Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 791–795, January 1990
Biochemistry

Murine β1,4-galactosyltransferase: Both the amounts and structure of the mRNA are regulated during spermatogenesis

(gene expression/testis/germ cell differentiation/down regulation)

NANCY L. SHAPER*, WILLIAM W. WRIGHT†, AND JOEL H. SHAPER*‡§

*Cell Structure and Function Laboratory, Oncology Center, and †Division of Reproductive Biology, Department of Population Dynamics, School of Hygiene and Public Health, and ‡Department of Pharmacology and Molecular Sciences, School of Medicine, The Johns Hopkins University, Baltimore, MD 21205

Communicated by Robert L. Hill, September 6, 1989

ABSTRACT **Previously we have shown that the gene encoding murine β1,4-galactosyltransferase (β1,4-GT; UDPgalactose:*N*-acetyl-D-glucosaminyl-glycopeptide 4-β-D-galactosyltransferase, EC 2.4.1.38) is unusual in that it specifies two sets of mRNAs of about 3.9 and 4.1 kilobases (kb). Translation of the 3.9- and 4.1-kb mRNAs results in the predicted synthesis of two related membrane-bound forms of the protein of 386 amino acids (short form) and 399 amino acids (long form), respectively. In this study we have examined the expression of β1,4-GT during murine spermatogenesis. Spermatogonia contain a 4.1-kb transcript that is comparable in size to the β1,4-GT mRNA identified in somatic cells. During differentiation from spermatogonia (2*n*) to pachytene spermatocytes (4*n*), the amount of β1,4-GT mRNA is reduced to barely detectable levels. Continued differentiation to round spermatids (*n*) is coincident with a renewed production of β1,4-GT mRNA to levels comparable with those detected in spermatogonia. However, the characteristic 4.1-kb mRNA detected in spermatogonia is replaced by two truncated transcripts of 2.9 and 3.1 kb. By S1 nuclease analysis, the 2.9- and 3.1-kb transcripts were shown to encode the same open reading frame as the 4.1-kb transcript found in somatic cells. The shorter round spermatid transcripts arise as a consequence of the use of alternative poly(A) signals. Lastly, we show that, in direct contrast to all somatic tissues and cell lines examined to date, male germ cells synthesize only the long form of the β1,4-GT polypeptide.**

UDP β1,4-galactosyltransferase (β1,4-GT; UDPgalactose:*N*-acetyl-D-glucosaminyl-glycopeptide 4-β-D-galactosyltransferase, EC 2.4.1.38) is one member of a functional family of intracellular, membrane-bound enzymes that participate coordinately in the biosynthesis of the carbohydrate moieties of glycoproteins and glycolipids. This enzyme catalyzes the transfer of galactose from UDPgalactose to the acceptor sugar *N*-acetylglucosamine (1).

Historically, β1,4-GT has served as a Golgi marker enzyme for cell fractionation procedures (reviewed in refs. 2 and 3). β1,4-GT has also been localized to the plasma membrane of a variety of cells and tissues by immunohistochemical (4–8) and biochemical procedures (9). Of particular interest is the observation that this enzyme can be localized by immunohistochemical procedures to a restricted domain on the murine sperm surface overlying the acrosome (5, 7). Furthermore, biochemical and immunological probes that can potentially interact with the cell surface β1,4-GT have been shown to block, *in vitro*, the ability of murine sperm to bind zona intact eggs (5). Collectively these observations have led to the hypothesis that the catalytic site of β1,4-GT functions in the binding of sperm to appropriate carbohydrate substrates on the zona pellucida (ref. 5; also see ref. 10).

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.

The full-length murine cDNA encoding this enzyme has recently been described (11, 12). The transcript is ≈4.1 kilobases (kb) in size and contains a 3′ untranslated region of 2582 base pairs (bp), a coding region of 1200 bp, and a 5′ untranslated region that is heterogeneous in size. The gene for β1,4-GT is unusual in that it specifies two sets of mRNA transcripts that differ in length by about 200 bp (11). The longer mRNA transcripts (4.1 kb) initiate upstream of the first two in-frame ATG codons and encode a protein of 399 amino acids. The shorter transcripts (3.9 kb) initiate between the first two in-frame ATG codons and encode a protein of 386 amino acids. The only difference between the two forms of the protein is that the long form contains an NH_2-terminal extension of 13 amino acids. The functional significance of the two forms remains an open question.

In this paper we have examined the expression of β1,4-GT during murine spermatogenesis. We show that both mRNA levels, the mRNA structure, and the form of the protein are regulated during this process. The potential functional significance of these observations is discussed.

MATERIALS AND METHODS

Animals. Specific-age male CD-1 mice (Charles River Breeding Laboratories) were used as the source of spermatogonia, pachytene spermatocytes, and round spermatids. Adult female BD_2F_1 mice were used as the source of the various mouse tissues.

Isolation and Purification of Testis Cell Fractions. Mixed dispersed testis cells were prepared by digesting testes for 10 min with 0.1% (wt/vol) collagenase/0.2% (wt/vol) hyaluronidase/0.03% (wt/vol) DNase/0.03% (wt/vol) soybean trypsin inhibitor in Ham's F-12/Dulbecco's modified Eagle's medium (DMEM), 1:1. Dispersed cells were filtered through nylon mesh (50 μm), centrifuged (400 × *g* for 5 min at 20°C), and resuspended in Ham's F-12/DMEM, 1:1 containing 2% bovine serum albumin. Single cells that remained in suspension after incubation for 30 min at 34°C were harvested by centrifugation (400 × *g* for 5 min at 20°C). Specific germ cell populations were isolated by centrifugal elutriation and Percoll density gradient centrifugation as described by Meistrich *et al.* (13). The homogeneity of each fraction was assessed by histological analysis. Spermatogonia (from 8- and 17-day-old mice), pachytene spermatocytes (from 17-day-old mice), and round spermatids (from mature mice) were at least 70%, 95%, and 95% pure, respectively.

RNA Isolation and RNA Blot Analysis. Total RNA was isolated as described (11). Tissue samples were homogenized immediately in 4 M guanidinium thiocyanate/25 mM sodium citrate/0.1 M 2-mercaptoethanol at 4°C. RNA blot analysis was carried out as described (11).

Abbreviation: β1,4-GT, UDP β1,4-galactosyltransferase.

§To whom reprint requests should be addressed at: The Johns Hopkins School of Medicine, Oncology Center, Room 1-127, 600 North Wolfe Street, Baltimore, MD 21205.

DNA Probes. The probes used for hybridization were derived from the C127 murine β1,4-GT cDNA clones MGT-1, MGT-2, and MGT-P4 (11). The probe for glyceraldehyde-3-phosphate dehydrogenase (14), a ubiquitously expressed "housekeeping" gene, was used to monitor the quantity and integrity of RNA. Probes were labeled with [α-^{32}P]dCTP by nick-translation as described (11).

S1 Nuclease Analysis. The probes used for S1 nuclease analysis were prepared as described (11). Briefly, DNA probes A, B, C, D, G, and H (see Fig. 1), which were complementary to the transcribed sequence, were prepared by primer extension of appropriate subclones in M13mp10 or M13mp11 in the presence of [^{32}P]dATP and DNA polymerase. After digestion with an appropriate restriction enzyme and purification, each probe was hybridized to 30 μg of total RNA or 5 μg of poly(A)$^+$ RNA overnight at the following temperatures: probe A, 59°C; probe B, 54°C; probe C, 47°C; probe D, 48°C; probe G, 46°C; and probe H, 41°C. After digestion with S1 nuclease (Boehringer Mannheim), the digestion products were analyzed on a 7% polyacrylamide/8 M urea gel.

RESULTS

Expression of the β1,4-GT Transcript in Various Mouse Tissues. The 1080-bp *Stu* I–*Stu* I fragment (probe C, Fig. 1 *Lower*) was used as a hybridization probe to examine total RNA (10 μg) isolated from various mouse tissues by RNA blot analysis. A broad band of ≈4.1 kb, which as discussed below contains both the 3.9- and 4.1-kb transcripts, was readily detected in total RNA from mouse spleen, duodenum, heart, liver, lung, and kidney (Fig. 1 *Upper*). In contrast, the amount of β1,4-GT transcript in total RNA from brain was barely detectable (Fig. 1 *Upper*); however, longer exposure times revealed a faint band at about 4.1 kb (data not shown). RNA isolated from mature mouse testis cells also contained low amounts of the β1,4-GT transcript (comparable to that from brain). However, when larger quantities of total RNA (30 μg) isolated from testis were analyzed (Fig. 1 *Upper*), the 4.1-kb transcript and three shorter transcripts of 2.6, 2.9, and 3.1 kb were observed. Of the three smaller size transcripts, the 2.9-kb transcript appeared to be the major species. The 2.6-kb species was not consistently seen in all testis RNA preparations.

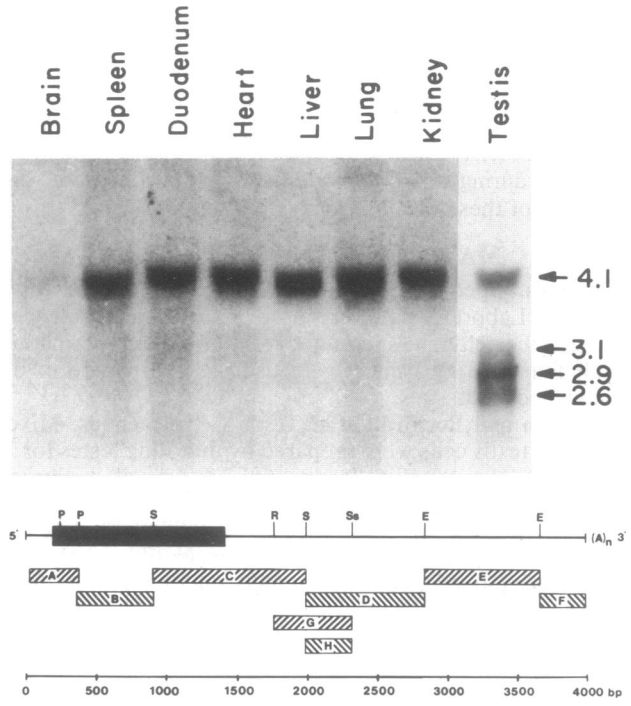

FIG. 1. Expression of β1,4-GT transcripts in various mouse tissues. (*Upper*) RNA blot containing 10 μg of total RNA isolated from mouse brain, spleen, duodenum, heart, liver, lung, and kidney and 30 μg of total RNA isolated from testis was hybridized with probe C. Sizes (in kb) were determined relative to an RNA ladder. (*Lower*) Partial restriction map of the murine β1,4-GT cDNA showing probes used for both RNA blot and S1 analysis. The solid bar indicates the protein coding region; the thin line indicates the 5′ and 3′ untranslated regions; the cross-hatched bars (with letters) represent the individual probes. Probe A, 321-bp fragment (nucleotides-157–144); probe B, 562-bp fragment (nucleotides 145–706); probe C, 1080-bp fragment (nucleotides 707–1786); probe D, 828-bp fragment (nucleotides 1787–2614); probe E, 839-bp fragment (nucleotides 2615–3453); probe F, 415-bp fragment [nucleotides 3454–3868 plus an 88-bp poly(A) tail]; probe G, 568-bp fragment (nucleotides 1571–2138); probe H, 352-bp fragment (nucleotides 1787–2138). For precise location of nucleotide positions see figure 2 in ref. 11. E, *Eco*RI; P, *Pst* I; R, *Rsa* I; S, *Stu* I; Ss, *Sst* I.

Selective Expression of β1,4-GT Transcripts in Specific Spermatogenic Cell Types. The adult mouse testis preparation contained predominantly pachytene spermatocytes and round spermatids as well as somatic cells. In order to establish which cell type(s) express the shorter transcripts, purified populations of germ cells were isolated and examined. Equal quantities of total RNA (10 μg), isolated from highly enriched populations of spermatogonia (2*n*), pachytene spermatocytes (4*n*), and round spermatids (*n*) were initially analyzed by blot hybridization with probe C. A single 4.1-kb transcript was detected in RNA prepared from spermatogonia (Fig. 2, lane 1). In contrast, no transcript was detected in RNA prepared from pachytene spermatocytes (Fig. 2, lane 2). [However, when 20–30 μg of total RNA, isolated from pachytene spermatocytes, was analyzed and exposure times were increased, faint bands at 2.9, 3.1, and 4.1 kb were seen (data not shown).] In round spermatids, the 2.9- and 3.1-kb transcripts, present in approximately equal amounts, were readily detected, whereas the 4.1-kb transcript was not detected (Fig. 2, lane 3). Thus in germ cell differentiation, β1,4-GT transcripts are regulated in two ways. During differentiation from spermatogonia to pachytene spermatocytes, the amount of β1,4-GT mRNA is significantly reduced. Continued differentiation of pachytene spermatocytes to round spermatids is coincident with a renewed production of β1,4-GT mRNA to levels comparable with those detected in spermatogonia. However, the characteristic 4.1-kb mRNA detected in spermatogonia is replaced by truncated transcripts of 2.9 and 3.1 kb.

The β1,4-GT Transcripts in Round Spermatids Have a Truncated 3′ Untranslated Region. To determine the regions in common between the 4.1-kb transcript and the 2.9- and 3.1-kb mRNA species, identical RNA blots were hybridized with a series of probes (probes A–F; see Fig. 1) that, starting from the 5′ end of the cDNA, represent the full-length 4.1-kb transcript for murine β1,4-GT in somatic cells. The essential features of this analysis can be summarized as follows: (*i*) the 4.1-kb transcript expressed by spermatogonia hybridizes to all six probes (A–F) and consequently is similar, if not

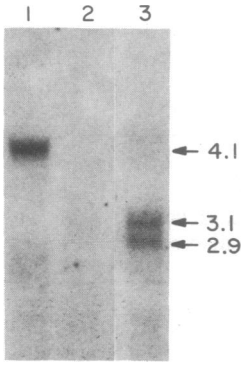

FIG. 2. Expression of the β1,4-GT transcript in specific germ cell populations. RNA blot analysis with probe C (see Fig. 1) was performed on total RNA (10 μg) derived from spermatogonia (lane 1), pachytene spermatocytes (lane 2), and haploid round spermatids (lane 3). Sizes (in kb) are indicated and were determined relative to an RNA ladder. When the blot was rehybridized with the probe for glyceraldehyde-3-phosphate dehydrogenase, the levels of hybridizing RNA were similar in all three lanes (data not shown).

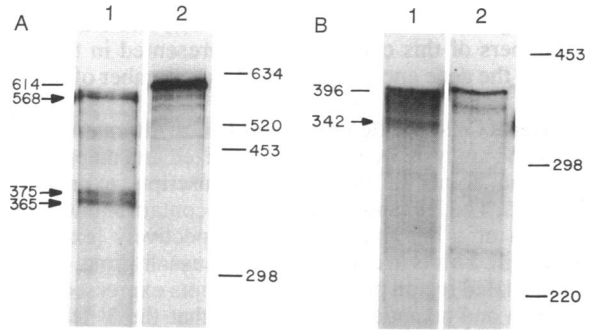

FIG. 3. S1 nuclease protection analysis. (*A*) The probe used for analysis was the 568-bp fragment, designated G in Fig. 1, subcloned into M13mp11. This fragment contains 23 bp of M13 polylinker DNA at its 3′ end and 23 bp at its 5′ end, giving a total probe length of 614 bp. Lanes: 1, 5 μg of round spermatid poly(A)$^+$ RNA; 2, 30 μg of yeast tRNA. (*B*) The probe used for analysis was the 352-bp fragment, designated H in Fig. 1, subcloned into M13mp11. This fragment contains 24 bp of M13 DNA at its 3′ end and 20 bp at its 5′ end, giving a total probe length of 396 bp. Lanes: 1, 30 μg of round spermatid RNA; 2, 30 μg of yeast tRNA. The specific S1 nuclease product is marked at the left with an arrow, and the band representing undigested probe is marked with a line. Molecular size markers (in bp) are marked at the right and are *Eco*RI–*Hin*fI-digested pBR327.

identical, to the 4.1-kb transcript expressed by various somatic cell types; (*ii*) the 2.9- and 3.1-kb mRNA species found in round spermatids hybridize to probes A–C, which represent the 5′ untranslated region, the entire protein coding sequence, and 589 bp of the 3′ untranslated sequence. Probes D–F however, which represent the 3′ terminal 1.5- to 2.0-kb region of the cDNA, detect no hybridizing species.

The Truncated Spermatid Transcripts Use Alternative Polyadenylylation Signals. To accurately define the termination points at the 3′ end of the 2.9- and 3.1-kb β1,4-GT transcripts expressed in spermatids, a series of S1 nuclease analyses were performed using probes D, G, and H (see Fig. 1). No protected species were seen when probe D (828 bp) was hybridized to RNA prepared from spermatids (data not shown). This was not unexpected since probe D did not detect a hybridizing species by RNA blot analysis. However, since short protected species may not have been detected with this probe, the S1 analysis was repeated with probe G (568 bp). As can be seen in Fig. 3*A*, lane 1, fragments of 365, 375, and 568 bp are protected by spermatid RNA. These species map approximately to nucleotides 1935, 1945, and 2138, respectively (see Fig. 4 for nucleotide positions).

Since the fully protected species mapped approximately to nucleotide 2138 and an alternative poly(A) signal, AATAAT (15), was noted at nucleotides 2111–2116 (Fig. 4), we examined this region in greater detail by using a shorter probe, to determine if termination was occurring immediately downstream of the AATAAT sequence motif. For this analysis probe H (352 bp) was used, which was 216 bp shorter than probe G. As seen in Fig. 3*B*, lane 1, a fragment of 342 bp was protected by spermatid RNA, but full-length protection (352 bp) was not observed. The protected species maps approximately to nucleotide 2128 (see Fig. 4), indicating that this AATAAT poly(A) signal is being used.

S1 Analysis with Probes That Represent the Coding Sequence. When probes B and C were used for S1 analysis, full-length protection was seen with total RNA prepared from spermatogonia and spermatids (data not shown). These data indicate that the corresponding nucleotide and predicted amino acid sequence of the spermatogonia and spermatid transcripts is identical to the C127 β1,4-GT transcript within the region extending from the 5′ end of probe B to the 3′ end of probe C (nucleotides 145–1786; see figure 2 in ref. 11). S1 analysis using probe A, which contains 5′ untranslated sequence and the remaining coding sequence for the NH_2-terminal region, also resulted in a fully protected species (Fig. 5 *Upper*, lanes 2 and 4). In summary, these results demonstrate that the coding sequence specified by the somatic and germ cell transcripts is identical. However, the results obtained with probe A, which includes the region spanning the first two in-frame ATG codons, gave an unanticipated result, which is discussed below.

Selective Translation Initiation at the Upstream AUG Codon: S1 Analysis with Probe A. As previously reported (11), S1 analysis of C127 mRNA using probe A revealed the presence of a fully protected species of 321 bp and a partially protected species of 145 bp (Fig. 5 *Upper*, lane 1). The 5′ boundary of the 145-bp species maps between the first two in-frame ATG codons, and the 321-bp species maps 157 bp upstream of the first in-frame ATG codon. These results demonstrate the presence of two mRNA transcripts, 3.9 and 4.1 kb in length, that, as a result of their initiation point with respect to the two ATG codons, specify the synthesis of two related forms of β1,4-GT of 386 and 399 amino acids. The two forms differ in that the longer form has an NH_2-terminal extension of 13 amino acids (11).

We have also carried out S1 analysis with probe A on total RNA isolated from various mouse somatic tissues, cells, or cell lines including spleen, duodenum, heart, lung, kidney, Sertoli cells (a somatic cell type in the seminiferous epithelium), F9 teratocarcinoma stem cells, PYS parietal endoderm cells, CTLL cytotoxic T cells, and L1210 lymphocytic leukemia cells. In each case, the two protected RNA species of 321

```
                                1220                    1240                    1260
TAGCATTTTTAGTACCAATAAGAGACCTGAGAATGGCCGGAGACCTCAGATATGTGTCTCTGCCAGTTGACTGGGCTGG
1280                    1300                    1320                    1340
TGCCTCTCATTTGTACAGTCTGAATGACAGTTCTTCTTATCATTCAGACGTCCCTCCAGATGCCCAGGGTGAGTGTAACA
1360                    1380                    1400                    1420
TTTACCCACAACCTGGCTCGGCACTGGATGAAATTCTACAAGGTGAGTGGAGTGTAAAACTGGTCAGCCCTTGGAGAGAC
1440                    1460                    1480                    1500
TTCTTGGTTGTGTCACCCCCAAAGAGTCAGAACTGTACACAGTTCAAAACTTAGTGACTGTGGGTCACATTCCACTGTTG
1520                    1540                    1560                    1580
AAACTGCTAAATTGTGACCTGGGGAAGGACTTTGCTTTAGTCGGTGATGTTCGTACTTGGTGACAAATTGAGACTGCTGC
1600                    1620                    1640                    1660
GTGATTCAGATTGACAAGATTTTCTTGGATTTTTTTTTTATACGAAAATCAAAATTTCAATCAGTCTCGTGCTCTGTCCC
1680                    1700                    1720                    1740
TTTACATCGGTATGCGACTATTACAATCACTGTGTGTGTGTCTTTTCTTAGCAAAGGCGTTTTAAAACTTGAGCCTGACC
1760                    1780                    1800                    1820
TTGGGGTCCTGTAGTGTGTGGATTCCAAGGCCTTGCCCTCAGAGCAGGGGCCTGGGCACTCTCACTCACGTGGCCTGTCT
1840                    1860                    1880                    1900
CCAGATCCCTGTCTGATTTCTGAATGTAAAGAGGCTTTTTGTTTTGTTTTTGTTTTTAGAAGCAGTTCGTAGTA
1920                    1940                    1960                    1980
TTTGAAAGAATAAATCAAGTTTTGATTATGCTATAGGTTGATTTTTGTGTTGATCCAAATCAGAATAGCTATTGAGTGTT
2000                    2020                    2040                    2060
TAAGTCATGACTTTATTTTTCTGGGCATGCTATATAAACTTGAATTTCCTATGTATTTTTATTGTGGTATTTTAAATGTG
2080                    2100                    2120                    2140
GGGGAAGGGATTGGGTATTTTTTAGGGGAAAAAAAATAATGTATGCTGTAGTCGTCAAAGGAGCTCTGATTGAGCTGGCA
2160                    2180                    2200                    2220
GGCCAGGTTTTGTAAAGAGCAAAATTACTTTCAGGTCCCTTGTTGAGGGAACAGGTCAGCGTTATCCTGCCAGACATCTG
```

FIG. 4. Nucleotide sequence of the region of the β1,4-GT transcript extending from nucleotide 1198 to 2236. The numbered nucleotide sequence is from ref. 11 (GenBank accession no. J03880). The sequence begins with the TAG translation termination codon at the 3′ end of the open reading frame and continues for an additional 1038 bp. The arrowheads indicate the approximate positions where the truncated species map. The putative poly(A) signals are boxed. The G+T sequence motifs downstream from the poly(A) site are overlined, and the A+T sequence motifs are underlined.

and 145 bp were readily detected (data not shown), demonstrating the presence of both the 3.9- and 4.1-kb transcript.

However, as seen in Fig. 5 *Upper*, when the S1 analysis with probe A was performed with total RNA isolated from spermatogonia (lane 2), pachytene spermatocytes (lane 3), or round spermatids (lane 4), only the 321-bp species was observed. Even when exposure times were increased, the 145-bp species was not detected. Thus murine male germ cells do not synthesize a mRNA transcript that initiates between the first two in-frame ATG codons. Of all the cells or cell lines examined to date, only the isolated male germ cells selectively use the upstream transcriptional start site. These results predict that only the long form (399 amino acids) of the β1,4-GT polypeptide is synthesized in these male germ cell populations.

DISCUSSION

Spermatogenesis is the ordered process whereby mitotically dividing diploid spermatogonia ultimately give rise to mature testicular spermatozoa. Our analysis of the expression of β1,4-GT during spermatogenesis has shown that both the levels and the structure of the mRNA are regulated with the state of maturation of these cells. The amount of mRNA decreases to barely detectable levels prior to meiosis as cells progress from spermatogonia to pachytene spermatocytes. Continued differentiation to round spermatids is coincident with a renewed production of β1,4-GT mRNA to levels comparable with those detected in spermatogonia. However, the 4.1-kb mRNA detected in spermatogonia is replaced by shorter transcripts of 2.9 and 3.1 kb.

It has previously been shown (16), by direct enzymatic assay, that β1,4-GT is present in enriched spermatogonia, pachytene spermatocytes, and round spermatid cell populations, indicating that the mRNA transcripts detected are actively translated. However, in contrast to the mRNA level, enzymatic activity (as determined by specific activity) was relatively constant in the spectrum of cells examined, from spermatogonia to round spermatids. This may reflect differences in the half-life between the mRNA and the protein.

An examination of gene expression during spermatogenesis suggests subdivision into three general classes (reviewed in ref. 17): (*i*) Genes that are expressed primarily in germ cells. Examples include the spermatid nuclear transition protein TP1 (18) and the protamines (19). (*ii*) Genes that encode a testis-specific isoform of a protein that is also found in somatic cells. Examples include lactate dehydrogenase X (20), cytochrome c_t (21), and actin (22). (*iii*) Genes whose expression is either qualitatively or quantitatively changed. α-Tubulin (23) and the protooncogenes *pim-1* (24), c-*mos* (25, 26), and c-*abl* (27–29) are members of this class. The data presented in this paper show that the gene encoding β1,4-GT is a member of this third class.

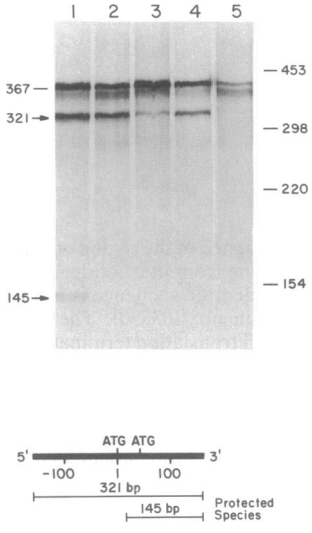

FIG. 5. (*Upper*) S1 nuclease protection analysis. The probe used was the 321-bp fragment designated A in Fig. 1 subcloned into M13mp11. This probe contains 46 bp of M13 polylinker DNA, giving a total length of 367 bp. Lanes: 1, 30 μg of C127 RNA; 2, 30 μg of spermatogonia RNA; 3, 30 μg of pachytene spermatocyte RNA; 4, 30 μg of spermatid RNA; 5, 30 μg of yeast tRNA. Specific S1 nuclease products are marked at the left with arrows, and the undigested probe is marked with a line. Molecular size markers (in bp) are marked at the right and are *Eco*RI–*Hin*fI-digested pBR327. (*Lower*) The schematic indicates the region of the transcript used for the S1 probe, the first two in-frame ATG codons, and the position of the protected species.

The Germ Cell-Specific Transcripts Utilize Alternative Poly(A) Sites. By S1 analysis we have characterized and mapped the 3′ termination points of the shorter transcripts expressed by spermatids. The 2.9- and 3.1-kb species contain 3′ untranslated sequences of ≈740 and ≈930 bp, respectively [excluding a poly(A) tail]. This is in contrast to the unusually long, ≈2600 bp 3′ untranslated region present in transcripts expressed by spermatogonia and somatic cells. We feel that the 3′ termination positions observed in spermatid mRNA arise as a result of alternative poly(A) and are not due to processing because (*i*) consensus poly(A) signals are found at the appropriate locations; (*ii*) the G+T sequences, which have been identified downstream of the poly(A) site (30, 31), are also present in the appropriate downstream location (see Fig. 4); and (*iii*) processing within the 3′ untranslated region is an extremely rare event (32).

The somatic cell 3.9- and 4.1-kb transcripts previously described (11) contain three potential poly(A) signals in the 3′ untranslated region. The distal AATAAA sequence located at nucleotide 3759 is used to generate the 3.9- and 4.1-kb transcripts found in all somatic cell types examined. This poly(A) signal is also apparently used by spermatogonia. The internal AATAAA sequence located at nucleotide 1925 and the AATAAT sequence located at nucleotide 2111, although utilized to a minor extent by somatic cells (N.L.S. and J.H.S., unpublished results), are utilized exclusively by the haploid spermatids as poly(A) signals. Therefore the 186-bp difference between these two sites most probably reflects the size difference between the 2.9- and 3.1-kb transcripts.

Two sequence motifs within the 3′ untranslated region, ATTT(A) (33) and TATT (34), have been identified, which appear to contribute to transcript instability. Within the 3′ untranslated region of the β1,4-GT cDNA characterized from somatic cells, the sequences ATTT(A) and TATT occur 16 and 11 times, respectively (11). Eight ATTT(A) and 2 TATT sequences occur upstream of the poly(A) site at nucleotide 1925. An additional 6 ATTT(A) and 6 TATT sequences are positioned between the poly(A) site at nucleotide 1925 and the poly(A) site at nucleotide 2111 (Fig. 4). It has been postulated that steady-state mRNA concentrations are regulated by posttranscriptional processes controlled by regions within the 3′ untranslated region (35). Perhaps the position of the A+T sequence motifs, relative to the poly(A) site, influence the stability of the β1,4-GT transcripts.

The Germ Cell-Specific Transcripts Apparently Utilize a Different Upstream Promotor. Based on the combined lengths of the S1 probes, the two truncated spermatid transcripts contain 2.1 and 2.3 kb of accountable sequence (see Fig. 1). However, these values do not agree with the mRNA sizes obtained from Northern blots (2.9 and 3.1 kb). This suggests that ≈650 bp of sequence is unaccounted for [assuming a poly(A) tail of ≈150 bp]. Although these transcripts may have an unusually long poly(A) tail, it is more likely that the size difference is due to the presence of additional sequence at the 5′ end of the transcript (see Fig. 6). In preliminary experiments, in which an S1 analysis was performed on spermatid RNA with a probe derived from a genomic fragment extending 57 bp beyond the 5′ end of the 4.1-kb transcript, we did detect a single fully protected species, thus supporting this conclusion (data not shown).

We have recently determined the genomic organization of β1,4-GT and have established that the first exon contains the complete 5′ untranslated region of the 4.1-kb somatic transcript (≈180 bp) and 415 bp of NH_2-terminal coding sequence (41). Since alternative splicing does not occur within the 5′ untranslated region, then expression of the 4.1-kb (and 3.9-

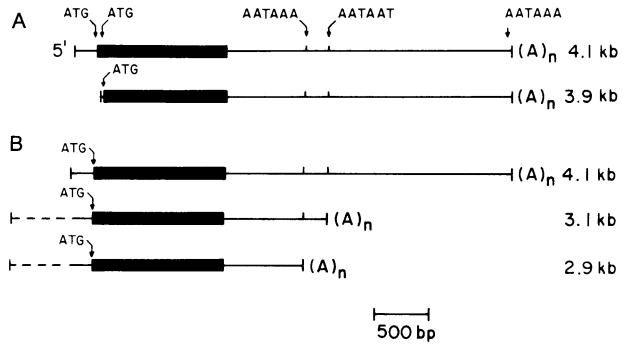

FIG. 6. A schematic representation of the β1,4-GT transcripts expressed in somatic and germ cells. The solid bar indicates the protein coding region; the thin line indicates the 5′ and 3′ untranslated regions. (*A*) The long (4.1-kb) and short (3.9-kb) forms of the somatic transcript, with the position of the three poly(A) signals, are shown. The sequence of the two forms is identical except that the short form lacks ≈180 bp of the 5′ untranslated sequence including the first in-frame ATG codon. (*B*) The three germ cell transcripts are shown. The inferred length of the 5′ untranslated region is indicated by the dashed line. A poly(A) tail of ≈150 bp is assumed.

kb) somatic cell transcripts (and presumably the 4.1-kb transcript in spermatogonia) is controlled by promoter elements that are predicted to be immediately upstream of the transcriptional start site. The 2.9- and 3.1-kb transcripts found in round spermatids contain a significant amount of additional 5′ untranslated sequence (estimated to be about 650 bp); therefore, they are most likely regulated by a different upstream promoter.

Male Germ Cells Synthesize Exclusively the Long Form of β1,4-GT. As pointed out in the Introduction, the gene for murine β1,4-GT specifies two sets of mRNA transcripts that predict the synthesis of two forms of the protein. The only difference in primary structure between the two proteins is that the long form contains an NH_2-terminal extension of 13 amino acids. With the exception of the murine male germ cells, transcripts that specify both the long and short forms of β1,4-GT have been detected in all of the murine tissues and cell lines that we have examined as well as bovine thymus and the MDBK cell line (37).

What is the functional significance of two forms of a membrane-bound protein that differ only in their NH_2-terminal domain? As previously discussed, the NH_2-terminal domain can specify both final cellular destination and topological orientation within a membrane (ref. 11; reviewed in ref. 38). Within this context, we have pointed out that the NH_2-terminal domain of the long form of β1,4-GT contains features of a cleavable signal sequence and consequently would be oriented in the membrane in a direction that is opposite to the short form of the enzyme (11). However recent studies (37) have indicated that both forms of β1,4-GT are oriented in the same direction, with their COOH-terminal domain positioned within the Golgi lumen.

The expression of only the long form of the β1,4-GT polypeptide in spermatogenic cells may be important to the final restriction of this molecule to the apical portion of the murine sperm head (5). During murine spermatogenesis, the cell surface β1,4-GT undergoes redistribution from being uniformly distributed on spermatocytes to a highly restricted distribution overlying the acrosome of mature spermatids (7). It is likely that the establishment of distinct plasma membrane domains on the sperm head involves the interaction of plasma membrane proteins with underlying cytoskeletal elements (39). Thus the extended NH_2-terminal domain of the long form of the protein, which is positioned in the cytoplasm, may be capable of specific interaction with cytoskeletal elements of the maturing spermatid, and this interaction may be responsible for the apparent restriction of β1,4-GT to the apical portion of the murine sperm head.

We thank Abdul Karzi for technical assistance in the preparation of spermatogenic cells and Ann Larocca for editorial assistance. This research was supported in part by National Institutes of Health Grants CA45799 and GM38310 (J.H.S.) and HD21648 (W.W.W.).

1. Hill, R. L., Brew, K., Vanaman, T. C., Trayer, I. P. & Mattock, P (1968) *Brookhaven Symp. Biol.* **21,** 139–154.
2. Farquhar, M. G. & Palade, G. E. (1981) *J. Cell Biol.* **91,** 77s–103s.
3. Roth, J. & Berger, E. G. (1982) *J. Cell Biol.* **93,** 223–229.
4. Bayna, E. M., Shaper, J. H. & Shur, B. D. (1988) *Cell* **53,** 145–157.
5. Lopez, L. C., Bayna, E. M., Litoff, D., Shaper, N. L., Shaper, J. H. & Shur, B. D. (1985) *J. Cell Biol.* **101,** 1501–1510.
6. Roth, J., Lentze, M. J. & Berger, E. G. (1985) *J. Cell Biol.* **100,** 118–125.
7. Scully, N. F., Shaper J. H. & Shur, B. D. (1987) *Dev. Biol.* **124,** 111–124.
8. Shaper, N. L., Mann, P. L. & Shaper, J. H. (1985) *J. Cell. Biochem.* **28,** 229–239.
9. Shur, B. D. (1982) *J. Biol. Chem.* **257,** 6871–6878.
10. Bleil, J. D. & Wassarman, P. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6778–6782.
11. Shaper, N. L., Hollis, G. F., Douglas, J. G., Kirsch, I. R. & Shaper, J. H. (1988) *J. Biol. Chem.* **263,** 10420–10428.
12. Nakazawa, K., Ando, T., Kimura, T. & Narimatsu, H. (1988) *J. Biochem.* **104,** 165–168.
13. Meistrich, M. L., Longtin, J., Brock, W. A., Grimes, S. R., Jr., & Mace, M. L. (1981) *Biol. Reprod.* **25,** 1065–1077.
14. Dugaiczyk, A., Haron, J. A., Stone, E. M., Dennison, O. E., Rothblum, K. N. & Schwartz, R. J. (1983) *Biochemistry* **22,** 1605–1613.
15. Wickens, M. & Stefenson, P. (1984) *Science* **226,** 1045–1051.
16. Letts, P. J., Meistrich, M. L., Bruce, W. R. & Schachter, H. (1974) *Biochim. Biophys. Acta* **343,** 192–207.
17. Willison, K. & Ashworth, A. (1987) *Trends Genet.* **3,** 351–355.
18. Heidaran, M. A., Showman, R. M. & Kistler, W. S. (1988) *J. Cell Biol.* **106,** 1427–1433.
19. Kleene, K. C., Distel, R. D. & Hecht, N. B. (1984) *Dev. Biol.* **105,** 71–79.
20. Tanaka, S. & Fujimoto, H. (1986) *Biochem. Biophys. Res. Commun.* **136,** 760–766.
21. Hennig, B. (1975) *Eur. J. Biochem.* **55,** 167–183.
22. Waters, S. H., Distel, R. J. & Hecht, N. B. (1985) *Mol. Cell. Biol.* **5,** 1649–1654.
23. Dobner, P. R., Kislauskis, E., Wentworth, B. M. & Villa-Komaroff, L. (1987) *Nucleic Acids Res.* **15,** 199–218.
24. Sorrentino, V., McKinney, M. D., Giorgi, M., Geremia, R. & Fleissner, E. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2191–2195.
25. Propst, F., Rosenberg, M. P., Iyer, A., Kaul, K. & Vande Woude, G. F. (1987) *Mol. Cell. Biol.* **7,** 1629–1637.
26. Propst, F., Rosenberg, M. P. & Vande Woude, G. F. (1988) *Trends Genet.* **4,** 183–187.
27. Ponzetto, C. & Wolgemuth, D. J. (1985) *Mol. Cell. Biol.* **5,** 1791–1794.
28. Meijer, D., Hermans, A., von Lindern, M., van Agthoven, T., de Klein, A., Mackenbach, P., Grootegoed, A., Talarico, D., Della Valle, G. & Grosveld, G. (1987) *EMBO J.* **6,** 4041–4048.
29. Oppi, C., Shore, S. K. & Reddy, E. P. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8200–8204.
30. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41,** 349–359.
31. McLauchlan, J., Gaffney, D., Whitton, J. L. & Clements, J. B. (1985) *Nucleic Acids Res.* **13,** 1347–1368.
32. Kozak, M. (1983) *Microbiol. Rev.* **47,** 1–45.
33. Shaw, G. & Kamen, R. (1986) *Cell* **46,** 659–667.
34. Reeves, R., Elton, T. S., Nissen, M. S., Lehn, D. & Johnson, K. R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6531–6535.
35. Pfarr, D. S., Rieser, L. A., Woychik, R. P., Rottman, F. M., Rosenberg, R. & Reff, M. E. (1986) *DNA* **5,** 115–122.
36. Hollis, G. F., Douglas, J. G., Shaper, N. L., Shaper, J. H., Stafford-Hollis, J. M., Evans, R. J. & Kirsch, I. R. (1989) *Biochem. Biophys. Res. Commun.* **162,** 1069–1075.
37. Russo, R. N., Shaper, N. L. & Shaper, J. H. (1990) *J. Biol. Chem.*, in press.
38. Geisow, M. J. (1986) *Bioessays* **4,** 149–151.
39. Eddy, E. M. (1988) in *The Physiology of Reproduction*, eds. Knobil, E. & Neill, J. D. (Raven, New York), Vol. 1, pp. 27–68.