MOLECULAR AND CELLULAR BIOLOGY, Apr. 1995, p. 2063–2070 Vol. 15, No. 4
0270-7306/95/$04.00+0

Nuclear Factor I and Mammary Gland Factor (STAT5) Play a Critical Role in Regulating Rat Whey Acidic Protein Gene Expression in Transgenic Mice

SHI LI AND JEFFREY M. ROSEN*

Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030

Received 30 November 1994/Returned for modification 3 January 1995/Accepted 11 January 1995

The rat whey acidic protein (WAP) gene contains a mammary gland-specific and hormonally regulated DNase I-hypersensitive site 830 to 720 bp 5′ to the site of transcription initiation. We have reported previously that nuclear factor I (NFI) binding at a palindromic site and binding at a half-site are the major DNA-protein interactions detected within this tissue-specific nuclease-hypersensitive region. We now show that point mutations introduced into these NFI-binding sites dramatically affect WAP gene expression in transgenic mice. Transgene expression was totally abrogated when the palindromic NFI site or both binding sites were mutated, suggesting that NFI is a key regulator of WAP gene expression. In addition, a recognition site for mammary gland factor (STAT5), which mediates prolactin induction of milk protein gene expression, was also identified immediately proximal to the NFI-binding sites. Mutation of this site reduced transgene expression by approximately 90% per gene copy, but did not alter tissue specificity. These results suggest that regulation of WAP gene expression is determined by the cooperative interactions among several enhancers that constitute a composite response element.

The tissue- and developmental stage-specific regulation of milk protein gene expression is influenced by a variety of factors, including various peptide and steroid hormones and cell-cell and cell-substratum interactions (2, 6, 11, 18, 25, 27, 42, 48). Our laboratory has employed the rat whey acidic protein (WAP) gene, a major whey protein gene expressed in rodents, as a model system to study milk protein regulation. Previous studies have demonstrated that a 3.0-kb rat WAP genomic transgene (designated +2020) with 949 bp of 5′- and 70 bp of 3′-flanking DNA is efficiently expressed specifically in the mammary glands of transgenic mice in a copy-number-dependent manner (8). Two mammary gland-specific DNase I-hypersensitive sites (HSS) have been identified in the 5′-flanking region of the endogenous rat WAP gene and the +2020 transgene during lactation, i.e., HSS I at −800 to −700 and HSS II at −150 to −50 bp relative to the transcriptional initiation site. Deletion analysis performed on the +2020 transgene revealed that 853 bp of the 5′-flanking region was capable of eliciting high-level expression but that a transgene containing only 729 bp of 5′-flanking sequences failed to produce detectable expression during lactation (28). Therefore, the sequences within the distal HSS I are essential for high-level expression of the WAP transgene.

By using in vitro DNA binding assays, the transcription factor nuclear factor I (NFI) was identified as the major factor binding to HSS I. In addition, genomic footprinting demonstrated that a palindromic NFI-binding site was preferentially occupied in the mammary glands of lactating rats in comparison with the liver, strongly suggesting that the in vivo interactions at these sites are important for WAP gene expression (28). Cell transfection experiments have also suggested that NFI-binding sites localized in the proximal promoter regions of the mouse WAP and sheep β-lactoglobulin genes are critical for the transcriptional activation of these milk protein genes (32, 45).

Although prolactin, glucocorticoids, and insulin act synergistically to regulate WAP gene expression, glucocorticoids in particular exert a rapid and dramatic effect (18, 36). Several glucocorticoid receptor (GR)-binding sites (GREs) have been identified within HSS I by using an in vitro footprinting assay with baculovirus-expressed GR, and these were able to confer glucocorticoid inducibility to a heterologous promoter. It was also demonstrated that HSS I was glucocorticoid dependent (29), indicating that this region appears to play an important role in glucocorticoid regulation of WAP gene expression. Interestingly, sequence comparison revealed that a conserved consensus gamma interferon-activated sequence similar to the mammary gland factor (MGF)-binding site located in the β-casein promoter (39) was located at −726 bp relative to the cap site, immediately proximal to the NFI- and GR-binding sites. It has been suggested that MGF-binding sites may mediate prolactin responsiveness of milk protein genes (4, 43). This site was disrupted in the 729-bp 5′ deletion construct that was not expressed in transgenic mice. Therefore, the HSS I region detected in the rat WAP gene appears to be similar to several tissue-specific DNase I HSS detected in other genes that contain multiple binding sites for various transcription factors.

Because of the presence of multiple binding sites for different transcriptional factors in close proximity within the HSS I region, the precise roles of the individual NFI- and MGF-binding sites in WAP transgene expression cannot be determined by the analysis of deletion constructs. Furthermore, it is necessary to determine the importance of these binding sites under conditions that do not alter transgene spacing (24). Accordingly, the function of these binding sites in the HSS I region of the rat WAP +2020 transgene has been studied following site-directed mutagenesis and introduction of mutant constructs into transgenic mice. These experiments have substantiated the importance of both the NFI- and MGF (STAT5)-binding sites in regulating the level and the tissue specificity of WAP gene expression.

* Corresponding author. Mailing address: Department of Cell Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. Phone: (713) 798-6210. Fax: (713) 798-8012. Electronic mail address: jrosen@mbcr.bcm.tmc.edu.

MATERIALS AND METHODS

Plasmid constructions. All constructs described in this report were subcloned into the pBluescript II SK vector (Stratagene). In order to make point mutations in +2020 without changing spacing or sequences except for the designated mutations, the corresponding site-directed mutagenesis was carried out in pSL103, a plasmid with a rat WAP gene fragment spanning bp −949 to −450 (WAP −949/−450 fragment) in the vector. After site-directed mutagenesis, the mutated pSL103, designated pSL103m1-4, was digested with *Xba*I, filled in with Klenow fragment, and then digested with *Bam*HI. Afterwards, the 0.6-kb *Eco*RI (filled in)-*Bam*HI fragment (WAP +200/−450) from +2020 was inserted into these two sites. The resulting plasmid was digested with *Ava*I and *Eco*RI, and the fragment (WAP −806/+200) was ligated with the 4.9-kb *Ava*I (partial digestion)-*Eco*RI fragment from +2020 to make +2020m1 through +2020m4. All cloning junctions and the directed mutation sites were sequenced.

Site-directed mutagenesis by recombination PCR. Site-directed mutagenesis was carried out by using recombination PCR as described previously (50). To reduce the possibility of making nondesired mutations by PCR, the recombination PCR was performed on pSL103 instead of the entire +2020, and the four mutants +2020m1 to -4 were generated afterwards as described above. The sequences of the primers used for PCR were as follows (the nucleotide changes are in lowercase): (i) RPCR common primer 1, 5′-AACAGCGGTAAGATCCTT GAG-3′; (ii) RPCR common primer 2, 5′-AAACTCTCAAGGATCTTAC-3′; (iii) WAPFP1m forward, 5′-ACACCGTTttaACAGCATGGGGCCC-3′; (iv) WAP FP1m reverse, 5′-CATGCTGTtaaAACGGTGTCATGGG-3′; (v) WAPFP2m forward, 5′-AATGttaACAGTtaaCAACAGGACATCCCATC-3′; (vi) WAP FP2m reverse, 5′-GTTGttaACTGTtaaCATTGCTCTCGGGACATTTC-3′; (vii) WAP mutated oligo (D) forward, 5′-TGGGGCCCTTCTGctAAGTGGGCTT-3′; and (viii) WAP mutated oligo (D) reverse, 5′-AAGCCCACTTagCAGAAG GGCCCCA-3′.

The above-described primers i and ii are within the ampicillin resistance gene in the vector and were the common primers for the PCR amplifications. Primers iii and iv were used to generate pSL103m1; the mutation introduced destroyed the *Xcm*I site at WAP gene bp −757. Primers v and vi were used to generate pSL103m2; two *Hpa*I sites were introduced by the mutations. Primers vii and viii were used to generate pSL103m4.

For PCR amplification, pSL103 was first digested with either *Bam*HI or *Sal*I. Then PCR was carried out in *Taq* buffer (50 mM KCl, 10 mM Tris-HCl [pH 8.4], 2.0 mM $MgCl_2$) with 200 μM each deoxynucleoside triphosphate, 25 pmol of each primer, 2 ng of digested pSL103, and 1.25 U of *Taq* polymerase (Promega) in 50 μl. Reactants underwent 25 cycles of denaturation (94°C, 1 min), annealing (45°C, 2 min), and extension (72°C, 2 min), followed by a final extension (72°C, 7 min). For the reactions with WAPFP1m forward primer or WAP mutated oligo (D) forward or reverse primer, the annealing step was carried out at 55°C instead of 45°C. For each mutation, two linear products were generated after separate amplifications. Two microliters each of the two products were combined, and the mixture was used to directly transform competent *Escherichia coli*. The resulting colonies were screened first by restriction enzyme analysis to determine if the mutation introduced changes in the enzyme digestion pattern, and then putative mutants were confirmed by sequencing. To make pSL103m3, amplifications were carried out with WAPFP1m primers and pSL103m2 as templates instead of pSL103.

EMSA. An electrophoretic mobility shift assay (EMSA) was carried out essentially as described previously (28). The following double-stranded oligonucleotides were used as probes in the assays: oligo (A) (WAP gene bp −809 to −777),

5′-gatcCCGAGAGCAATGGGCACAGTGCCCAACAGGACA-3′
3′-GGCTCTCGTTACCCGTGTCACGGGTTGTCCTGTctag-5′,

and oligo (D) (WAP gene bp −740 to −720),

5′-gatccGGGCCCTTCTGAGAAGTGGGCa-3′
3′-gCCCGGGAAGACTCTTCACCCGtctag-5′.

Production and screening of transgenic mice. Transgenes were separated from the vector sequence by digestion with *Bss*HII and *Bgl*I, purified by preparative electrophoresis through agarose gels, and recovered by adsorption to glass beads (Bio 101, Inc., La Jolla, Calif.). Transgenic mice were generated, and mouse tail DNA was isolated as described previously (26). PCR was employed to screen for positive transgenic mice (8).

Southern analysis of transgenes. Positive lines were examined by Southern hybridization to determine the copy number of the transgenes. Five micrograms of tail DNA was digested with 40 U of *Bam*HI and 40 U of *Kpn*I overnight. Digested products were separated by electrophoresis on a 0.7% agarose gel and transferred to a Zeta-probe membrane (Bio-Rad) by the alkaline transfer method. Hybridization was performed as suggested by the manufacturer. The insert of the pX32 clone, a 420-bp partial rat WAP cDNA (18), was labeled by random priming and used as a probe. Data were quantitated with a Molecular Dynamics PhosphorImager system. Besides the 2.3-kb transgene, a 3.1-kb endogenous mouse WAP fragment was observed because of the extensive homology between the mouse and rat WAP cDNA sequences in this region (approximately 80%). The endogenous mouse WAP signal was used to normalize for loading and transfer efficiency differences among the samples. By comparing the transgene signal with the endogenous mouse WAP signal for each sample, the copy numbers could be determined. The endogenous mouse signal was determined to be approximately 75% of the hybridization signal observed for a line carrying two copies of the rat transgene.

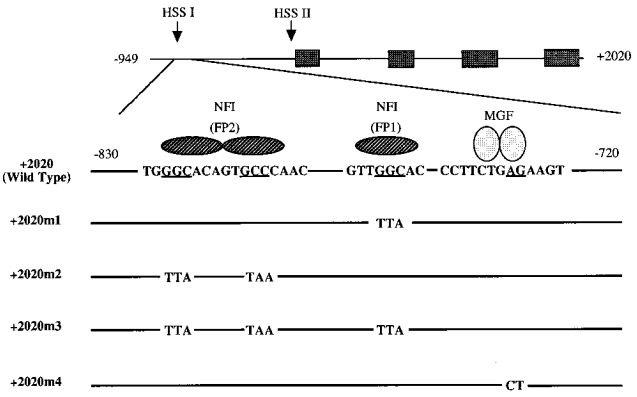

FIG. 1. Schematic representation of point mutations in WAP +2020 transgenes. The DNase I HSS are indicated by the vertical arrows, and the exons are represented by the shaded boxes. The sequences of two NFI-binding sites and one MGF (STAT5)-binding site in HSS I are shown. The base changes in the respective mutants are indicated.

To determine the arrangement of the transgenes, 5 μg of tail DNA was digested with 30 to 40 U of *Hinc*II, *Eco*RI, or *Pvu*II. Southern analysis was performed as described above with the pX32 probe. For most transgene constructs, these three enzymes only had one recognition sequence within the transgenes, and a head-to-tail or tail-to-tail integration pattern of the transgenes resulted in fragments with different expected lengths. The appearance of fragments of unexpected length was considered evidence for rearrangement of the transgenes.

RNA isolation and primer extension assay. RNA isolation and primer extension analysis were performed as described previously (28). Data were quantitated with the Molecular Dynamics PhosphorImager. The assays were repeated usually 2 to 3 times for each RNA sample in all studies.

RESULTS

Generation of mutated transgene constructs. NFI interacts with two sites in the WAP gene HSS I with different affinities in vitro (28). The footprint FP1 binding site at about bp −750 contains an NFI half-site, TGGCA, and exhibits a weak DNA-protein interaction; the FP2 site centered around bp −790 is a palindromic sequence sharing high homology with the NFI consensus binding sequence TGGC/AN_5GCCAA (10, 16, 17). This site binds to an NFI protein(s) with a much higher affinity than the NFI half-site. Although such binding activities were observed with both mammary and liver extracts, a strong mammary gland-specific interaction at FP2 in vivo has been demonstrated by genomic footprinting. To test the function of both FP1 and FP2 sites, mutations were introduced into the +2020 transgene by PCR-mediated site-directed mutagenesis, and three constructs with mutated NFI-binding sites, individually or in combination, were generated (Fig. 1). Three conserved bases in each of the NFI half-binding sites, which have been shown by contact point analysis to be essential for binding of NFI to the rat WAP promoter (28), were changed as indicated. All of these mutated constructs were based on the rat WAP genomic transgene +2020, which displays high-level, tissue-specific, and position-independent expression in transgenic mice (8). This permitted the functional definition of individual regulatory elements in mutated transgenes.

To demonstrate that these mutations indeed prevented NFI binding, an EMSA using a whole-cell extract prepared from the mammary glands of rats at day 2 of lactation was employed, and oligo (A), corresponding to the FP2 site, was used as a

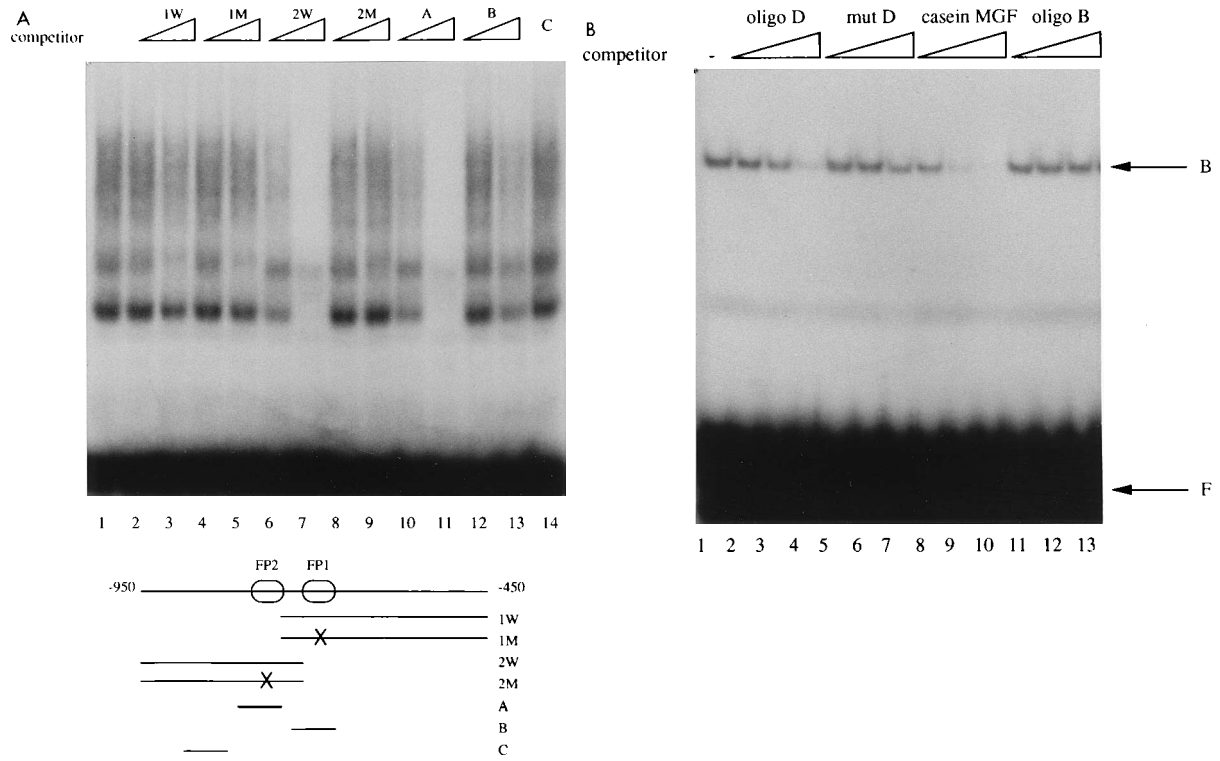

FIG. 2. Inhibitory effect of site-specific mutations on protein-DNA interactions at the NFI- and STAT5-binding sites in HSS I. (A) Mutations in the NFI-binding sites. The EMSA was performed with a whole-cell extract prepared from the mammary gland of a rat at day 2 of lactation, and oligo (A), corresponding to the FP2 site, was used as the probe. Competitors were as shown, and 1M and 2M stand for DNA fragments with mutated FP1 and FP2 sites, respectively. The molar excesses of competitor were 5- and 50-fold. (B) Mutations at the MGF (STAT5) site. EMSAs were performed with a double-stranded ^{32}P-labeled oligonucleotide [oligo (D)] incubated with 10 μg of whole-cell extract similar to that described for panel A in the absence (lane 1) or presence of increasing amounts of unlabeled competitor oligonucleotides as indicated (lanes 2 through 13). The molar excesses of the competitors were 10-, 50-, and 200-fold. The positions of the specifically retarded complex (B) and the unbound fraction of the probe (F) are shown by the arrows.

probe (Fig. 2A). As shown previously (28), multiple specific NFI-DNA complexes were formed with an oligonucleotide encompassing the FP2 site, and wild-type DNA fragments containing the FP2 site [2W and oligo (A)] were more efficient competitors than fragments containing the FP1 site [1W and oligo (B)]. Compared with the corresponding wild-type competitors, the oligonucleotides containing mutated sites failed to compete effectively, suggesting that each mutation indeed disrupted NFI binding. Thus, three constructs with mutated NFI sites in the context of an entire +2020 gene were generated and transgenic mice were produced by microinjecting these constructs into fertilized mouse eggs.

NFI-binding sites are essential for WAP transgene expression in the lactating mammary gland. Two independent transgenic lines were generated for construct +2020m1 (mutation of the NFI half-site), four were generated for +2020m2 (mutation of the NFI palindromic site), and six were generated for +2020m3 (mutation of both the NFI half-site and the NFI palindromic site). DNA Southern blot analysis (Fig. 3) revealed that all the transgenic mice contained multiple transgene copies (Table 1). Analysis of the integrity and arrangement of the transgenes by Southern analysis revealed that most transgene copies in each transgenic line were intact and present in a head-to-tail tandem array (data not shown). The expression level for each transgene was determined by a primer extension assay with a rat WAP-specific primer, and the expression of the endogenous mouse WAP gene was measured simultaneously with a mouse-specific primer as an internal control. Results of such an assay are shown in Fig. 4A. For most transgenic lines, two or more F_1 offspring were analyzed and the levels of transgene expression were very similar. In order to compare the expression levels of the different mutant constructs and the wild-type +2020, RNA from four independent lines of the +2020 transgene was analyzed in parallel. For all transgenic lines in which we failed to detect any transgene expression, the primer extension assays were repeated with much higher levels of RNA input (up to 30 μg) to ensure that we could detect extremely low levels of expression. The sensitivity of this assay is about 0.05% of the endogenous rat WAP gene expression level during lactation.

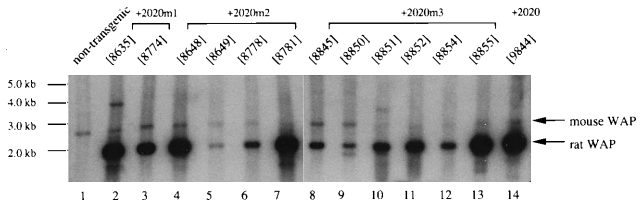

FIG. 3. Southern blot analysis for determining transgene copy number. DNA (5 μg) was digested with *Bam*HI and *Kpn*I, separated by agarose gel electrophoresis, transferred to a charged nylon membrane, and hybridized with a rat WAP cDNA probe. Quantitation of the hybridization was performed as described in Materials and Methods. The number at the top of each lane indicates the line number of each transgenic mouse. Normal mouse genomic DNA provided a negative control (lane 1), and tail DNA from the +2020 line (9844) (8) was included for comparison (lane 14). The positions of molecular weight markers and the transgene and endogenous gene signals are shown.

TABLE 1. Effects of NFI- and STAT5-binding site mutations on WAP transgene expression

Construct and line no.	Copy no.[a]	Relative expression level (%)[b]	Level of expression per gene copy (% of +2020)[c]
+2020m1			
8635	52	8.44	16.2
8774	21	1.21	5.8
+2020m2			
8648	53	~0.003	<0.006
8649	6	~0.006	<0.1
8778	14	ND[d]	
8781	48	ND	
+2020m3			
8850	4	ND	
8851	26	ND	
8852	31	ND	
8854	20	ND	
+2020m4			
9605	31	3.07	9.9
9867	1	ND	
9876	5	0.58	11.5
9877	2	ND	
1037	4	0.17	4.3

[a] The copy numbers of the constructs +2020m1 to +2020m4 were determined from Southern blots as described in Materials and Methods. One DNA sample from +2020 was reexamined by the same method, and the copy number estimated was similar to that determined previously (8).

[b] RNA expression levels were quantitated by using a primer extension assay and the Molecular Dynamics PhosphorImager, and the transgene expression levels were normalized to the endogenous mouse WAP mRNA levels.

[c] The expression level for each line was divided by the copy number, and it is shown relative to the average level of expression per gene copy for the +2020 transgene, which is designated 100%.

[d] ND, not detectable.

As summarized in Table 1, transgene +2020m1 was expressed in the mammary gland in both lines of mice at day 2 of lactation, although the average level of expression per gene copy was only about 10% of that of the +2020 gene (Fig. 4B). This result indicated that the NFI binding activity at FP1 plays a significant role in determining the overall activity of the WAP gene in vivo. Although a higher level of expression was observed for the transgenic line with the higher copy number (line 8635), it is not possible to determine if the expression was position independent, since only two independent, expressing transgenic lines were assayed.

Much more dramatic effects were observed for mutations in the NFI-binding site at FP2. Of the four lines of mice containing the +2020m2 transgene, two did not express the transgene during lactation and the other two exhibited levels of expression just above the detection limit of the primer extension assay (Table 1). The mutations in both NFI-binding sites of HSS I essentially eliminated transgene expression, as indicated by the lack of detectable expression in all four +2020m3 transgenic lines analyzed. Thus, both of the NFI-binding sites in HSS I are required for maximal WAP gene expression during lactation, and the NFI interactions at FP2 appear to be essential.

The expression of the +2020 transgene is essentially restricted to the mammary gland. To test whether the tissue specificity of WAP gene expression is altered by mutations in NFI sites, several nonmammary tissues from lactating transgenic mice containing +2020m1 (line 8635) and +2020m2 (line 8648) were analyzed. In all of the tissues examined, including kidney, lung, spleen, and liver, no transgene expression was observed (data not shown). A trace amount of transgene expression was detected in the thymus, but a low level of the endogenous mouse WAP RNA was also observed in the thymus. These data suggest that the tissue specificity of WAP transgene expression is not affected by mutations in these NFI-binding sites.

Mutation of the WAP MGF site. The binding site for the transcription factor MGF was first detected in the β-casein promoter, and it is highly conserved in other milk protein gene promoters (39, 44, 45). Although MGF activity is not mammary gland specific, it is regulated by lactogenic hormones and is thought to mediate prolactin induction of milk gene expression. It has also been suggested that MGF-binding sites are a subset of gamma interferon-activated sequences, and MGF is now known to be a member of the STAT (signal transducers and activators of transcription) family, designated STAT5 (37, 43). By sequence comparison with the consensus MGF-binding site TTCNNNGAA, a potential MGF-binding site, TTCTGAGAA, centered at −726 bp 5′ to the rat WAP gene transcription start site, was found (Fig. 1).

To test whether any specific DNA-protein interactions occur within this region, an EMSA using oligo (D), which spans bp −740 to −720, was performed. As shown in Fig. 2B, a complex with a retarded mobility was formed when the probe was incubated with a whole-cell extract prepared from the mammary gland of a rat at day 2 of lactation. This complex was inhibited specifically by increasing amounts of unlabeled oligo (D) or a double-stranded oligonucleotide corresponding to the MGF-binding site of the rat β-casein gene (37) but not by oligo (B), spanning WAP gene bp −763 to −729. Therefore, the interaction observed was specific for the MGF-binding site. A similar but weaker binding activity was detected when a whole-cell liver extract was used in the EMSA (data not shown).

In order to determine the role of this MGF-binding site in regulating WAP gene expression in vivo, +2020 containing a mutated MGF-binding site (Fig. 1) was analyzed in transgenic mice. The A-to-C and G-to-T changes were introduced into the MGF-binding site, since previous studies of the β-casein gene have demonstrated that these two contact sites are critical for MGF binding and activity (39). The mutant site competed less effectively with wild-type oligo (D) in the EMSA (Fig. 2B, lanes 5 through 7), indicating that binding to this site was markedly reduced by these specific point mutations.

Five independent lines of transgenic mice with construct +2020m4 were established, with copy numbers of transgenes ranging from 1 to 31. Except for one line carrying a single copy of the transgene, all lines showed a head-to-tail tandem array integration pattern (data not shown). The F_1 offspring of these five lines were analyzed for transgene expression levels. As shown in Fig. 5 and Table 1, during lactation three of five transgenic lines with +2020m4 expressed the transgene, although the average level of expression was only about 10% of the +2020 wild-type level when normalized for gene copy number. The two lines without detectable expression contained either one or two copies of the transgene. This result was similar to those obtained previously for the +2020 wild-type transgene (8, 24), i.e., when the copy number of the +2020 transgene was very low, the transgene was inactive. On the basis of the results for the three lines expressing +2020m4, mice with higher copy numbers appeared to express the transgene at higher levels, indicating that copy number-dependent expression may not depend upon the presence of an intact MGF-binding site (Fig. 5).

Although MGF binding activity was enriched in the mam-

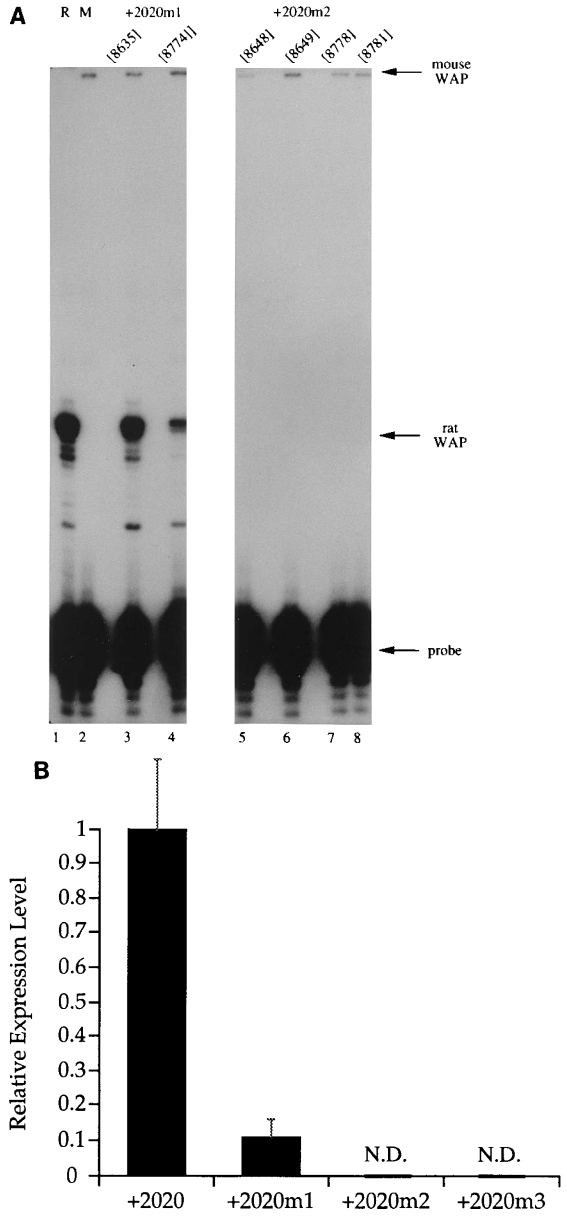

FIG. 4. Effect of NFI-binding site mutations on rat WAP transgene expression. (A) Primer extension assay of +2020m1 and +2020m2 transgenic mice. RNAs (5 μg) isolated from the mammary glands of mice at day 2 of lactation were incubated with a mixture of primers specific for the rat WAP and mouse WAP genes. The positions of the probe and the extension products for mouse or rat WAP mRNA are as designated in the figure. The line number for each transgenic mouse sample is indicated at the top of the gel. RNAs from a lactating rat mammary gland (lane 1) and a nontransgenic mouse mammary gland (lane 2) were included as controls. (B) Effect of NFI-binding site mutations on +2020 transgene expression. The bar graph shows the average levels of expression per gene copy for +2020 and for specific binding site mutants in the mammary gland at day 2 of lactation. N.D., not detected.

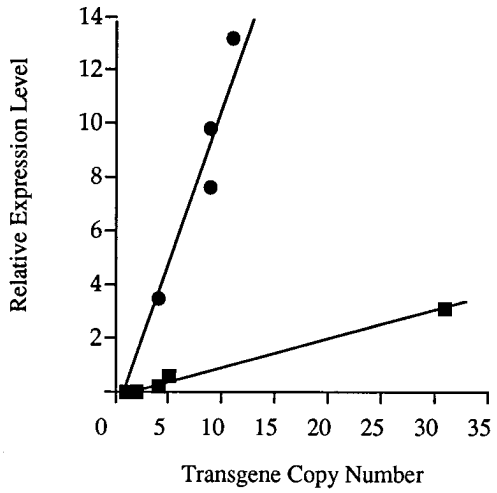

FIG. 5. Effect of STAT5-binding site mutations on WAP expression levels. The relative WAP transgene expression levels in the mammary gland at day 2 of lactation versus copy number are depicted for independent lines of +2020m4 or +2020 transgenic mice. ●, +2020; ■, +2020m4.

mary gland during late pregnancy and early lactation, it was also detected in several other tissues and cell lines (37). To examine whether the tissue specificity of transgene expression was affected by the mutation within the MGF-binding site, several other tissues were removed from line 9605, which contained 31 copies of the transgene, and analyzed for transgene expression levels (Fig. 6). None of the five nonmammary tissues displayed any level of detectable transgene expression. A trace amount of endogenous mouse WAP expression was, however, observed, as before, in thymus tissue. Therefore, the tissue specificity of WAP gene expression is not dependent solely on the presence of an intact MGF-binding site in the +2020 transgene.

DISCUSSION

NFI is a family of proteins generally regarded as ubiquitous transcriptional factors for various cellular and viral genes (22, 38, 40) and is required for virus replication (33). Recently, however, the role of NFI family members in tissue-specific gene regulation has received more attention, with the observation that different isoforms of NFI proteins vary with cell type and growth conditions (1, 14, 30, 34). NFI proteins are critical for the expression of liver-specific (5, 20) and brain-specific genes (41), for activation of an adipocyte-specific enhancer (15), and for epithelial cell-specific transcription (1). In most cases, NFI proteins cooperate with other ubiquitous or tissue-specific transcription factors whose binding sites are in close proximity. This appears also to be the situation for several milk protein genes, including the WAP gene.

In this study, mammary gland-specific WAP transgene expression was totally abolished in vivo by site-directed mutagenesis of two NFI-binding sites in the distal promoter region of the rat WAP gene. Both of these sites (FP1 and FP2) are within a mammary gland-specific DNase I-hypersensitive region, and the in vivo protein-DNA interactions at the high-affinity binding site (FP2) displayed a clear mammary gland preference. The DNA-protein interactions detected at the FP1 site, however, appeared to be weak not only in vitro, but also in vivo, since genomic footprinting failed to demonstrate clear occupancy at this site in either mammary or nonmammary nuclei (28). The results of our functional analyses of these two sites in the mutated +2020 transgene are in agreement with the affinities of these two sites. The disruption of the FP1 site, which contains a half NFI-binding site, resulted in a 90% reduction in expression, while mutations within the FP2 palindromic sequence essentially eliminated transgene expression. Although NFI-binding sites have been located in several milk protein gene promoters and these sites have been reported to

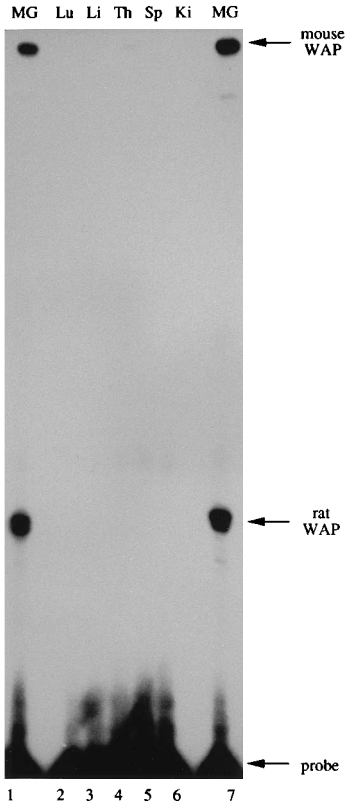

FIG. 6. Tissue-specific expression is retained in the presence of STAT5-binding site mutations. RNA primer extension assays were performed with RNA isolated from the mammary gland (2 μg) or from nonmammary tissues (10 μg) of the +2020m4 transgenic mouse line (9605). MG, mammary gland; Lu, lung; Li, liver; Th, thymus; Sp, spleen; Ki, kidney. Lanes 1 and 7 represent duplicate samples from the mammary gland. The exposure time was 10 h. Exposure for 24 to 40 h revealed a trace amount of mouse WAP RNA in the thymus (data not shown).

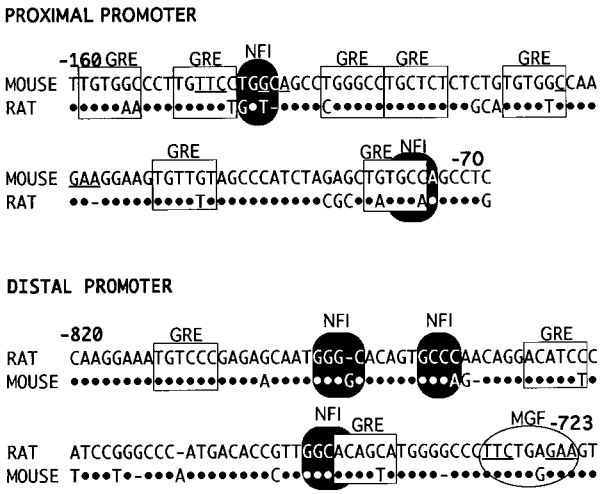

FIG. 7. Composite regulatory elements in the rat and mouse WAP gene promoters. The potential regulatory elements are as depicted. The locations of potential NFI-binding sites and GREs in the proximal promoter region are as described in references 32 and 47. The locations of NFI-binding sites, GREs, and MGF (STAT5)-binding sites in the distal promoter region are as described in references 28 and 29. The underlined nucleotides indicate the palindromic nature of the MGF-binding sites. These sites are not conserved in the proximal promoter. The numbers indicate positions relative to the start sites of transcription in the mouse and rat promoters, as indicated for the proximal and distal promoter regions.

be critical for promoter activity in cell transfection experiments (32, 45), this is the first evidence for NFI regulation of milk protein gene transcription in vivo. The dramatic effect of site-specific interactions at these NFI-binding sites was unexpected.

Although this study has concentrated on the HSS I region of the rat WAP gene, the HSS II region near the transcriptional initiation site may still play an important role in regulating WAP gene expression. In the mouse WAP gene, several GR- and NFI-binding sites are also present in a cluster in the proximal promoter region (32, 47), but these binding sites are not well conserved in the rat WAP gene (Fig. 7). Thus, it is possible that composite response elements have been duplicated in the flanking regions of the milk protein genes and that these have not been precisely conserved during mammalian evolution. This may explain the differences observed in the expression of the mouse, rat, and rabbit WAP gene constructs in transgenic mice (3, 8, 9). Other regions of the WAP gene proximal promoter may be important for the correct temporal regulation of expression during mammary development. For example, deletion of a conserved Ets-binding site in the proximal region of the mouse WAP promoter has been reported not to affect the activity of a WAP transgene, but it may influence the developmental timing of WAP gene expression (31).

As illustrated in Fig. 7, the NFI sites in the distal promoter region are flanked by GREs, and these binding sites are highly conserved in both the rat and mouse WAP genes. Considering the multiplicity of these sites, it is surprising that the mutation of a single NFI half-site should dramatically reduce the level of WAP gene expression. It has been suggested that a variant NFI present in the mammary gland recognizes a half-site better than the palindromic NFI consensus (32, 45). However, the palindromic FP2 site displayed a higher binding affinity in vitro and mammary gland-specific occupancy in vivo (28). It also appears functionally to be more important than FP1, but the roles of the individual NFI half-sites in the palindromic FP2 sites in transgenic mice were not evaluated individually. These results suggest that interactions of NFI at all three binding sites in the distal promoter region are critical for WAP gene regulation. These results also suggest that interactions of NFI with other transcription factors binding proximally to the NFI sites are important determinants of WAP gene expression.

Recently it has been reported that NFI has an essential functional motif within its proline-rich transcriptional activation domain and that this motif shares extensive homology with the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II (49). The CTD of polymerase II interacts directly with TFIID, suggesting that the CTD may be important for modulating the association of polymerase with the TFIID-TFIIB complex (7). Hyperphosphorylation of the CTD blocks incorporation of polymerase into the initiation complex, indicating that hyperphosphorylation may facilitate elongation by preventing competition with the initiation complex (35). It has been proposed that the CTD-like sequence of NFI may act through the same mechanism and interact with the general transcription machinery to stimulate transcription (49). This could explain why mutations of the NFI sites in the distal region of the rat WAP promoter exert such a dramatic effect on the level of WAP gene expression.

Site-specific mutation of the single MGF (STAT5)-binding site in the rat WAP distal promoter region resulted in a dramatic reduction in the level of transgene expression, but it did not alter the pattern of tissue-specific expression. No other

functional MGF (STAT5)-binding sites have been detected in the 949 bp of the rat WAP promoter. Several vestigial MGF (STAT5)-binding sites have been reported to exist in the proximal WAP promoter region (Fig. 7), but recent results suggest that these are not capable of binding MGF (46). Likewise, a recent study in which the individual MGF sites in the proximal region of the sheep β-lactoglobulin gene promoter were mutated both individually and collectively also reported a severalfold reduction in transgene expression while tissue-specific expression in the mammary glands of transgenic mice was maintained during lactation (4). These two studies performed with two different whey protein genes in transgenic mice suggest that while MGF may play a critical role in prolactin induction of milk protein gene expression, it alone may not be sufficient for mammary gland-specific expression. However, as with any mutagenesis study, a weak interaction of MGF, other related STATs, or even unrelated proteins with the mutated binding site may still be sufficient to elicit some transactivation activity. Therefore, as suggested by in vitro studies (4, 13, 43, 46), the role of MGF may be to modulate the prolactin induction of milk protein gene expression and to facilitate the synergistic induction of whey protein gene expression by prolactin and glucocorticoids. Interestingly, several NFI sites have been detected in close proximity to the MGF (STAT5)-binding sites in the β-lactoglobulin promoter, suggesting that cooperative interactions between factors binding to these sites may also be important for mammary gland-specific gene expression of whey protein genes.

Why then are these milk protein genes expressed specifically in the mammary gland during late pregnancy and lactation, and what role do NFI and MGF (STAT5) play in this process? One possibility is the existence of a yet-unidentified mammary gland-specific factor(s) or a modified form(s) of this factor which may determine the pattern of tissue-specific expression. For example, there may be alternative forms of NFI that are preferentially expressed in the mammary gland. A recent report concerning the function of NFI in determining the epithelial specificity of the human papillomavirus type 16 enhancer lends support to the hypothesis that NFI proteins play an essential role in cell-type-specific gene expression and that the composition of specific subsets of NFI proteins might vary in a cell-specific manner as a function of their differential expression, splicing, or heterodimerization (1). Recent results from our laboratory have also suggested that there may be additional STAT proteins that are closely related to MGF and could conceivably be preferentially expressed in the mammary gland (23).

However, a more likely possibility is that tissue specificity is controlled not by a single factor or family of factors that are exclusively expressed in mammary epithelial cells but by multiple factors whose cooperative interactions dictate the specific pattern of expression. For example, on the basis of some elegant experiments designed to elucidate the mechanisms controlling pattern formation in the *Drosophila* embryo, it has been proposed that enhancers can act as templates to bring weakly interacting regulatory factors into close proximity, so that they function combinatorially to regulate transcription. Furthermore, these enhancers can respond to signaling pathways to elicit the precise temporal and spatial patterns of gene expression (12, 21). In our studies of the rat WAP gene promoter multiple binding sites for the GR and for NFI and MGF proteins have been detected within a 100-bp region of the 5′-flanking sequence which displays a glucocorticoid-dependent and mammary gland-specific DNase I hypersensitivity (28, 29) indicating an alteration of local chromatin structure. Mutation of either the NFI or MGF sites had a dramatic effect on the level of WAP expression. Thus, a model in which the interplay among different transcriptional factors controls both the hormonal induction and tissue specificity of WAP gene expression appears to fit the available data. During late pregnancy and lactation, glucocorticoid stimulation may result in GR-mediated changes in chromatin structure that relieve the repressive role of nucleosomes and create an "open window" for the interaction of other nonhistone DNA-binding proteins. This could increase the accessibility of nearby NFI- and MGF-binding sites and the subsequent binding and transactivation by NFI and MGF, resulting in the activation of WAP gene expression. Binding of MGF may either directly facilitate NFI function because of direct protein-protein interactions or help to maintain open chromatin structure. This model can account for the dramatic effect of glucocorticoids on WAP gene expression and the requirement for the synergistic interaction with prolactin mediated by the induction of MGF tyrosine phosphorylation through the JAK-STAT signaling pathway (13, 19). Further experiments will be required to confirm the validity of this model.

ACKNOWLEDGMENTS

These studies were supported by grant CA16303 from the National Institutes of Health and grant 92-37205-8263 from the U.S. Department of Agriculture.

We thank Joyce Pike for the generation of transgenic mice; Shu-Wen Sun for help with animal husbandry; and Norman Greenberg, Darryl Hadsell, and Brian Raught for helpful comments on the manuscript. We also thank Lothar Hennighausen and John Clark for providing preprints of their manuscripts prior to their publication.

REFERENCES

1. **Apt, D., Y. Liu, and H.-U. Bernard.** 1994. Cloning and functional analysis of spliced isoforms of human nuclear factor I-X: interference with transcriptional activation by NFI/CTF in a cell-type specific manner. Nucleic Acids Res. **22:**3825–3833.
2. **Blum, J. L., M. E. Zeigler, and M. S. Wicha.** 1987. Regulation of rat mammary gene expression by extracellular matrix components. Exp. Cell Res. **173:**322–340.
3. **Burdon, T., L. Sankaran, R. J. Wall, M. Spencer, and L. Hennighausen.** 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.
4. **Burdon, T. G., K. A. Maltland, A. J. Clark, R. Wallace, and C. J. Watson.** 1994. Regulation of the sheep β-lactoglobulin gene by lactogenic hormones is mediated by a transcription factor that binds an interferon-γ activation site-related element. Mol. Endocrinol. **8:**1528–1536.
5. **Cereghini, S., M. Raymondjean, A. G. Carranca, P. Herbomel, and M. Yaniv.** 1987. Factors involved in control of tissue-specific expression of albumin gene. Cell **50:**627–638.
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. **Conaway, R. C., J. N. Bradsher, and J. W. Conaway.** 1992. Mechanism of assembly of the RNA polymerase II preinitiation complex. J. Biol. Chem. **267:**8464–8467.
8. **Dale, T. C., M. J. Krnacik, C. Schmidhauser, C. L. Yang, M. J. Bissell, and J. M. Rosen.** 1992. High-level expression of the rat whey acidic protein gene is mediated by elements in the promoter and 3′ untranslated region. Mol. Cell. Biol. **12:**905–914.
9. **Devinoy, E., N. R. Malienou, D. Thepot, C. Puissant, and L. M. Houdebine.** 1991. Hormone responsive elements within the upstream sequences of the rabbit whey acidic protein (WAP) gene direct chloramphenicol acetyl transferase (CAT) reporter gene expression in transfected rabbit mammary cells. Mol. Cell. Endocrinol. **81:**185–193.
10. **de Vries, E., W. van Driel, S. J. L. van den Heuvel, and P. C. van der Vliet.** 1987. Contactpoint analysis of the Hela nuclear factor I recognition site reveals symmetrical binding at one side of the DNA helix. EMBO J. **6:**161–168.
11. **Ganguly, R., N. Ganguly, N. M. Mehta, and M. R. Banerjee.** 1980. Absolute requirement of glucocorticoid for expression of the casein gene in the presence of prolactin. Proc. Natl. Acad. Sci. USA **77:**6003–6006.
12. **González-Crespo, S., and M. Levine.** 1994. Related target enhancer for dorsal and NF-κB signaling pathways. Science **264:**255–258.
13. **Gouilleux, F., H. Wakao, M. Mundt, and B. Groner.** 1994. Prolactin induces phosphorylation of tyr694 of Stat5 (MGF), a prerequisite for DNA binding

and induction of transcription. EMBO J. **13:**4361–4368.
14. **Goyal, N., J. Knox, and M. Gronostajski.** 1990. Analysis of multiple forms of nuclear factor I in human and murine cell lines. Mol. Cell. Biol. **10:**1041–1048.
15. **Graves, R. A., P. Tontonoz, S. R. Ross, and B. M. Spiegelman.** 1991. Identification of a potent adipocyte-specific enhancer: involvement of an NF-1-like factor. Genes Dev. **5:**428–437.
16. **Gronostajski, R. M.** 1986. Analysis of nuclear factor I binding to DNA using degenerate oligonucleotides. Nucleic Acids Res. **14:**9117–9131.
17. **Gronostajski, R. M., S. Adhya, K. Nagata, R. A. Guggenheimer, and J. Hurwitz.** 1985. Site-specific binding of nuclear factor I: analysis of cellular binding sites. Mol. Cell. Biol. **5:**964–971.
18. **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.
19. **Ihle, J. N., B. A. Witthuhn, F. W. Quelle, K. Yamamoto, W. E. Thierfelder, B. Kreider, and O. Silvennoinen.** 1994. Signaling by the cytokine receptor superfamily: JAKs and STATs. Trends Biochem. Sci. **19:**222–227.
20. **Jackson, D. A., K. E. Rowader, K. Stevens, C. Jiang, P. Milos, and K. S. Zaret.** 1993. Modulation of liver-specific transcription by interactions between hepatocyte nuclear factor III and nuclear factor I binding DNA in close apposition. Mol. Cell. Biol. **13:**2401–2410.
21. **Jiang, J., and M. Levine.** 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell **72:**741–752.
22. **Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian.** 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48:**79–89.
23. **Kazansky, A., L.-Y. Yu-Lee, and J. M. Rosen.** Unpublished observations.
24. **Krnacik, M., S. Li, J. Liao, and J. M. Rosen.** Position-independent expression of whey acidic protein transgenes. Submitted for publication.
25. **Lee, E. Y.-H., G. Parry, and M. J. Bissell.** 1984. Modulation of secreted protein of mouse mammary epithelial cells by collagenous substrata. J. Cell Biol. **98:**146–155.
26. **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.
27. **Levine, J. F., and F. E. Stockdale.** 1985. Cell-cell interactions promote mammary epithelial cell differentiation. J. Cell Biol. **100:**1415–1422.
28. **Li, S., and J. M. Rosen.** 1994. Distal regulatory elements required for rat whey acidic protein gene expression in transgenic mice. J. Biol. Chem. **269:**14235–14243.
29. **Li, S., and J. M. Rosen.** 1994. Glucocorticoid regulation of rat whey acidic protein gene expression involves hormone-induced alterations of chromatin structure in the distal promoter region. Mol. Endocrinol. **8:**1328–1335.
30. **Lichtsteiner, S., J. Wuarin, and U. Schibler.** 1987. The interplay of DNA-binding proteins in the promoter of the mouse albumin gene. Cell **51:**963–973.
31. **McKnight, R. A., T. Burdon, M. Spencer, R. J. Wall, and L. Hennighausen.** A transgenic mouse study on a mammary specific gene defines the in vivo significance of transcription factor binding sites and upstream insulator sequences. Submitted for publication.
32. **Mink, S., E. Hartig, P. Jennewein, W. Doppler, and A. C. B. Cato.** 1992. A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor. Mol. Cell. Biol. **12:**4906–4918.
33. **Nagata, K., R. A. Guggenheimer, T. Enomoto, and J. H. Lichy.** 1983. Adenovirus DNA replication *in vitro*: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc. Natl. Acad. Sci. USA **79:**6438–6442.
34. **Paonessa, G., F. Gounari, R. Frank, and R. Cortese.** 1988. Purification of a NFI-like DNA-binding protein from rat liver and cloning of the corresponding cDNA. EMBO J. **7:**3115–3123.
35. **Peterson, M. G., and R. Tjian.** 1992. Transcription. The tell-tail trigger. Nature (London) **358:**620–621.
36. **Puissant, C., and L. M. Houdebine.** 1991. Cortisol induces rapid accumulation of whey acid protein mRNA but not of αsl and β-casein mRNA in rabbit mammary explants. Cell Biol. Int. Rep. **15:**121–129.
37. **Raught, B., B. Khursheed, A. Kazansky, and J. Rosen.** 1994. YY1 represses β-casein gene expression by preventing the formation of a lactation-associated complex. Mol. Cell. Biol. **14:**1752–1763.
38. **Rossi, P., G. Karsenty, A. B. Roberts, N. S. Roche, M. B. Sporn, and B. de Crombrugghe.** 1988. A nuclear factor I binding site mediates the transcriptional activation of collagen promoter by transformation growth factor-β. Cell **52:**405–414.
39. **Schmitt-Ney, M., W. Doppler, R. K. Ball, and B. Groner.** 1991. β-Casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary-gland-specific nuclear factor. Mol. Cell. Biol. **11:**3745–3755.
40. **Shaul, Y., R. Ben-Levy, and T. De-Medina.** 1986. The high affinity binding site for nuclear factor I next to the hepatitis B virus S gene promoter. EMBO J. **5:**1967–1971.
41. **Tamura, T., M. Masayuki, K. Ikenaka, and K. Mikoshiba.** 1988. Analyses of transcription control elements of the mouse myelin basic protein gene in Hela cell extracts: demonstration of a strong NF-I binding motif in the upstream region. Nucleic Acids Res. **16:**11441–11459.
42. **Topper, Y. J., and C. S. Freeman.** 1980. Multiple hormone interactions in the developmental biology of the mammary gland. Physiol. Rev. **60:**1049–1106.
43. **Wakao, H., F. Gouilleux, and B. Groner.** 1994. Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. EMBO J. **13:**2182–2191.
44. **Wakao, H., N. M. Schmitt, and B. Groner.** 1992. Mammary gland-specific nuclear factor is present in lactating rodent and bovine mammary tissue and composed of a single polypeptide of 89 kDa. J. Biol. Chem. **267:**16365–16370.
45. **Watson, C. J., K. E. Gorden, M. Robertson, and A. J. Clark.** 1991. Interaction of DNA-binding proteins with a milk protein gene promoter *in vitro*: identification of a mammary gland-specific factor. Nucleic Acids Res. **19:**6603–6610.
46. **Welte, T., K. Garimorth, S. Philipp, and W. Doppler.** 1994. Prolactin-dependent activation of a tyrosine phosphorylated DNA binding factor in mouse mammary epithelial cells. Mol. Endocrinol. **8:**1091–1102.
47. **Welte, T., S. Philipp, C. Cairns, J. A. Gustafsson, and W. Doppler.** 1993. Glucocorticoid receptor binding sites in the promoter region of milk protein genes. J. Steroid Biochem. Mol. Biol. **47:**75–81.
48. **Wiens, D., C. S. Park, and F. E. Stockdale.** 1987. Milk protein gene expression and ductal morphogenesis in the mammary gland *in vitro*: hormone-dependent and -independent phases of adipocyte-mammary epithelial cell interaction. Dev. Biol. **120:**245–258.
49. **Xiao, H., J. T. Lis, H. Xiao, J. Greenblatt, and J. D. Friesen.** 1994. The upstream activator CTF/NFI and RNA polymerase II share a common element involved in transcriptional activation. Nucleic Acids Res. **22:**1966–1973.
50. **Yao, Z., D. H. Jones, and C. Grose.** 1992. Site-directed mutagenesis of herpesvirus glycoprotein phosphorylation sites by recombination polymerase chain reaction. PCR Methods Appl. **1:**205–207.