Novel splice variants of prostate-specific antigen and applications in diagnosis of prostate cancer

Georgios Pampalakis a, Andreas Scorilas b, Georgia Sotiropoulou a,⁎

a Department of Pharmacy, University of Patras, Rion-Patras 26500, Greece
b Department of Biology, University of Athens, Athens, Greece

Received 30 August 2007; received in revised form 6 December 2007; accepted 29 December 2007
Available online 11 January 2008

Abstract

Objectives: We aimed to identify novel splice variants of prostate-specific antigen/or human kallikrein 3 (PSA/KLK3), the most widely used serum biomarker for case-finding, screening and monitoring of prostate cancer.

Design and methods: The full-length sequences of splice variants were assembled as contigs from human ESTs that displayed homology to the cDNA sequence encoding PSA. Expression of variants in clinical samples was analyzed by semi-quantitative RT-PCR.

Results: EST database mining led to the identification of seven previously unidentified splice variants encoding PSA-like proteins that are predicted to contain epitope sequences recognized by PSA-specific antibodies, therefore, expression of these isoforms may affect the amount of total PSA measured by established immunoassays. Analysis of the differential expression profile of isoform PSA-SV5 in patients with benign prostate hyperplasia and prostate cancer showed that it is specifically expressed in prostate cancers.

Conclusions: A novel splice variant of PSA was identified, PSA-SV5, that may be exploited in clinical diagnosis to distinguish prostate cancer from benign prostate hyperplasia.

© 2008 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.

Keywords: Prostate-specific antigen; Splice variants; Isoforms; Differential diagnosis; Benign prostate hyperplasia; Prostate cancer; Biomarkers

Introduction

Human KLK3/PSA (KLK3/PSA for gene; KLK3 or PSA for protein) is a member of the tissue kallikrein family [1], and the first tumor marker approved by the Food and Drug Administration as an aid in case-finding and monitoring of prostate cancer in population screening. Using PSA tests, many prostate cancers can be detected when the tumor is clinically localized and potentially curable. The clinical utility of PSA rests on its prostate tissue-specific expression, and the fact that it is a circulating biomarker appropriate for monitoring therapeutic response and relapse [2]. In normal prostate, PSA is highly expressed by luminal epithelial cells and only minor amounts of PSA leak into the bloodstream. However, prostate cancer is usually associated with an increase in the serum concentration of PSA (>3–4ng/mL), due to increased numbers of PSA-secreting circulating cancer cells and destruction of prostatic architecture that results in leakage of PSA into blood circulation [2]. In addition, elevations in PSA are sometimes indicative of benign prostatic hyperplasia (BPH), prostatic inflammation, ejaculation or urologic manipulations [3]. To overcome the problem that elevated PSA is not a specific marker of malignancy, various expedients were applied, such as PSA density, PSA velocity, ratios of free/total PSA [2].

In this study, EST sequences available in public databases were analyzed in silico with the aim to identify all transcripts encoded by the KLK3/PSA gene. Seven previously unknown splice variants of classical KLK3/PSA transcript that all encode for putative PSA-like proteins with homology at their N-terminal

Abbreviations: PSA, prostate-specific antigen; PSA-SV, splice variant of PSA; PSA-RP, PSA-related protein; BPH, benign prostate hyperplasia; KLK, kallikrein gene; SV, splice variant; EST, expressed sequence tag; ORF, open reading frame; UTR, untranslated region.

The sequence of PSA-SV5 was deposited in GenBank™ with accession number DQ641251.

⁎ Corresponding author. Fax: +30 2610 969940.
E-mail address: gsotiro@upatras.gr (G. Sotiropoulou).
domains, were identified and analyzed. Interestingly, five of these splice variants displayed prostate-specific expression. These PSA isoforms contain epitope sequences known to be recognized by PSA-specific antibodies. Therefore, it is possible that these isoforms interfere with immunoassays currently used for the determination of PSA levels in clinical samples [4], and should be taken into account for the development of improved PSA-based methods applied for the molecular diagnosis of prostate cancer. More importantly, we identified experimentally a novel splice variant of PSA, named PSA-SV5, that was shown to be specifically expressed in prostate cancers, but not in patients with BPH. In addition to classical PSA, PSA-SV5 isoform could represent a useful biomarker that distinguishes PSA-expressing BPHs from prostate cancers.

Materials and methods

Database search

ESTs displaying homology with the cDNA encoding for the classical form of PSA were obtained from the EST database at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). Information on the KLK3/PSA gene was obtained using Map Viewer (http://www.ncbi.nlm.nih.gov/mapview). Model Maker was applied to predict gene structure including intron–exon junctions. Using the tool “expand ESTs” clones that display high similarity to KLK3/PSA were identified and graphically depicted. Homologous EST clones characterized by differences in predicted gene structures were obtained and aligned with the PSA/KLK3 gene sequence that was retrieved from the tissue kallikrein gene family locus (AF243527). EST clones spanning different introns of PSA/KLK3 were excluded from this analysis, since they may originate from genomic DNA contamination. We focused on long EST clones (>300bp) with high homology (>95%) to PSA/KLK3, in order to exclude EST clones derived from genes other than PSA/KLK3. More specifically, homology at the central region of ESTs was approximately 99–100%, while some differences were allowed at both ends that usually reflect sequencing errors [5]. The high homology filter excludes EST clones derived from paralogous genes, as for example the human KLK2 gene that is 80% identical with PSA/KLK3 [1].

Cell lines

LNCaP and PC-3 were obtained from the American Type Culture Collection (Rockville, MD) and maintained in RPMI supplemented with 10% FBS (Gibco BRL, Gaithersburg, MD).

Population study and prostatic tissues

Included in this study were prostate tissues from 16 patients who had undergone transrectal ultrasound guided needle biopsy. Patient ages ranged from 54 to 80 years with a mean of 66 and a median of 64 years. After biopsy, patients were subdivided into two groups on the basis of their histopathological diagnosis: a) patients with benign prostatic hyperplasia (BPH, 7 males) and b) patients with prostate cancer (9 patients with disease stages

![Fig. 1. Schematic representation of known isoforms encoded by the human KLK3/PSA gene. Gray and white boxes represent coding and untranslated exons, respectively. Numbers in boxes indicate lengths of exons in nt. A ? indicates that the full-length sequence was not determined. H, D, S denote the catalytic amino acid residues of PSA. Genbank™ accession numbers are given on the left side.](image-url)
T1c–T2a). Tissue samples from which RNAs were isolated for this experiment were verified to consist of either only cancer cells or no cancerous tissue (BPH). All specimens were residuals from routine testing and were stored frozen in liquid nitrogen until analyzed. Investigations were carried out in accordance with the ethical standards of the Helsinki Declaration of 1975, as revised in 1983.

**RNA extraction and RT-PCR**

RNA was extracted from cell lines with RNeasy (Qiagen, Valencia, CA). For RNA extraction from clinical samples, tissue specimens were minced with a small scalpel on dry-ice, immediately transferred to 500 μL polypropylene tubes and homogenized. Total RNA was extracted using the TRI total RNA isolation system (Ambion Inc, TX), according to the manufacturer’s instructions. RNAs were treated with DNase. The concentration and purity of RNA were determined spectrophotometrically. Total RNA (1 μg) was reverse-transcribed into first strand cDNA in a 10 μL reaction using the Takara RNA pre-amplification system (Takara Bio Inc, Shiga, Japan). RNA integrity and equal loading were confirmed based on amplification of the housekeeping gene β-actin. Subsequently, 1 μL of the cDNA template was PCR-amplified in a 100 μL reaction containing 2.5 pmol of each primer, 200 μM dNTPs, and 2.5 U of Taq polymerase (New England Biolabs, Beverly, MA). Cycling conditions were as follows: 95°C for 5 min, followed by 28 cycles for PSA or 40 cycles for PSA-SV5 of 95°C for 1 min, 62°C for 1 min, 72°C for 1 min, and a final extension step at 72°C for 10 min. Primers for KLK3/PSA were: 5′-CAC CGG AA TCG CTG TGT CAC CAT GTG GG-3′ (PSA-S) and 5′-TGA CTC GAG TGC TCA GGG GTT GGC CAC GA-3′ (PSA-AS), and for PSA-SV5: 5′-GGA CCC TGG GGA GCG AAG TGG AGG ATA-3′ (IN2S) and PSA-AS. Equal volumes of PCR reaction products were resolved on 1.5% agarose gels and visualized by ethidium bromide staining. The product that corresponds to splice variant PSA-SV5 (~1200 bp) was extracted and sequenced (Lark Technologies Inc., Essex, UK). The

---

Fig. 2. Novel isoforms encoded by the human KLK3/PSA gene. A. The structure of seven EST clones representing novel splice variants of KLK3/PSA is shown schematically. PSA-SV1, SV2, SV3 and SV4 contain a complete ORF that encodes for truncated isoforms of classical PSA; the remaining three variants represent partial sequences of presently unknown PSA isoforms. B. Partial structure of four EST clones indicating the presence of cryptic introns inside exon V of the PSA/KLK3 gene. C. EST clones that contain partial sequences of PSA/KLK3 and KLK2 genes. All other symbols as indicated in Fig. 1.
sequence of PSA-SV5 was determined and found identical in RNAs from LNCaP cells, one tissue specimen from BPH and one from prostate cancer.

**Results**

**Identification of novel splice variants of PSA**

We analyzed in silico expressed sequences deposited in EST databases [5] with the aim to identify unknown splice variants of PSA. Here, splice variants refer to gene isoforms. Conventionally, however, PSA isoforms refer to different forms of PSA present in the bloodstream (complexed PSA, pro-PSA, mature/active PSA) or to differentially glycosylated PSA molecules [6,7]. Fig. 1 summarizes all known PSA isoforms produced by alternative splicing [8–11]. Analysis of EST sequences that display high homology to PSA and contain a complete ORF resulted in the identification of four previously unknown transcripts, i.e. PSA splice variants 1–4 (PSA-SV1–4) shown in Fig. 2.

Splice variant 1 (PSA-SV1) is represented by EST clone BF965220 that was derived from a library prepared from skeletal muscle and enriched for full-length cDNAs. PSA-SV1 is characterized by a cryptic intron inside exon III, which results in a significantly shorter exon III of 42nt, and an internal splice junction GC...AG that represents a relatively rare type of splicing event, but was previously also implicated in splicing out intron III of the KLK10/NES1 gene [12,13]. Lack of a large part of exon III causes a frameshift and results in a new ORF encoding for a putative protein of 78aa, with homology to the N-terminal domain of classical PSA. Splice variant 2 (PSA-SV2) is represented by EST clone CA488071 that was derived from breast cancer cell lines ZR-75-1, MCF-7, MDA-MB-231 and SK-BR-3 and prostate cancer cell line LNCaP. PSA-SV2 variant contains conserved splice junctions (GT...AG) but lacks part of exon III, while it uses an alternative exon V. The inferred protein sequence consists of 199aa that differs from classical PSA in the last exon of KLK3/PSA, as depicted in Fig. 2. Interestingly, two bicistronic RNAs are expressed, corresponding to two bicistronic RNAs that contained partial sequences corresponding to the last exon of KLK3/PSA, as depicted in Fig. 2. The stretch of six adenine residues is likely to represent part of a polyA tail, downstream of a non-conserved polyadenylation signal.

In addition to PSA-SV1–4, we identified three partial cDNA sequences, which based on high (> 95%) nucleotide sequence homologies, represent PSA isoforms, shown schematically in Fig. 2. Two identical EST clones (BX111160, AI669704) were derived from a cDNA library prepared from prostate tissue. These novel transcripts skip exon IV, but retain part of exon III and a polyadenylation signal AATAAA, followed by a polyA tail located 25nt downstream. They utilize a non-conserved splice junction AA...AG [12]. EST clone AI685510 skips part of exon III, the entire exon IV and a large part of exon V, but contains the same polyadenylation signal as the classical KLK3/PSA transcript. The splice junction between the first part of exon III and the last part of exon V, is AC...GC. The 5' sequence and the full-length ORF of this transcript variant were not determined. EST clone AA522842 contains conserved splice junctions and a different 3' sequence that extends further 3' downstream as compared to the classical KLK3/PSA transcript. The stretch of six adenine residues is likely to represent part of a polyA tail, downstream of a non-conserved polyadenylation signal. The utilization of unfavorable splice junctions in some of these splice variants, probably explains their observed low abundance (Table 1). In addition, four EST clones were identified that were derived from cryptic intronic sites within the last exon of KLK3/PSA, as depicted in Fig. 2. Interestingly, two bicistronic RNAs that contained partial sequences corresponding to KLK3/PSA and KLK2 genes were derived from ESTs isolated from normal prostate tissue. Bicistronic RNAs are extremely rare in mammalian systems, as for example in GDF-1.

**Table 1**

<table>
<thead>
<tr>
<th>Genbank™ accession number</th>
<th>Tissue</th>
<th>Length (nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF965220 (PSA-SV1)</td>
<td>Skeletal muscle</td>
<td>866</td>
</tr>
<tr>
<td>CA488071 (PSA-SV2)</td>
<td>Breast and prostate cancer cell lines</td>
<td>928</td>
</tr>
<tr>
<td>BF673243 (PSA-SV3)</td>
<td>Prostate</td>
<td>849</td>
</tr>
<tr>
<td>BE840701 (PSA-SV4)</td>
<td>Prostate</td>
<td>552</td>
</tr>
<tr>
<td>AA177005 (PSA-SV4, partial ORF)</td>
<td>Prostatic intraepithelial neoplasia 2 (PIN2)</td>
<td>435</td>
</tr>
<tr>
<td>BX111160</td>
<td>Prostate</td>
<td>395</td>
</tr>
<tr>
<td>AI669704 (same as BX111160)</td>
<td>Prostate</td>
<td>367</td>
</tr>
<tr>
<td>AA522842</td>
<td>Prostate cancer, bone metastasis</td>
<td>310</td>
</tr>
<tr>
<td>AI685510</td>
<td>Prostate</td>
<td>454</td>
</tr>
<tr>
<td>BF372485</td>
<td>Prostate tumor</td>
<td>309</td>
</tr>
<tr>
<td>BP326135</td>
<td>Prostate</td>
<td>581</td>
</tr>
<tr>
<td>AA618514</td>
<td>Prostate</td>
<td>419</td>
</tr>
<tr>
<td>AA526297</td>
<td>Prostate</td>
<td>345</td>
</tr>
<tr>
<td>BF056717</td>
<td>Breast</td>
<td>220</td>
</tr>
<tr>
<td>AA595348</td>
<td>Prostate</td>
<td>463</td>
</tr>
<tr>
<td>BF372485</td>
<td>Prostate</td>
<td>309</td>
</tr>
<tr>
<td>BP326135</td>
<td>Prostate</td>
<td>581</td>
</tr>
</tbody>
</table>
Recently, a bicistronic RNA containing sequences from KLK8 and KLK9 transcripts was reported (AY566267). Our finding of KLK3/PSA and KLK2 bicistronic RNAs indicates that the observed coordinated expression of multiple tissue kallikrein genes, may be regulated by bicistronic RNAs.

Identification of a novel PSA isoform differentially expressed in prostate cancer and benign prostatic hyperplasia

Because exploitation of intronic sequences for the transcription of multiple isoforms by alternative promoters is common to tissue kallikrein genes, we designed a primer that specifically anneals in intron II of KLK3/PSA gene, with the purpose to search for as yet unidentified splice variants of KLK3/PSA transcribed from intron II [16]. We focused our analysis on intron II because PSA-SVs containing sequences that correspond to introns I, III, and IV were identified (Fig. 1). Reverse-transcription and subsequent PCR-amplification of total RNA isolated from LNCaP yielded a product of the anticipated length. Sequencing of this product showed that it represents a novel splice variant encoded by the PSA/KLK3 gene that we named PSA-SV5, and was deposited in Genbank™ with accession number DQ641251. PSA-SV5 contains a full-length ORF (Fig. 3A) and an ATG located in a Kozak-like sequence (CCCATGTTC) [17]. Furthermore, the expression of PSA-SV5 was studied in prostate tissue specimens isolated from patients diagnosed with prostate cancer or BPH. As shown in Fig. 3 (upper, representative clinical samples), the classical KLK3/PSA transcript was indistinguishably expressed in both patients with prostate cancer or BPH. Expression analysis of the PSA-SV5 variant in prostate samples, revealed that the variant is highly expressed in 6/9 (67%) of patients with prostate cancer but only in 1/7 (14%) patients with BPH (Fig. 3B, middle, clinical samples) ($p = 0.036$ according the Fisher’s Exact Test). Expression of PSA-SV5 isoform is restricted to prostate cancer and was neither detected in normal breast cells nor in breast and ovarian cancer cell lines (data not shown). The differential expression of PSA-SV5 isoform in BPH may be used to distinguish BPH from prostate cancer and should be assessed in combination with classical PSA as an exclusion marker for prostate cancer to reduce the number of false-positive cancer diagnoses.

Discussion

The discrimination of benign prostatic hyperplasia (BPH) from prostate cancer (CaP), which many times is present in minuscule amounts within the needle biopsy tissue, often gives false-negative or false-positive results. The most common mimickers, giving rise to false-positive cancer diagnosis, are atrophy, atypical adenomatous hyperplasia and seminal vesicle-type tissue. A number of histoanatomical structures such as Cowper’s gland, metaplastic and hyperplastic processes, as well as, inflammatory processes may confuse BPH with CaP. Routine microscopy coupled with the careful use of immunocytochemistry will lead to a correct diagnosis and avoid a false-positive cancer interpretation [18,19]. Although, prostate-specific antigen (PSA) is the most widely used biomarker for the diagnosis,
prognosis and management of prostate cancer, its efficiency in distinguishing BPH from CaP with this test is not possible, due to the lack of its specificity. Novel tissue biomarkers will be able to improve the discrimination between BPH and CaP within the needle biopsy of prostate gland [20,21].

Recent evidence suggests that several human tissue kallikrein genes are regulated by alternative splicing [1,16,22,23]. Interestingly, splice variants of KLK5 and KLK7 genes display cancer-specific expression profiles, enabling their potential application as cancer biomarkers [22], while a splice variant of the prostate-specific KLK2 gene was proposed as a tumor marker for metastatic prostate cancer [23]. KLK2 also represents a new prostate cancer biomarker [24].

In this study, in an attempt to identify novel prostate cancer biomarkers that will aid the differential diagnosis of BPH and prostate cancer, seven new splice variants were found and analyzed, as well as two bicistronic RNAs containing sequences from KLK3/PSA and KLK2 genes. Although we analyzed numerous EST clones covering the entire sequence of KLK3/PSA, we focused our study on those clones that have either insertions or deletions relative to KLK3/PSA mRNA, in order to exclude sequences derived from genomic DNA contamination [5]. The physiological significance of the identified KLK3/PSA splice variants is currently unknown.

ESTs are partial cDNA sequences obtained by random sequencing of cDNA libraries, that usually correspond to the 3' or 5' end of a cDNA. EST sequences from numerous individuals, multiple tissues and varying conditions are accumulated in a high-throughput manner, thus, allowing biocomputational identification of genes, analysis of tissue distribution of mRNAs, identification of new transcript or splice variants and their positional mapping within a genome [25,26]. Alignment of EST clones with genomic sequences or known mRNAs can lead to the identification of previously unidentified splice variants derived from cryptic introns, splicing out of exons, usage of alternative promoters or polyadenylation signals [5]. Alternative splicing is a critical post-transcriptional event that accounts for the observed increased transcriptome diversity. Notably, recent estimates indicate that at least 60% of human genes encode for alternatively spliced variants [26]. It should be noted that ESTs usually correspond to the 3' region of the gene, because they are generated from oligo(dT)-primed cDNA libraries. Therefore, ESTs provide for a rather confident prediction of long 3' UTRs. More recently, EST clones are enriched for full-length mRNAs by a cap site-based selection, thus, enabling in silico cloning of 5' UTRs [5]. However, EST mining should be used with caution to avoid false-positive data that reflect “splice-noise” or transcripts derived from splicing errors. For this, splice junctions of mRNAs and the abundance of splice isoforms should be carefully analyzed. Furthermore, ESTs cannot provide data on whether variant transcripts are translated in vivo.

Free serum PSA mainly consists of multiple pPSA zymogen forms differing in the lengths of their propeptide, i.e. [-7], [-5], [-4], [-2]-pro-PSA, benign PSA (BPSA, clipped form of PSA present in cases of BPH) and forms of intact non-native PSA (iPSA). Intact inactive PSA, namely iPSA and pPSA, are referred to as iPSA [6]. KLK3/PSA splice variants, such as PSA-RP1 that shares epitopes with classical PSA are expected to contribute to measurements of iPSA concentration [4]. PSA-SV2 that was described here for the first time will additionally contribute to iPSA concentration, as it shares high homology with PSA-RP3 and will possibly be recognized by PSA-specific antibodies currently used in established immunoassays. Importantly, we identified the novel variant, PSA-SV5 that may be produced by the activity of alternative promoters located inside intron II, since, as shown in Fig. 3, in PC-3 cells that do not express PSA, expression of PSA-SV5 was demonstrated. Usage of alternative promoters within the kallikrein gene family was reported for KLK4 [27], KLK6 [28,29], and KLK11 [30].

PSA-SV5 may have applications in clinical diagnosis of prostate cancer based on its expression in a much higher percentage of prostate cancers compared to BPHs. Analysis of a large panel of clinical samples will be required to assess PSA-SV5 as a marker for the differential diagnosis of prostate cancer and BPH. Furthermore, in vitro transcription and translation experiments are required in order to find out whether this splice variant expresses a protein or it represents a non-coding RNA. However, analysis of the inferred amino acid sequence using SignalP (http://www.cbs.dtu.dk/services/SignalP/) identified a signal peptide, suggesting that the encoded protein is secreted. Production of specific antibodies that specifically recognize the PSA-SV5 isofrom in the presence of wild-type PSA will allow for the detection and differential quantification in prostatic tissues and in biological fluids.

Acknowledgments

This study was funded by the Greek Secretariat of Research and Technology and European Union through research projects PENED2001 (01ΕΔ557) and PENED2003 (903ΕΔ430) of operational program competitiveness.

References