Volume 214, number 2, 317–322 FEB 04620 April 1987

Molecular cloning of human prostate specific antigen cDNA

Åke Lundwall and Hans Lilja

Department of Clinical Chemistry, University of Lund, Malmö General Hospital, S-214 01 Malmö, Sweden

Received 12 February 1987

A λgt11 clone encoding prostate specific antigen has been isolated from a human prostate cDNA library. The cDNA insert of 1415 nucleotides hybridizes specifically to a prostate mRNA species of 1.5 kb. The nucleotide sequence codes for part of a signal peptide, a short propiece and the mature protein of 237 amino acid residues. The M_r for the non-glycosylated protein was 26089. One potential site for *N*-linked carbohydrate attachment was identified. The primary structure shows extensive homology with proteases of the kallikrein family.

cDNA clone; cDNA library; Kallikrein-like; (Human prostate)

1. INTRODUCTION

The digestive enzymes trypsin and chymotrypsin are among the best characterized proteins. Their structures have become the model for several related proteolytic enzymes that constitute the superfamily of serine proteases. In this group of proteins are found enzymes of such diverging specificity as collagenase, elastase and the highly specific enzymes of the complement and coagulation systems. Characteristic of the serine proteases are the active-site amino acid residues His, Asp and Ser. The substrate specificity is mainly determined by the unique microenvironment provided by the amino acid residues lining the enzymes' catalytic pockets.

The glandular kallikreins constitute a subfamily of the serine proteases, originally identified by the ability of the proteases to release kinins from high-M_r substrates (review [1]). More recently, proteases such as the epidermal growth factor-binding protein [2] and the γ-subunit of the 7 S nerve growth factor complex [3] have been included in the category due to their high degree of sequence homology to pancreatic kallikrein. The glandular kallikreins have been most intensively studied in the mouse, where 25–30 genes are concentrated to a locus on chromosome 7 [4]. There are also several kallikrein genes in the rat [5]. Recently, by means of low-stringency hybridization to a pancreatic kallikrein cDNA probe, it was shown that the human genome also contains several kallikrein genes [6].

Human prostate specific antigen (PSA) is a kallikrein-like protease [7] present in seminal plasma at a concentration of 0.7 mg/ml [8]. It is a single-chain glycoprotein of M_r ~33000 [8]. Although the biological function of the enzyme has yet to be fully determined, recent experiments have shown it to be involved in the liquefaction of seminal coagulum by proteolysis of its predominant protein component – a high-M_r seminal vesicle protein (HMW-SV-protein) or semenogelin [7,9]. The primary structure of PSA was recently determined, confirming its likeness to kallikrein [10]. Here we present the molecular cloning of a PSA encoding cDNA.

Correspondence address: Å. Lundwall, Department of Clinical Chemistry, University of Lund, Malmö General Hospital, S-214 01 Malmö, Sweden

Published by Elsevier Science Publishers B.V. (Biomedical Division)

2. MATERIALS AND METHODS

2.1. *Construction of the cDNA library*

Human prostate tissue was obtained from patients undergoing surgical treatment for benign prostatic hyperplasia. RNA was purified from the prostate tissue by the guanidine thiocyanate method of Chirgwin et al. [11]. The mRNA fraction was recovered by two passages of total RNA on oligo(dT)-cellulose (Pharmacia).

Single-stranded cDNA was synthesized from 10 μg mRNA with 75 units AMV reverse transcriptase (Boehringer-Mannheim), following the protocol of Maniatis et al. [12]. Second strand synthesis was done with DNA polymerase (Boehringer-Mannheim), RNase H (Pharmacia) and *E. coli* DNA ligase (New England Biolabs), according to Gubler and Hoffman [13]. Excess nucleotides were removed by ethanol precipitation from 2 M ammonium acetate. The recovered ds-cDNA (520 ng) was methylated with 28 units *Eco*RI methylase (New England Biolab) in 80 μM *S*-adenosylmethionine (Sigma) containing buffer (0.1 M Tris-HCl, 10 mM EDTA, pH 8). Linkers were added to the cDNA by overnight incubation with 1 μg *Eco*RI linker (Pharmacia) and 5 units T_4 DNA ligase (Amersham) at 12°C. Following phenol/chloroform and chloroform extractions and precipitation from 2 M ammonium acetate, the resolubilized cDNA was digested with 250 units *Eco*RI for 3 h. The digested material was size-selected on a 0.6 × 21 cm column of Sepharose CL-4B, run in 10 mM Tris-HCl, pH 8.0, 0.3 M NaCl and 1 mM EDTA. The material from the break-through fraction, approx. 140 ng cDNA, was ligated into 14 μg *Eco*RI-digested and phosphatased λgt11 DNA. One-third of the ligated material was packed in vitro using a commercial extract (Promega Biotecniques).

2.2. *Isolation and analysis of cDNA clones*

A rabbit antiserum raised against PSA purified from pooled human seminal plasma has been described previously [7]. Antibodies were affinity-purified on a column with PSA immobilized to Sepharose 4B, following standard methods. Recombinant phages were screened with the affinity-purified antibodies on duplicate nitrocellulose filters, using a modified procedure of Young and Davies [14,15]. Immunoreacted antibodies were detected with a commercial alkaline phosphatase-conjugated second antibody (Promega Biotecniques), following the manufacturer's recommendations. Recombinant phages were prepared by the plate lysate method and subsequent banding in a CsCl gradient. Approx. 400 μg phage DNA, readily digestable by restriction enzymes, was obtained from five 'large' (13 cm diameter) plates. The cDNA inserts were subcloned into the *Eco*RI site of the plasmid vector, pUC18. Plasmid preparations were done by the alkaline lysis method in [12]. Nucleotide sequences were determined with the modified dideoxy chain terminator method of Biggin et al. [16] on subclones in bacteriophage M13 mp8, generated from sonicated material as described by Bankier and Barrell [17]. The nucleotide sequences were computed with programmes described by Staden [18,19].

2.3. *Miscellaneous*

RNA was analysed by electrophoresis in 1.3% agarose gels containing formamide and blotted onto nitrocellulose paper as described [20,21]. Nick translation was done with a commercial kit from Amersham, and following the manufacturer's recommendations. Hybridizations were done with 5×10^6 cpm nick-translated DNA per ml at 42°C for 16 h in 50 mM sodium phosphate, pH 7.4, 0.75 M NaCl, 0.1% SDS, 0.05% (w/v) sodium pyrophosphate and 100 μg/ml of herring sperm DNA. Excess radioactivity was washed away by several changes of 30 mM sodium citrate, pH 7.0, containing 0.3 M NaCl and 0.05% (w/v) pyrophosphate at room temperature, and then in 7.5 mM sodium citrate and 75 mM NaCl at 68°C for 0.5 h. The filters were exposed to Kodak X-AR-5 film between enhancer screens at −70°C.

3. RESULTS

3.1. *Isolation of the cDNA clone*

A human prostate cDNA library with 3×10^5 members was constructed in bacteriophage λgt11. Affinity-purified antibodies to human PSA were used to screen 50000 recombinants of the library by a modified Young and Davies procedure. One clone producing recombinant PSA was identified. The clone, designated λHPSA-1, was isolated and the DNA analysed. By agarose gel electrophoresis

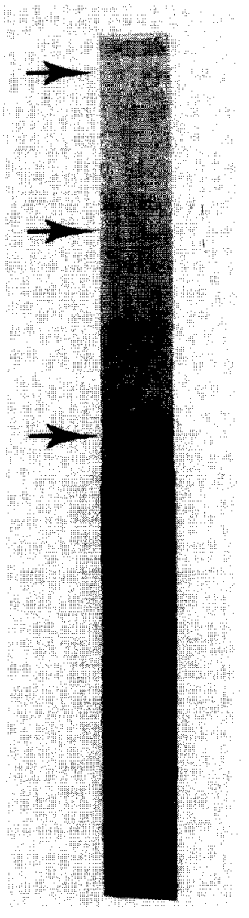

Fig.1. Northern blot. 20 μg prostate RNA was electrophoresed in a 20 cm agarose gel containing formaldehyde. The material was blotted onto nitrocellulose filter and probed with nick-translated cDNA from λHPSA-1. The arrows indicate the origin of electrophoresis and the positions of 28 S and 18 S RNA.

of an *Eco*RI digest, the size of the cDNA insert was estimated to be approx. 1.5 kb. A labelled probe was prepared by nick translation of the cDNA. Upon Northern blot analysis of total prostate RNA, this probe hybridized specifically to a

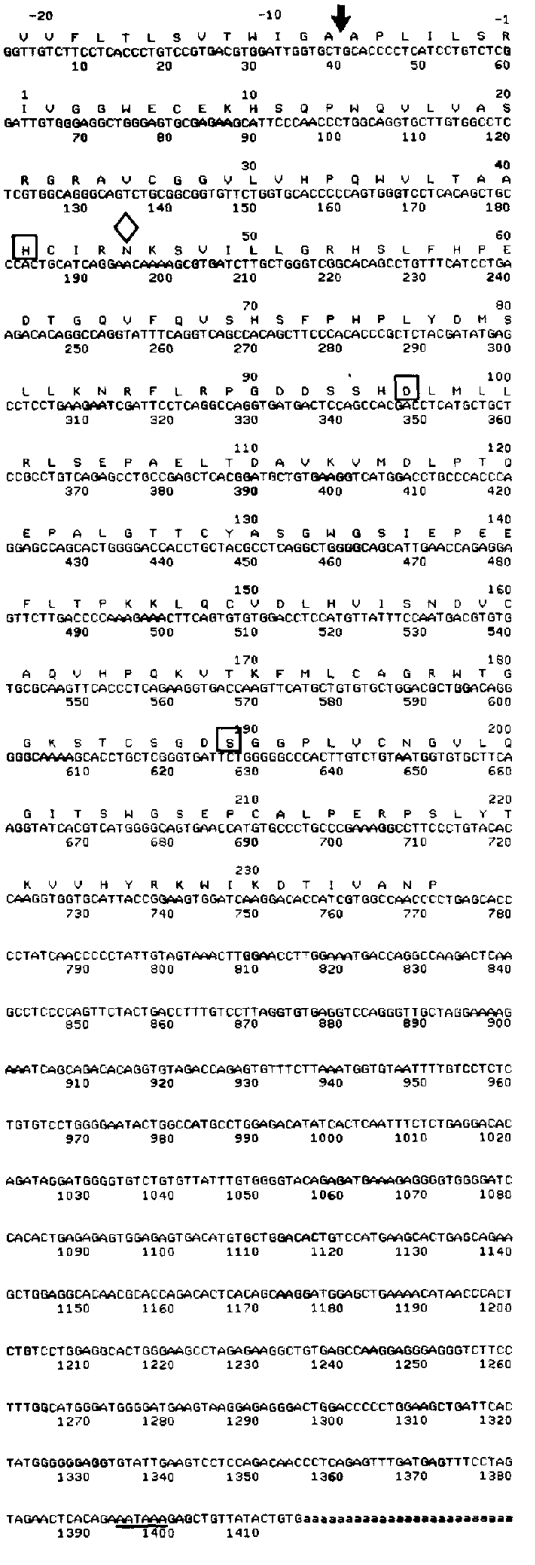

Fig.2. Nucleotide and derived amino acid sequence of a cDNA encoding human prostate specific antigen. Amino acid numbering starts with residue 1 of the mature protein. The arrow denotes the suggested cleavage site for signal peptidase. The active site amino acid residues are boxed. Position for possible *N*-linked carbohydrate attachment is indicated by the lozenge. The polyadenylation signal is underlined.

message of approx. 1.5 kb (fig.1). It was therefore concluded that λHPSA-1 carried an almost full-length cDNA insert. The labelled fragment was also used to rescreen 16000 recombinants of the library. 36 of these hybridized to the probe, and thus approx. 0.2% of the prostate mRNA encodes PSA.

3.2. *Human prostate specific antigen cDNA sequence*

The structure of the cDNA insert in λHPSA-1 was determined by sequencing 48 random subclones in M13 mp8. Each nucleotide was determined on average 3.96 times. The sequence of the insert is shown in fig.2. The length is 1415 nucleotides plus a poly(A) tail of undetermined size. An open reading frame encodes 257 amino acid residues. The amino acid sequence of PSA is coded for by nucleotides starting at position 62 to a stop codon, TGA, at position 773–775. The active-site amino acid residues were identified, confirming the serine esterase nature of the protein (fig.2). One potential site for *N*-glycosylation was identified at amino acid residue 45. The polypeptide chain of PSA consists of 237 amino acid residues and the M_r of the unglycosylated protein is 26089. In the 5′-end, the cDNA codes for a signal peptide and a short propiece. These residues were identified by their homology with other members of the kallikrein family and with trypsinogen [2–4,6,22]. At the 3′-side of the stop codon are 640 non-translated nucleotides, followed by a poly(A) tail. A polyadenylation signal, AATAAA, is located 16 nucleotides upstream from the poly(A). Interestingly, there is also a putative polyadenylation signal, AGTAAA, located at position 800–806. This is the same signal and the same location as for the polyadenylation signal of human pancreatic kallikrein [6]. However, with Northern blot analysis we were unable to detect an mRNA of a size around 1.0 kb, so the proximal polyadenylation signal is probably not used.

4. DISCUSSION

PSA has been identified as one of the predominant proteins secreted by the human prostate gland [9]. The abundance of messages encoding PSA in prostatic mRNA is consistent with this finding. It has previously been reported that PSA is strongly homologous to members of the glandular kallikrein family [7,10]. Comparison of the PSA amino acid sequence with that of other serine proteases confirms this. The highest degree of homology is obtained with human pancreatic kallikrein (62%), γ-NGF (55%) and EGF-BP (53%), while the figures for bovine trypsin and chymotrypsin are 41 and 36%. That it is a member of the glandular kallikrein family is further confirmed by the finding that PSA is synthesized with a leader peptide that is highly conserved among kallikreins. It can in fact be predicted that the mRNA encodes an initiator methionine plus three more hydrophobic amino acid residues that are not represented in the cDNA clone.

Trypsinogen is converted to β-trypsin by a specific cleavage of the peptide bond between Lys_6 and Ile_7 [23]. A similar zymogen activation also takes place with the kallikreins. Urinary prokallikrein is activated to kallikrein through specific cleavage of the peptide bond between Arg_7 and Ile_8 [24,25]. We suggest that PSA is converted from an inactive proform to an active enzyme by a similar mechanism. From the homology with human urinary prokallikrein [25], we also think it likely that the amino-terminal amino acid residue of Pro-PSA is Ala_{-7}.

Several of the glandular kallikreins cleave at the carboxy-terminal side of a Lys or Arg residue. That is because they have an Asp residue at the bottom of the catalytic pocket, capable of attracting the positively charged side chains electrostatically [26]. In PSA, this position is occupied by Ser_{183} which is equivalent to Ser_{189} in chymotrypsinogen. The Gly and Ser residues in positions 208 and 219 have small side chains, and thus allow access to the catalytic pocket of amino acid residues with bulky side chains [26]. It was reported recently from studies with synthetic substrates that PSA had no detectable activity against N^a-benzoyl-L-tyrosine ethyl ester, but instead had activity against Arg/Lys containing substrates [7]. The activity was low with all substrates tested and did not exclude a restricted chymotryptic activity. Indeed, elsewhere it has been reported that PSA displays low but detectable activity towards chymotryptic substrates [10]. The predominant protein in the seminal gel structure (semenogelin) is cleaved by PSA during liquefac-

tion of semen. Peptides that are degradation products of semenogelin have been isolated from liquefied semen, and the primary structure has been determined for one fragment [27]. The carboxy-terminal amino acid residue was reported to be tyrosine, a residue that is typically recognized by chymotrypsin.

Recently the primary structure of human PSA was determined by Watt et al. [10], using protein chemical methods. The sequence presented here is deduced from a cDNA clone. Therefore, the cDNA might express a rare genetic variant, while only the commonest variants are to be expected in the protein from pooled seminal plasma. The difference at position 70, where the cDNA encodes His but Watt et al. report Thr, might thus be explained by point mutation. There is another disagreement at amino acid residue 151, where Watt et al. report Gln; the cDNA encodes Asp at this position, and Asp was also identified at this position by Edman degradation of a proteolytic fragment present in PSA preparations from seminal plasma. The amino-terminal sequence of the fragment was Lys, Leu, Gln, Cys, Val, Asp, Leu, His (Lilja, H., unpublished). However, the differences starting at position 141 where the cDNA encodes Phe, Leu, Thr and Pro, while Watt et al. report His, Leu, Leu, Tyr, Asp, Gln and Met, are probably too extensive to be explained by point mutations. Although the sequence difference might reflect an allelic variation, we find it more likely that it is caused by either two non-allelic variants, as has been reported for trypsinogen [22] or the occurrence of two forms of PSA due to alternative splicing of a single primary transcript. These questions will probably be answered by further study of the gene structure and biosynthesis of PSA.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Medical Research Council (project B85-03X-04487-11C and project B87-13X-07903-01A), the Faculty of Medicine, University of Lund and the Swedish Society of Medicine.

REFERENCES

[1] Schachter, M. (1980) Pharmacol. Rev. 31, 1–17.
[2] Lundgren, S., Ronne, H., Rask, L. and Peterson, P.A. (1984) J. Biol. Chem. 259, 7780–7784.
[3] Thomas, K.A., Baglan, N.C. and Bradshaw (1981) J. Biol. Chem. 256, 9156–9166.
[4] Mason, A.J., Evans, B.A., Cox, D.R., Shine, J. and Richards, R.I. (1983) Nature 303, 300–307.
[5] Ashley, P.L. and MacDonald, R.J. (1985) Biochemistry 24, 4520–4527.
[6] Fukushima, D., Kitamura, N. and Nakanishi, S. (1985) Biochemistry 24, 8037–8043.
[7] Lilja, H. (1985) J. Clin. Invest. 76, 1899–1903.
[8] Wang, M.C., Kuriyama, M., Watt, K.W.K., Loor, R. and Chu, T.M. (1982) Methods Cancer Res. 19, 179–197.
[9] Lilja, H., Oldbring, J., Rannevik, G. and Laurell, C.-B. (1987) J. Clin. Invest., submitted.
[10] Watt, K.W.K., Lee, P.-J., M'Timkulu, T., Chan, W.-P. and Loor, R. (1986) Proc. Natl. Acad. Sci. USA 83, 3166–3170.
[11] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[12] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
[13] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
[14] Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194–1198.
[15] Young, R.A. and Davis, R.W. (1983) Science 222, 778–782.
[16] Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.
[17] Bankier, A.T. and Barrell, B.G. (1983) Techniques in Nucleic Acid Biochemistry, vol.85 (Flavell, R.A. ed.) pp.1–73, Elsevier, Limerick.
[18] Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961.
[19] Staden, R. (1982) Nucleic Acids Res. 10, 4731–4751.
[20] Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) Biochemistry 16, 4743–4751.
[21] Goldberg, D.A. (1980) Proc. Natl. Acad. Sci. USA 77, 5794.
[22] MacDonald, R.J., Stary, S.J and Swift, G.H. (1982) J. Biol. Chem. 257, 9724–9732.
[23] Davie, E.W. and Neurath, H. (1955) J. Biol. Chem. 212, 515–529.

[24] Takada, Y., Skidgel, R.A. and Erdös, E.G. (1985) Biochem. J. 232, 851–858.
[25] Takahashi, S., Irie, A., Katayama, Y., Ito, K. and Miyake, Y. (1986) J. Biochem. 99, 989–992.
[26] Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G. and Bartunik, H. (1983) J. Mol. Biol. 164, 237–282.
[27] Lilja, H. and Jeppson, J.-O. (1985) FEBS Lett. 182, 181–184.