Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein

HANS LILJA* AND ÅKE LUNDWALL

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

Communicated by Jan Waldenström, February 12, 1992 (received for review December 20, 1991)

Freshly ejaculated human semen has the ap-ABSTRACT pearance of a loose gel in which the predominant structural protein components are the seminal vesicle-secreted semenogelins (Sg). The primary structure of the 439-residue SgI has previously been obtained by cDNA cloning. This cDNA crosshybridizes to a larger transcript coding for a second secretory protein, SgII. Here we report the almost complete structure of a precursor of SgII established by λ gt11 clones isolated from epididymal and seminal vesicular cDNA libraries. The deduced amino acid sequence of the 559-residue mature protein has a molecular weight of 62,931 but an increase in weight may be provided by asparagine-linked oligosaccharide attachment at residue 249. SgII, which has 78% overall identity with SgI, contains eight 60-residue regions that display conspicuous internal sequence similarity, whereas SgI only contains six of these regions. The SgII structure is translated from an open reading frame in a polyadenylylated 2.4-kilobase transcript. The message is abundant in the seminal vesicles but rare in the epididymis.

After spermatogenesis, testicular spermatozoa mature in the epididymis to become fully fertile. This process takes ≈ 2 weeks and depends on a variety of secretory products provided by the epithelium of the epididymis. Although the epididymal spermatozoa have fertilizing capacity in vitro, they do not display the vigorous propulsive mobility of ejaculated spermatozoa. Therefore, it may be assumed that final maturation or activation of the spermatozoa takes place after the ejaculatory mixing of spermatozoa with the secretions of the prostate and the seminal vesicles (1). The ejaculatory mixing results in immediate formation of a spermentrapping gel. Predominant structural proteins of the gel are the major secretory proteins of the seminal vesicles, the 52-kDa semenogelin (Sg) and two antigenically related 71and 76-kDa components (2-5). Within 5-15 min, progressively motile spermatozoa are released when the gel dissolves to give free-flowing liquid. Dissolution of the gel parallels limited proteolysis of Sg and Sg-related proteins (2-5) by an abundant prostate-secreted serine proteinase commonly referred to as the prostate-specific antigen (PSA) (5-11).

We have previously described the primary structure of the precursor to the 52-kDa Sg (11), coined here as SgI. It consists of a 23-amino acid signal peptide and a mature protein of 439 amino acid residues. The protein structure has no similarity with other proteins but contains consecutive regions that display extensive internal sequence similarity (11).

A SgI coding mRNA of ≈ 1.8 kilobases (kb) is abundant in the seminal vesicles but absent in the epididymis, the prostate, and the testis (11). However, a larger message crosshybridizes to the SgI cDNA in RNA from both the epididymis and the seminal vesicles. A cDNA clone (λ ESRP-V) corresponding to this RNA species was isolated from an epididymis cDNA library and shown to carry a 1.4-kb insert that

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

had extensive structural similarity with that of SgI (11). We have now determined the primary structure of a Sg-related protein precursor, coined here as SgII, which is the translation product of an ≈ 2.4 -kb transcript.

EXPERIMENTAL PROCEDURES

Reagents. Deoxynucleotides, dideoxynucleotides, pUC18, and the replicative forms of M13mp18 and M13mp19 were from Pharmacia; Hybond N filters, $[\gamma^{-32}P]ATP$, $[\alpha^{-32}P]dCTP$, and the multiprime labeling kit were from Amersham. Human seminal vesicular and epididymal cDNA libraries constructed in $\lambda gt11$ have been described (11).

Probes. Two oligonucleotide constructs (18-mers) based on the nucleotide sequence of λ VSRP-XXI were synthesized on an Applied Biosystems model 381A DNA synthesizer. SRP-1 (ACCTAGGTGTTGTTTTGA) was complementary to nucleotides 280-297 and SRP-2 (CTGCTCAATTATAAACAA) corresponded to nucleotides 310-327. The oligonucleotides were 5'-end-labeled with [γ^{-32} P]ATP and polynucleotide kinase to a specific activity of >10⁸ cpm/ μ g. The following cDNA probes were used: ES146, comprising 146 base pairs (bp) from the artificial 5' *Eco*RI site to the *Sac* I site in λ ESRP-V; BP1028, comprising 1028 bp from *Bam*HI (at position 560) to *Pst* I (at position 1588) in λ VSRP-XXI; XE613, comprising 613 bp from *Xba* I (at position 1857) to the 3' *Eco*RI site in λ VSRP-XII. The probes were ³²P-labeled by random priming to a specific activity of >10⁹ cpm/ μ g.

Molecular Cloning. Screening of λ gt11 libraries with cDNA probes was performed under standard conditions (12), except for the final wash of the hybridized filters that was done at 68°C for 20 min in 15 mM sodium citrate, pH 7.0/0.15 M NaCl $(1 \times SSC)$ with 0.1% SDS and 0.05% sodium pyrophosphate. Hybridizations with oligonucleotide probes were performed at 37°C for 16 h in 6× SSC with 0.1% SDS/0.05% sodium pyrophosphate/100 μ g of herring sperm DNA per ml. Excess radioactivity was washed off at 37°C for 1 h in 6× SSC with 0.1% SDS/0.05% sodium pyrophosphate, followed by a final wash at 47°C for 10 min. Recombinant phages were purified from plate lysates by ultracentrifugation in CsCl gradients and the DNA was extracted by standard procedures (12). The cDNA inserts were subcloned into the plasmid vector pUC18 and plasmid preparations were done by the alkaline lysis method as described (12).

Nucleotide Sequence Determinations. Fragments for nucleotide sequencing were isolated from restriction enzyme digests or generated by sonication and subcloned in bacteriophages M13mp18 or M13mp19. Nucleotide sequences were determined by the dideoxynucleotide chain-termination method of Sanger *et al.* (13) as modified by Biggin *et al.* (14). Specific priming of sequence reactions was in some instances performed with the probes SRP-1 and SRP-2 as well as

Abbreviations: PSA, prostate-specific antigen; SgI, semenogelin I; SgII, semenogelin II.

^{*}To whom reprint requests should be addressed.

[†]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M81652).

oligonucleotides SRP-3 (GGACCCGAAAAAGTCAGC) corresponding to λ VSRP-XXI nucleotides 227–244, SRP-4 (GGTCGGGTGACACCTTGC) complementary to λ VSRP-XXI nucleotides 1783–1797, SRP-5 (TTTGTCTGGTGTT-GATTC) complementary to λ VSRP-XXI nucleotides 788– 806, and SRP-6 (CACCTTTCTGTACATCCT) complementary to λ VSRP-XXI nucleotides 1262–1279. The nucleotide sequences were aligned and analyzed by computer programs (15–17).

Northern Blotting. Tissue specimens were obtained at operation from patients receiving surgical treatment for carcinoma of the urinary bladder (seminal vesicles) or prostate cancer (epididymis). In all cases, the Helsinki Declaration regarding the use of human tissue was strictly observed. The RNA was isolated by the guanidinium thiocyanate method (18). Electrophoresis was run in 1.0% agarose gels containing formaldehyde with 20 μ g of RNA per slot. Blotting to nylon membranes and hybridizations with cDNA probes in 50% (vol/vol) formamide were carried out as described (19, 20).

RESULTS

Isolation and Analysis of cDNA Clones. The isolation and preliminary characterization of an epididymal cDNA clone $(\lambda ESRP-V)$ encoding SgII has been described (11). The clone carried a polyadenylylated insert of ≈1.4 kb. Since Northern blots implied the size of the SgII message to be larger, an additional round of screening was performed to isolate a full-length clone. For this purpose, the probe ES146 was isolated and used to screen both the epididymis and the seminal vesicular cDNA libraries. The rescreening of $\approx 5 \times$ 10⁴ members of the epididymis library did not yield any clones that extended the sequence of λ ESRP-V. In contrast, the screening of the seminal vesicular cDNA library resulted in isolation of a strongly hybridizing clone (λ VSRP-XII) that carried a 2.2-kb insert with one internal EcoRI site. Restriction enzyme digests and partial sequence analysis revealed that $\lambda VSRP$ -XII extended the 5' end of $\lambda ESRP$ -V by ≈ 0.4 kb. However, compared to λ ESRP-V the insert of λ VSRP-XII contained a unique nonpolyadenylylated 3' end.

Based on the sequence of λ VSRP-XII, two oligonucleotides (SRP-1 and SRP-2) were designed to provide optimal mismatching with SgI transcripts and were used in simultaneous screening of 2.0×10^5 members of the seminal vesicular cDNA library yielding 0.1% hybridizing clones. A strongly hybridizing clone (λ VSRP-XXI) was isolated and demonstrated to carry an insert of ≈ 1.9 kb with one internal *Eco*RI site. Partial sequence analysis showed that the λ VSRP-XXI insert contained a polyadenylylated 3' end, identical to that of λ ESRP-V, except for the location of the poly(A) tail, which was 1 nucleotide closer to the polyadenylylation signal. The 5' end of λ VSRP-XXI extended the λ ESRP-V sequence ≈ 0.6 kb, thereby also extending the 5' end of λ VSRP-XII by 234 nucleotides.

An Identical Transcript in Seminal Vesicles and Epididymis. The nucleotide sequence of the λ VSRP-XXI insert was established in completion on both strands, with each nucleotide being determined, on average, 2.9 times. The sequencing strategy is outlined in Fig. 1. The length of the insert is 1953 nucleotides plus a poly(A) tract (Fig. 2). An open reading frame comprising 1740 nucleotides encodes an almost full-length precursor of a secretory protein. At the 3' side of the stop codon, there are 210 nontranslated nucleotides.

Restriction enzyme mapping and partial nucleotide sequence analysis of λ ESRP-V, as outlined in Fig. 1, suggests complete structural identity to nucleotides 628–1953 of the seminal vesicular λ VSRP-XXI insert. The minor discrepancy in the location of the poly(A) tail is caused by an additional 3' thymine nucleotide in λ ESRP-V. Thereby, the position of the poly(A) tail in λ ESRP-V is identical to that previously described for the SgI transcript (11).



FIG. 1. Sequencing strategy and partial restriction endonuclease map. Translated nucleotides are represented by boxes with signal peptide encoding nucleotides in hatched area. The 3' nontranslated nucleotides are shown as lines of different patterns, indicating identical or unique sequences. Indicated restriction enzymes were used to generate probes for hybridization and fragments for sequencing. Arrows indicate the direction and extent of sequence determinations. Arrows with open circles represent priming of sequence reactions with specific oligonucleotides.

Structure of λ VSRP-XII. Partial sequence analysis on the λ VSRP-XII insert was performed as outlined in Fig. 1. The sequence was established for 992 of the nucleotides contained in the 1.35-kb region upstream of the Pst I site. The sequence of this region was identical to the corresponding parts of λ VSRP-XXI (nucleotides 235–1588). The structure 3' to the Pst I site was obtained from 21 random subclones and each nucleotide was determined on average 5.33 times. The cDNA insert downstream of the Pst I site has a total length of 882 nucleotides. The sequence of 223 nucleotides downstream from the Pst I site is identical to that of λ VSRP-XXI (nucleotides 1589-1811 in Fig. 2). However, 69 nucleotides downstream from the stop codon the sequences diverge to yield 659 unique 3' nucleotides in λ VSRP-XII (data not shown) and 142 unique 3' nucleotides in λ VSRP-XXI (Fig. 2, shaded area).

Northern Blot Analysis. Because λ VSRP-XXI and λ VSRP-XII share 1576 5' nucleotides and only differ in the 3' end, we investigated the possibility that they derive from alternative splicing of a common primary transcript. The probe BP1028, recognizing both SgII-t1 and SgII-t2, was hybridized to RNA from the male genital glands on Northern blots. Hybridizing message with an approximate size of 2.4 kb was abundant in the seminal vesicles and rare in the epididymis (Fig. 3). No hybridizing signal was detected in the testis or the prostate (data not shown). The XE613 probe, unique to the 3' part of

Biochemistry: Lilja and Lundwall

ser liei lieuwithevai liewei tuei leui lieuwit luewei tuei leui lieuwit lieuwi	-20 -15 -10 -5			
 J. J. J. J. J. J. Schweizer Strate Str	SerIleIleLeuPheValLeuSerLeuLeuLeuIleLeuGluLysGlnAlaAlaValMet TCCATCATCCTCTTTGTCCTTTCCCTGCTCCTTATCTTGGAGAAGCAAGC	60	300 305 HisGlyLysSerGlnAsnGlnValThrIleHisSerGlnAspGlnGluHisGlyHisLys CATGGCAAGTCTCCAAAACCAGGTAACAATTCATAGTCAAGAGCATGGCCATAAG	1020
Citycinyreilycinyetrycelrycinyetrycelrycinwetrocom32032033033020253035330330330202530353303403453303302025303535034034535035035020253035350 <td>-1 1 5 10 15</td> <td></td> <td></td> <td></td>	-1 1 5 10 15			
20 updived mil 22 updived mil 25 upd	GlyGlnLysGlyGlySerLysGlyGlnLeuProSerGlySerSerGlnPheProHisGly GGACANANAGGTGGATCANAAGGCCANTGCCAAGCGGATCTTCCCAATGGA	120	320 325 330 335 GluasnLysIleSerTyrGlnSerSerStrThrGluGluArgHisLeuAsnCysGlyGlu GAAAATAAAATATCATACCAATCTTCAAGTACAGAAGAAAGA	1080
Giniyadgi QiniistyrPhelolyGiniyaAppelinghiistriyyser 340 345 350 355 MaddecouckCharthTroGatcharthTroGatcharthTrodatCoucharthC	20 25 30 35		•	
10 36 37 375 10 36 370 375 10 10 10 10 385 390 395 10	GlnLysGlyGlnHisTyrPheGlyGlnLysAspGlnGlnHisThrLysSerLysGlySer CAAAAGGGCCAGCACTATTTTGGACAAAAAGACCAACAACAACAACAAACA	180	340 345 LysGlyIleGlnLysGlyValSerLysGlySerIleSerIleGlnThrGluGluGlnIle AAGGGCATCCAGAAAGGTGTATCCAAAGGCAGCATATTCGATCCAAACTGAAGAGCAAATA	1140
Phese:lieldinH:stm:ry:HisValAsplieAsplitestr:rtx:rd:rt:rt:rt:rd:rtx:rd:rtx:rd:rt:rt:rd:rt:rt:rt:r	40 45 50 55			
B0 30 38 30 35 30 35 CACCANTATCATTRONTCOCCTACTANGEGGGAAAACCANACCANACCANACCACTAGEGGGGA 300 36 36 30 35 30 35 30 35 30 35 30 35 30 35 30 36 36	PheSerIleGlnHisThrTyrHisValAspIleAsnAspHisAspTrpThrArgLysSer TTTTCTATTCAACACACATATCATGTAGACTATGATCATGACTGGACCCGAAAAAGT	240	360 365 370 375 HisGlyLySerGlnAsnGlnValArglleProSerGlnAlaGlnGluTyrGlyHisLys CATGGCAAGTCTCAAAACCAGGTAAGAATTCCTAGTCAAGCTCAAGAGTATGGCCATAAG	1200
Clinkintyraspleusania iz euliis sysäl artintysser lysäl nitis euli yöly caccomatrochartmacharthasocananovania arteuliis sysäl artingen sysäl artingen sysäl artingen sysäl artingen sysäl artingen sysäl artingen system system sysäl artingen system system sysäl artingen system sy	60 65 70 75			
30 30 40 410 415 100 105 110 115 115 115 115 126 127 130 135 127 130 135 410 435 120 125 130 135 121 130 155 410 455 120 125 130 155 121 150 155 155 1210 155 170 175 1230 155 170 175 1230 155 170 175 1230 155 170 175 1230 155 170 175 1230 155 190 155 130 155 100 155 130 155 170 175 1230 125 130 155 130 155 100 155 130 155 100 155 131 150 155 100 132 100 155 100 155 131 135 100 155 130 135 100 155 <td>GInGInTyrAspLeuAsnAlaLeuHisLysAlaThrLysSerLysGInHisLeuGiyGiy CAGCAATATGATTGAATGCCCTACATAAGGCGGCAAAAATCAAAACAACACCTAGGTGGA</td> <td>300</td> <td>380 385 GluasnLysIleSerTyrGInSerSerShrfGluGluArgArgLeuAsnSerGlyGlu GAAAATAAAATATCATACCAATCTTCGAGTACAGAAGAAAGA</td> <td>1260</td>	GInGInTyrAspLeuAsnAlaLeuHisLysAlaThrLysSerLysGInHisLeuGiyGiy CAGCAATATGATTGAATGCCCTACATAAGGCGGCAAAAATCAAAACAACACCTAGGTGGA	300	380 385 GluasnLysIleSerTyrGInSerSerShrfGluGluArgArgLeuAsnSerGlyGlu GAAAATAAAATATCATACCAATCTTCGAGTACAGAAGAAAGA	1260
Sercincinculuculeukantyriyscinciluciyargaphilsasplysserlysciynis40541041141110010511011510101010	80 85 90 95			
100100100100100105100105110110110100 </td <td>SercincinceuleuAsnTyrLysGinGiuGiyArgAspHisAspLySSerLySGiyHis AGTCAACAACTGCTCAATTATATAACAAGAAGGCGGAGACCATGATAAATCAAAAGGCCAT</td> <td>360</td> <td>400 405 LysAspValGlnLysGlyValSerLysGlySerIleSerIleGlnThrGluGluLysIle AAGGATGTACAGAAAGGTGTATCCAAAGGCAGTATTTCTATCCAAACTGAAGAGAAAATA</td> <td>1320</td>	SercincinceuleuAsnTyrLysGinGiuGiyArgAspHisAspLySSerLySGiyHis AGTCAACAACTGCTCAATTATATAACAAGAAGGCGGAGACCATGATAAATCAAAAGGCCAT	360	400 405 LysAspValGlnLysGlyValSerLysGlySerIleSerIleGlnThrGluGluLysIle AAGGATGTACAGAAAGGTGTATCCAAAGGCAGTATTTCTATCCAAACTGAAGAGAAAATA	1320
Phenelised FileValleValleVisitisty GilyGilAnanishisGilyTinGinAshron420420420420435120125130135SerGinAspCinGilyAshrortZAGGCANATCCCACCAACCTACCCACCAAAACCAA140445450140145150155140145150155160165170175180185190195180185190195180185190195200205210215201205210215202225230235203225230235204225230235204225230235205210215230206225230235207225230235208225230235208225230235208225230235208225230235208225230235208225230235208235230235219235230235210215240245220225230235230235530535240245250555240245250555240245250555250255230535240 <td></td> <td></td> <td>100 105 100 105</td> <td></td>			100 105 100 105	
120125135226 [139]135227 [130]135228 [14]135238 [14]150140145150155155240145241150245170246155247145248156249155240155240155241156242156243156244156245245246245246245247245248245249245250255251251252250255250255250255250255250255250255250255250255250255250255250255250255250255250255250255250255250260255270275280255280255280255280255280255280255280255280255280255280255280255280255280255 <td< td=""><td>Phenismet IlevalilenishisLysGiyGlyGlanahishisGlyThTGInAShPro TTTCACATGATAGTTATACATCATATAAGGAGGCAAGCTCATCATGGGACACAAAATCCT</td><td>420</td><td>420 425 430 435 435 HisGlyLysSerGlnAspGlnGluHisGlyHisLys CATGGCAAGTCTCAAAACCAGGTAACAATTCCTAGTCAAGAATCAAGAGCATGGCCATAAG</td><td>1380</td></td<>	Phenismet IlevalilenishisLysGiyGlyGlanahishisGlyThTGInAShPro TTTCACATGATAGTTATACATCATATAAGGAGGCAAGCTCATCATGGGACACAAAATCCT	420	420 425 430 435 435 HisGlyLysSerGlnAspGlnGluHisGlyHisLys CATGGCAAGTCTCAAAACCAGGTAACAATTCCTAGTCAAGAATCAAGAGCATGGCCATAAG	1380
SerGinAspcincipAspcincipAssnerProserGipUpSerGinCySerAsintri440450450655140145150155440445450655140145150155155156170175140165170175175175175140165170175175175175140165170175175175175140165170175175175175140185190195195195195195180185190195500510515515200205210215500505510515210215210215520525530535210225230235510515555555210225230235510555555555210225230235510555555555555210225230235550555	120 125 130 135			
140145150150GlulysargLeuTrpValHisGlyLeuSerLysGluGInAlaSerAlaSerGlyAlaGIn460465470475GAAAAAAGGCTAATGGGTCATCAGGCATAAGTAAAGAACAAGCTTCAGGCTCAGGCACAA540465470475160165170175175180485490495AAAGGCAGGCAGAAACAAGGTGGATCCCAAAGCAGTATGTAT	Sercinaspeinelyashservroserelyuyselyueuserserelincysserashinr TCTCAAGATCAGGGGAATAGCCCATCTGGAAAGGGATTATCCAGTCAATGTCCAAACACA	480	440 455 GluasnLysMetSerTyrGlnSerSerSerThrGluGluArgArgLeuAsnTyrGlyGly GAAAATAAAATGTCATACCAATCTTCAAGTACAGAAGAAAGA	1440
$ \begin{array}{c c c y argleut Ty Valhis Gi y Leu Ser Ly so Gi u Gi na la Ser Al a $	140 145 150 155			
160165170173LysGlyArgThrGlnGlyGlySerGInSerSerTyrValLeuGInThrGluGluLeuVal AAAGGTAGAACAAAGGTGGATCCCAAAGGGTGATCCCAAAGGAGTTATGTTCTCCAAAACTAGAAGTATGGT480485490495GluGlyLysSerGInTLeGInThrGluGlySerGInTrpSerGlyGlnAsn GAAGGCAAGTCCAAACCAAAGGACCAAACCAAGGTGTGGCCAAAACTAGGTA600600600600180185190195500505510515ValAsnLysGinGinArgGluGluHisSerSerLysLeuGinThrSerLeuHisProAlaHis GTGGTGACGAAGAGGAGCAACTACCAAGCTACCCAAGCTCCACACT500505530535200205210215520525530535535ValValAspValArgGluGluHisSerSerLysLeuGinThrSerLeuHisProAlaHis GGTGGTGACGAAGAGGCCCCAAAGTCCAAGCTCCAACTCCACTCCTCCGCACAT7205255305355352202252302355305355505555555505555555505555555505555555555505555555505	GluLysArgLeuTrpValHisGlyLeuSerLysGluGlnAlaSerAlaSerGlyAlaGln GAAAAAAGGCTATGGGTTCATGGACTAAGTAAAGAACAAGCTTCAGGCTCTGGTGCACAA	540	460 465 470 475 LysSerThrGlnLysAspValSerGlnSerSerIleSerPheGlnIleGluLysLeuVal AAGAGCACGCAGAAAGATGTATCCCAAAGCAGTATTTCTTTC	1500
LysGigArgThrGinGiyGiySerGinSerSerTyrValLeuGinThrGinGiuGeuVal480485490495AAAGGTAGAACACAGGTGACCCCAAGGTGAGTCTCCCCAAACTGAGACACAGGTAGTACTGATCGTCCCAAACTGAGACAAGGTGAGCACCAAGGTGAGCACCAAATGGACAGAGTGAGCAACTGGTCCGAAATGGACAAGGGCAACTGCAAGAGGCAACTGAGCCAAATGGACAAATGAAGGCAAAGTGAACGAGAAATGCCAAAAGGAAAATGCAAAAGGCAAAGTGACAACAA480485490495AAAGGTAGAACACAGGTGAGCACGATAGCCAAGCAGTATGCCCAAACTGACCAAACTGCCCAAGACGGACCAAATGGACAAGGGCACCCAAATGGCCAAGCAAATGGCCAAGCGAAATGCCAAAAGGCAAATGCAAAAGGCAAATGCAAAAGGCAAATGCAAAAGGCAAACTGCACGATAGCAAAAGGCAAACTGCAGCAGACAAA600600505510515ValAaspvalArgGiuGiuHisSerSerLiyEuGinThrScinAspGluEuLeuSerHisGluGin GCGAAAAGGAAAGGAAACTCCAACGAAACTCCAACGAAACTCCAACGGAAACTCCAACAGGAAACCAGGAAACCAGGAAATCCAAAACAGGAAATCCAAAAAGGCAAACGGAAATCCAAAAACAGGAAATCCAAAAACTGCAACAGAAAACTCCAACAGAAAACTCCAACAGAGAAACTCCAACAGAGAAACCCAAGACAAAAAGCCAACAAAAAAAA	160 165 170 175			
180185190195ValAsnLysGinGinArgGluhrtysAsnSerHisGinAsnLysGiyHisTyrGinAsn GTTAACAACAACGTGAGACTAAAAATTACCACAAAATAAAGGGCATTACCAAAAAT660505510515ValValAspValArgGluGluHisSerSerLysLeuGinThrSerLeuHisProAlaHis GTGGTGACGGAGAGACATCTCAGGAGAAACTTCCACTCCACTCCACCACCACCAT C20205210215ValValAspValArgGluGluHisSerSerLysLeuGinThrSerLeuHisProAlaHis GTGGTGACGTGAGAGAGACATCTCAGAAAACTACAAACATCCACTCCCACCCCCCACCAT C20225230535220225230235540545550555ChAGACAGGACCCCAACATGGACCCAAAGAGACCACATTTTTACTACCCAAGATGAGCCTCCTAGTA CAACACAGACTCCAACATGGACCCCAAAGAGACAATCCAACAGGACATCCAACGAGACCACACTGAGGCAGAACCACACACA	LysGIyArgThrGInGIyGIySerGInSerSerTyrValLeuGInThrGIUGIULeUVal AAAGGTAGAACACAAGGTGGATCCCAAAGCAGTTATGTTCTCCCAAACTGAAGAACTAGTA	600	480 495 GluGlyLysSerGlnIleGlnThrProAsnProAsnGlnAspGlnTrpSerGlyGlnAsn GAAGGCAAGTCTCAAATCCAGACACCAAATCCTAATCAAGATCAATGGTCTGGCCAAAAT	1560
ValkanLysGlnGlnArgGluThrLysAsnSerHisGlnAsnLysGlyHisTyrGlnAsn GTTAACAAACAGCGGGGGCATTAACAAATCTCATCAAAATGGGCATTACCAAAAT500505510515GTAACAACAGCGGGGACAAACTCCACCAAAAATCCCAAAAATCCCAAAAAA500505510515510515200205210215520525530535535535535536535536 <td< td=""><td>180 185 190 195</td><td></td><td></td><td></td></td<>	180 185 190 195			
2002052102152002052102152102152102152102152102152102152302352102252302352102252302352102452502557yrAsnLysAsnGlnHisGlnThrLysAsnLeuSerGlnAspGlnGluHisGlyArgLys5405452402452502557yrAsnLysAsnGlnHisGlnThrLysAsnLeuSerGlnAspGlnGluHisGlyArgLys54054526026527027511aHistys1LeSerTyrProSerArgThrCiluGluArgGInLeuHisHisGlyGlu GCACATAAAAAACGCAGACACCCCAACGAACACACACACA	ValAsnLysGlnGlnArgGluThrLysAsnSerHisGlnAsnLysGlyHisTyrGlnAsn GTTAACAAACAACGTGAGACTAAAAATTCTCATCAAAATAAAGGCATTACCAAAAT	660	500 505 510 515 AlalysGlyLySSerGlyGinSerAlaAspSerLySGlnAspLeuLeuSerHisGluGln GCAAAAGGAAASTCTGGTCAATCTGCAGATAGCAAACAAGACCTACTCAGTCATGAACAA	1620
ValvalaspvalargGluGluHisSerSerLysLeuGinThrSerLeuHisFrohlaHis 520 525 530 535 GTGGTTGACGTGAGAGGAACATTCAAGTAACTACAAACTTCACTCCATCCTGCACAT 720 225 230 235 CalaspargLeuGlnHisGlyProLysAspIlePheThrThGlnAspGluLeuLeuVal 540 545 550 555 Z40 245 220 255 250 555 540 540 545 550 555 YalasplandLuHisGlnAntractacacatagacacatatacatacacacatagacacacac	200 205 210 215			
GinaspargLeuGinHisGlyProLysAspIlePhethrThrGInAspGluLeuLeuVal 540 545 550 555 ValAlaGInAspAspHisLeuThrGInGInTyrAsnGluAspArgAsnProIleSerThr 780 780 780 780 240 245 250 255 777AsnLysAsnGlnHisGInThrLysAsnLeuSerGInAspGlnGluHisGlyArgLys 780 780 780 780 780 780 780 780 780 780 781 780 780 780 780 780 780 782 780 7	ValAspValArgGluGluHisSerSerLysLeuGlnThrSerLeuHisProAlaHis GTGGTTGACGTGAGAGAGGAACATTCAAGTAAACTACAAACTTCACTCCATCCCTGCACCT 220 225 230 235	720	520 535 536 LysGlyArgTyrLysGlnGluSerSerGluSerHisAsnIleValIleThrGluHisGlu AAAGGCAGATACAAACAGGAATCCAGTGAGTCACATAATATTGTAATTACTGAGCATGAG	1680
CAAGACAGAGACTCCAACATGGACCCAAAGACATTTTTACTACCCAAGATGAGCTCCTAGTA 780 240 245 250 255 TyrasnLysAsnGlnHisGlnThrLysAsnLeuserGInAspGInGluHisGlyArgLys 840 TAGCCCTGTGCTTAGCAACCACAACATGGACCAATATCCAATATCTAGCA 1 260 265 270 275 1	GinAspArgLeuGinHisGivProLysAspTiePheThrThrGinAspGluLeuLeuVal		540 545 550 555	
240 245 250 250 250 260 265 270 275 AlaHisLysIleSerTyrProSerSerArgThrGluGluArgGlnLeuHisHisGlyGlu 840 TAGCCCTGTGGCTAGAGAGGCAGGACCCAATAGCAAGGTGTCACCCGACC 1 280 285 290 295 295 245 200 295 240 240 240 245 1 280 285 290 295 295 240 245 240 245 290 295 1 LysSerValGinLysAspValSerLysGlySerIleSerIleSerIleGinThrGluGluLysIle 960 1 TCAAACTTCCAATAAAAGGGGATCATTTCTGCTTGGCTTaaaaaaaaaa	CAAGACAGACTCCAACATGGACCCAAGAGACTTTTTACTACCCAAGATGAGCTCCTAGTA	780	ValAlaGlnAspAspHisLeuThrGlnGlnTyrAsnGluAspArgAsnProIleSerThr GTTGCCCAAGATGATCATTTGACACAACAATATATATGAAGACAGAAATCCAATATCTACA	1740
Tyrashuysashcininistianininitysashleuberuinaspoindiumistyarguys ************************************			•	
260 265 270 275 TCAGTGAAGTQTTTGATGGTTCTGAGAGGCAGACCCCCATGTGGTCCCAGATCCTTGGTC 1 AlaHisLysIleSerTyrProSerSerArgThrGluGluArgGInLeuHisHisGlyGlu 900 CATGGAAGACCTCCCATGCTCCCATGCGCCCCAGATCCTTGGTC 1 280 285 290 295 TCAATGGAAGCCACCATCTCCCATGCCCTTGCCATAGAGAGAG	TYTASHLYSAShUTHI SUITITLYSAShLeuserginaspundiuni Suiyarguys TATAACAAGAATCAACACCAGACAAAAAATCTCAGTCAAGATCAAGAGCATGGCCGGAAG	840	TAGCCCTGTTGCTTAGCAACCACTTGAAAAGCTGGACCAATAGCAAGGTGTCACCCGACC	1800
AlaHisLysIleSerTyrProSerSerArgThrGluGluArgGlnLeuHisHisGlyGlu	260 265 270 275		TCAGTGAAGTCTTTGATGTTTCTGAGAGGCAGACTCCCATGTGGTCCCAGATCCTTGGTC	1860
280 285 290 295 TCAAACTTCCAATAAAGGGATCATTTTCTGCTTaaaaaaaaaa	AlaHisLysIleSerTyrProSerSerArgThrGluGluArgGlnLeuHisHisGlyGlu GCACATAAAATATCATACCCGTCTTCACGTACAGAAGAAAGA	900	CATGGATGACACCACCTTCCCATGCTTCCTTGCATTAGGCTTTCTAAACCCGGAGCCCCT	1920
LysSerValGlnLysAspValSerLysGlySerIleSerIleGlnThrGluGluLysIle	280 285 290 295		TCAAACTTCCAATAAAGGGATCATTTTCTGCTTaaaaaaaaa	1962
	LysSerValGlnLysAspValSerLysGlySerIleSerIleGlnThrGluGluLysIle			
	AGAGTGTACAGAAAGATGTATCCAAAGGCAGCATTTCTATCCAAACTGAAGAGAAAATA	960		

FIG. 2. Structure of SgII transcript. Nucleotide sequence of the λ VSRP-XXI insert with translated amino acids indicated above. Amino acid numbering starts with residue 1 of the predicted mature 559-residue protein. Diamonds at Asn-118 and Asn-249 indicate two sites of potential oligosaccharide attachment. Shaded nucleotides represent the 3' nontranslated structure absent from λ VSRP-XII. The polyadenylylation signal is underlined.

 λ VSRP-XII, hybridized to a moderately abundant message with an approximate size of 2.4 kb in RNA from the seminal vesicles (data not shown). The size determined for the mRNA species hybridizing to the BP1028 probe is consistent with the size expected for a regularly polyadenylylated message. Assuming that \approx 240 nucleotides are lacking from the 5' end of λ VSRP-XII yields a predicted size of almost 2.5 kb for the message but does not allow a further increase in size by the addition of a 3' extending poly(A) tail. Therefore, the possibility that the unique 3' part of λ VSRP-XII constituted a cloning artifact was investigated by nuclease S1 mapping of RNA from the seminal vesicles. As $\lambda VSRP-XII$ protected two transcripts specifically hybridizing to probes BP1028 and XE613, but failed to protect a message hybridizing to both probes (data not shown), we conclude that the 659 unique 3' nucleotides of λ VSRP-XII represent a cloning artifact unrelated to the SgII message.

Properties of SgII. The deduced translation product of the λ VSRP-XXI insert contains the almost complete primary structure of the SgII precursor. As shown in Fig. 4, the entire structure of the SgII precursor is very similar to that of the SgI precursor (11). A calculation of the number of identical amino acid residues in the two protein precursors reveals an overall identity of 78%. From the comparison with the SgI precursor (11), it is possible to predict that the incomplete signal sequence of the SgII precursor lacks the initiating

methionine residue and another charged residue and that the most probable site for signal peptidase cleavage follows Gly-1 (Fig. 2). It is therefore suggested that the secreted protein comprises 559 amino acid residues. This polypeptide has a calculated isoelectric point of 9.79 and a molecular weight of 62,931. Asparagine-linked oligosaccharide attachment at residue 249 may increase the molecular mass of SgII but from the comparison with SgI it can be assumed that asparagine-linked carbohydrate attachment is not likely to occur at residue 118, which is not glycosylated in SgI (11).

The SgII structure contains eight internally homologous regions. These regions of 60 amino acid residues cluster in three distinct groups of repetitive units that also display an internal relationship (Fig. 5). An optimal alignment is provided by four consecutive type I repeats, having 53–87% conserved amino acids, two consecutive type II repeats, with 33% conserved amino acids, and two type III repeats with 27% conserved amino acids. An identical alignment done for SgI demonstrated six repeats. From this, the smaller size of SgI compared to SgII is explained by the lack of two type I repeats.

As previously described for SgI (11), significant sequence similarity exists between the nontranslated nucleotides at the 3' end of SgII-t1 and the nontranslated nucleotides at the 3' end of both the rat and the murine seminal vesicle secretory protein IV (21, 22). As recently described for SgI (23), the signal peptide of SgII is also similar to that of the major



FIG. 3. Northern blot analysis. Twenty micrograms of RNA from seminal vesicles (lane 1) and epididymis (lane 2) was separated in a 1% agarose gel, blotted to nylon membranes, and hybridized with the BP1028 probe. Membranes were exposed to Kodak XAR-5 film for 4 or 70 h. Arrows indicate positions of 28S and 18S rRNA.

seminal clotting protein of guinea pig seminal vesicles, GP1 (23, 24).

DISCUSSION

The secretion produced by human seminal vesicles has been shown to contain three predominant protein components of 52, 71, and 76 kDa (3). These three proteins, commonly referred to as the Sg, are degraded during semen liquefaction by PSA (5–7) and are distributed on the surface of ejaculated spermatozoa (11). The Sg share common antigenic epitopes, as demonstrated by immunological cross-reactivity with polyclonal rabbit IgG raised against a peptide fragment encompassing residues 85–136 of the 52-kDa component, SgI (3, 11, 25). We have previously described the sequence of the 439 amino acids of SgI (11). The close structural similarity between SgI and the 559 amino acids of SgII, reported in the



FIG. 4. Structural similarity between SgI and SgII. The computer program COMPARE (17) was used to compare the primary structures of SgI and SgII precursors. The size of the sliding window was set to 21 and the stringency was set to 14. The result was displayed by the program DOTPLOT (17).

present investigation, suggests a high degree of immunological cross-reactivity between the Sg. In particular, recognition of the 52-kDa, as well as the 71- and 76-kDa proteins by polyclonal rabbit IgG against the peptide fragment, encompassing residues 85-136 of SgI, receives structural support from the fact that 44 of the 52 amino acids in the SgI peptide are conserved in SgII. A similar immunological crossreactivity has also been seen by a monoclonal antibody. Analysis of freshly ejaculated human semen with the monoclonal antibody MHS-5 identified a common epitope of the three major proteins of 58, 69, and 71 kDa (26, 27). Similar to the properties described for Sg, the three MHS-5 reacting proteins, commonly referred to as the seminal vesiclespecific antigen (SVSA), are degraded during semen liquefaction by PSA (27, 28) and distributed on the surface of ejaculated spermatozoa (29). The similar properties of the SVSA and Sg suggest that these proteins are identical, although this remains to be confirmed.

The molecular mass of 49.6 kDa calculated for the polypeptide backbone of SgI closely agrees with the mass of 52 kDa estimated by SDS/PAGE (3, 11). However, the difference is greater between the mass of 62.9 kDa calculated for the mature SgII and the 71 and 76 kDa estimated by SDS, PAGE. This suggests an aberrant mobility on SDS/PAGE that may be due to the relatively high positive net charge of SgII (pI 9.79) or to posttranslational modifications. It is likely that the 71- and 76-kDa components have the same amino acid sequence and differ in glycosylation only. The evidence for this is the demonstration that the human genome contains two Sg genes, one for SgI and one for SgII and that only the 76-kDa component stains for carbohydrate (unpublished work). Attempts to confirm the N-terminal sequences of the 71- and 76-kDa proteins electroblotted onto poly(vinylidene difluoride) membranes have been unsuccessful (results not shown). This has previously been reported for SgI (11) and may be due to cyclization of the N-terminal glutamineresidue in SgII. Confirmatory deglycosylation and N-terminal deblocking with pyroglutamate aminopeptidase is limited by the inability to isolate the 71- and 76-kDa proteins from freshly ejaculated semen and by the available amounts of purified proteins obtained from SDS/PAGE of seminal vesicle secretion. In conclusion, we suggest that the 71-kDa component corresponds to the nonglycosylated 559-residue SgII, whereas carbohydrate side-chain attachment at asparagine residue 249 explains the size difference between the 71and 76-kDa components on SDS/PAGE.

Sg occur in disulfide-linked high molecular mass complexes in seminal vesicular secretion and in freshly ejaculated semen (4). A single cysteine residue at position 216 in SgI (11) and two cysteine residues at positions 136 and 337 in SgII provide the structural background to the occurrence of Sg in these disulfide-linked intermolecular complexes. Homodimerization of SgI and homomerization of SgII is possible, but heteromer complexes engaging two 52-kDa SgI, one 71-kDa SgII, and one 76-kDa SgII would significantly increase the size of the formed complexes and provide a relatively symmetrical molecule. However, the lack of clarifying protein data makes the issue speculative.

Proteolytic fragmentation of Sg has only been shown to occur C terminal of certain tyrosine and leucine residues in SgI (11, 30, 31). This is consistent with the demonstrated restricted chymotrypsin-like substrate specificity of PSA (7, 32). As yet, no data have been reported on the PSA-mediated proteolytic fragmentation of SgII and only one of the cleavage sites in SgI is conserved in SgII (leucine residue at position 84).

As reported previously, the expression of SgI seems to be restricted to the seminal vesicular tissue, whereas Sg have been detected in the secretory epithelium of the epididymis by polyclonal rabbit IgG against the 52-residue SgI fragment

	HKIQYPSSRTEERQLHHGERSUQKDUSKCSISIQTEEKIHGKSQHQUTIH <u>B</u> SQDQEHGHKE	261-320
Ť.	NKISYOSSSTEERHINCGEKGIQKGUSKGSISIQTEEQIHGKSONQURIPSQAQEYOHKE	321-380
1	NKISYDSSSTEERRLNSGEKDVOKGUSKGSISIOTEEKIHGKSONDUTIPSODOEHGHKE	381-440
	NKMSYDSSSTEERR_NYGGKSTQKDUSQSSISFQIEK_UEGKSQIOTPHPNODQUSGQNA	441-500
	RLUVHGLSKEDASASGADKGATOGGSOSSYVLDTEELVUHKOORETKHSHUHKGHYONV	141-200
11	UDUREEHSSKLOTSLHPAHQDRLQHGPKDI TTODELLUYNKNOHOTKNLSQDQEHGRKR	201-260
	GKSGOSADSKODLESHEOKGRYKOESSESHNIVIITEHEUAODDHLTOOYNEDRHPIST	501-559
111	ATKSKOHLGGSOQLLNYKOEGROHDKSKGH HNIVIHHKGGQAHHGTQNPSODQGNSPSG	70-129

FIG. 5. Internal repeats of SgII. Based on top scores for sequence similarity, the primary structure of SgII was divided into three groups of repeats as indicated by roman numerals. Shaded amino acids are conserved between groups. Residues are presented in the standard one-letter code for amino acids.

(11). The present investigation shows evidence in strong support of SgII expression in both the epididymis and the seminal vesicles, although cDNA cloning and Northern blots showed a higher abundance of SgII messages in the seminal vesicular than in the epididymal tissue. This can be explained by tissue-specific regulation of transcription or perhaps confinement of SgII transcription to a limited number of specialized cells in the secretory epithelium of the epididymis.

We thank Ingrid Wigheden for her expert technical assistance. This investigation was supported by the Swedish Medical Research Council (Grants B92-13X-7903-6B and B92-13X-8660-4B), the Faculty of Medicine at the University of Lund, the Research Fund at Malmö General Hospital, the Magnus Bergvall Foundation, the Alfred Österlund Foundation, and the Fundacion Federico S.A.

- 1. Mann, T. & Lutwak-Mann, C. (1981) Male Reproductive Function and Semen, eds. Mann, T. & Lutwak-Mann, C. (Springer, New York).
- Lilja, H. & Laurell, C.-B. (1984) Scand. J. Clin. Lab. Invest. 44, 2. 447-452.
- 3. Lilja, H., Laurell, C.-B. & Jeppsson, J.-O. (1984) Scand. J. Clin. Lab. Invest. 44, 439-446.
- Lilja, H. & Laurell, C.-B. (1985) Scand. J. Clin. Lab. Invest. 45, 635-641.
- Lilja, H., Oldbring, J., Rannevik, G. & Laurell, C.-B. (1987) J. 5. Clin. Invest. 80, 281-285.
- Lilja, H. (1985) J. Clin. Invest. 76, 1899-1903.
- Christensson, A., Laurell, C.-B. & Lilja, H. (1990) Eur. J. 7. Biochem. 194, 755-763.
- Watt, K., Lee, P.-J., M'Timkulu, T., Chan, W.-P. & Loor, R. 8. (1986) Proc. Natl. Acad. Sci. USA 83, 3166-3170.
- Lundwall, Å. & Lilja, H. (1987) FEBS Lett. 214, 317-322. 0
- 10. Schaller, J., Akiyama, K., Tsuda, R., Hara, M., Marti, T. & Rickli, E. (1987) Eur. J. Biochem. 170, 111-120.
- 11. Lilja, H., Abrahamsson, P. A. & Lundwall, Å. (1989) J. Biol. Chem. 264, 1894-1900.

- 12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 14. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
- 15. Staden, R. (1982) Nucleic Acids Res. 10, 2951-2961.
- Staden, R. (1982) Nucleic Acids Res. 10, 4731-4751. 16.
- Devereux, J., Haeberli, P. & Smithies, O. (1984) Nucleic Acids 17. Res. 12, 387-395.
- 18. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochemistry 18, 5294-5299.
- 19. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) Biochemistry 16, 4743-4751.
- 20. Goldberg, D. A. (1980) Proc. Natl. Acad. Sci. USA 77, 5794-5798.
- 21. Mansson, P.-E., Sugino, A. & Harris, S. E. (1981) Nucleic Acids Res. 9, 935–946.
- Chen, Y. H., Pentecost, B. T., McLachlan, J. A. & Teng, C. T. (1987) Mol. Endocrinol. 1, 707-716.
- 23. Hagstrom, J. E., Harvey, S., Madden, B., McCormick, D. & Wieben, E. D. (1989) Mol. Endocrinol. 3, 1797-1806.
- 24. Moore, J. T., Hagström, J., McCormick, D., Harvey, S., Madden, B., Holicky, E., Stanford, D. R. & Wieben, E. D. (1987) Proc. Natl. Acad. Sci. USA 84, 6712-6714.
- 25. Lilja, H. & Jeppsson, J. O. (1985) FEBS Lett. 182, 181-184.
- Herr, J. C., Summers, T. A., McGee, R. S., Sutherland, W. M., Sigman, M. & Evans, R. J. (1986) Biol. Reprod. 35, 773-784.
- McGee, R. S. & Herr, J. C. (1987) *Biol. Reprod.* 37, 431–439. McGee, R. S. & Herr, J. C. (1988) *Biol. Reprod.* 39, 499–510. 27.
- 28.
- Evans, R. J. & Herr, J. C. (1986) Anat. Rec. 214, 372-377. 29
- Schneider, K., Kausler, W., Tripier, D., Jouvenal, K. & 30. Spiteller, G. (1989) Biol. Chem. Hoppe-Seyler 370, 353-356.
- Lilja, H. (1990) Andrologia 22, Suppl. 1, 132-141. 31
- Akiyama, K., Nakamura, T., Iwanaga, S. & Hara, H. (1987) 32. FEBS Lett. 225, 168-172.