Blackwell Publishing

Protease inhibitor locus on chromosome 20
DOI: 10.1111/j.1745-7262.2007.00303.x

AJA
www.asiaandro.com

·*Review*·

A locus on chromosome 20 encompassing genes that are highly expressed in the epididymis

Åke Lundwall

Department of Laboratory Medicine, Lund University, Clinical Chemistry, University Hospital MAS, Malmö S-205 02, Sweden

Abstract

During liquefaction of the ejaculate, the semen coagulum proteins semenogelin I (SEMG1) and semenogelin II (SEMG2) are degraded to low molecular mass fragments by kallikrein-related peptidase 3 (KLK3), also known as prostate-specific antigen. Semenogelin molecules initiate their own destruction by chelating Zn^{2+} that normally would completely inhibit the proteolytic activity of KLK3. In a similar way, semenogelins might regulate the activity of kallikrein-related peptidases in the epididymis, something that might be of importance for the maturation of spermatozoa or generation of anti-bacterial peptides. Studies on the evolution of semen coagulum proteins have revealed that most of them carry an exon that displays a rapid and unusual evolution. As a consequence, homologous proteins in rodents and primates show almost no conservation in primary structure. Further studies on their evolution suggest that the progenitor of the semen coagulum proteins probably was a protease inhibitor that might have displayed antimicrobial activity. The semenogelin locus on chromosome 20 contains at least 17 homologous genes encoding probable protease inhibitors with homology to semen coagulum proteins. All of these are highly expressed in the epididymis where they, similar to the semenogelins, could affect the maturation of spermatozoa or display antibacterial properties. *(Asian J Androl 2007 July; 9: 540–544)*

Keywords: antimicrobial; inhibitor; kunitz; proteolysis; semen; semenogelin; zinc; whey acidic protein four disulphide core

1 Introduction

Following spermatogenesis, spermatozoa are transported through the tubular system of the epididymis, where they acquire motility and fertilizing capability. The destination is the caudal part of the epididymis, which serves as a reservoir of mature spermatozoa. During their transport and storage, spermatozoa are exposed to epididymal fluid containing various components that could affect both the maturation process and provide protection against microbes.

At ejaculation, epididymal spermatozoa are transported through vas deferens and mixed with secretions of the accessory sex glands. The ejaculatory mixing of human epididymal sperm and the fluids provided by the prostate and the seminal vesicles gives rise to a semi-solid mass, known as the seminal coagulum. Within minutes, the seminal coagulum liquefies to yield seminal plasma containing spermatozoa that now display propulsive motility. The major structural components of the human seminal coagulum are semenogelin I (SEMG1) and semenogelin II (SEMG2), which are both secreted at very high concentrations by the seminal vesicles [1, 2]. The semenogelins are also synthesized in the secretory epithelium of the epididymis, but presumably in lower quantities than in the seminal vesicles [3, 4]. During liquefaction, the semenogelins are degraded to peptide fragments by proteolytic enzymes.

The predominant semenogelin-degrading enzyme in semen is kallikrein-related peptidase 3 (KLK3), which is

Correspondence to: Dr Åke Lundwall, Department of Laboratory Medicine, Lund University, Clinical Chemistry, University Hospital MAS, Malmö S-205 02, Sweden.
Tel: +46-40-333-156 Fax: +46-40-929-023
E-mail: Ake.Lundwall@med.lu.se

an abundant protein secreted by the prostate gland, also known as prostate-specific antigen [5]. This review will address the function of the semenogelins and describe the unique evolution of semen coagulum proteins, which have also led to the discovery of a new locus of homologous genes expressed in the epididymis. The potential role in epididymal biology of proteins expressed by genes from this locus is also discussed.

2 Function of the semenogelins

Recent studies have cast new light on how the semenogelins might function during semen liquefaction [6]. Studies on the enzymatic properties of KLK3 have demonstrated that the protease is sensitive to Zn^{2+} and that micromolar concentrations of the ion could inhibit the proteolysis of chromogenic peptide substrate [7]. The prostate secretion that delivers KLK3 to semen also contains a high concentration of Zn^{2+} ions, which effectively blocks the proteolytic activity of the enzyme. In seminal plasma, the Zn^{2+} concentration of approximately 2 mmol/L would completely inhibit KLK3 and, therefore, prevent semen liquefaction. The reason that semen liquefaction still takes place is sequestration of Zn^{2+} by seminal plasma proteins. The molecules that sequester the Zn^{2+} are the semenogelins, as was concluded from experiments that demonstrated that both SEMG1 and SEMG2 bind at least 10 molecules of Zn^{2+} each with an average affinity of 5 μmol/L [6]. Therefore, by chelating Zn^{2}, the semenogelins activates KLK3 and, thereby, initiate their own destruction.

The regulation of KLK3 by Zn^{2+} need not be confined to semen liquefaction, as overlapping expression of KLK3 and SEMG1 and SEMG2 has been demonstrated in several tissues [4, 8]. This suggests that the semenogelins might function as general regulators of the extracellular Zn^{2+} concentration, something that in turn might affect proteolytic enzymes and other Zn-sensitive molecules or biological systems. Among the potential targets are 15 proteases, including KLK3, from the kallikrein locus on the long arm of chromosome 19 [9–11]. Several of them have been shown to be sensitive to Zn^{2+} and studies have clearly demonstrated that most of them are expressed in the epididymis [8, 12]. It is reasonable to believe that these enzymes are active during the maturation of spermatozoa and that the semenogelins are important as regulators of their activity. However, many more studies are required to clarify the role of Zn^{2+} and semenogelins in the modulation of these enzymes' activities.

Another potential function of the epididymal semenogelins could be in the innate immunity of the tissue. Peptide fragments of semenogelins possess antibacterial activity, which might be important as a defense mechanism both for the epithelial cells and for the spermatozoa [13]. The latter is presumably less likely, given the much higher semenogelin concentration in seminal plasma that comes from the seminal vesicle secretion.

3 Rapid evolution of semen coagulum proteins

Large parts of the semenogelins consist of repeats of 60 amino acid residues. The repeats are poorly defined, but owing to how well the overall structure is conserved, they have been named type I, type II and type III repeats [2]. Of these, the type I repeats are the most conserved. In the common allelic variant of SEMG1 there are two repeats of each type. SEMG2 is 79% similar in sequence to SEMG1, but owing to two extra type I repeats, the size of the molecule is 63 kDa compared with 50 kDa for SEMG1 [1, 2]. Recently, a rare *SEMG1* allele was described, which is carried by 3–6% of the population in both Europe and Japan [14, 15]. The novel allele is lacking one type I repeat and gives rise to a smaller molecule with a mass of 43 kDa.

Studies on primate semenogelins have revealed some very interesting findings regarding their evolution. Most conspicuous is the extension of semenogelin molecules by species-specific additions of type I repeats (Figure 1). In the chimpanzee and the orangutan this has created very large SEMG1 molecules containing eight and nine type I repeats [16, 17]. Similar duplications that extend SEMG1 have also occurred in the gorilla, but in this species there are frequent alleles that carry premature stop codons that give rise to small SEMG1 molecules [16]. This type of evolution, with exon extension, is not confined to apes and SEMG1, as SEMG2 from the rhesus monkey, an Old World Monkey, contains six type I repeats and the cotton-top tamarin, a New World Monkey, carries five type I repeats in its SEMG1 [18, 19]. It has been proposed that the extension of the molecules could be the result of sexual selection caused by polyandry. Species with promiscuous females, such as the chimpanzee tend to have large SEMG1 and perhaps produce a firmer coagulum, which could function as a barrier against spermatozoa from a second mating [17]. However, there is still no experimental evidence to support this hypothesis and, therefore, it must be considered as mere speculation.

The seminal vesicles of murine rodents (i.e. rats and mice) secrete six proteins denoted SVS1–SVS6, at high concentrations. Of these, SVS1 is related to copper amine oxidases, such as histaminase, but because of substitutions it is presumably enzymatically inactive [20]. The remaining components, SVS2–SVS6, are homologous with the

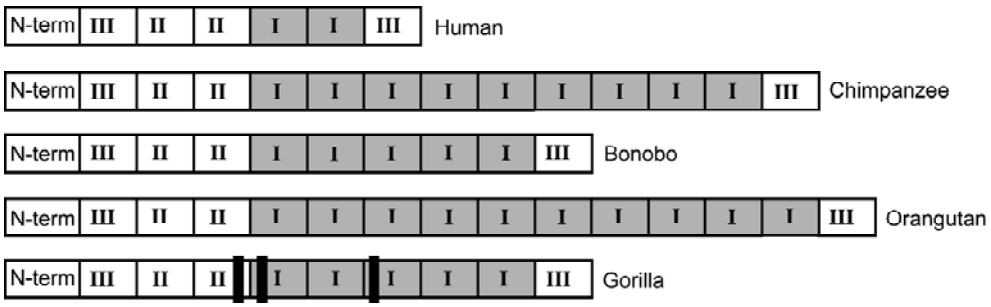

Figure 1. Extension of semenogelin I (SEMG1) in hominoids. SEMG1 is illustrated with boxes that represent different conserved parts of the molecule. The repeats of 60 amino acid residue are indicated with Roman numerals: the type I repeat that varies in copy number is also shaded in grey. The vertical bars in the illustration of SEMG1 from gorilla denote the locations of the C-terminus in allelic variants that are shorter as a result of premature stop codons.

semenogelins in spite of very limited similarity at the amino acid level [21]. However, the overall gene structure is similar and consists of three exons separated by two relatively short introns. The first exon codes for the signal peptide and the two amino terminal residues of the mature protein. The second exon codes for the remainder of the secreted protein and also carries a few 3' non-translated nucleotides. The third exon has no coding nucleotides, but carries the poly-adenylation signal. Studies have shown that the first and the third exons are conserved between the semenogelin and the SVS genes, whereas the second exon displays major differences, which suggests rapid evolution. A comparison of the second exon from *SEMG1* and *SVS2*, also known as semenoclotin, shows that the two termini of the exon are conserved [22]. Therefore, both genes appear to have evolved from a gene with a fairly small second exon that has expanded in size since the separation of the murine rodents and the primates (Figure 2). The reason for the size expansion is duplications of repeats, which in the primates are 180 bp and in the murine rodents are between 21 and 39 bp [2, 20, 23]. *SEMG2* and *SVS3* also seem to have evolved in this way. The second exons of *SVS4–SVS6* are smaller and have evolved in a different manner owing to selection of a new splice site [20]. The same mechanism also yields the unique structure of the gene that codes for the polyprotein that after processing yields three of the major proteins in guinea pig seminal vesicle secretion [24]. Comparison of nucleotide sequences shows that approximately 0.5 kb in the first intron of the guinea pig gene and the second exon of *SEMG2* are conserved. Owing to mutation, the first splice site, which is homologous with that in *SEMG2*, is skipped and instead a second splice site located 4 360 bp further downstream is used. This generates a new and unique gene product in the guinea pig, which has no similarity to either the semenogelins or SVS proteins.

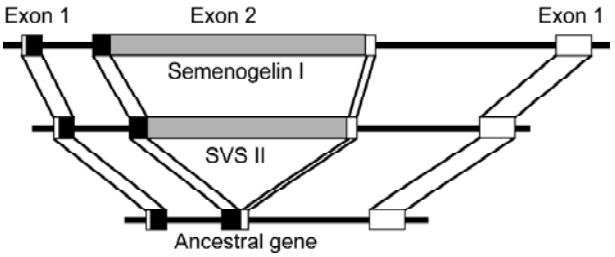

Figure 2. Evolution of semen coagulum proteins. The human semenogelin I (*SEMG1*) gene and the mouse *SVS2* gene are illustrated by horizontal bars with boxes indicating the location of exons. Filled or shaded boxes represent translated nucleotides sequences and open boxes the non-translated sequences. The shading in exon 2 indicates nucleotide sequences that have evolved after the separation of the lineages leading to primates and murine rodents. The hypothetical ancestral gene is illustrated with exons consisting of nucleotide sequences that are conserved between the human and mouse genes.

4 Identification of a protease inhibitor locus

To identify additional genes with homology to the semen coagulum proteins, sequence databases were searched using the conserved first exon of semenogelin and SVS genes. It was discovered that the genes of elafin (*PI3*) and secretory leukocyte proteinase inhibitor (*SLPI*) are homologous to those of the semen coagulum proteins [25]. As with the above described similarity between different genes of semen coagulum proteins, the similarity of the *PI3* and *SLPI* genes to those of the semen coagulum proteins is confined to the first and the last exon. The translation product has no similarity at all to any of the known semen coagulum proteins. Instead, both elafin and SLPI carry the structural motif of whey acidic protein four disulphide core (WFDC) domains. There are two such domains in SLPI and one in elafin. The WFDC domain binds and inhibits serine proteinases and both elafin and SLPI are strong inhibitors of elastase, a serine

protease secreted by neutrophil granulocytes [26, 27].

Investigations into the chromosomal localization by *in situ* hybridization had shown that the genes of *SEMG1*, *SEMG2* and *PI3* were located in the same region on the human chromosome 20q12-13.1 [28, 29]. When DNA sequences from the human genome project were made available, it became clear that *SLPI* was also located close to the semenogelin genes. This inspired a closer inspection of the semenogelin locus for additional genes. In this way, we first discovered a gene that was highly expressed in the prostate and that also carried a WFDC domain [30]. Eventually, 14 new genes that had between 1 and 4 WFDC domains were discovered at the locus [31]. Three of the genes also encode kunitz domains, another serine protease inhibitor domain, in addition to the WFDC domains. Recently, we have also identified 3 novel genes at the locus that carry a single kunitz domain, but no WFDC domain (unpublished observation). In all, this increases the number of functional protease inhibitor genes at the locus to 17 (Figure 3). All the genes are basically organized as the *PI3* and *SLPI* genes, with one exon for the signal peptide, and one exon for each protease inhibitor domain, followed by an exon with 3' non-translated nucleotides that has no or very little coding information.

5 Possible function of inhibitor molecules in the epididymis

Studies on the expression of WFDC-encoding genes show that several of them (e.g. *PI3, SLPI, WFDC2, WFDC3* and *WFDC10B*) are expressed in almost every tissue, whereas others have a more restricted pattern of expression [31]. All of them generate transcripts in the epididymis and most of them also in the testis, with the trachea as the third most common site of expression. The presence of WFDC and Kunitz domains in the genes suggests that they might function as inhibitors of locally produced proteases, which could be either endogenous in origin or the products of micro-organisms. Potential endogenous target proteases are members of the kallikrein-related peptidases on chromosome 19, as was also suggested above for the Zn^{2+}/semenogelin system. Prime candidate targets among exogenous proteases are serine peptidases secreted by bacteria, which might serve as virulence factors. However, the novel genes might also have a direct toxic effect on micro-organisms, as there are several reports showing that SLPI possesses antimicrobial activity and that this property is independent of the protease inhibitor function [32]. Therefore, it is possible that the novel proteins are an important part of the natural defense against infection.

Except for the information gained by studies on SLPI, and to a lesser extent elafin, very little is known regarding the function of other proteins from the locus. There are a few scattered studies, which, among other things, have shown that both human eppin and mouse WFDC12, also named SWAM1, possess anti-bacterial activity and, therefore, support the natural defense hypothesis [33, 34]. Why innate immunity should be so important in the epididymis is harder to understand. Perhaps the epididymis serves as the equivalent of a weapon store for the spermatozoa, where they can arm themselves before they encounter microorganisms in the female genital tract. This is a very attractive idea, but definite answers have to await future research into this new and interesting aspect of epididymal function.

Acknowledgment

This work was supported by grants from Österlunds trust, MAS cancer and the Medical Faculty at Lund

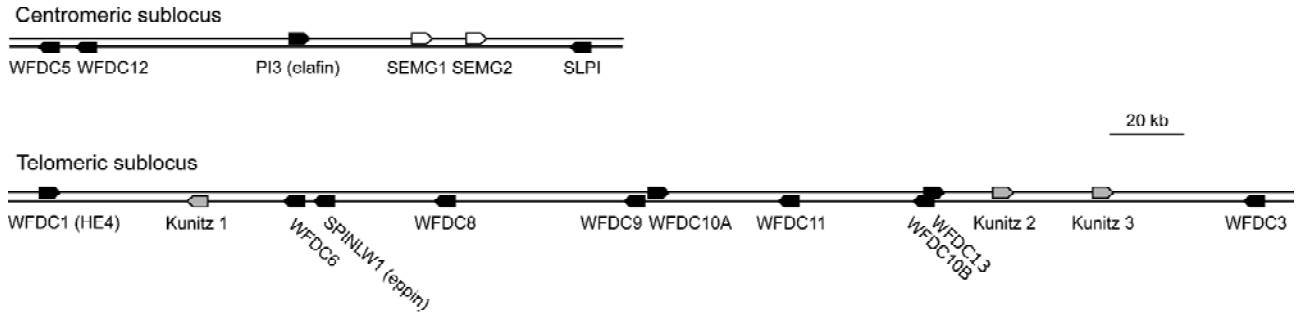

Figure 3. The protease inhibitor locus on human chromosome 20q12-13.1. The two subloci, on the chromosome separated by 215 kb, are shown separately. The double horizontal bars represent the two DNA strands. The arrow heads indicate genes and direction of transcription and the gene symbols are written below. The newly discovered genes that encode kunitz domains, denoted Kunitz 1–3, are shaded in grey.

University.

References

1 Lilja H, Abrahamsson PA, Lundwall Å. Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa. J Biol Chem 1989; 264: 1894–900.
2 Lilja H, Lundwall Å. Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein. Proc Natl Acad Sci U S A 1992; 89: 4559–63.
3 Bjartell A, Malm J, Möller C, Gunnarsson M, Lundwall Å, Lilja H. Distribution and tissue expression of semenogelin I and II in man as demonstrated by *in situ* hybridization and immunocytochemistry. J Androl 1996; 17; 17–26.
4 Lundwall Å, Bjartell A, Olsson AY, Malm J. Semenogelin I and II, the predominant human seminal plasma proteins, are also expressed in non-genital tissues. Mol Hum Reprod 2002; 8: 805–10.
5 Lilja H. A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. J Clin Inves 1985; 76**:** 1899–903.
6 Jonsson M, Linse S, Frohm B, Lundwall Å, Malm J. Semenogelins I and II bind zinc and regulate the activity of prostate-specific antigen. Biochem J 2005; 387: 447–53.
7 Malm J, Hellman J, Hogg P, Lilja H. Enzymatic action of prostate-specific antigen (PSA or hK3), substrate specificity and regulation by Zn^{2+}, a tight-binding inhibitor. Prostate 2000; 45: 132–9.
8 Olsson AY, Bjartell A, Lilja H, Lundwall Å. Expression of prostate-specific antigen (PSA) and human glandular kallikrein 2 (hK2) in ileum and other extraprostatic tissues. Int J Cancer 2005; 113: 290–7.
9 Gan L, Lee I, Smith R, Argonza-Barrett R, Lei H, McCuaig J, *et al*. Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region. Gene 2000; 257: 119–30.
10 Harvey TJ, Hooper JD, Myers SA, Stephenson SA, Ashworth LK, Clements JA. Tissue-specific expression patterns and fine mapping of the human kallikrein (KLK) locus on proximal 19q13.4. J Biol Chem 2000; 275: 37397–406.
11 Yousef GM, Chang A, Scorilas A, Diamandis EP. Genomic organization of the human kallikrein gene family on chromosome 19q13.3-q13.4. Biochem Biophys Res Commun 2000; 276: 125–33.
12 Petraki CD, Papanastasiou PA, Karavana VN, Diamandis EP. Cellular distribution of human tissue kallikreins, immunohistochemical localization. Biol Chem 2006; 387: 653–63.
13 Bourgeon F, Evrard B, Brillard-Bourdet M, Colleu D, Jegou B, Pineau C. Involvement of semenogelin-derived peptides in the antibacterial activity of human seminal plasma. Biol Reprod 2004; 70: 768–74.
14 Lundwall Å, Giwercman A, Ruhayel Y, Giwercman Y, Lilja H, Hallden C, *et al*. A frequent allele codes for a truncated variant of semenogelin I, the major protein component of human semen coagulum. Mol Hum Reprod 2003; 9: 345–50.
15 Miyano S, Yoshida K, Yoshiike M, Miyamoto C, Furuichi Y, Iwamoto T. A large deletion of the repeat site in semenogelin I is not involved in male infertility. Int J Mol Med 2003; 11: 435–40.
16 Jensen-Seaman MI, Li WH. Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. J Mol Evol 2003; 57: 261–70.
17 Kingan SB, Tatar M, Rand DM. Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. J Mol Evol 2003; 57**:** 159–69.
18 Lundwall Å. The cotton-top tamarin carries an extended semenogelin I gene but no semenogelin II gene. Eur J Biochem 1998; 255: 45–51.
19 Ulvsbäck M, Lundwall Å. Cloning of the semenogelin II gene of the rhesus monkey. Duplications of 360 bp extend the coding region in man, rhesus monkey and baboon. Eur J Biochem 1997; 245: 25–31.
20 Lundwall Å, Malm J, Clauss A, Valtonen-Andre C, Olsson AY. Molecular cloning of complementary DNA encoding mouse seminal vesicle-secreted protein SVS I and demonstration of homology with copper amine oxidases. Biol Reprod 2003; 69: 1923–30.
21 Lundwall Å, Lazure C. A novel gene family encoding proteins with highly differing structure because of a rapidly evolving exon. FEBS Letters 1995; 374: 53–6.
22 Lundwall Å. The cloning of a rapidly evolving seminal-vesicle-transcribed gene encoding the major clot-forming protein of mouse semen. Eur J Biochem 1996; 235: 424–30.
23 Harris SE, Harris MA, Johnson CM, Bean MF, Dodd JG, Matusik RJ, *et al*. Structural characterization of the rat seminal vesicle secretion II protein and gene. J Biol Chem 1991; 265: 9896–903.
24 Hagstrom JE, Fautsch MP, Perdok M, Vrabel A, Wieben ED. Exons lost and found. Unusual evolution of a seminal vesicle transglutaminase substrate. J Biol Chem 1996; 271: 21114–9.
25 Lundwall Å, Ulvsbäck M. The gene of the protease inhibitor SKALP/elafin is a member of the REST gene family. Biochem Biophys Res Commun 1996; 221: 323–7.
26 Thompson RC, Ohlsson K. Isolation, properties, and complete amino acid sequence of human secretory leukocyte protease inhibitor, a potent inhibitor of leukocyte elastase. Proc Natl Acad Sci USA 1986, 83: 6692–6.
27 Wiedow O, Schroder JM, Gregory H, Young JA, Christophers E. Elafin, an elastase-specific inhibitor of human skin. Purification, characterization, and complete amino acid sequence. J Biol Chem 1990; 265: 14791–5.
28 Molhuizen HO, Zeeuwen PL, Olde Weghuis D, Geurts van Kessel A, Schalkwijk J. Assignment of the human gene encoding the epidermal serine proteinase inhibitor SKALP (PI3) to chromosome region 20q12–>q13. Cytogenet Cell Genet 1994; 66: 129–31.
29 Ulvsbäck M, Lazure C, Lilja H, Spurr NK, Rao VV, Loffler C, *et al*. Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20. J Biol Chem 1992; 267: 18080–4.
30 Lundwall Å, Clauss A. Identification of a novel protease inhibitor gene that is highly expressed in the prostate. Biochem Biophys Res Commun 2002; 290: 452–6.
31 Clauss A, Lilja H, Lundwall Å. A locus on human chromosome 20 contains several genes expressing protease inhibitor domains with homology to whey acidic protein. Bioch J 2002; 368: 233–42.
32 Hiemstra PS. Novel roles of protease inhibitors in infection and inflammation. Bioch Soc Trans 2002; 30: 116–20.
33 Hagiwara K, Kikuchi T, Endo Y, Huqun, Usui K, Takahashi M, *et al*. Mouse SWAM1 and SWAM2 are antibacterial proteins composed of a single whey acidic protein motif. J Immunol 2003; 170: 1973–9.
34 Yenugu S, Richardson RT, Sivashanmugam P, Wang Z, O'Rand MG, French FS, *et al*. Antimicrobial activity of human EPPIN, an androgen-regulated, sperm-bound protein with a whey acidic protein motif. Biol Reprod 2004; 71: 1484–90.