Production of SVP-1/3/4 in Guinea Pig Testis
CHARACTERIZATION OF NOVEL TRANSCRIPTS CONTAINING LONG 5'-UNTRANSLATED REGIONS AND MULTIPLE UPSTREAM AUG CODONS

Michael P. Fautsch, Monique M. Perdok and Eric D. Wieben
From the Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, Minnesota 55905

The GP1G gene of the guinea pig codes for three of the four abundant seminal vesicle secretory proteins produced in this species. This gene is expressed at highest efficiency in the seminal vesicle (SV) from a promoter that contains a canonical TATA box and CCAAT box. However, GP1G gene transcripts and proteins have also been identified in other tissues. To investigate the structure of GP1G transcripts produced in the testis, cDNA clones were isolated by screening a testis library. Three unique cDNAs (TSM1–3) were isolated. Each of these clones contained a 3'-untranslated region (UTR) and coding region identical to that of the seminal vesicle transcript. However, the 5'-UTRs of the testis transcripts were significantly longer than that found on the SV mRNA (416–646 nucleotides compared with only 23 nucleotides for the SV). Each of these alternatively spliced 5'-UTRs incorporated the SV promoter elements into transcribed sequence, and each contained multiple upstream AUG codons predicted to abolish translation of the major open reading frame. Nevertheless, each of the testis transcripts was capable of directing the synthesis of a signal peptide, the second exon codes for the functional homologs of SVP-1 are human semenogelin I and II (9), rat SVS II (10), and mouse semenoclotin (11). All of these genes are organized into a three exon-two intron structure in which the first exon contains sequences responsible for synthesis of a signal peptide, the second exon codes for the protein product, and the third exon is exclusively 3'-untranslated region (3'UTR) (9, 12, 13). Although there is clear homology between the 5'-flanking regions and first exons of these genes, their protein coding regions have experienced rapid evolutionary change. Hence, this gene family is sometimes referred to as the REST gene family (rapidly evolving seminal vesicle transcribed) (12).

Although most members of the REST gene family have been reported to be expressed only in the seminal vesicle, transcripts for some of the REST proteins have been identified in other tissues. Semenogelin II is produced in the epididymis (14), rat SVS II transcripts are found in the prostate (15), and rat SVS IV mRNA is abundant in skeletal muscle (16). The finding that SVP-1 also has extensive homology to SKALP/elafin led to the discovery that SVP-1-related transcripts were expressed outside the seminal vesicle in tissues such as liver, lung, kidney, and testis (13). Of particular interest was the finding that the testis produced longer SVP-related mRNAs that were not found in any of these other tissues.

Here we report the characterization of three novel SVP cDNAs from guinea pig testis (TSM1–3). Each cDNA maintains the coding and 3'-untranslated region of the seminal vesicle transcript (called GP1) but varies in length and sequence of the 5'-untranslated region (5'-UTR). In vitro translation results demonstrate that the TSMs are functional despite the presence of multiple upstream AUG codons in each of the mRNAs. In each case, multiple antibody-precipitable proteins are synthesized, suggesting that the 5'-UTR extensions found in TSM1, TSM2, and TSM3 may be responsible for translational regulation.

EXPERIMENTAL PROCEDURES

Library Screening—A guinea pig testis cDNA library (custom made by Stratagene in Lambda-Zap II) was screened with either a 32P-labeled random primed 1400-bp GP1 probe (3) or a 175-bp 5'-end fragment from TSM1. Southern blot analysis of isolated clones revealed 13 positives. Four of the 13 clones represented TSM2 sequence. TSM1 and TSM3 were represented by individual clones. The remaining clones were shorter versions of the other cDNAs. TSM1–3 were sequenced in both directions with either an Applied Biosystems or a Pharmacia LKB Biotechnology A. L. F. DNA sequencer.

Northern Blot Analysis—Total RNA was isolated from guinea pig testis or seminal vesicle (17). mRNA was isolated from total guinea pig testis RNA using mRNA isolation kit III (Promega; No. 5300). RNAs were separated on a 1% agarose, 2.2 M formaldehyde gel and transferred to MSI nylon by diffusion using 10 x SSC. The transferred RNA was probed with a 32P]dATP random primed 1400-bp GP1 cDNA. After washing in 0.1 x SSC and 0.1% SDS at 50 °C for 20 min, the blot was exposed to Kodak XAR film.

Western Blot Analysis—Guinea pig testis was homogenized in 125 mM Tris pH 8, 2% SDS, 10% glycerol, 5% 2-mercaptoethanol, 0.01% bromophenol blue. The homogenate was boiled for 5 min. Then 10% SDS-PAGE was run and transferred to nitrocellulose by electroblotting with 0.1 x SSC. The transferred membrane was probed with 32P-labeled random primed 5300 bp GP1 probe (3) and exposed to Kodak XAR film.
formed at 37 °C for 60 min in the presence of cap analogue m 7GpppG, 35299, 5
TSM1–3 (35122 and 35479). Amplified DNA was cloned into pCR2.1 anchor primer and nested oligonucleotides that would recognize
Technologies, Inc., No. 18374–025). Oligonucleotides specific for TSM1
identical to GP1, suggesting that they originated from the
2). Both of these clones had 3’-UTRs and coding sequences identical to GP1, suggesting that they originated from the
GPIG locus. The additional length of TSM1 and TSM2 was due to extensions of the 5’-UTRs on each clone. TSM1 contained an extra 640 bp while TSM2 was extended by 422 bp relative to GP1. The size difference between these two clones reflects both a truncated 5’-terminus on TSM2 and an internal deletion of 153 nts within the 5’-UTR of the TSM2 cDNA. The heterogeneity in the 5’-UTR sequence prompted us to rescreen the testis cDNA in an attempt to isolate additional clones that might correspond to the third size class of testis GPIG-related mRNAs. Using the first 175 bp of TSM1 as a probe, an additional screening of the library yielded TSM3, which is identical to the other GPIG transcripts in the 3’-UTR and coding regions, but had an additional 77-bp internal deletion immediately upstream of the 153-bp deletion found in the 5’-UTR of TSM2 (Fig. 2).

Since the TSM1, TSM2, and TSM3 clones have heterogeneous 5’ ends, 5’-RACE was performed to determine if these three cDNAs originated from mRNAs having a common 5’ terminus. Using three different oligonucleotides specific for TSM1, TSM2, and TSM3 to prime first strand synthesis (see “Experimental Procedures”), 16 different clones were identified and their 5’ ends mapped (Fig. 3). Although there was substantial heterogeneity in the 5’ termini of the resulting RACE clones, all three TSM transcripts have larger 5’ ends compared with their cDNAs. This suggests that TSM1–3 transcripts originate from a common promoter.

As noted above, the major ORF of the TSM clones can code for the same secretory proteins found in the seminal vesicle. However, the AUG codon that initiates translation of this ORF in the SV mRNA (AUG2) is preceded by an additional in-frame AUG codon in the testis mRNAs (Fig. 2, AUG1). Translation beginning at this first AUG would yield an SVP-1/3/-4 precursor protein with an additional 15-amino acid extension to the signal peptide that is used to direct secretion of the seminal vesicle protein. Significantly, use of either of these two AUG codons would require the skipping of numerous upstream AUG codons (uAUG) present in the 5’-UTR of the testis clones. The shortest member of this group, TSM3, has 10 uAUGs before AUG1 while TSM1 has 17 uAUGs spread throughout all 3 forward reading frames before AUG1 (Table I).

Genomic Structure—Sequence comparisons of the 5’ extensions on the TSM clones to the previously determined genomic sequence of GPIG reveals that the proximal 237 nts of the extended 5’-UTR is derived by transcription of 5’-flanking sequences that lie immediately upstream of the first exon used in the seminal vesicle. Thus, this extended exon is 342 nts long as it is used in the testis compared with 105 nts as used in the SV (Fig. 4; exon 1 and exon D). As a consequence of the extension of this exon, several conserved promoter elements that are

![Fig. 1. Novel GPIG transcripts accumulate in guinea pig testis.](image-url)

3 μg of testis mRNA and 10 ng of total seminal vesicle RNA were separated by electrophoresis on a 1% agarose, 2.2 M formaldehyde gel and transferred to nylon membrane. The blot was then probed with a GPIG 32P-labeled random primed cDNA probe. The three novel testis transcripts are identified by arrows. The source of RNA is indicated below each lane. The lane marked Testis/SV contains a mixture of 3 μg of testis mRNA and 10 ng of seminal vesicle RNA.
likely to contribute to the expression of GP1G in the seminal vesicle (including a CCAAT box at -41, a TATA box at -241, and the 290 element found in most members of the REST gene family) become part of the 5'-UTR of the testis transcripts. Further analysis of the genomic sequence reveals that the internal sequence differences between TSM1–3 are due to the inclusion or exclusion of all or part of a 230 nt alternatively spliced exon located 1.5 kb upstream of the 5'-extended first coding exon (Fig. 4; exon C). TSM2 contains only the first 77 bp of exon C while TSM3 does not contain this exon at all. Two additional short exons are required to code for the 5'-termini of the TSM clones as defined by the RACE clones shown in Fig. 3 (Fig. 4, and data not shown). All of the intron-exon junctions shown in Fig. 4 conform to the GT-AG rule. From the sizes of the genomic clones containing the six exons required to code for TSM1 and TSM2, we estimate that the testis-specific promoter lies at least 30 kb upstream of the seminal vesicle promoter.

Translation—The scanning hypothesis for translation initiation dictates that upstream AUG triplets should constitute a significant barrier to the translation of a downstream ORF (20). As noted previously, each of the TSM clones has multiple uAUGs predicted to code for small peptide products upstream of the ORF that codes for SVP-1/-3/-4. To determine if the TSM transcripts were capable of directing the synthesis of SVP-1/-3/-4 precursors, full-length TSM constructs were used as templates for the synthesis of synthetic, capped mRNAs. Translation of each of these TSM mRNAs in vitro yielded three products that were immunoprecipitable with antisera directed against SVP-1 or SVP-4 (Fig. 5). The middle translation product in each case migrates with SVP-1/-3/-4 precursor produced by in vitro translation of GP1. Although the identity of the products that migrate slower and faster than the GP1 precursor has not been established, the size of the longest product is consistent with the hypothesis that it results from the initiation of translation from an upstream AUG codon. Further analysis of the genomic sequence reveals that the internal sequence differences between TSM1–3 are due to the inclusion or exclusion of all or part of a 230 nt alternatively spliced exon located 1.5 kb upstream of the 5'-extended first coding exon (Fig. 4; exon C). TSM2 contains only the first 77 bp of exon C while TSM3 does not contain this exon at all. Two additional short exons are required to code for the 5'-termini of the TSM clones as defined by the RACE clones shown in Fig. 3 (Fig. 4, and data not shown). All of the intron-exon junctions shown in Fig. 4 conform to the GT-AG rule. From the sizes of the genomic clones containing the six exons required to code for TSM1 and TSM2, we estimate that the testis-specific promoter lies at least 30 kb upstream of the seminal vesicle promoter.

TABLE I

| cDNA  | Length of mRNA (in bp) | Length of 5'-UTR (in bp) | Number of uAUG codons |
|-------|------------------------|--------------------------|-----------------------|
| TSM1  | 2071                   | 646                      | 17                    |
| TSM2  | 1918                   | 493                      | 13                    |
| TSM3  | 1841                   | 416                      | 10                    |
| GP1   | 1403                   | 23                       | 0                     |
The smallest product may be a translation product from an internal AUG such as AUG3 or AUG4. We cannot rule out the possibility that the smaller translated product is a degradation product of one of the larger products. Mutational analysis of these AUGs will help delineate if multiple AUG initiation codons are used. However, these results clearly demonstrate that all three testis-specific mRNAs are capable of producing SVP-1/-3/-4 products in vitro despite the presence of multiple upstream AUGs.

To determine if SVP-1 proteins are synthesized by the testis in vivo, we used both polyclonal mouse anti-SVP-1 (MCM2) or rabbit anti-SVP-1 (081694) antisera to probe Western blots of total testis proteins (Fig. 6). Both of these antisera reacted with proteins the size of SVP-1 on these blots (Fig. 6, lanes 3 and 5). In parallel experiments, proteins co-migrating with SVP-4 were specifically detected by a rabbit anti-SVP-4 antibody (Fig. 6, lane 1). Thus, antibody reactive secretory proteins co-migrating with the SV proteins are synthesized in the testis of mature guinea pigs.

**DISCUSSION**

The data presented here demonstrate that the GP1G gene is expressed in the testis as well as the seminal vesicle, and highlight new regulatory mechanisms that are involved in the production of proteins from this locus. In particular, the structure of the novel testis cDNAs described here reveals that the GP1G gene has a second promoter that is active in the testis. All three of the TSM cDNAs described here appear to originate from a common GC-rich region located at least 30 kb upstream of the promoter used in the seminal vesicle. The occurrence of multiple transcription start sites from this upstream promoter contrasts with the existence of a single, evolutionarily conserved start site for transcripts initiating from the downstream seminal vesicle promoter. The unique start site in the SV probably reflects the presence of a consensus TATA box located at position −27 in the downstream promoter. This TATA box, as well as a CCAAT box and a promoter element conserved in REST genes (the −90 element), are all included in the first coding exon (Fig. 4, exon D) of each of the TSM cDNAs. A similar incorporation of somatic promoter sequences into the
5'-UTR of a testis-specific mRNA has been described for the carbonic anhydrase gene in the chicken (21).

The testis-specific exon that lies to the 5' side of the first coding exon is alternatively spliced in TSM1–3. At present, we do not know the functional consequences of these alternative splicing events. The absence of a long ORF upstream of the SVP reading frame suggests that the resulting sequence differences are restricted to 5'-UTR. Thus, there are apparently no direct consequences for the protein coding capacity of these mRNAs. It remains possible that the different 5'-UTRs could influence either the efficiency of translation and/or choice of translation start site from the TSM messages in vivo.

The scanning model of translation suggests that the 43 S ribosome preinitiation complex will interact at or near the 5' cap structure of an mRNA and migrate in the 3' direction, sequentially evaluating the sequence until it reaches an AUG that is in an appropriate context (22). Ninety percent of all mRNAs follow this dogma (23). However, many proto-oncogenes and other potent regulatory proteins contain transcripts whose 5'-UTR contains multiple upstream AUGs (20). It has been postulated that these transcripts contain upstream AUGs as a way of regulating the expression of their protein product, and this seems likely in the case of the TSM RNAs.

Translation of a synthetic mRNA corresponding to the GP1 mRNA expressed in the SV is efficient and produces a single protein (4) as predicted by the scanning model. Although translation of the TSM RNAs is not as efficient as that of GP1, each of the TSM constructs directs the synthesis of at least three different products precipitable with antisera directed against SVP proteins (Fig. 5). Since GP1 yields only a single stable protein product, the accumulation of multiple products from the TSM RNAs is not easily explained by instability of a single primary translation product. An alternative trivial possibility is that the multiple translation products derive from partially degraded or truncated TSM transcripts. This hypothesis would also predict the accumulation of translation products initiated at a number of other internal AUGs within the SVP reading frame.However, no such products are detected. Thus, we favor the hypothesis that the extended 5'-UTRs of the TSM RNAs lead to a lack of precision in the initiation of translation from the SVP ORF. Specifically, it appears likely that translation of the TSM RNAs initiates at more than one of the four in-frame AUG codons near the start of the SVP ORF (marked AUG1-AUG4 in Fig. 2). The first AUG lies 15 codons upstream of the initiation codon used in the seminal vesicle while the third and fourth in-frame AUGs lie 13 and 44 codons downstream of the SV translation start site.

In principle, the alternative initiation of translation could lead to alternative targeting of the translation products. Proteins initiated at the upstream AUG1 would have an extended signal sequence. Such amino-terminal extensions to functional signal sequences have been shown not to interfere with the secretion of β-thyrotropin or β-1,4-galactosyltransferase isoforms (24, 25). However, signal sequences have also been postulated to function as modulators of protein folding (26). Since stably folded precursor proteins are more likely to be translocated incompetent, the addition of 15 amino acids to the amino terminus may cause or reduce premature folding of SVP-1/3/4 protein and influence its ability to be secreted (27, 28). Although the effect of initiating translation at AUG1 is difficult to predict, it seems likely that proteins initiated at AUG3 or AUG4 would lack a cleavable signal peptide altogether and would be predicted to accumulate intracellularly. Alternative translation initiation as a mechanism for regulating protein targeting has been reported previously for probasin (29). In this case, translation initiation at a downstream AUG results in the synthesis of a probasin isoform that accumulates in the nuclei of prostatic epithelial cells. In the case of TSM1–3, it is tempting to speculate that intracellular forms of the SVPs could become cross-linked by transglutaminase to form part of the cytoskeleton or nuclear matrix.

Of course, any use of internal AUG start codons implies a mechanism for the internal initiation of translation. Several possible mechanisms have been advanced to account for the translation of internal reading frames from cellular mRNAs (30, 31). Several cellular mRNAs have been shown to contain internal ribosome entry sites that direct the translation of downstream reading frames without scanning the entire 5'-UTR (32). An alternative possibility is translational initiation using a discontinuous scanning mechanism (ribosome shunting) such as that described for the initiation of translation from cauliflower mosaic virus and adenovirus mRNAs (33, 34). Further experiments will be required to determine how translation is initiated from the SVP reading frame of the TSM RNAs.

Acknowledgments—We thank the Mayo Molecular Biology Core Facility for efforts in sequencing the TSM clones and Debbie Pearson for assistance with the manuscript.

REFERENCES
1. Aumuller, G., and Steitz, J. A. (1990) Int. Rev. Cytol. 121, 127–231
2. Moore, J. T., Hagstrom, J., McCormick, D. J., Harvey, S., Madden, B., Holicky, E., Stanford, D. R., and Wieben, E. D. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 6712–6714
3. Hagstrom, J. E., Harvey, S., Madden, B., McCormick, D., and Wieben, E. D. (1989) Mol. Endocrinol. 3, 1797–1806
4. Norvitch, M. E., Harvey, S., Moore, J. T., and Wieben, E. D. (1988) Biol. Reprod. 38, 1155–1160
5. Williams-Ashman, H. G. (1983) Curr. Top. Cell. Regul. 22, 201–275
6. Netides, A. C., and Williams-Ashman, H. G. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 1991–1995
7. Williams-Ashman, H. G., Netides, A. C., Pahalan, S. S., and Lorand, L. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 2322–2325
8. Venezie, C. M., and Deering, N. C. (1976) Andrologia 8, 75–82
9. Ulvsback, M., Lazure, C., Lilja, H., Spurr, N. G., Rao, V. V. N. G., Loffler, C., Hansmann, I., and Lundwall, A. (1992) J. Biol. Chem. 267, 18980–18984
10. Harris, S. E., Harris, M. A., Johnson, C. M., Bean, M. F., Dodd, J. G., Matusik, R. J., Carr, S. A., and Crab, J. W. (1990) J. Biol. Chem. 265, 9896–9903
11. Lundwall, A. (1990) Eur. J. Biochem. 205, 424–430
12. Lundwall, A., and Ulvsback, M. (1996) Biochem. Biophys. Res. Commun. 221, 323–327
13. Hagstrom, J. E., Fautsch, M. P., Perdek, M., Vrabl, A., and Wieben, E. D. (1996) J. Biol. Chem. 271, 11114–11119
14. Lilja, H., Abrahamsson, P.-A., and Lundwall, A. (1989) J. Biol. Chem. 264, 1894–1900
15. Dodd, J. G., Kreis, C., Sheppard, P. C., Hamel, A., and Matsukis, R. J. (1986) Mol. Cell. Endocrinol. 37, 121–129
16. Yoo-Warren, H., Willee, A. G., Hull, J., and Brownell, E. (1993) J. Exp. Zool. 265, 651–658
17. Cahahla, G. G., Savouret, J., Mended, B., West, B. L., Karin, M., Martial, J. A., and Baxter, J. D. (1983) DNA 2, 329–335
18. Horton, R. M., Cai, Z., Ho, S. N., and Pease, L. R. (1990) BioTechniques 8, 528–533
19. Jackson, R. E., and Hunt, T. S. (1983) Methods Enzymol. 96, 50–74
20. Kozak, M. (1991) J. Cell. Biol. 115, 887–903
21. Mezquita, J., Pau, M., and Mezquita, C. (1994) (Mol. Gene) 147, 231–236
22. Kozak, M., and Shatkin, A. J. (1978–1979) J. Biol. Chem., 253, 6568–6577
23. Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8147
24. Russo, R. N., Shaper, N. L., Taatjes, D. J., and Shaper, H. J. (1992) J. Biol. Chem. 267, 9241–9247
25. Januszek, M. M., Gabriel, J. L., Shennan, K. I., Doeherty, K., and Gurr, J. A. (1994) Endocrinology 134, 1859–1867
26. Pfanner, N., and Neupert, W. (1990) Annu. Rev. Biochem. 59, 331–353
27. Eilers, M., and Schatz, G. (1986) Nature 322, 228–232
28. Pluckthun, A., and Knowles, J. R. (1987) J. Biol. Chem. 262, 3951–3957
29. Spence, A. M., Sheppard, P. C., Davie, J. R., Matuo, Y., Nishi, N., McKeehan, W. L., Dodd, J. G., and Matusik, R. J. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 7843–7847
30. Sonenberg, N. (1994) Curr. Opin. Genet. Dev. 4, 310–315
31. Geballe, A. P., and Morris, D. R. (1994) Trends Biochem. Sci. 19, 159–164
32. Jackson, R. J. (1996) in Translational Control (Hershey, J. W. B., Matthews, M. B., and Sonenberg, N., eds) pp. 71–112, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
33. Futterer, J., Kiss-Lazlo, Z., and Hohn, T. (1993) Cell 74, 789–802
34. Yueh, A., and Schneider, R. J. (1996) Genes Dev. 10, 1537–1567