin-like activity and is very likely to be a membrane-anchored serine proteinase, it may play important roles in biological processes occurring on cell surfaces, including processing of protein hormones, growth factors and receptors and activation of precursors of proteinases.
Fig. 14. Structure of human prostasin gene. The prostasin gene consists of six exons and five introns. Solid boxes represent the coding regions of exons whereas the shaded boxes stand for noncoding regions of exons. Lines represent introns and 5' and 3' flanking regions. The positions of the GC box, initiator element and polyadenylation signal are indicated with bars. Triangles indicate positions of codons for the catalytic triad, His$^{53}$, Asp$^{102}$ and Ser$^{206}$. 
Fig. 15. Nucleic acid and amino acid sequence of human prostasin gene. The sequences of exons and 5' and 3' flanking regions are shown with uppercase letters whereas the sequences of introns are represented with lowercase letters. The amino acid sequence of prostasin is denoted by single-letter codes. The arrows indicate the cleavage site to remove the signal peptide and the activation site. Triangles indicate the catalytic triad and the solid circle indicates the N-linked glycosylation site. The star stands for a stop codon. Nucleotide numbering begins with the transcription initiation site, the G residue. Amino acid numbering starts with the first amino acid of the propeptide of prostasin. Putative regulatory elements in the 5' flanking region are double underlined, including a sterol regulatory element (SRE), an API binding site, two erythroid-specific promoter elements (ESPE), an AP2 binding site, a GC box and an initiator element. The polyadenylation signal, which is 12 bases upstream of the poly(A), is underlined.
Fig. 16. Determination of the transcription initiation site. Primer extension assay was conducted with a radiolabeled 21-mer oligonucleotide complementary to positions 149-169. The neighboring wells contain the dideoxy chain-termination ladders of a sequencing reaction using the same primer. The interpretation of the sequence is shown on the left side.
Fig. 17. Genomic Southern blot analysis. Human genomic DNA was digested with 10 restriction enzymes, and generated DNA fragments were resolved on an agarose gel. The DNA fragments were hybridized with the labeled prostasin cDNA after transferred to an Immobilon-N membrane.
Fig. 18. Chromosomal localization. The 9 kb fragment of the prostasin gene was labeled with biotin and used as a probe for fluorescence *in situ* hybridization with human chromosome spreads. Panel A shows the hybridization signals from one of the 25 metaphase spreads examined. Panel B shows a montage of chromosomes hybridized with the prostasin gene and a chromosome 16 ideogram. The prostasin gene is localized at 16p11.2.
V. GENERAL DISCUSSION

Serine proteinases are the most important and best studied family of proteolytic enzymes. They are involved in many biological and physiological processes. Most (if not all) of the enzymes involved in blood coagulation and fibrinolysis are serine proteinases. In this dissertation, I purified a new serine proteinase, designated as prostasin, from human seminal fluid, isolated and characterized its full-length cDNA and gene. I also studied its tissue-specific expression, cellular localization and determined its chromosomal location at 16p11.2 by in situ hybridization.

The purified prostasin is an acidic protein with a pI ranging from 4.5 to 4.8 and a molecular mass of 40 kDa. The catalytic triad of prostasin, His^{53}, Asp^{102} and Ser^{206}, are located on different exons, similar to other serine proteinases. Prostasin displays trypsin-like activity by cleaving synthetic arginine and lysine substrates. This result is consistent with the finding that, in the deduced prostasin from the cDNA, an Asp residue is present at position 200, which is located six residues before the active site Ser^{206}. This Asp residue is located at the bottom of the substrate-binding pocket in trypsin and is involved in an interaction with a Arg or Lys residue at the position of P1 in its substrate (Ruhlmann et al., 1973). In addition, two glycine residues, Gly^{227} and Gly^{237}, are conserved in the deduced prostasin. These two glycine residues are present at the entrance of the substrate-binding pocket in trypsin and permit entry of large amino acid side chains. These features in the primary structure of prostasin foster its trypsin-like activity. Compared with human tissue kallikrein, the kinetic constants, Km, of prostasin toward several synthetic substrates are much higher and the specificity constants, kcat/Km, are much lower, suggesting that the best substrates for prostasin activity have yet to be identified. It is also likely that prostasin poorly cleaves synthetic substrates, but efficiently cleaves specific natural substrates. The
activity of prostasin toward synthetic substrates is inhibited by several trypsin inhibitors. Prostasin has high affinity to aprotinin, a potent inhibitor of human tissue kallikrein. This biochemical property of prostasin provided a simple way for its purification by employing an aprotinin-affinity chromatography.

Prostasin is present at high levels in human prostate gland and seminal fluid. Immunologically reactive prostasin and its mRNA have been localized in prostatic epithelial cells by immunohistochemistry and in situ hybridization. It is believed that prostasin in semen is synthesized in the prostatic epithelial cells and secreted into the prostatic ducts. Prostasin is then injected into seminal fluid together with the prostatic secretion. Human semen coagulates a few minutes after ejaculation and then goes through liquefaction within 30 min. Proteolysis occurs during this liquefaction process. It is known that a variety of proteolytic enzymes are present in seminal fluid, such as PSA, angiotensin converting enzyme, acrosin and tissue kallikrein. Prostasin may play a role in these proteolytic processes due to its arginine and lysine amidase activities, especially in posttranslational proteolytic processes to regulate the activities of other proteinases. Acrosin, which shares high sequence identity with prostasin, is a serine proteinase present in acrosomes (Klemm et al., 1991). Acrosin has important functions in recognition, binding and penetration of the zona pellucida of the ovum during the interaction of the sperm and egg. It is secreted as an inactive zymogen which needs activation by a trypsin-like proteinase to cleave a peptide bond between Arg23-Ile24 (Baba et al., 1989). Considering that prostasin exists at a high level in seminal fluid and possesses trypsin-like activity, it may be involved in the process of acrosin activation. The potential role of prostasin in controlling the maturation of acrosin remains to be explored.

In contrast to the major seminal fluid enzyme, PSA, which is exclusively expressed in the prostate gland (Wang et al., 1979; Morris, 1989), prostasin has been detected in many other tissues, suggesting that its biological roles not be restricted to the prostate gland and seminal fluid. Since it is likely to be a membrane-bound serine proteinase, prostasin
may be involved in some important processes on the surface of cell membranes, such as removal of propeptides from hormones and growth factors and the activation of proenzymes associated with membranes or close to the membrane. In addition, we have detected prostasin at similar levels in both male and female urine. The fact that prostasin is present in female urine and in kidney extracts suggests that prostasin is also synthesized in the kidney and then secreted into the urine. Although the level of prostasin in urine samples is low compared to that in seminal fluid, it would appear that the expression of prostasin is very high in the kidney because of the large volume of urine excretion. In fact, the level of prostasin in urine is similar to that of human tissue kallikrein. Immunohistochemistry demonstrates that in the kidney prostasin is expressed in the proximal and distal tubular cells (data not shown). Biological function of prostasin in these cells remains to be explored.

Prostasin mRNA encodes a protein of 343 amino acids including a 32-amino acid signal peptide. It is secreted as a zymogen which requires a process of maturation to become active. The active prostasin consists of a 12-amino acid light chain and a 299-amino acid heavy chain. These two chains are predicted to be linked together by an interchain disulfide bond between Cys\(^5\) in the light chain and Cys\(^{122}\) in the heavy chain based on comparison with plasma kallikrein, coagulation factor XI and acrosin. The heavy chain of the deduced prostasin shares 34-42% identity with plasma kallikrein, coagulation factor XI, \(\beta\)-tryptase, hepsin, plasminogen and acrosin. A potential glycosylation site, Asn\(^{127}\), has been identified in a sequence of Asn-Ala-Ser. The Asn\(^{127}\) residue could be glycosylated because this residue could not be detected in the amino acid sequencing of tryptic peptide T77 although signals before and after this residue were very clear. This explains why the purified prostasin has an apparently larger molecular weight (40 kDa) than the deduced one (32 kDa). It is very likely that O-linked glycosylation also contributes to the discrepancy since there are many serine and threonine residues in the deduced prostasin (Marshall, 1974). A highly hydrophobic portion which contains 19 amino acids
is present at the COOH-terminus. The length of this segment is sufficient to span the membrane and the two charged residues flanking this segment make it perfect to anchor prostasin in the membrane. Microsequencing at the COOH terminus indicates that this hydrophobic portion is removed from the secreted form of prostasin by a cleavage between Arg<sup>290</sup> and Pro<sup>291</sup>, just upstream of the hydrophobic portion. This result strongly suggests that prostasin is a membrane-anchored serine proteinase and its activity is focused close to the cell surface where it may have important functions. The release of prostasin from the membrane seems to be conducted by a proteinase which has trypsin-like activity. Considering that prostasin displays trypsin-like activity toward synthetic substrates, it raises the possibility that it releases itself from the membrane. Further experiments have to be performed to prove this hypothesis.

The prostasin gene consists of six exons which are interrupted by five introns and the coding sequence is spread over all exons. Consistent with other serine proteinases, each of the three components of the catalytic triad is encoded by sequences in different exons. Prostasin gene structure is similar to the trypsin gene except that the prostasin gene has an extra exon (Craik et al., 1984). The small exon seems to be coming from interruption of the second exon of trypsin-like serine proteinases by insertion of an intron during divergent evolution since it is believed that all serine proteinases share a common ancestor. This insertion results in producing a small additional exon which codes for a cleavage site for removal of the signal peptide during the translocation of the nascent prostasin into the endoplasmic reticulum. In addition, prostasin has a larger exon VI than trypsin, which corresponds to the unique protruding COOH terminus. This protruding portion contains 19 hydrophobic amino acids, which are removed from the secreted prostasin by a cleavage between Arg<sup>290</sup>-Pro<sup>291</sup>. This hydrophobic portion does not seem important in enzymatic activity since secreted prostasin shows trypsin-like activity and other serine proteinases do not have this portion. However it may have an important role in regulating prostasin secretion and focusing prostasin activity close to the membrane.
The prostate gland has the highest level of prostatin, 20-fold higher than other tissues examined in this study, suggesting that cis-acting elements and tissue-specific trans-acting factors are involved in the regulation of the expression of the prostatin gene. Evaluation of the 5'-flanking region of the prostatin gene reveals a unique promoter structure compared with other serine proteinases. A TATA box is present in most genes, which directs basal transcription by interacting with the RNA polymerase II complex. The TATA box can not be identified in the upstream sequence of the transcription initiation site. Instead, an initiator element is identified in the promoter region of the prostatin gene, which is present from -5 to +4, and overlaps with the transcription initiation site. The initiator element is believed to be present in genes which do not have a TATA box and has similar functions to the TATA box. It can interact with RNA polymerase II complex and has the ability to direct basal transcription from a specific position which is located within this element. This element also has the ability to be activated by an upstream activator element, such as a GC box (Smale et al., 1990). In the prostatin gene, the transcription initiation site starts with a G residue or its adjacent A residue within the initiator element. Furthermore, a variant GC box is present in the promoter region. The initiator element and the GC box are separated by a highly AC-repetitive sequence, which spans over 300 bp. It is known that the GC box is a Sp1 binding site and the Sp1 can activate transcription by interacting with TAF110, a component of the TFIID complex (Hoey et al., 1993). These results indicate that the highly AC-repetitive region may form a loop in order to make the Sp1 interact with the TFIID complex during the activation of prostatin gene transcription.

In addition, there is an AP2 binding site and a variant AP1 binding site in the 5'-flanking region of the prostatin gene. AP1 and AP2 are ubiquitous factors which respond to phorbol esters and kinase C, indicating that prostatin gene expression may be regulated by phorbol esters and kinase C. It is well known that phorbol esters are potent tumor promoters and a AP1 binding site is present in the promoters of the collagenase gene and the stromelysin gene (Whitham et al., 1986; Matrisian et al., 1986). Both proteinases play
important roles in tumor invasiveness and metastasis. This phenomenon raises the possibility that prostasin might have a role in tumor development, tumor cell invasion and metastasis, especially considering that it is a putative membrane-anchored serine proteinase. Further experiment are required to prove this hypothesis. In addition, there is a negative element, the sterol regulatory element, present in the prostasin promoter region. This sterol regulatory element has been found to be involved in repression of gene transcription, suggesting that prostasin gene transcription may be regulated by sterol hormones.

Whether the potential regulatory elements present in the promoter region are involved in the regulation of prostasin gene expression remains to be established by further promoter studies. Immunohistochemistry and in situ hybridization experiments demonstrated that prostasin is localized in the mucous cells in the salivary gland and in the proximal and distal tubular cells in the kidney. The localization of prostasin in these tissues is different from tissue kallikrein. Tissue kallikrein is synthesized in the connecting tubule and the cortical collecting duct of the nephron in the kidney (Orstavik et al., 1979; Simson et al., 1979) and in granular convoluted tubules and triated duct cells in the salivary gland (Simson et al., 1979; Schachter et al., 1980; Orstavik et al., 1980). Prostasin can cleave synthetic kallikrein substrates and has similar sensitivity to trypsin inhibitors, but it displays no kininogenase activity toward either high molecular weight kininogen or low molecular weight kininogen. These results indicate that prostasin has different functions from kallikrein in the human body. It will be important to determine trans-acting factors which are involved in the regulation of the cell-specific expression of prostasin and to elucidate its physiological functions in these tissues.

In this dissertation, prostasin has been purified and characterized and its cDNA and gene have been cloned and sequenced. The posttranslational proteolytic processes have been established and the gene has been localized on human chromosome 16p11.2. However, its physiological substrate and biological function have yet to be established. Since prostasin has a trypsin-like activity and is a putative membrane-anchored serine
proteinase, it may play important roles in biological processes occurring on cell surfaces, including processing of protein hormones, growth factors and receptors and in the activation of precursors of proteinases. Only a limited amount of prostasin is present in human semen, therefore expression of recombinant prostasin in \( E. coli \) or mammalian cells seems necessary to obtain a sufficient amount of enzyme for substrate screening. Prostasin activity on natural substrates such as prohormones, growth factors and proteinase precursors can also be conducted by coexpression in mammalian cells. Since prostasin has the highest levels in the prostate gland and in seminal fluid, it might play important roles in the development of the prostate gland and in reproduction. It has long been known that there are various proteolytic activities in the prostate gland and its secretions. Studies have reported that the protein patterns of prostatic secretions from patients with prostatitis are different from those of healthy subjects (Balerna \textit{et al.}, 1982; Tsai \textit{et al.}, 1984), suggesting that an abnormality in proteolysis might occur in those patients. Whether prostasin is involved in this abnormal proteolysis occurring in the prostatic disease remains to be investigated. To explore this matter, evaluation of the difference in prostasin levels between normal healthy subjects and prostatic patients needs to be conducted. This experiment can be conducted by measuring prostasin levels in seminal fluid by either prostasin-specific ELISA or radioimmunoassay, methods that were established in this study. It can also be approached by immunohistochemistry and \textit{in situ} hybridization to detect prostasin expression in specific cells and tissues.
VI. LIST OF REFERENCES


Beaumont, W. (1834) "Experiments and observations on the gastric juice, and the physiology of digestion." Boston, MA.


87


94


VII. Appendix: Experimental Procedures

DFP Labeling

1. To 2-5 µg protein, add 10 µl of 20 µCi diisopropylfluorophosphate [1,3-14C]. The total volume is kept at 30 µl.
2. Incubate at 37 °C for 2 hr, then overnight either at 4 °C or 37 °C.
3. Add 15 µl of 3 x SDS sample buffer containing DTT.
4. Load on a SDS-PAGE with a gradient from 7.5-15%.
5. Stain the gel for one hr.
6. Distain it from 1 hr to overnight.
7. Wash the gel briefly with water and put it into EN3HANCE solution for one hr at RT with gentle shaking.
8. Put it into water for 30 min with gentle shaking with several changes of water.
9. Dry the gel and expose to an X-ray film for several days or longer at -70 °C.
Preparation of Protein for NH2-terminal Sequencing

1. At least 5 μg of protein is required for each lane. After adding sample buffer, incubate at 50 °C for one hr.

2. Load on a mini-gel and electrophoresis is run at 160 V.

3. Wet an Immobilon-P membrane with the highest grade of methonal for a few minutes. Then put the membrane in transferring buffer for a while.

4. Transfer proteins in the SDS-PAGE onto the membrane at just under 0.5 A for one hr. (Note: The current keep going up during the process. Don't let it be above 0.5 A to be overheated. The transferring buffer can be used for 3 times.)

5. Stain the membrane with 0.2% Ponceau S in 3% TCA for 1 min. Then distain with 1 M HoAc with several changes until the background become clear. Wash it twice with water and air-dry it. Thus it is ready for protein sequencing.

Mini-gel transferring buffer:

<table>
<thead>
<tr>
<th>Caps</th>
<th>6.639 g (M.W. 221.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol</td>
<td>300 ml</td>
</tr>
<tr>
<td>Add distilled water to</td>
<td>3,000 ml</td>
</tr>
<tr>
<td>Adjust pH to 11 with NaOH.</td>
<td></td>
</tr>
</tbody>
</table>
ELISA for Prostasin (Avidin-Biotin Enzyme Immunoassay)

1. The IgG of rabbit anti-prostasin is biotinylated with BNHS:
   1) Prepare a solution containing 2 mg of IgG anti-prostasin in 1 ml of 0.1 M sodium bicarbonate buffer (pH 9.5) and dialyze overnight at 4 °C against the same buffer.
   2) Add 10 µl of 0.1 M BNHS (a fresh solution of BNHS in dimethylformamide is prepared each time immediately before use. 0.1 M BNHS solution is made by dissolving 1 mg of biotinyl-N-hydroxysuccinimide ester in 30 µl of dimethylformamide).
   3) Incubate it for 1 hr at R.T.
   4) Dialyze it against PBS buffer at 4 °C, during which the buffer is changed several times.
   5) Add an equal volume of double distilled glycerol for storage.
   6) Store it in refrigerator or freezer.

2. Enzyme Immunoassay for Prostasin
   1) Coat a plate with 100 µl/well of rabbit IgG anti-prostasin in PBS (1 µg/ml). Incubate the plate at 4 °C overnight or longer but no more than 1 week.
   2) Block the plate with 1% BSA in PBS for 1 hr at 37 °C (200 µl/well).
   3) Wash the wells with PBS containing 0.1% Tween 20, three times.
   4) Add 100 µl of prostasin or samples (diluted in PBS containing 0.05% tween 20 and 0.5% gelatin. The solution is called as dilution solution.) and incubate at 37 °C for 90 min.
   5) Wash the wells three times with PBS containing 0.1% of Tween 20.
   6) Add 100 µl of 1 µg/ml biotinylated rabbit antibody anti-prostasin, which is dissolved in dilution solution and incubate the plate at 37 °C for 60 min.
   7) Wash the wells three times with PBS containing 0.1% of Tween 20.
   8) Add 100 µl of 1 µg/ml peroxidase-avidine (1:2500 diluted in dilution solution) at 37 °C for 30 min.
9) Wash the wells five times with the PBS containing 0.1% of Tween 20 and one time with 1xPBS.

10) Add 100 µl of substrate solution to each well, incubate at room temperature and watch the color change. The substrate solution is 0.1 M citric acid containing 0.03% of ABT [2, 2'-azinobis(3-ethylbenzthiazoline sulfonyl acid)], and 3 x 10^{-3} % of H_{2}O_{2}. This solution is prepared by adding 3 mg of ABT and 10 µl of 3% of H_{2}O_{2} to 10 ml of 0.1 M of citric acid (pH 4.37).

The standard is prepared by diluting prostasin from 10 ng/ml to 0.15 ng/ml in dilution buffer (total 7 points). Meanwhile, dilution buffer is used as a negative control. The plate is read several times at interval of 5 - 10 min and find the best one. Usually the reading is best at 30 min.
Immunohistochemistry

1. Human prostate gland is sectioned in 5 μm thickness and the sections are mounted onto gelatin-coated glass slides.

2. Rehydration:
   - Xylene 5 min x 2
   - 100% ETOH 10 min
   - 95 % ETOH 2 min
   - 70% ETOH 2 min
   - 50% ETOH 2 min
   - dH2O 5 min
   - Wash in 50 mM Tris-Cl, pH 7.6 (Buffer A) 5 min
   - 1% H2O2 in Buffer A 20 min
   - Wash in Buffer A 5 min x 3

3. The sections were incubated with normal goat serum (1:100, 1:500, or 1:1000) diluted in Buffer B for 1 hr.

4. The sections are incubated with rabbit anti-prostasin serum (1:100, 1:500, 1:1000 in Buffer B, namely Buffer A with 0.5 % Triton X-100 and 0.7% lambda-carrageena) for several hours to overnight at RT. Wash the sections in Buffer A with gentle shaking, 10 min x 3.

5. The sections are incubated with IgG of goat anti-rabbit IgG F(ab')2 (1:50 in Buffer B) for 1 hour. Wash in Buffer A with gentle shaking, 10 min x 3.

6. The sections are incubated with peroxidase anti-peroxidase complex (rabbit, 1:100 in Buffer B) for 1 hour. Wash in Buffer A with gentle shaking, 15 min x 4.
7. Stain the sections with 0.1% DAB (3,3' diaminobenzidine) with 0.01% H2O2 in Buffer A. Monitor the staining process under microscopy. Wash in buffer A with gentle shaking, 5 min x 3.

8. Wash the sections for 5 min in dH2O. Then dehydrate them by dipping briefly in the following steps:
   - 50% ethanol
   - 70% ethanol
   - 95% ethanol
   - 100% ethanol
   then, 100% ethanol 2 min
   xylene 2 min
   xylene 5 min

Then immediately put a drop of permount on the section and cover it with a upper glass. Stand at 40 °C for quickly drying or at room temperature for longer period of time.

8. A negative control is set up by treating a section with normal rabbit serum instead of the antiserum of prostasin. Other steps are identical.
**In Situ Hybridization**

1. Formalin-fixed and Paraffin-embedded human tissue sections (5 μm in thickness) are deparaffinized and rehydrated as follows:

<table>
<thead>
<tr>
<th>Solution</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xylene</td>
<td>10 min x 2</td>
</tr>
<tr>
<td>100% ETOH</td>
<td>5 min x 2</td>
</tr>
<tr>
<td>95% ETOH</td>
<td>2 min</td>
</tr>
<tr>
<td>70% ETOH</td>
<td>2 min</td>
</tr>
<tr>
<td>50% ETOH</td>
<td>2 min</td>
</tr>
<tr>
<td>DEP-dH2O</td>
<td>2 min</td>
</tr>
</tbody>
</table>

2. Treat the sections with proteinase K (1 μg/ml in 0.1 M Tris-HCl, pH 8.0, 50 mM EDTA. The solution need to be prewarmed before adding proteinase K) at 37 °C for 30 min.

3. Incubate in freshly prepared 0.25% (v/v) acetic anhydride in 0.1 M Triethanolamin-HCl (pH 8.0) for 10 min.

4. Rinse with DEP-dH2O and dehydrated the sections from DEP-dH2O, 50% to 70% to 95% ethanol and dry them by air.

5. Apply each section with hybridization solution, which contains riboprobe at a concentration of 1.5 μg/ml. Cover each tissue section by a piece of parafilm. Put sections in a humid chamber and seal the chamber. Incubate at 42 °C overnight.

6. Wash in 2 x SSC at RT, 2 x 30 min.

7. Wash in 0.2 x SSC at RT, 2 x 30 min.

8. Rinse in buffer 1.

9. Incubate in buffer 1 containing 2% normal sheep serum and 0.3% Triton X-100 for 30 min.
10. Dilute anti-digoxigenin antibody conjugate (Vial 3) by 1:500 with buffer 1 containing 1% normal sheep serum and 0.3% Triton X-100. Apply the solution on each section and incubate at RT for 3 hr.

11. Wash with Buffer 1 for 10 min.

12. Wash with Buffer 2 for 10 min.

13. Apply color reaction mixture on each section by adding 4.5 μl NBT solution (Vial 4), 3.5 μl X-phosphate solution (Vial 5) and 0.24 mg levamisole to 1 ml Buffer 2. This solution should be prepared before use. Incubate in dark overnight.

14. Stop the color reaction by incubating in Buffer 3 for a while.

15. Dehydrate from 50% to 70%, 95% 100%, 100% ethanol and incubate in Xylene, 2 x 5 min. Apply Permount on each section and lay a cover glass on it.

Note:

1. Hybridization solution:
   - 0.6 M NaCl
   - 40 mM Tris-HCl (pH 8.0)
   - 2 x Denhardt's
   - 10 mM EDTA
   - 20% Dextran Sulfate

2. Preparation of DEP-dH2O: Add 0.5 ml DEP in 1 liter dH2O and mix completely.
   Incubate at 37 °C overnight and autoclave it.

3. Buffer 1: 0.1 M Tris-HCl (pH 7.5), 0.15 M NaCl.
   Buffer 2: 0.1 M Tris-HCl (pH 9.5), 0.1 M NaCl, 50 mM MgCl2.
   Buffer 3: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA.

4. All the glass jars need to be baked at 130 °C overnight. All solutions used in this experiment are prepared with DEP-dH2O.
Protein Digestion by Trypsin and Separation by RP-HPLC

A. Digestion of prostasin by trypsin:
   1. 1 nmolar (~40 μg) prostasin in low salt solution (10 mM KPO₄, pH 7.4) is dried by speedvac. Add 50 μl of a solution containing 8 M urea, 0.4 M NH₄HCO₃ to the dry protein. Vortex gently and spin shortly.
   2. Check the pH by spotting 1-2 μl on pH paper. The pH should be between 7.5-8.5.
   3. Add 5 μl of 45 mM dithiothreitol (dissolved in H₂O). Incubate at 50 °C for 15 min under nitrogen.
   4. Add 5 μl 100 mM iodoacetic acid in 0.4 M NH₄HCO₃ (prepared freshly) after cooling to room temperature. Incubate for 30 min at room temperature in dark. Then add 140 μl water.
   5. Add TPCK-treated trypsin (1:100, dissolved in 0.1 mM HCl-10 mM CaCl₂, prepared freshly) in a volume of 5 μl and incubate at 37 °C overnight.
   6. Stop reaction by freezing or injecting directly onto a reverse phase HPLC system.

B. Separation by RP-HPLC

   All of the solutions and buffers have been filtered with 0.20 micron filters.
   Column: 3.9 mm x 30 cm μBONDAPAK C18 column from WATERS
   Buffer A: 0.06% TFA, H₂O
   Buffer B: Methanol
   Buffer C: 0.052% TFA, 80% Acetonitrile, H₂O
   Injection volume = 200 μl
   UV Detection = 215 nm
   Chart speed = 2 mm/min
   Recorder requirement: 1 mV full-scale potentiometric strip-chart
Gradient:

0 - 60 min       2.0 - 38% C
60 - 90 min      38 - 75% C
90 - 105 min     75 - 98% C

1. The column is washed with buffer B.
2. The column is equilibrated with the solution containing 98% A + 2% C, namely the initial gradient solution.
3. Before loading, the loading syringe and loading loop are washed with filtered water several times. The sample is spin to remove pellet. Then load the sample and run the gradient.
4. After completion of the gradient, the column is washed with 100% buffer C, then wash it with buffer B.
5. Reduce the volume of each peak collection to 20 µl or less which is best for sequencing. Don’t let the sample be dry. TFA and acetonitrile don’t interfere with sequencing.
Southern Blot

Denaturing solution: 20 g NaCl and 5 g NaOH in 250 ml dH2O.

Neutralizing solution: 22 g NaCl and 125 ml of 1 M Tris-HCl (pH 8.0) dissolved in dH2O to a total volume of 250 ml.

Prehybridization solution: 50 ml

- 5 x SSPE
- 5 x Denhardt’s
- 0.5% SDS
- 100 μg/ml ssDNA or hsDNA

Then add dH2O to 50 ml.

Denhardt’s solution:

- 1% Bovine serum albumin
- 1% Ficol (400)
- 1% Polyvinyl pyrrolidone

1. Electrophoresize DNA in 0.8% agarose gel containing EB. Put the gel upside down into a steel dish containing some water. Then replace the water with 250 ml denaturing solution and incubate for 30’ with gentle shaking.

2. Pour out the denaturing solution and rinse the gel with dH2O. Then add neutralizing solution (250 ml). Incubate for 30’ with gentle shaking.

3. During the time, rinse a piece of membrane, Immobilon-N, with 95% ethanol (reusable). Wash the membrane with dH2O for 30 sec and then put the membrane in 10 x SSC solution and incubate for up to 2’.

4. Put the gel still upside down on the preequilibrated package of glass, 3 mm Wattman filters and filter bridge. Let the middle of gel touch the filters first, then lay down the two sides of the gel. Remove bubbles between gel and filters with gentle pressing of
hands. Then put the Immobulon-N membrane on the gel. Remove bubbles by peeling up one side of the membrane and then lay it down slowly. Then place three pieces of 3 mm wattman filters (11 x 14 cm), which have been prewetted with 10 x SSC solution, on the membrane. Smooth out bubbles by rolling a test tube on it slowly. Put a stack of paper towels on the filters, cover them with Saran Wrap. Finally, put a bottle on the top and let is stand overnight.

5. Next day, remove the paper towels on the gel. Turn over the gel, the Immobilon-N filter and 3 pieces of 3 mm Wattman filters together and lay them down with the gel side up on a dry sheet of 3 MM Wattman paper. Mark the positions of the gel slots on the filter with a pencil. Peel the gel from the filter and discard it. The filter is then treated by UV crosslinker 2 times, 1200 counts/each, namely 1200 x 100 microjoules/each.

6. Put the filter into prewarmed (at 55 °C) prehybridization solution (50 ml) with DNA side up. Wrap the box with Saran Wrap and incubate at 42-50 °C for 1.5 hours with gentle shaking (around 35 rotations/min).

7. Making probe:

\[
\begin{align*}
25 \text{ pmole primer} & \quad x \ \mu l \\
\text{dH}_2\text{O} & \quad y \ \mu l \\
5 \times \text{ forward buffer} & \quad 5 \ \mu l \\
^{32}\text{P}-\gamma-\text{ATP} & \quad 5 \ \mu l \\
\text{T4 polynucleotide kinase} & \quad 1 \ \mu l \ (10 \ \text{units/ml}) \\
\text{Total volume} & \quad 25 \ \mu l
\end{align*}
\]

Incubate the reaction mixture for 1 hr at 37 °C. Desalt the probe with G-25 column to remove free \(^{32}\text{P}-\gamma-\text{ATP}\). Spot 1 \mu l of the end-labeling reaction mixture on a glass membrane and dry the membrane under infra-red light for 5 min. Then put it on the bottom of a scintillation bottle, add 5 ml scintillation fluid in it, and count the probe.
Add the labeled probe to prehybridization solution and adjust the temperature to 37 °C. Incubate for 4 hr with gentle shaking.

8. Wash the filter with 2 x SSPE solution containing 0.1% SDS three times, 10 min/each at room temperature with gentle shaking. During the time, monitor the radioactivity with Counter. Pick up the filter with a blunt end forcep and drop off water as much as possible. Put the filter on a piece of Saran Wrap with DNA side toward Saran Wrap and wrap it. Then put the wrapped filter on a 3mm Wattman paper and tape it. Make some markers on the paper and expose it to an X-ray film.
Primer Extension Assay

1. End-labeling reaction:
   - Dep-dH2O                     x µl
   - Oligo-Primer                25 pmole
   - 5 x Kinase buffer           5 µl
   - γ-32P-ATP                   5 µl
   - T4 Kinase                   1 µl
   - Total volume               25 µl

   Incubate at 37 °C for 30 min. Add 25 µl TE buffer (pH 8.0) and load on a G-25 column to remove free isotopes.

2. Annealing Reaction:
   - Total RNA from a cell line   10 µg to 20 µg
   - 5 x Annealing buffer         2 µl
   - End-labeling oligoprimer     x µl (2-5 µl)
   - Dep-dH2O                     y µl to final volume of 10 µl

   Mix and incubate at 70 °C for 3 min and allow it slowly cool down to 42 °C and keep it for 3 hr at 42 °C.

3. Primer Extension Reaction:
   - Annealed Reaction Mixture  10 µl
   - 5 x Reverse Transcriptase Buffer 10 µl
   - 0.1 M DTT                   5 µl
   - dNTP (2 mM)                 8 µl
   - RTase                      1 µl
   - Dep-dH2O                   to final volume of 50 µl

   Mix well and incubate at 42-45 °C (depending on the specificity of the reaction) for 1 hr.

   Recover the DNA by ethanol precipitation (or isopropyl alcohol) and rinse the pellet
with cold 70% ethanol, dry the pellet by S.V. Dissolve the pellet in 10 µl dH2O, then add 5 µl sequencing stop solution, mix and is ready for loading gel.

Note: 5 x Annealing Buffer:

125 mM Pipes
2 M NaCl
5 mM EDTA
Adjust the pH to 6.8.
Isolation of Lambda DNA

1. Start with 5 ml of cells grown in LB at a concentration of A550 = 0.15-0.20
2. Pellet the cells at 8,000 rpm for 5 min.
3. Resuspend the cell pellet in 100 µl of NZCYM.
4. Add 10 µl of 1 x 10^8 phage stock and incubate at 37 °C for 20 min.
5. Add the total content to 25 ml of NZCYM in a 250 ml flask.
6. Shake the flask at 37 °C until lysis occur (usually take 4-6 hr).
7. Add 0.5 ml of chloroform to each flask and shake quickly at 37 °C for 15 min (300 rotation/min).
8. Spin the lysates in 50 ml polypropylene centrifuge tubes at 12,000 rpm for 15 min to clear the lysate.
9. RNase A and DNase I are added to a final concentration of 1 µg/ml each and the solution is incubated at 37 °C for one hr or longer/overnight.
10. An equal volume of 20% PEG-8000 and 2.5 M NaCl in H2O is added and the solution is incubated at 0 °C (on ice) for one hr.
11. The precipitate is recovered by centrifugation at 10,000 g for 30 min at 4 °C.
12. The bacteriophage pellet is resuspended in 0.5 ml of TM buffer.
13. Extract with chloroform twice to eliminate the remaining PEG.
14. RNase A and DNase I are added again to a final concentration of 1 µg/ml and the solution is incubated at 37°C for 1 hr or longer/overnight.
15. 5 µl of 10% SDS and 5 µl of 0.5 M EDTA are added and the solution is incubated at 68 °C for 15 min. Add more SDS to make it clear if necessary.
16. RNase A is added once more to a final concentration of 1 µg /ml and the solution is incubated at 37 °C for an hr (optional).
17. The mixture is extracted once with phenol/chloroform, then with chloroform once.
18. An equal volume of isopropanol is added to the final aqueous phase and the solution is incubated at -70 °C for at least 20 min.

19. The precipitate is collected by centrifugation, washed with 70% ethanol, dried under vaccum and dissolved in 50 μl of TE solution or H2O.
Colony Hybridization

1. Pick up potential transformants to a plate of LB agar containing an appropriate antibiotic and grow overnight.

2. Label the membrane (either nitrocellulose or nylon - S&S nitrocellulose, 0.45 μm membrane) with a pencil and slowly lay the membrane on the plate by using forceps.

3. Mark the membrane and the agar plate by making three asymmetric holes with a needle.

4. Allow 5-10 min for colonies to stick to the membrane.

5. Cut some pieces of Whatman 3 MM paper a little larger than the membrane and place them on the bottom of the first three dishes in the following steps.

6. Wet the Whatman paper with the solution in each dish and pour off any excess:
   - Dish 1: 10% SDS
   - Dish 2: Denaturing solution (0.5 M NaOH, 1.5 M NaCl prepared by combining 4 g NaCl, 1 g NaOH, and adding distilled water to a final volume of 50 ml)
   - Dish 3: Neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 8.0, prepared by combining 4.4 g NaCl, 25 ml 1 M Tris-HCl and adding distilled water to a final volume of 50 ml.)
   - Dish 4: 2 x SSPE (200 ml)

7. Using blunt-ended forceps, peel the membrane off the agar plate and lay on the SDS-impregnated whatman paper with colony-side up for 5 min.

8. Treat the membrane through following steps:
   a. Transfer the membrane to the dish 2 and leave for 5 min.
   b. Transfer to dish 3 and leave for 5 min.
   c. Transfer to dish 4 and wash the membrane for a while.

9. Lay the membrane colony side up on a dry sheet of 3 MM paper to let excess liquid dry. Then treat the membrane with cross-linker (2400 count).

10. Rinse the membrane in 50 ml of prewashing solution at 50 ºC for 1 h.
11. Transfer the membrane to pre-hybridization solution which is prepared as follows:
   12.5 ml of 20 X SSPE, 5 ml of 50 X Denhardts solution, 25 ml of H₂O, 0.25 ml of 20% SDS and 0.5 ml of freshly heated hsDNA (10 mg/ml). Add water to a final volume of 50 ml.
12. Incubate the membrane for 1 h at 40 °C with shaking.
13. Prepare End-labelled oligoprobe as follows:
   dH₂O 13 µl
   Oligoprobe (25 mM) 1 µl
   5 X Forward buffer 5 µl
   ³²P gamma ATP 5 µl
   T4 Kinase (20 units/ml) 1 µl
   Incubation the mixture at 37 °C for 1 h. The reaction product is treated with a G-25 column to remove free isotope.
14. Add the probe to the pre-hybridization solution and incubate at 37 °C overnight.
15. Wash the membrane with 6 X SSC at RT. The washing solution is changed several times (the stringency may change depending on the specificity of the probe). After dried on a 3MM Whatman paper, the membrane is exposed to an X-ray film.
Direct DNA Sequencing from Colonies

1. When a colony grows to 1-2 mm in diameter, use a yellow tip to suck the colony, touch the tip on a master plate, resuspend the colony in 20 µl of a lysis buffer (10 mM TrisHCl, pH 7.5; 1 mM EDTA and 50 µg/ml proteinase K) and vortex.

2. Incubate the tubes at 55 °C for 15 min.

3. Incubate at 80 °C for 15 min.

4. Spin the tubes and the supernatant is subjected to dsDNA sequencing system. For each sequencing reaction, 10 µl is enough for DNA sequencing.
RT-PCR Amplification

1. Reverse transcription:
   
   Dep H2O 2 µl
   5 x first strand buffer 4 µl
dNTP (2 mM) 8 µl
Antisense primer (10 µM) 2 µl
DTT (0.1 M) 2 µl
Total RNA from a cell line (1 mg/ml) 1 µl
RTase 1 µl

Total 20 µl

Mix the reaction mixture and spin briefly. Incubate the tube at 37 °C for 1 h. Then heat it at 95 °C for 5 min to stop the reaction.

2. To the above reaction mixture, add following reagents:
   
   dH2O 14.5 µl
   10 x PCR buffer 4 µl
Sense primer (10 µM) 2 µl

Mix well and spin it briefly. Overlay it with 40 µl of light mineral oil and put in PCR machine to run program 7 (94 °C/60 °C/72 °C). When the temperature reach 94 °C, hold the temperature for 2 min. Then add 10 µl of a solution containing following reagents to each tube quickly.

   dH2O 8 µl
   10 x PCR buffer 1 µl
Tag polymerase (5 units/µl) 1 µl

As soon as adding above reagents, release the hold key to resume the PCR program.
3. After finishing the PCR program, add 40 µl of chloroform to each tube and mix well by vortexing. Spin the tubes in microcentrifuge and the PCR product will be present in the upper phase.