MOLECULAR AND CELLULAR BIOLOGY, Mar. 1992, p. 905–914
0270-7306/92/030905-10$02.00/0

Vol. 12, No. 3

High-Level Expression of the Rat Whey Acidic Protein Gene Is Mediated by Elements in the Promoter and 3′ Untranslated Region

TREVOR C. DALE,[1]† MICHAEL J. KRNACIK,[1] CHRISTIAN SCHMIDHAUSER,[2] CLAUDIA L.-Q. YANG,[2] MINA J. BISSELL,[2] AND JEFFREY M. ROSEN[1]*

Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030,[1] and Cell and Molecular Biology Division, Lawrence Berkeley Laboratory, Berkeley, California 94720[2]

Received 27 June 1991/Accepted 19 November 1991

The high-level expression of the rat whey acidic protein (WAP) gene in transgenic mice depends on the interaction of 5′-flanking promoter sequences and intragenic sequences. Constructs containing 949 bp of promoter sequences and only 70 bp of 3′-flanking DNA were expressed at uniformly high levels, comparable to or higher than that of the endogenous gene. Although this WAP transgene was developmentally regulated, it was expressed earlier during pregnancy than was the endogenous WAP gene. Replacement of 3′ sequences, including the WAP poly(A) addition site, with simian virus 40 late poly(A) sequences resulted in an approximately 20-fold reduction in the expression of WAP mRNA in the mammary gland during lactation. Nevertheless, position-independent expression of the transgene was still observed. Further deletion of 91 bp of conserved WAP 3′ untranslated region (UTR) led to integration site-dependent expression. Position independence was restored following reinsertion of the WAP 3′ UTR into the deleted construct at the same location, but only when the insertion was in the sense orientation. The marked differences observed between the expression levels of the 3′-end deletion constructs in transgenic mice were not seen in transfected CID 9 mammary epithelial cells. In these cells, expression of the endogenous WAP gene was dependent on the interaction of these cells with a complex extracellular matrix. In contrast, the transfected WAP constructs were not dependent on extracellular matrix for expression. Thus, both the abnormal expression of WAP in cells cultured on plastic and the precocious developmental expression of WAP in transgenic mice may reflect the absence of a negative control element(s) within these recombinant constructs.

During the progression from virgin through pregnancy and lactation, mammary alveolar epithelial cells proliferate and differentiate in a hormone-dependent process that leads to the high-level, tissue-specific expression of a number of milk protein genes. The regulation of several of these milk protein genes, notably β-casein, has been studied in mammary cell lines and in primary mammary epithelial cells (30). Whey acidic protein (WAP) is the major whey protein in rodent milk. Although WAP mRNA accounts for 10 to 15% of polyadenylated RNA in the lactating rodent mammary gland, it is not usually expressed in cell culture (6, 21, 31). Recent studies indicate that WAP mRNA can be expressed in vitro if primary mammary epithelial cells are cultured on a complex extracellular matrix derived from the Engelbreth-Holm-Swarm mouse tumor, (EHS), on which they are able to organize themselves into hollow spherical structures termed mammospheres (6). However, the definition of *cis*-acting elements required for WAP gene expression has been accomplished primarily with transgenic mice owing to the lack of cell culture models that can be transfected efficiently and accurately reproduce the complex regulatory phenomena present in the pregnant and lactating gland (3, 4).

Many genes, including that for rat β-casein, are expressed at low levels in transgenic mice (24, 26). Studies on the human β-globin and CD2 genes have shown transgene expression at levels equivalent to those of the endogenous gene through the action of locus control regions (LCRs) that lie up to 40 to 60 kb away and may define the boundaries of the gene clusters (14, 16, 37). It has been hypothesized that high-level expression of the β-casein gene could be mediated by such far-upstream or far-downstream transcriptional elements, since a 14-kb region containing the rat β-casein gene is expressed at only 0.01 to 1% of the level of the endogenous gene (24, 29). It is known, however, that the expression of several small milk protein transgenes that are not localized in gene clusters can approach or equal that of the endogenous gene (3, 34, 38). In most of these studies, however, transgene expression varied over at least a 100-fold range in independent lines of mice. For example, a 4.3-kb region encompassing the rat WAP gene was expressed at levels between 1 and 95% of that of the endogenous gene (3). This is in marked contrast to the β-globin and CD2 studies, in which LCR sequences direct site-independent, high-level expression, possibly by abolishing the effects of the chromatin structure surrounding the site of integration. We show here that deletion of 3′-flanking sequences from the rat WAP transgene allows site-independent, high-level expression.

Previous studies with expression constructs containing up to 2.5 kb of WAP promoter sequences fused to a variety of heterologous genes have generally resulted in mammary-specific expression, but absolute levels are low compared with those of the entire WAP gene construct. This suggests that elements within the gene body or 3′ untranslated or flanking regions are involved in high-level expression. Interestingly, the WAP 3′ untranslated region (3′UTR) is more highly conserved than the coding region (5, 36). The 3′UTRs of a number of mRNAs have been shown to be important for their stability (9). Therefore, it has been hypothesized that the rat WAP 3′UTR may have a role in the high-level expression of the gene.

High-level expression of the WAP gene is the result of an

* Corresponding author.
† Present address: Institute of Cancer Research, Sutton, Surrey SM2 5NG, United Kingdom.

almost 10^4-fold induction of WAP mRNA from virgin levels during pregnancy and lactation (21, 27). Endogenous WAP is induced later in pregnancy than the caseins (21, 27). However, a number of WAP transgenic constructs have been described in which the WAP transgene is expressed before the endogenous gene (3, 4, 17, 27), suggesting the lack or disruption of a developmentally regulated repressor element in these constructs.

WAP gene expression is more sensitive to the extracellular environment than β-casein gene expression. For example, β-casein but not WAP is expressed in primary mammary epithelial cells cultured on floating collagen (23). Recently, a selected population of COMMA-1D cells, designated CID 9, has been isolated that, like primary cells, are able to form mammospheres and express WAP when plated on a complex extracellular matrix (EHS) (33). However, we report here that the transfected rat WAP gene constructs in CID 9 cells, as well as transgenes in primary cell cultures derived from rat WAP transgenic lines, are expressed on both plastic and EHS.

MATERIALS AND METHODS

Cells and medium. CID 9 cells (33), a WAP-expressing subpopulation of COMMA-1D cells (11), were grown in Dulbecco's modified Eagle's medium–F12 medium (1:1; GIBCO) containing 5% heat-inactivated fetal calf serum (FCS), gentamicin (50 μg/ml), and insulin (Sigma; 5 μg/ml). The medium was changed every other day. For expanding cultures, the cells were split at a ratio of 1:5. For hormonal and matrix-dependent induction of differentiation, the cells were plated on plastic dishes or a reconstituted basement membrane EHS (22) at 8×10^4 cells per cm^2 in the presence of 2% FCS with insulin (5 μg/ml) and/or hydrocortisone (Sigma; 1 μg/ml) and/or prolactin (ovine, 3 μg/ml; National Institutes of Health) as indicated. At 24 h after plating, the cells were removed to medium without FCS. The cell cultures were then allowed to differentiate for 5 additional days; the medium was changed each day.

Transfection protocol and RNA isolation. CID 9 cells were plated at 4×10^4 cells per cm^2 24 h prior to transfection. For calcium phosphate-mediated transfections, WAP constructs (40 μg/100-mm dish) and pSV2neo (2 μg/100-mm dish) were coprecipitated as described before (13). The cells were incubated for 4 h, then shocked for 90 s with 25% glycerol in Hanks's balanced salt solution (pH 7.4), washed, and supplied with fresh medium containing butyrate at a final concentration of 10 mM. G418 (GIBCO; 400 μg/ml) selection was started 48 h after the transfection; 200 to 500 surviving colonies were pooled and expanded for stocks and assays. Total RNA was isolated by a single-step guanidinium-thiocyanate-phenol-chloroform extraction protocol as described before (7).

Plasmids and enzymes. The majority of techniques used for the isolation and purification of molecular clones were those described by Sambrook et al. (32). Restriction enzymes were purchased from Pharmacia-LKB (Piscataway, N.J.), with the exception of DNA ligase (New England BioLabs, Beverley, Mass.) and exonuclease III (Stratagene, La Jolla, Calif.). All plasmids and constructs described were made or subcloned into the pSKII+ vector (Stratagene).

The β-casein–WAP plasmid was made by subcloning an *Xba*I fragment containing the −511 to −12 β-casein promoter sequences upstream of a −30 to +3250 *Dra*I-*Acc*I fragment described by Bayna and Rosen (3).

Exonuclease III deletions of the +3250 rat WAP genomic construct (−949 to +3250, where +1 is the site of transcription initiation) were performed essentially as recommended by the manufacturer (Stratagene; exonuclease, mung bean deletion kit, catalog number 212205). The *Sal*I-*Acc*I fragment containing the entire +3250 region was subcloned into the *Acc*I site of the pSKII+ vector. The plasmid was digested with *Kpn*I and *Acc*I and then digested with exonuclease III for 5 to 8 min at 30°C. Clones that appeared to be of the desired length by restriction enzyme analysis were sequenced to confirm the extent of the deletion. The UT/SV construct was prepared by subcloning the simian virus 40 (SV40) *Bam*HI-*Bcl*I fragment (kindly provided by S. Berget) into the vector *Bss*HII site present 3′ of the UT/SV deletion indicated in Fig. 2B. The ΔUT/SV construct was prepared by subcloning a *Bam*HI-*Eco*RI fragment (+200 to +1819; see Fig. 2B, ΔUT/SV) of rat WAP upstream of the SV40 *Bam*HI-*Bcl*I fragment. The −949/+200 promoter sequences were then reconstructed exactly with an *Eco*RI fragment from the +3250 construct. The remaining SV40-containing plasmids were derived from the ΔUT/SV plasmid by reinserting an *Eco*RV-*Bss*HII (+1796 to +1919) fragment containing the 3′UTR from the UT/SV deletion into the *Bam*HI site between the SV40 and WAP sequences (+UT/SV and −UT/SV). The 5′UT/SV construct was made by inserting the *Eco*RV-*Bss*HII (+1796 to +1919) fragment upstream of the −949 sequence in a polylinker *Bam*HI site.

Generation and screening of transgenic mice. Transgenic mice were generated by microinjection of linear sequences prepared by digestion of the constructs cloned into pSKII+ with *Bss*HII, which excises the entire polylinker region. The microinjection and transfer of mouse embryos were carried out as described previously, as was the isolation of genomic tail DNA (24). Initial characterization of transgenic lines was performed by amplification of tail DNA by the polymerase chain reaction (PCR), essentially as described previously (15). The oligonucleotides used were specific for rat WAP and were made to the first exon and the 3′-flanking region, resulting in a 2-kb product (5′ primer [5′ to 3′], ATGTCGA CATCAGTCATCACTTGCCTGCCGCCG; 3′ primer [5′ to 3′], GAGGTACCTCATTCTGTCAAGAGCTCAGGACAG). Positive lines were screened by Southern hybridization to determine the copy number of the transgenes, as described previously (3).

Primary mammary epithelial cell cultures from transgenic mice. Male and female transgenic mice derived from WAP transgenic mouse line 5394 carrying the +3250 transgene (3) were bred initially between themselves or with nontransgenic CD1 mice (Charles River, Wilmington, Mass.) to produce F_1 offspring. Positive offspring were then bred with nontransgenic CD1 mice to give rise to F_2 offspring. Primary mammary epithelial cell cultures were generated from positive F_2 offspring. Epithelial cells from pregnant (12 to 15 days of gestation) mice were prepared as described previously (2, 23) and plated at 1.2×10^6 cells per cm^2 on to the appropriate substrata in F12 medium with lactogenic hormones as indicated (5 μg of insulin, 1 μg of hydrocortisone, 3 μg of prolactin, and 1 mg of fetuin per ml, plus 10% FCS). After 36 h, the medium was replaced with serum- and fetuin-free F12 medium containing lactogenic hormones. Cells were cultured in the serum-free medium with daily medium changes for another 4 days before they were lysed for RNA extraction.

RNA isolation and primer extension analysis. Mammary gland biopsies were performed on the female founder mice and on the F_1 females from the male founder mice at day 10 of lactation. The fourth abdominal mammary gland was

isolated under anesthesia, and total RNA was isolated by homogenization of tissues in a guanidium isothiocyanate solution and fractionated on a CsCl gradient (12).

The amounts of total RNA described were annealed at 30°C to 2 ng (1:1) of ^{32}P-labeled rat- and mouse-specific primers essentially as described before (32). The oligonucleotide primers both hybridize within the first exon and identify multiple start sites of 56 and 58 nucleotides (nt) for the rat WAP and 41, 42, 43, and 44 nt for the mouse WAP transcripts. The primer sequences are 5'-CCGCCGCCGA CACCATGCGCTGTTCGATCAGCC-3' for the rat primer and 5'-ACCCGGTACCCATGAAGTTGCCTC-3' for the mouse primer. The full-length rat and mouse extension products map 4 and 6 nt and 3 to 6 nt, respectively, upstream of their published start sites (5). Following primer extension, the reaction products were separated on a denaturing 12% polyacrylamide gel and dried onto Whatman 3MM paper. To quantify expression, rat and mouse bands from at least three different RNA input reactions per sample were localized by autoradiography, excised from the gel, and counted with Cerenkov radiation. The expression of the transgenes is reported as a percentage of rat WAP mRNA levels in the mammary gland at day 10 of lactation, since the rat WAP transcripts gave more reproducible titers than the mouse WAP transcripts in multiple experiments (data not shown).

Quantitative PCR analysis of WAP mRNA levels. Total RNA (1 μg) from transfected CID 9 cells or from primary mouse mammary epithelial cell cultures derived from transgenic mice was annealed to 0.5 μg of oligo(dT)$_{12-18}$ (Pharmacia-LKB) and reverse-transcribed in a 10-μl reaction mix containing 50 mM Tris (pH 8.3), 75 mM KCl, 1 mM dithiothreitol (DTT), 3 mM $MgCl_2$, and 0.5 mM deoxynucleoside triphosphates (dNTPs) for 1.5 h at 37°C. The reaction mix was diluted to 250 μl with double-distilled H_2O, and 25 μl was used for PCR amplification. Two separate sets of amplifications were performed for the rat and mouse WAP primer sets, with glycerol-3-phosphate dehydrogenase (G3PD) primers as an internal control. The PCR reaction mixes, in a final volume of 50 μl, contained 50 mM KCl, 10 mM Tris (pH 8.4, 25°C), 1 mM $MgCl_2$, 1 μM each primer, 200 μM each dNTP, and 2.5 U of *Taq*I polymerase (Cetus). Rat primers were made to the first and fourth exons and were designed to amplify rat WAP RNA from all the transfected constructs, resulting in a 450-bp product. The 5' primer was as described for the first-exon probe for transgenic screening above. The 3' primer was TCACTGAAAGGATATCACTG TAGGAG. Mouse primers were made to the first and third exons and give rise to a 331-bp product. The 5' primer was 5'-CTGACCCGGTACCATGAGTTGCCT-3', and the 3' primer was 5'-GGTCGCTGGAGCATTCTATCTTCA-3'. G3PD primers were made to the sixth and eighth exons and give rise to a 213-bp product. The 5' primer was 5'-GTGGT TCTGCAGACAGCGGATGAAT-3', and the 3' primer was 5'-AGAGGCCTTTGCTCGAACTGGAAAG-3'.

CID 9 RNA was amplified for 23 cycles under the following conditions (1 min at 94°C, 2 min at 59°C, 3 min at 72°C) for both rat and mouse primer sets, while RNA from transgenic lines was amplified for 14 and 19 cycles for the rat and mouse WAP primer sets, respectively. These cycle numbers were chosen to be within the linear range of the assay. Expression of the endogenous mouse WAP mRNA is considerably higher in primary cell cultures than in CID 9 cells, and rat WAP levels are even higher than mouse WAP levels in primary cells, possibly owing to a precocious expression shown previously in the transgenic lines (3).

To quantify expression levels, 12 μl of each reaction mix (<5 ng of product) was separated on a 2.1% (2:1, NuSeive-normal) agarose gel (ICN Pharmaceuticals) and transferred onto a Zetaprobe membrane (Bio-Rad, Richmond, Calif.) according to the manufacturer's protocols. The membranes were hybridized with ^{32}P-labeled rat WAP and mouse G3PD cDNAs generated with random hexanucleotide primers (32) and washed in an NaH_2PO_4-sodium dodecyl sulfate (SDS) solution according to the Zetaprobe protocols. Following autoradiography, rat and mouse WAP-specific bands were excised and counted by Cerenkov counting. Counts were normalized to average G3PD levels for cells grown on plastic or on EHS, since G3PD itself is induced two- to fourfold on EHS compared with plastic.

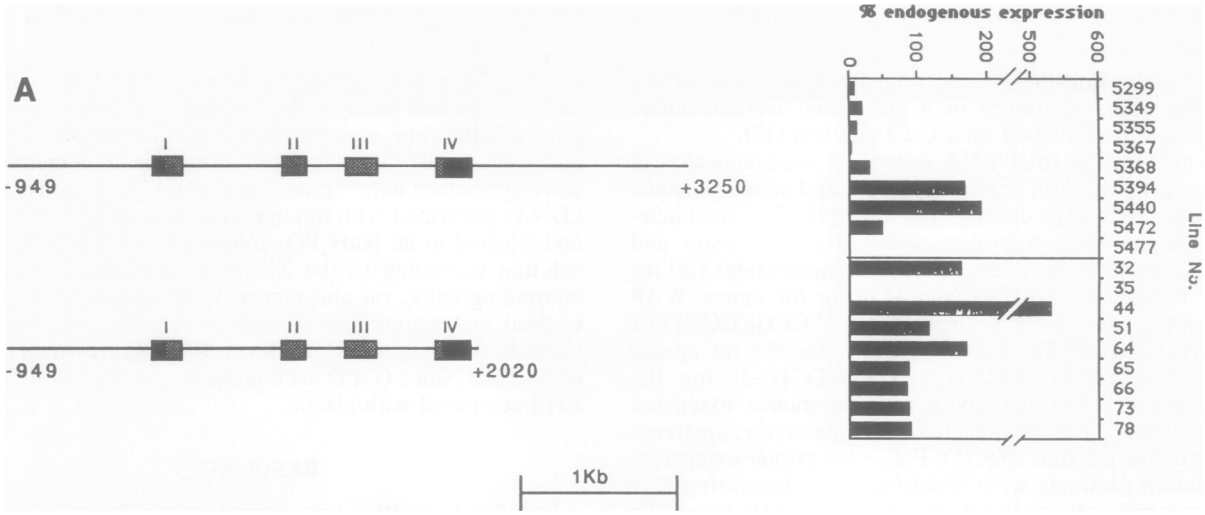
RESULTS

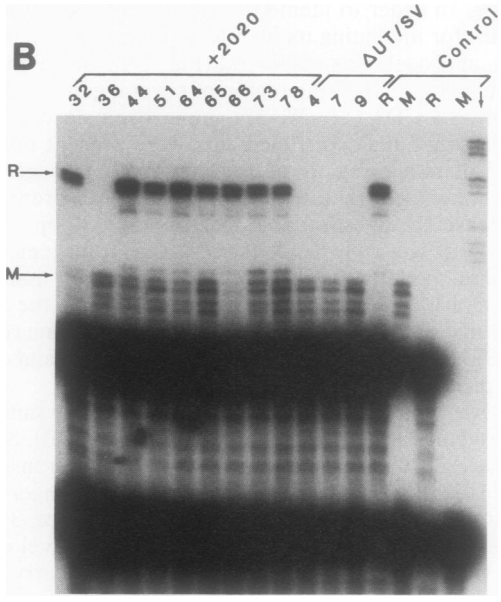
High-level, position-independent expression of the WAP transgene. In order to identify regions of the rat WAP gene important for mediating its high-level expression, a deletion analysis approach was employed (Fig. 1A). Since enhancer elements in the 3'-flanking region of several genes, including β-globin and CD2 (14, 28, 37), have been described, the possibility that analogous elements were present in the 1.3 kb of 3'-flanking region of the rat WAP transgene was initially examined. The +2020 construct was prepared by exonuclease III digestion and contains only 70 bp 3' of the last exon (Fig. 1A). Endogenous and transgenic WAP expression was measured in a primer extension assay with rat and mouse WAP-specific primers that enable the 5' ends of the transcripts to be accurately mapped with one reaction (Fig. 1B). Essentially identical results were obtained in an RNase protection assay (data not shown).

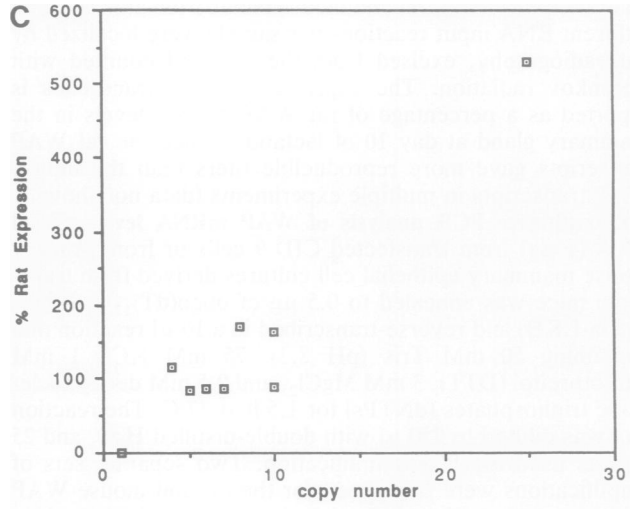
RNA was isolated from mammary gland biopsy samples of the transgenic lines at day 10 of lactation (Fig. 1B). Surprisingly, the +2020 transgenic lines expressed the transgene at higher levels than did most lines carrying the longer +3250 construct (reassayed in Fig. 1A from reference 3). This indicates that the elements necessary for high-level expression lie within the +2020 construct. One +2020-carrying line that did not express the transgene contained only a single copy of it (line 35). The line that expressed it at 500% of the endogenous level (line 44) had the highest transgene copy number.

While a statistically significant correlation can be drawn between copy number and expression level (Fig. 1C), more data are needed from high-copy-number lines to validate copy number dependence. The other seven +2020 lines that have similar copy numbers (4 to 10 copies) show little variation in expression, showing that expression is integration site independent. Since the average level of +2020 transgene expression relative to that of the endogenous rat WAP gene during lactation is approximately 40% per gene copy, it is conceivable that additional regulatory elements are missing from the +2020 construct or that the level of critical *trans*-acting factors may be limiting (20). No evidence for mosaic integration of the transgenes was obtained after examining F_1 transmission for the various transgenic lines (data not shown) (3).

To try to further localize the regions in the WAP gene allowing high-level expression, 506 bp of the WAP promoter was fused to the chloramphenicol acetyltransferase (CAT) reporter gene. Of the 12 lines derived from this construct, none were found to express significant CAT activity (data not shown). It is possible that sequences between −949 and −536 bp may be involved in high-level expression, as these were not included in this construct. However, analogous







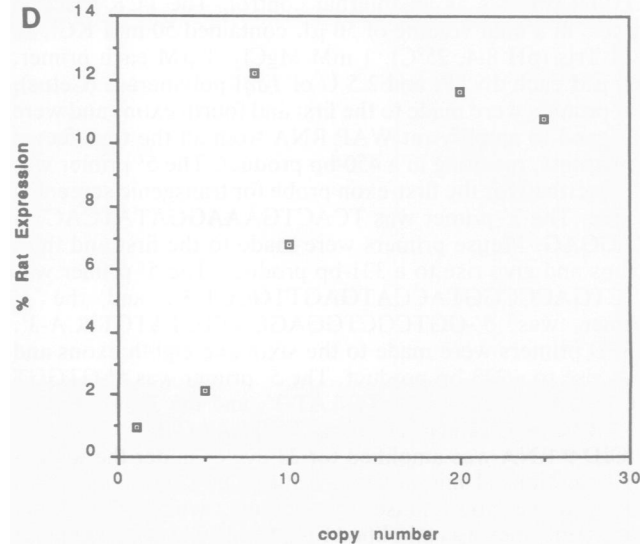
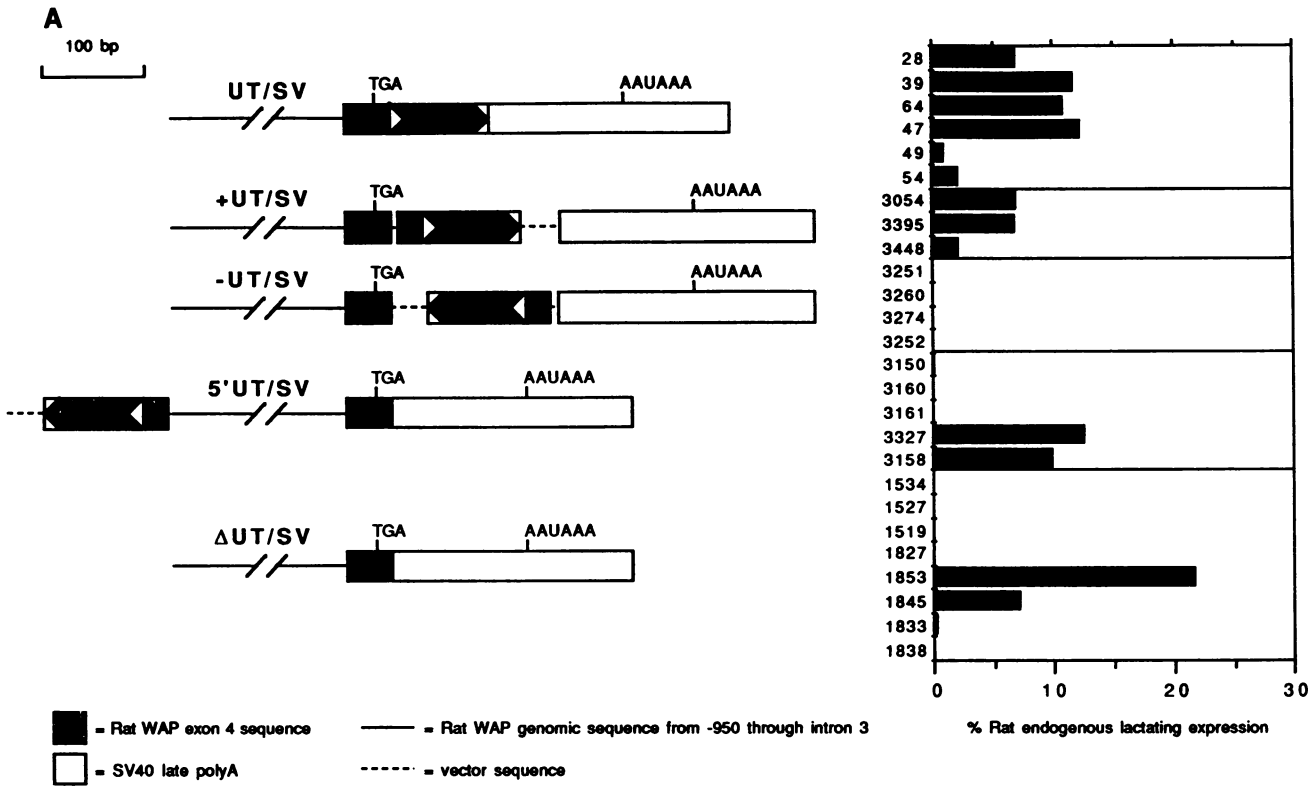


FIG. 1. Expression of rat WAP recombinant constructs in transgenic mice. (A) Summary of expression of WAP gene constructs. Exons are numbered I, II, III, and IV, with the shaded areas indicating coding regions and the solid areas indicating noncoding regions. The structures of various clones are shown next to a summary of expression data. RNA isolated from mammary gland biopsy samples of the +3250 lines described previously (3) at day 10 of lactation was reassayed by primer extension to enable direct comparison with the +2020 lines shown. Quantitative data for the +2020 lines was derived from titrations of 0.5, 0.25, and 0.125 μg of this RNA by primer extension. Expression levels are displayed as a percentage of the expression rat WAP determined in RNA isolated from the mammary gland of rats at day 12 of lactation, because the different extension efficiencies of the rat and mouse primers allowed better estimates of absolute RNA concentration between samples containing rat WAP than within a sample containing rat and mouse WAP. (B) Primer extension assay of +2020 lines. RNA (10 μg) isolated from nine positive F_0 female transgenic lines at day 10 of lactation was assayed by primer extension. RNAs from three female F_0 ΔUT/SV lines are also included (see Fig. 2A). Control tracks (left to right): RNAs obtained at day 10 of lactation from a rat and a nontransgenic mouse, respectively, extended with rat and mouse primers, and rat and mouse primers, respectively, alone. ↓, size markers (G sequencing track from pBluescript with the M13 primer). The positions of the rat (R)- and mouse (M)-specific bands are indicated. (C) Copy number dependence of transgene expression for rat WAP +2020 lines. The copy number of +2020 F_0 transgenic lines was determined three times in independent Southern analyses as described before (3). The expression level is presented as a percentage of diploid endogenous rat WAP expression and was determined as described above and in Materials and Methods. A linear correlation for the data can be derived (R^2 = 0.914) but is obviously dependent on the single high-copy-number line 44. (D) Copy number dependence of transgene expression for UT/SV lines. The determination was performed and is presented as in panel C.



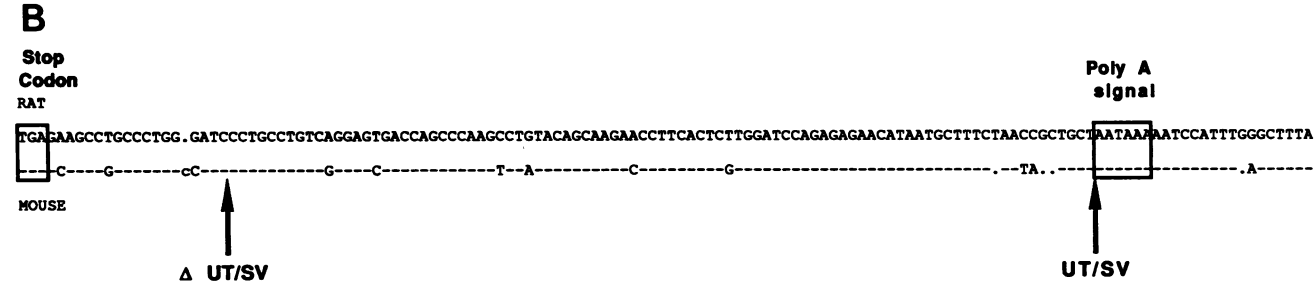


FIG. 2. Effect of rat WAP 3'UTR on expression. (A) Summary of constructs designed to investigate the function of the rat WAP 3'UTR. Expression levels of the transgenic mouse lines generated with each construct are also shown. Total RNA was isolated from lactating mammary glands and assayed by primer extension analysis as described in the legend to Fig. 1. (B) Comparison of rat and mouse 3'UTRs. The precise locations of the 3' deletions for the constructs ΔUT/SV and UT/SV are indicated. Dashes indicate conserved nucleotides, dots indicate gaps, and lowercase letters indicate an insertion.

promoter-fusion constructs containing up to 2.5 kb of the mouse WAP promoter were generally expressed at low levels relative to the endogenous gene (19). Thus, additional elements necessary for high-level expression are likely to lie within the WAP gene body.

Involvement of 3' elements in high-level expression. High-level expression elements within the WAP gene could include sequences involved in both transcriptional and post-transcriptional regulation. Since the WAP gene has a strongly conserved 3'UTR (Fig. 2B), deletions through this region were prepared in order to determine their effects on the level of expression. The first deletion removed the WAP poly(A) addition sequence. This was replaced with a fragment of DNA containing the SV40 late poly(A) addition sequence (Fig. 2A, UT/SV). All six of these UT/SV lines expressed the transgene, but at an average 23-fold-lower level than the +2020 construct (Fig. 1A and 2A), indicating that endogenous WAP 3' processing signals are more effective in these transgenic mice than SV40 sequences in allowing high-level expression. The 3' ends of transcripts from every UT/SV line were shown to map to the expected site of SV40 poly(A) addition by RNase protection studies (data not shown).

A further deletion of an additional 91 bp of the WAP 3'UTR (Fig. 2A, ΔUT/SV) resulted in only two of eight transgenic lines which expressed the transgene at levels equivalent to that of the longer UT/SV construct. To confirm the significance of this result, and to ask whether the deleted region could be functioning as a transcriptional enhancer or as an RNA stability element, the deleted 3'UTR region was reinserted into the ΔUT/SV construct, either upstream of the WAP promoter (5'UT/SV) or between the WAP and SV40 sequences in the sense (+UT/SV) and antisense (−UT/SV) orientations. Only two of the five 5'UT/SV lines expressed the transgene at levels comparable to the UT/SV construct, mimicking the ΔUT/SV results and suggesting that the UTR

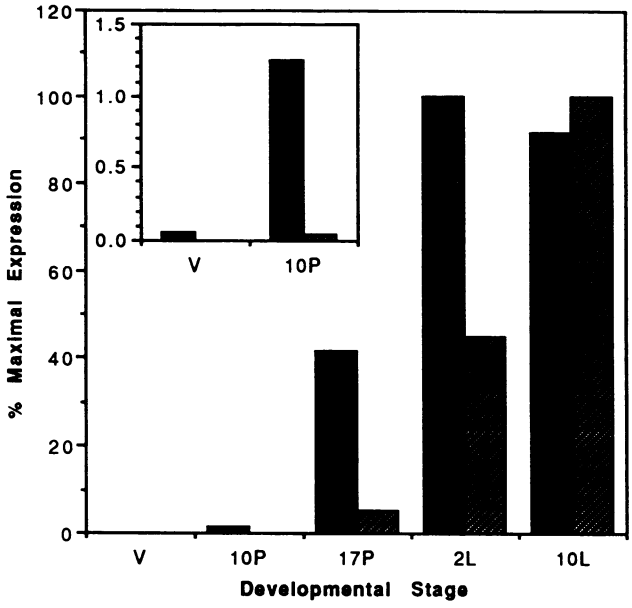


FIG. 3. Developmental regulation of endogenous and transgenic WAP mRNA. Expression levels of rat (solid bars) and mouse (hatched bars) WAP RNA were assayed by primer extension in mammary biopsy samples of virgin, pregnant (10 and 17 days), and lactating (2 and 10 days) glands. Induction is expressed as a percentage of maximal expression for rat or mouse WAP levels. The inset is a magnification of the scale to show the detectable low levels of WAP expression in the virgin and 10-day-pregnant mice. The mice were F_1 siblings from +2020 line 44.

does not function as an enhancer. All of the sense orientation +UT/SV lines (three of three) expressed the gene at levels comparable to those of the UT/SV lines, while none of the four antisense −UT/SV lines gave detectable levels of expression. As with lines carrying the +2020 transgene, transgenic mice with the UT/SV construct showed not only position-independent expression of the rat WAP transgene, but also a trend toward copy number dependence (Fig. 1D). Taken together, these data support the conclusion that the 91-bp region of the WAP UTR is important for position-independent expression, possibly in a copy number-dependent fashion, if not for the absolute level of expression.

Deregulation of transgene expression during early pregnancy. In rodents, the endogenous WAP is normally not expressed at significant levels until day 16 of pregnancy. However, precocious expression of rat and mouse WAP transgenes at days 7 to 10 of pregnancy has been described previously (17, 27). In order to determine how the high-expressing +2020 transgene was regulated during mammary development, we assayed WAP expression in five F_1 transgenic siblings (line 44) at different times of pregnancy and lactation (Fig. 3). Although detectable, virgin levels of WAP expression were over 1,000-fold below the maximal level. At various developmental stages (days 10 and 17 of pregnancy), the transgene was expressed at 31- and 8-fold-higher relative levels than the endogenous gene. Transgene expression peaked at day 2 of lactation but did not decline significantly by day 10 of lactation, as has been described for a genomic mouse WAP transgene (4). A second independent +2020 line (line 73) also showed no decline in expression between days 2 and 10 of lactation (data not shown). Since the elements involved in high-level WAP expression appear to lie largely within the +2020 construct, these data suggest that restriction of endogenous WAP expression in early pregnancy is likely to be mediated by negative control elements that lie outside the regions contained in the transgene.

Deregulation of cell-substratum control. Cell-substratum regulation of the WAP transgenes was also investigated. Expression was examined in primary cells derived from the highest-expressing +3250 transgenic line (3) and in transfectants of the CID 9 cell line that can be induced to form mammospheres and express endogenous WAP mRNA (Fig. 4B). Cells were plated either on a complex extracellular matrix (EHS) or on plastic culture dishes, on which the level of endogenous WAP expression is much reduced.

The levels of WAP expression in the primary cells and more particularly in the CID 9 cells were not sufficient for detection by the primer extension assay, so a rapid quantitative PCR-based assay was used to distinguish the rat and mouse RNAs. In order to obtain quantitative data, RNA concentrations and cycle numbers were selected to fall within the logarithmic range of the assay (Fig. 4A). To control for sample-to-sample variation, primers for the G3PD mRNA were included in each sample. Similar quantitative responses were obtained with rat WAP primers and for variable-input/constant-cycle titrations (data not shown).

In primary cells derived from a +3250 transgenic line, endogenous WAP RNA was expressed at an 18-fold-higher level in cells plated on EHS than in cells grown on plastic. By contrast, rat WAP RNA levels were only 3.7-fold greater in cells plated on EHS than in those cultured on plastic. To confirm this result, a variety of rat WAP constructs were tested in CID 9 cells (Fig. 4B and C). In all of the transfected cell lines, the rat WAP gene was induced only two- to fourfold by culture on EHS, while the endogenous gene was strongly induced (30- to 50-fold) (Fig. 4C). By estimating the absolute levels of rat and mouse WAP expression with primer efficiency titrations (data not shown), it was concluded that the low EHS-plastic induction ratio for rat WAP was mainly due to an increase in rat WAP levels in cells grown on plastic and not to decreased expression of the transgenes on EHS. This suggests that control elements involved in suppressing WAP expression on plastic may lie outside the region of the +3250 transgene. All of the transfected constructs tested in Fig. 4C for induction on EHS and plastic were expressed at similar levels on both EHS and plastic (Fig. 4A and data not shown). This contrasts with the dramatic effects of the same 3′-end deletions in transgenic mice (Fig. 1A and 2A). This result indicates that while the cells may be good models for substratum regulation, they cannot be used to study the effects of 3′ sequences on expression in the lactating gland.

DISCUSSION

Position-independent WAP gene expression in transgenic mice. The level of expression of transgenes normally varies from mouse to mouse and is often much less than that of the endogenous gene (26). The reasons for this variation are often attributed to "position effects" resulting from differences in the chromatin structure surrounding the integrated transgene. A new class of element that allows copy number-dependent, high-level expression has been characterized upstream of the α- and β-globin and CD2 genes (14, 20, 25, 37). These elements, termed LCRs, were initially found 40 to 60 kb upstream of the β-globin gene and are thought to define the boundaries of the gene cluster. By contrast, the results of this study show that elements located within a 2.9-kb frag-

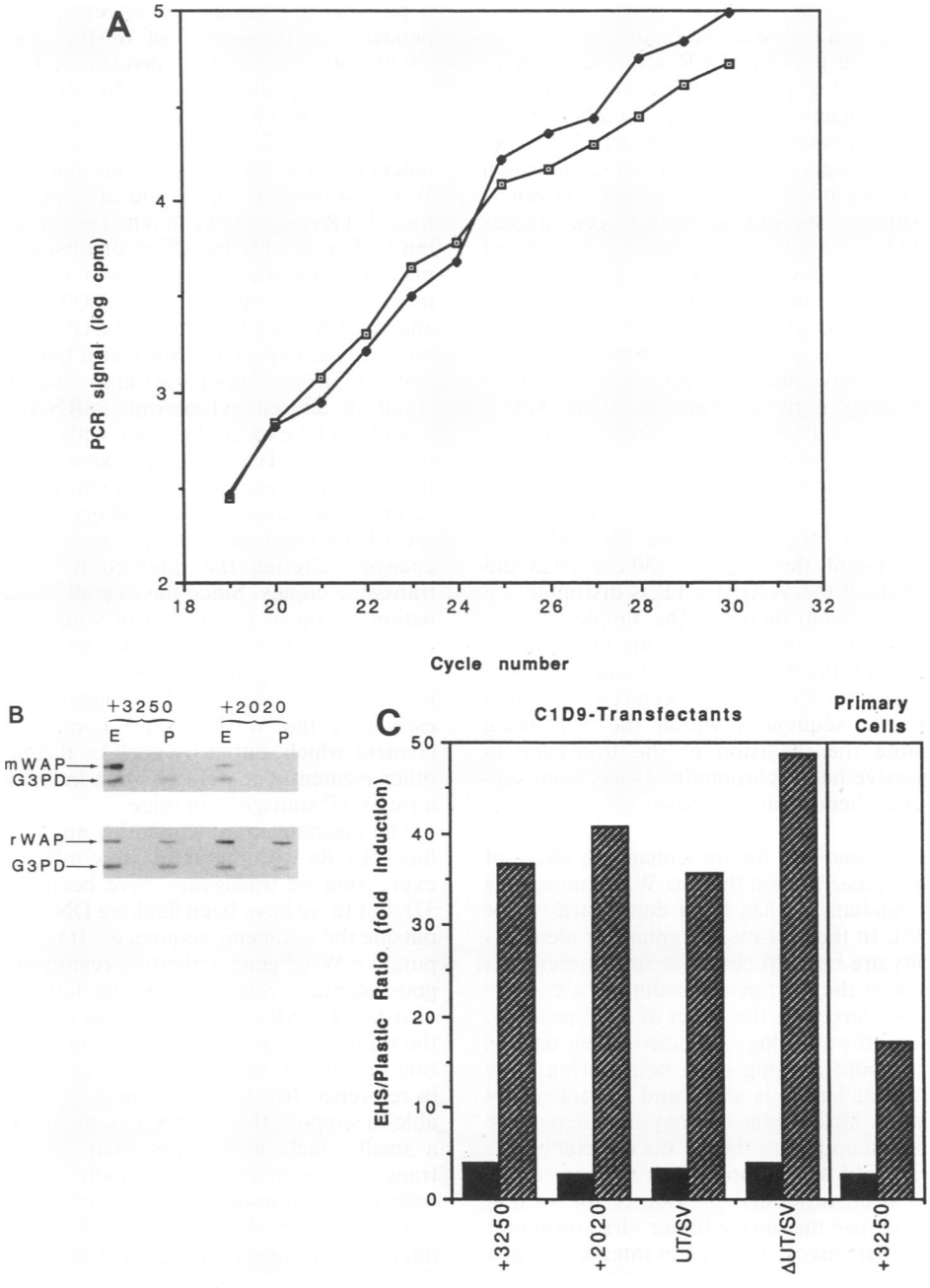


FIG. 4. Cell-substratum control of transfected and transgenic constructs. (A) Cycle dependence of mouse WAP amplification. A 0.1-μg equivalent of reverse-transcribed RNA from a +3250-transfected CID 9 cell line was amplified with mouse WAP and G3PD primers for the cycle numbers shown. Mouse WAP (◆) and G3PD (□) product levels were directly quantitated following gel electrophoresis and Southern transfer. (B) Cell-substratum regulation of endogenous and transfected genes in CID 9 cells. A 0.1-μg equivalent of reverse-transcribed RNA from +3250- and +2020-transfected CID 9 cells was amplified with mouse WAP and G3PD primers (top) or rat WAP and G3PD primers (bottom) for 23 cycles. Amplification products were visualized following transfer to a nylon membrane by hybridization with ^{32}P-labeled rat WAP and mouse G3PD cDNA probes. Lanes, E, cells plated on EHS; lanes P, cells plated on plastic. (C) Differential regulation by extracellular matrix of exogenous and endogenous WAP genes. G3PD-normalized counts for rat (solid bars) and mouse (hatched bars) WAP are expressed as a ratio of cpm on EHS to cpm on plastic. Mouse and rat WAP levels were determined by reverse transcription followed by amplification. Owing to the variability of endogenous WAP levels on plastic, data are only presented for rat and mouse WAP mRNA levels from those cultures in which the endogenous gene was strongly regulated by EHS. For the CID 9 transfectants, data for +3250 are the mean of three experiments (mouse, 36.6 ± 10.6; rat, 3.9 ± 1.2), data for +2020 are the mean of three experiments (mouse, 40.7 ± 10.5), data for UT/SV are for one experiment, and data for ΔUT/SV are the mean of two experiments. For the primary cells, data are for +3250 in one experiment. Data are combined from two independent sets of transfections. Rat WAP levels from the transgenic line could not be normalized to G3PD levels because of the large differences in transcript levels; however, Northern (RNA blot) analysis confirmed the induction ratio.

ment encompassing the rat WAP gene are sufficient for high-level, position-independent expression in the mammary gland of transgenic mice.

It is useful to distinguish between the related phenomena initially described for the β-globin LCR sequences. High-level expression is defined as mRNA levels which equal or exceed those of the endogenous gene. Copy number dependence occurs when expression varies linearly with the number of integrated transgenes. Finally, a transgene which is consistently expressed among independent lines is generally considered position or site independent. Recent studies with the chicken and human globin genes have demonstrated that copy number dependence is not obligatorily linked to high-level expression or to far-upstream elements (16, 28). Our data support the conclusion that site-independent expression as well as high-level expression can be mediated by elements in close proximity to a gene. Position site-independent and apparent copy number-dependent expression has also been observed for relatively small constructs containing another whey protein gene, the entire sheep β-lactoglobulin gene, in transgenic mice (8).

While high-level expression of the +2020 construct and the reduced expression of the UT/SV and +UT/SV constructs are site independent, the longer +3250 construct and WAP constructs in which the WAP 3'UTR is disrupted are expressed in a site-dependent manner. The simplest explanation for the appearance of apparent site-independent expression observed with the deletion of 3'-flanking DNA is that some form of repressor element has been removed. It is possible that repressor sequences within the 3'-flanking region could promote the inclusion of the transgene in transcriptionally inactive heterochromatin, as has been suggested for the genetic phenomenon of position effect variegation (18).

Two alternative explanations for the enhancing effect of the deletion rest on the assumption that the WAP transgenes have integrated in tandem, as has been demonstrated for other transgenes (26). In the first model, enhancer elements within one gene copy are brought closer to similar elements within the gene body of the next gene, leading to a cooperative interaction that overcomes the effect of local position. In the second model, the phasing of nucleosomes on one gene copy could promote phasing on a neighboring copy when the tandem repeat length is small and the spacing is correct. The phasing or lack of phasing may then determine whether the gene is susceptible to the effects of neighboring chromatin. The phasing of nucleosomes over transcriptional control regions has been described previously for several genes, including the mouse mammary tumor virus promoter (1). Both of these models involving tandem integration may also explain the lack of expression observed for the sole single-copy +2020 line.

Function of the WAP 3'UTR. The endogenous WAP gene accounts for 10 to 15% of total mRNA expression during lactation (21). Although the +3250 and +2020 constructs (Fig. 1A) are expressed at levels similar to the endogenous rat WAP gene, WAP promoters fused to various heterologous genes have shown mammary-specific expression but at relatively low levels (17, 19). These results suggest that combined elements within the WAP promoter and intragenic sequences are necessary for efficient WAP gene expression.

A potential intragenic determinant of high-level expression in the rat WAP gene is the highly conserved 3'UTR. Many 3'UTRs have been shown to influence posttranscriptional regulation. Therefore, constructs were designed to investigate the function of the rat WAP 3'UTR (Fig. 2A). It is important to note that these experiments were done in the context of the SV40 late poly(A) addition sequence, making it possible to examine the effect of the WAP 3'UTR independently of the control of WAP 3' processing. Use of the SV40 poly(A) sequence instead of the endogenous WAP sequences resulted in a >20-fold decrease in absolute expression levels in transgenic mice (Fig. 1A and 2A). However, the SV40 processing signals did not affect the site independence and possible copy number dependence of rat WAP expression (Fig. 1 and 2). Unexpectedly, deletion of the 3'UTR resulted in a much lower frequency of expression but had no discernible effect on absolute levels of rat WAP mRNA when the transgene was expressed. Only 4 of 17 transgenic lines with disrupted 3'UTRs (5'UT/SV, −UT/SV, and ΔUT/SV) expressed rat WAP mRNA, compared with nine of nine expressing lines with the 3'UTR intact (UT/SV and +UT/SV). Though striking, the effect may not be a result of altered cytoplasmic mRNA stability, since this would not be expected to vary with the site of integration of the transgene. However, we cannot exclude the possibility that certain integration sites might result in increased transcription or transcripts with alternative 3' ends. Disrupting the 3'UTR might also cause nucleosome phasing or spacing changes, altering the interaction of tandemly integrated transgene copies. Since the overall distance from the termination codon to the 3' end of some of these constructs is comparable (Fig. 2A, UT/SV and +UT/SV versus −UT/SV), such an explanation seems unlikely. Alternatively, the results can be interpreted to suggest a rather novel hypothesis: i.e., the WAP 3'UTR contains an RNA-mediated element which, either by itself or through cooperation with other elements, confers position-independent expression to a rat WAP transgene in mice.

The mechanism by which this might occur remains undefined. LCRs with the ability to confer position-independent expression on transgenes have been identified (14, 16, 28, 37), but these have been flanking DNA elements located well outside the intragenic sequences. It is unlikely, then, that the putative WAP gene activating region functions in an analogous manner. Studies with the human immunodeficiency virus TAT-TAR complex provide a biological precedent for the transcriptional rate affecting the interaction of transcription machinery with the newly transcribed RNA (reviewed in reference 10). Based on these observations, it is reasonable to suppose that an RNA element within the WAP gene, a small, single gene locus, may be able to interact with transcription machinery in such a way as to overcome negative position-dependent effects.

The effect of WAP 3' sequences observed in transgenic mice was not observed in transfected CID 9 cells, in which these constructs were all expressed at similar low levels. The failure to observe differences among these 3' constructs in the CID 9 cells suggests that factors involved in mediating the effect of the 3'UTR are not present in these cells or have no detectable effect at the low level of WAP expression observed early in development. Since CID 9 cells were initially derived from a mouse in midpregnancy (11), they may not contain some developmentally regulated factor(s) involved in the regulation of WAP gene expression. Alternatively, they might have lost such factors owing to culture in vitro.

Cell-substratum and developmental control. In transgenic animals, the effects of deregulation of transgene constructs can only be measured as a loss in overall expression or as aberrant temporal or spatial expression. For example, constructs containing the promoter of the fatty acid-binding

protein gene are expressed in a wider variety of cell types within the small intestine than the endogenous gene (35). By using an enriched population of COMMA-1D cells that are able to correctly regulate endogenous WAP expression in culture (CID 9 cells), we have been able to study the cell-substratum regulation of the WAP gene in pooled, stable transfectants. When CID 9 cells or primary mammary epithelial cells are plated on a complex extracellular matrix (EHS), they organize into hollow spheroid structures that secrete milk proteins into the lumen (2). The same cells, when plated on plastic, fail to organize and usually do not express WAP RNA. The differences in WAP expression on plastic and EHS may be due to both positive factors in the extracellular matrix that allow the epithelial cells to polarize as well as negative signals, synthesized by cells on plastic, that inhibit WAP expression (6).

The transgenic and transfected WAP genes largely fail to be regulated by cell-substratum, compared with the endogenous gene, which is expressed at 30- to 50-fold-higher levels on EHS. The lack of regulation of the transfected constructs and transgenes results from an increase in the level of rat WAP expression in cells cultured on plastic, not from their decreased expression on EHS, suggesting that the endogenous gene on plastic is normally maintained in a repressed state by negative control mechanisms that are localized outside the region of the transgene. Essentially the same argument regarding the presence of negative control regions present outside of the transgene region is used to account for the precocious expression of the rat WAP transgene during development. This raises the possibility that cell-substratum control may be the in vitro manifestation of the late induction of endogenous mouse WAP during development. It is tempting to speculate that the plastic and EHS phenotypes of the cells in culture may share some characteristics with mammary epithelial cells at different stages of development in vivo.

While the CID 9 cell line is a poor model for studying high-level WAP expression in the mammary gland during lactation, it can be used to dissect the mechanisms by which WAP is regulated by the extracellular matrix. Recent results (data not shown) indicate that the transfected rat WAP genes in CID 9 cells are incompletely regulated by lactogenic hormones, e.g., the endogenous WAP mRNA levels are induced about 60-fold by lactogenic hormones, while induction of transfected genes is usually reduced, ranging from 2- to 4-fold and, in one experiment, up to 10-fold). As with cell-substratum regulation, the decreased induction ratios may be solely the result of an increased basal level of expression. This raises the prospect that repressor elements necessary for normal regulation may be functionally dissected in vitro.

The mechanisms that lead to high-level, site-independent expression in transgenic animals are poorly understood. The small size of the WAP gene (2.9 kb) and the fact that WAP expression is not detrimental to the embryo should help in the elucidation of some of these phenomena. As elements within 185 bp of the 3′ end of the WAP gene are responsible for a large part of the optimal expression of WAP, WAP-based constructs for high-level expression should include these endogenous elements in preference to the heterologous processing signals. The precise mechanism by which the WAP 3′ sequences influence both the frequency and level of expression in transgenic mice remains to be established.

ACKNOWLEDGMENTS

We thank Jennifer Liao and Norman Greenberg for helpful comments on the manuscript, Tracey Duffy and Farnam Farzam for excellent technical assistance, April Kilburn for help in generating the 3′ deletion constructs, and Kathy Tucker for secretarial help.

This work was supported in part by NIH grant CA 16303 and USDA grant 88-37266-3951 to J.M.R. and by a Health Effects Research Division, U.S. Department of Energy (contract DE-AC03-76SF00098), award to M.J.B. C.S. is a fellow of the Schweizerische Nationalfonds.

REFERENCES

1. **Archer, T. K., M. G. Cordingley, R. G. Wolfrod, and G. L. Hager.** 1991. Transcription factor access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoter. Mol. Cell. Biol. **11**:688–698.
2. **Barcellos-Hoff, M. H., J. Aggler, T. G. Ram, and M. J. Bissell.** 1989. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. Development **105**:223–235.
3. **Bayna, E. M., and J. M. Rosen.** 1990. Tissue-specific, high level expression of the rat whey acidic protein gene in transgenic mice. Nucleic Acids Res. **18**:2977–2985.
4. **Burdon, T., L. Sankaran, R. J. Wall, M. Spencer, and L. Henninghausen.** 1991. Expression of a whey acidic protein transgene during mammary development: evidence for different mechanisms of regulation during pregnancy and lactation. J. Biol. Chem. **266**:6909–6914.
5. **Campbell, S. M., J. M. Rosen, L. G. Hennighausen, U. Strech-Jurk, and A. E. Sippel.** 1984. Comparison of the whey acidic protein genes of the rat and mouse. Nucleic Acids Res. **12**:8685–8697.
6. **Chen, L.-H., and M. J. Bissell.** 1989. A novel regulatory mechanism for whey acidic protein gene expression. Cell Regul. **1**:45–54.
7. **Chomczynski, P., and N. Sacchi.** 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.
8. **Clark, A. J.** 1991. Personal communication.
9. **Cleveland, D. W., and T. J. Yen.** 1989. Multiple determinants of eukaryotic mRNA stability. New Biol. **1**:121–126.
10. **Cullen, B. R.** 1990. The HIV TAT protein: a RNA sequence-specific processivity factor? Cell **63**:655–657.
11. **Danielson, K. G., C. J. Oborn, E. M. Durban, J. S. Butel, and D. Medina.** 1984. An epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro. Proc. Natl. Acad. Sci. USA **81**:3756–3760.
12. **Glisin, V., R. Crkvenjakov, and C. Byus.** 1974. Ribonucleic acid isolation by cesium chloride centrifugation. Biochemistry **13**: 2633–2637.
13. **Gorman, C.** 1986. High efficiency gene transfer into mammalian cells, p. 143–211. *In* D. M. Glover (ed.), DNA cloning. IRL Press, Oxford.
14. **Greaves, D. R., F. D. Wilson, G. Lang, and D. Kioussis.** 1989. Human CD2 3′-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. Cell **56**:979–986.
15. **Greenberg, N. M., J. D. Anderson, A. J. W. Hsueh, K. Nishimori, D. Avilla, J. Reeves, D. M. Ward, and J. M. Rosen.** 1991. Expression of biologically active, heterodimeric bovine follicle stimulating hormone in the milk of transgenic mice. Proc. Natl. Acad. Sci. USA **88**:8327–8331.
16. **Grosveld, F., G. B. van Assendelft, D. R. Greaves, and G. Kollias.** 1987. Position-independent, high level expression of the human β-globin gene in transgenic mice. Cell **51**:975–985.
17. **Gunzburg, W. H., B. Salmons, B. Zimmermann, M. Muller, V. Erfle, and G. Brem.** 1991. A mammary-specific promoter directs expression of growth hormone not only to mammary gland, but also to Bergman glia cells in transgenic mice. Mol. Endocrinol. **5**:123–133.
18. **Henikoff, S.** 1990. Position-effect variegation after 60 years. Trends Genet. **6**:422–426.
19. **Hennighausen, L.** 1990. The mammary gland as a bioreactor:

production of foreign proteins in milk. Protein Expression Purification **1**:3–8.

20. **Higgs, D. R., W. G. Wood, A. P. Jarman, J. Sharpe, J. Lida, I.-M. Pretorius, and H. Ayyub.** 1990. A major positive regulatory region located far upstream of the human α-globin gene locus. Genes Dev. **4**:1588–1601.
21. **Hobbs, A. A., D. A. Richards, D. J. Kessler, and J. M. Rosen.** 1982. Complex hormonal regulation of casein gene expression. J. Biol. Chem. **257**:3598–3605.
22. **Kleinman, H. K., M. L. McGarvey, J. R. Hassell, V. L. Star, F. B. Cannon, G. W. Lauriem, and G. R. Martin.** 1986. Basement membrane complexes with biological activity. Biochemistry **25**:312–318.
23. **Lee, E. Y.-H., W.-H. Lee, C. S. Kaetzel, G. Parry, and M. J. Bissell.** 1985. Interaction of mouse mammary epithelial cells with collagenous substrata: regulation of casein gene expression and secretion. Proc. Natl. Acad. Sci. USA **82**:1419–1423.
24. **Lee, K.-F., F. J. DeMayo, S. H. Atiee, and J. M. Rosen.** 1988. Tissue-specific expression of the rat β-casein gene in transgenic mice. Nucleic Acids Res. **16**:1027–1041.
25. **Orkin, S. H.** 1990. Globin gene regulation and switching: circa 1990. Cell **63**:665–672.
26. **Palmiter, R. D., and R. L. Brinster.** 1986. Germ-line transformation of mice. Annu. Rev. Genet. **20**:465–499.
27. **Pittius, C. W., L. Sankaran, Y. J. Topper, and L. Hennighausen.** 1988. Comparison of the regulation of the whey acidic protein gene with that of a hybrid gene containing the whey acidic protein gene promoter in transgenic mice. Mol. Endocrinol. **2**:1027–1032.
28. **Reitman, M. L. E., H. Westphal, and G. Felsenfeld.** 1990. Site-independent expression of the chicken beta A-globin gene in transgenic mice. Nature (London) **348**:749–752.
29. **Rosen, J. M., E. M. Bayna, and K.-F. Lee.** 1989. Analysis of milk protein gene expression in transgenic mice. Mol. Biol. Med. **6**:501–509.
30. **Rosen, J. M., T. Dale, S. Gavigan, and T. Bühler.** Molecular biology of milk proteins, in press. *In* M. E. Picciano and B. Lönnerdal (ed.), Mechanisms regulating lactation and infant nutrition (Contemporary Issues in Clinical Nutrition Series). John Wiley & Sons, Inc., New York.
31. **Rosen, J. M., J. R. Rodgers, C. H. Couch, C. A. Bisbee, Y. David-Inouye, S. M. Campbell, and L.-Y. Yu-Lee.** 1986. Multihormonal regulation of milk protein gene expression. Ann. N.Y. Acad. Sci. **478**:63–76.
32. **Sambrook, J., E. F. Frisch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Schmidhauser, C., M. J. Bissell, C. A. Myers, and G. F. Casperson.** 1990. Extracellular matrix and hormones transcriptionally regulate bovine β-casein 5′ sequences in stably transfected mouse mammary cells. Proc. Natl. Acad. Sci. USA **87**:9118–9122.
34. **Simons, J. P., M. McClenaghan, and A. J. Clark.** 1987. Alteration of the quality of milk by expression of β-lactoglobulin in transgenic mice. Nature (London) **328**:530–532.
35. **Sweetser, D. A., E. H. Birkenmeier, P. C. Hoppe, D. W. McKeel, and J. I. Gordon.** 1988. Mechanisms underlying generation of gradients of gene expression within the intestine: an analysis using transgenic mice containing fatty acid binding protein-human growth hormone fusion genes. Genes Dev. **2**:1318–1332.
36. **Thepot, D., E. Devinoy, M. L. Fontaine, C. Hubert, and L. M. Houdebine.** 1990. Complete sequence of the rabbit whey acidic protein gene. Nucleic Acids Res. **18**:3641.
37. **Townes, T. M., and R. R. Behringer.** 1990. Human globin locus activation region (LAR): role in temporal control. Trends Genet. **6**:219–223.
38. **Vilotte, J.-L., S. Soulier, M.-G. Stinnakre, M. Massoud, and J.-C. Mercier.** 1989. Efficient tissue-specific expression of bovine α-lactalbumin in transgenic mice. Eur. J. Biochem. **186**:43–48.