Transcriptional Regulation of Murine β1,4-Galactosyltransferase in Somatic Cells

ANALYSIS OF A GENE THAT SERVES BOTH A HOUSEKEEPING AND A MAMMARY GLAND-SPECIFIC FUNCTION*

Bhanu Rajput, Nancy L. Shaper, and Joel H. Shaper‡

From the Cell Structure and Function Laboratory, Oncology Center and the Department of Pharmacology and Molecular Sciences, School of Medicine, Johns Hopkins University, Baltimore, Maryland 21287-8937.

β1,4-Galactosyltransferase (β4-GT) is a constitutively expressed enzyme that synthesizes the β4-N-acetylgalactosamine structure in glycoconjugates. In mammals, β4-GT has been recruited for a second biosynthetic function, the production of lactose which occurs exclusively in the lactating mammary gland. In somatic tissues, the murine β4-GT gene specifies two mRNAs of 4.1 and 3.9 kilobases (kb), as a consequence of initiation at two different start sites—200 base pairs apart. We have proposed that the region upstream of the 4.1-kb start site functions as a housekeeping promoter, while the region adjacent to the 3.9-kb start site functions primarily as a mammary gland-specific promoter (Harduin-Lepers, A., Shaper, J. H., and Shaper, N. L. (1993) J. Biol. Chem. 268, 14348–14359).

Using DNase I footprinting and electrophoretic mobility shift assays, we show that the region immediately upstream of the 4.1-kb start site is occupied mainly by the ubiquitous factor Sp1. In contrast, the region adjacent to the 3.9-kb start site is bound by multiple proteins which include the tissue-restricted factor AP2, a mammary gland-specific form of CTF/NF1, Sp1, as well as a candidate negative regulatory factor that represses transcription from the 3.9-kb start site. These data experimentally support our conclusion that the 3.9-kb start site has been introduced into the mammary gland-specific promoter (β4-GT) gene to accommodate the recruited role of β4-GT in lactose biosynthesis.

β1,4-Galactosyltransferase (β4-GT) is a trans-Golgi resident, type II membrane-bound glycoprotein that is widely distributed in the vertebrate kingdom. It catalyzes the transfer of galactose to N-acetylgalcosamine residues, forming the β4-N-acetylgalactosaminyl (Galβ4-GlcNAc) or poly-N-acetylgalactosamine structure found in glycolipids and the N- and O-linked side chains of glycoproteins and proteoglycans. Since glycoconjugate biosynthesis occurs in essentially all tissues, it can be considered a housekeeping function. In mammary glands, β4-GT has been recruited for an additional tissue-specific biosynthetic function, which is the production of lactose (Galβ4-Glc) in the lactating mammary gland (LMG).

The synthesis of lactose is catalyzed by the protein heterodimer, lactose synthetase (EC 2.4.1.22), which is assembled from β4-GT and α-lactalbumin. The net result of this association is to lower the Km of glucose for β4-GT about three orders of magnitude, thus making glucose an effective acceptor substrate at physiological concentration. α-Lactalbumin is synthesized exclusively in the epithelial cells of the mammary gland beginning in late pregnancy. Enzymatic levels of β4-GT also increase in the mammary gland beginning in mid-pregnancy, in preparation for lactose biosynthesis (3). The expression of both α-lactalbumin and β4-GT is positively influenced by the lactogenic hormones, insulin, hydrocortisone, and prolactin (3).

We have shown that the murine (4) and bovine (5) β4-GT genes specify two mRNAs of ~4.1 and ~3.9 kb in somatic cells. The two transcripts are generated as a result of initiation at two different start sites located on exon 1, and separated by ~200 bp. The main difference between the two mRNAs is the length and extent of predicted secondary structure present in the respective 5'-untranslated region (6). Because each start site is positioned either upstream of the first two in-frame ATGs (4.1 kb) or between these two in-frame ATGs (3.9 kb), translation of the two mRNAs results in the synthesis of two functional, structurally related protein isoforms that differ only in the lengths of their NH2-terminal cytoplasmic domain (reviewed in Shaper and Shaper (7)).

The 4.1-kb start site is predominantly used in all somatic cells and tissues examined. An exception is found in the mid-to-late pregnant and lactating mammary gland, where the 3.9-kb start site is preferentially utilized (6). This switch to the predominant use of the 3.9-kb start site is coincident with the cellular requirement for increased levels of β4-GT enzyme for lactose biosynthesis. These observations, combined with a promoter deletion analysis using β4-GT/CAT hybrid constructs, led us to propose a model for transcriptional and translational regulation of the β4-GT gene in which the distal region upstream of the 4.1-kb start site functions as a housekeeping promoter in all somatic cells, while the proximal region upstream of the 3.9-kb start site serves primarily as a mammary gland-specific promoter. In addition, we proposed that a putative negative regulatory region identified adjacent to the 3.9-kb start site, down-regulates transcription from this start site in all somatic tissues except the mid-to-late pregnant and lactating mammary gland.
ing mammary gland. The key feature of our model is that mammals have evolved a two-step mechanism to generate the elevated levels of β4-GT enzymatic activity required for lactose biosynthesis. First, there is an up-regulation of the steady state levels of β4-GT mRNA by the predominant synthesis of the transcript (3.9 kb) that is regulated by mammary gland-specific factors. Second, the 3.9-kb β4-GT transcript with its short (~20 nucleotides), less structured 5′-untranslated region is translated more efficiently compared to its housekeeping counterpart (4.1 kb) which has a long (~200 nucleotides), highly structured 5′-untranslated region (6).

In this study, we have focused on verifying those predictions of our model pertaining to the transcriptional regulation of the β4-GT gene. We have used DNase I protection and electrophoretic mobility shift assays (EMSAs) to identify specific cis-acting elements and the corresponding trans-acting factors potentially involved in the expression of the 4.1-kb and the 3.9-kb β4-GT transcripts. We show that the distal promoter region immediately upstream of the 4.1-kb start site is bound primarily by the ubiquitous transcription factor Sp1. In contrast, the proximal promoter region adjacent to the 3.9-kb start site is a target for binding by multiple proteins which include a candidate negative regulatory factor, Sp1, a mammary gland-specific form of CTF/NF1 and the tissue-restricted factor, AP2.

EXPERIMENTAL PROCEDURES

Materials—Reagents for molecular biology and tissue culture were from Life Technologies, Inc. 32P-Labeled radiolabeled were from Amersham Corp. All protease inhibitors, formic acid (99%), and piperidine were from Sigma. Protein assay dye reagent was from Bio-Rad. Poly(dI-dC) proteinase K, and calf intestinal alkaline phosphatase were from Boehringer Mannheim. DNase I was from Cooper Biomedical. Mid-pregnant Swiss Webster mice were obtained from Harlan Sprague-Dawley Laboratory Animals. Purified Sp1 protein and anti-Sp1 and anti-AP2 antibodies were from Santa Cruz Biotechnology Inc. Anti-CTF/NF1 antisera was a kind gift from Dr. N. Tanese (New York University Medical Center).

Cells and Cell Culture—Mouse L-cells were obtained from ATCC and maintained in Dulbecco’s modified Eagle’s medium supplemented with 10% horse serum, 100 units/ml penicillin, and 50 μg/ml streptomycin at 37 °C in 5% CO2.

Preparation of Nuclear Extracts—Nuclear extracts from mouse L-cells (~90% confluent) were prepared according to the method of Din- nam et al. (8) and from mouse brain and LMG by the combination of methods of Roy et al. (9) and Dinnam et al. (8). Briefly, frozen tissue (2 g) was pulverized under liquid nitrogen to a fine powder, using a mortar and pestle, and transferred to an ice-cold Dounce homogenizer (type B pestle) containing 20 ml of NE2 buffer (20 mM Tris-HCl, pH 7.9, 140 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 25 mM KCl, 2 mM MgCl2, 0.15 mM spermine, 0.5 mM spermidine, and 1 mM dithiothreitol). The number of strokes required to lyse the cells depended on the tissue type, and this step was monitored by checking aliquots of the lysate with a phase-contrast microscope. The nuclei were centrifuged at 1000 × g for 10 min. The nuclear pellet was washed once with the same buffer and resuspended in 1 packed cell volume of NE2 buffer (NE1 buffer containing 350 mM KCl). The extracted nuclei were centrifuged at 180,000 × g for 90 min, and the supernatant (nuclear extract) was collected, dialyzed against buffer D (8), aliquoted, and stored at −70 °C. All of the steps were carried out at 4 °C, and the buffers were supplemented with a mixture of the following protease inhibitors: 0.5 mM phenylmethylsulfonyl fluoride (added from anhydrous stock immediately before use), 1 μg/ml each of leupeptin, chymostatin, and pepstatin, 2 μg/ml antipain, 10 μg/ml benzamidine, and 1 unit/ml aprotinin. Protein concentrations of the extracts, which ranged from 0.5 to 2 mg/ml, were estimated by the method of Bradford (10).

Oligonucleotide Probes for Electrophoretic Mobility Shift Assays—Single-stranded oligonucleotides were synthesized by Integrated DNA Technologies, and complementary strands were annealed before use. Each double-stranded oligonucleotide contained a recessed 3′-end which was filled in with [α-32P]dATP and the remaining dNTPs using the Klenow enzyme. The 32P-labeled probes were separated from the unincorporated nucleotides by chromatography on Sephadex G-25 (fine) packed in a 9-inch disposable Pasteur pipette and equilibrated with 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA. The DNA sequence of the oligonucleotides is shown in Table I.

Electrophoretic Mobility Shift Assays—EMSAs were performed essentially as previously described (11). Briefly, 5 μg of each nuclear extract was incubated with 20,000 cpm of 32P-labeled, double-stranded probe (5–10 fmol) and 1 μg poly(dI-dC) in a 20-μl reaction mixture containing 20 mM Hepes-NaOH, pH 7.9, 50 mM KCl, 1 mM MgCl2, 0.5 mM dithiothreitol, and 10% glycerol, and the mixture was incubated at room temperature for 20 min. For competition experiments, a 50–500-fold molar excess of unlabeled, double-stranded probe was incubated with the nuclear extract for 20 min, prior to the addition of the labeled probe. To identify specific transcription factors in a protein-DNA complex, 2 μl (1 mg/ml IgG) of antibodies against a known transcription factor were included in the binding reaction, and the mixture was incubated at room temperature for 60 min. The samples were subjected to electrophoresis on a 5% nondenaturing, polyacrylamide gel in 40 mM Tris acetate, pH 8.0, 1 mM EDTA at 10 V/cm at room temperature. The gel was dried and exposed to x-ray film at −70 °C with intensifying screens.

Probes for DNase I Protection Assays—Restriction digests, fragment isolation and purification, and 3′- and 5′-end labeling with Klenow enzyme or T7, polynucleotide kinase, respectively, were performed using standard techniques (12, 13). The cDNA clone, MGT-P5 (4), harboring the mouse β4-GT sequence from −172 to +187, was digested with HindIII and BamHI, or EcoRV and BamHI, and the respective fragment was isolated. The former was labeled at the 3′-end and digested with MaelI, and the 299-bp single-end-labeled HindIII–MaelI fragment (containing 17 bp of the vector sequence and β4-GT region from −172 to +110) was purified on a 4% polyacrylamide gel. To generate a probe labeled on the complementary (coding) strand, the EcoRV–BamHI fragment was digested with MaelI and labeled at the 3′-end, and the 293-bp EcoRV–MaelI fragment (with 11 bp of the vector sequence and the β4-GT sequence from −172 to +110) was isolated. Two additional probes prepared from the −474/−53 CAT-EN construct (6) were labeled at the 3′-end on the noncoding strand: (i) HindIII–HindIII fragment (containing the β4-GT sequence from −295 to +53 and 13 bp of vector DNA) and (ii) EcoRI–HindIII fragment (with β4-GT sequence from −474 to +55 and 20 and 10 bp of vector sequence at each end). Finally, a 379-bp AvaiI–EcoO109I fragment containing the β4-GT sequence from −828 to −449 was isolated from the 47197/55 CAT plasmid (6) and 5′-end-labeled on the noncoding strand.

DNase I Protection Assays—The protein-DNA binding reactions were performed as described for EMSA above, except that 10,000 cpm of 32P-labeled, single-end-labeled DNA fragment was incubated with 25 μg of bovine serum albumin (BSA) or 25–50 μg of nuclear extract in the presence of 2 μg of poly(dI-dC). Following incubation at room temperature for 20 min, 1 μl of DNase I, diluted from a 5 mg/ml stock solution in 10 mM Hepes–NaOH, pH 7.6; and 25 μg CaCl2, was added to the binding mixture, and digestion was allowed to proceed for 2 min at room temperature. Dilutions of DNase I used were 1:1500 for BSA, 1:150 for L-cell nuclear extract, and 1:20 for brain and LMG nuclear extracts. DNase digestion was stopped by the addition of 80 μl of 10 mM Tris-HCl, pH 7.9, 400 mM NaCl, 2 mM EDTA, and exposed to x-ray film at 45 °C for 20 min, 20 μg of Tris–HCl, pH 8.0, 20 mM EDTA, 250 μg NaCl, 0.5% SDS, 10 μg of sonicated salmon sperm DNA, and 10 μg of proteinase K. The samples were incubated at 45 °C for 60 min, extracted once with phenol/chloroform (1:1), and precipitated with ethanol. The pellets were resuspended in 80% formamide dye and electrophoresed on an 8% polyacrylamide, 8% urea sequencing gel. An aliquot of the same end-labeled DNA fragment was also subjected to the A–G chemical sequencing reaction (14) and electrophoresed on the same gel to determine the position and the sequence of the protected regions. The gel was dried and exposed to x-ray film at −70 °C with an intensifying screen.

RESULTS

We have previously shown that the cellular requirement for β4-GT enzymatic activity correlates with the transcriptional start site used (6). In the majority of mouse somatic tissues, including the mammary gland from virgin mice,2 and established cell lines derived from somatic tissues (e.g. L-cells), the 4.1-kb start site is predominantly used (the ratio of the 4.1- to the 3.9-kb transcript is ~5:1). However, in brain tissue, the N18TG2 neuroblastoma cell line, and spermatogonia, the steady state levels of β4-GT mRNA are ~10-fold lower relative to most somatic tissues and L-cells, and the 4.1-kb start site is exclusively used. Additionally, in the mid- to late pregnant and

2 B. Rajput and N. L. Shaper, unpublished data.
lactating mammary gland, the steady state β4-GT mRNA levels are 10-fold higher compared to most somatic tissues and L-cells, and the 3.9-kb start site is preferentially used (the ratio of the 4.1- to the 3.9-kb transcript is 1:10). This differential utilization of the two start sites suggested that housekeeping and mammary gland-specific transcription factors, binding to different promoter elements, regulated the use of the 4.1- and the 3.9-kb start sites, respectively. Therefore, to experimentally verify this prediction, the DNA sequence flanking the two start sites was analyzed for protein binding by DNase I footprinting and EMSAs using nuclear extracts prepared from L-cells, brain tissue, and LMG, which represent the three patterns of β4-GT mRNA expression described above.

Identification of Nuclear Factor Binding Sites in the Region Adjacent to the 3.9-kb Transcriptional Start Site (−2172 to −110): Evidence for Tissue-specific Binding—Promoter deletion analysis using β4-GT-CAT hybrid constructs transfected into L-cells showed that the DNA fragment just upstream of the 3.9-kb start site (−172 to −13) had promoter activity. However, inclusion of additional sequence from −13 to +55 in this construct reduced this activity about 90-fold, suggesting the presence of a negative regulatory element within this 68-bp region. An examination of the sequence from −172 to +55 revealed potential binding sites for positive ubiquitous and mammary gland-specific transcription factors, as well as a putative negative element (6).

To determine whether these, or other, sequence elements do in fact bind nuclear factors, and if this binding is tissue-specific, a single end-labeled DNA fragment containing the β4-GT sequence from −172 to +110 was subjected to DNase I footprinting analysis using nuclear extracts from mouse L-cells, brain tissue, and LMG. Five protected regions, designated FP-1 to FP-5, were seen on the noncoding strand (Fig. 1A), and four protected regions corresponding to FP-1 to FP-4, were observed on the coding strand (Fig. 1B). The sequence of each protected region was subsequently compared against the entries in the transcription factor database (15). The combined results of these analyses are summarized in Fig. 2.

FP-1, a rather weak footprint seen with all three extracts, is located between −36 to +60. It contains a GC-rich element (5′-GGGCGCG-3′) which is similar to a sequence motif (5′-GGGCGGC-3′) found just upstream (+24 to +30) of FP-1. Although this upstream region was not protected, a hypersensitive site indicative of a protein-DNA interaction, was seen at position +29 (Fig. 1B). Footprints FP-2 and FP-4 were also obtained with all three extracts but were more clearly observed on the coding strand (Fig. 1B) compared to the noncoding strand (Fig. 1A), where the interactions were primarily char-
characterized by the presence of hypersensitive sites (indicated by the arrows). FP-2 (−34 to +2) contains an inverted GT box (5′-CCACCC-3′) and FP-4 (−119 to −87) an inverted GA box (5′-GCCCTCCC-3′). In contrast to footprints FP-1, FP-2, and FP-4, footprint FP-3 was most prominent with the LMG extract and footprint FP-5 was clearly LMG-specific (Fig. 1A, lane 4). FP-3 (−70 to −42) is a complex region that contains multiple overlapping protein binding motifs: a CTF/NF1 half-site (5′-TGGC-3′), a GC-rich element (5′-GGCCGGG-3′) identical to that found at +24 to +30, an AP2 site (5′-GCCTGCCGG-3′), and an Sp1 site (5′-GGGCGGG-3′). The only motif noted within FP-5 (−162 to −140) is a perfect AP2 site (5′-GCGCCAGGC-3′). Because FP-5 extended to the extreme 5′-end of the DNA probe, an overlapping fragment (−295 to +55) was used to more precisely map its 5′-boundary (Fig. 1C). This analysis revealed an additional LMG-specific, DNase I-protected region (FP-6, −185 to −165) that also contains an AP2 site (5′-TCCGGCGG-3′).

The above data obtained from the DNase I footprinting analysis corroborates our previous studies (6) and shows that the region adjacent to the 3.9-kb start site is recognized by mammary gland-specific as well as ubiquitous factors. In order to characterize the nuclear proteins interacting with the protected sites, double-stranded oligonucleotides corresponding to the footprinted regions were analyzed by EMSA using nuclear extracts from mouse L-cells, brain tissue, and LMG.

Characterization of the Nuclear Protein Binding to the FP-1 Site
Identification of a Putative Negative Regulatory Factor—Based on our previous data, a negative regulatory element involved in repressing transcription from the 3.9-kb start site is predicted to reside between −13 and +55 (6). A potential candidate for the negative regulatory factor is the protein(s) that interacts at the FP-1 site (+36 to +60) and the hypersensitive site at +29. To characterize this factor, oligo 1 (+20 to +59, Table I), containing both GC-rich elements, was analyzed by EMSA. An equal amount of total protein (5 μg) was used per reaction in order to compare the relative binding activity of this factor in each of the three nuclear extracts.

A protein-DNA complex (Fig. 3A, indicated by the solid arrow) of similar mobility was seen with all three extracts (lanes 2–4), with the brain extract giving the most intense band. It should be noted that, even though footprint FP-1 was rather weak, the protein-DNA complex as visualized by EMSA was quite strong. This is due to the fact that EMSA is a more sensitive DNA-protein binding assay than the DNase I protection assay (13). The formation of the complex was extract-dependent, as it was not seen in the control reaction performed in the absence of the nuclear extract (Fig. 3A, lane 1). The specificity of binding was demonstrated by competition assays in which unlabeled oligo 1 was preincubated with L-cell nuclear extract followed by the addition of labeled oligo 1. As seen in Fig. 3B, preincubation with a 100-fold molar excess of unlabeled oligo 1 greatly diminished complex formation (lane 2) and preincubation with a 500-fold molar excess abolished complex formation (lane 3).

Since the GC-rich elements (GGGGCGGC and GGCCCGG) contained within oligo 1 are similar to the Sp1 recognition sequence (GGGCGG), an oligonucleotide containing the con-
sensus Sp1 site (Table I) was also tested in competition assays with labeled oligo 1 as the probe. Oligo Sp1 was not an effective competitor, even at a 500-fold molar excess (Fig. 3B, lanes 4 and 5), indicating that the protein recognizing oligo 1 was not Sp1 or an Sp1 family member. This conclusion was verified by showing that polyclonal antibodies against human Sp1, which cross-react with the mouse protein, neither inhibited nor caused a supershift (retard the mobility) of the specific protein-DNA complex (Fig. 3C, lane 3). The anti-Sp1 antibodies were shown to supershift authentic Sp1 in a control experiment (data not shown, also see Fig. 4). Analogous experiments performed using oligo 1 and nuclear extracts from brain and LMG gave results similar to those described for L-cells (data not shown).

We had previously identified a sequence motif between 215 and 269 with a weak similarity to a negative element described by Kageyama and Pastan (16). However, EMSA using an oligonucleotide containing this sequence motif failed to demonstrate any protein binding (data not shown). Therefore, the protein binding to oligo 1, which we term GC binding factor (GCBF), is the candidate for the negative regulatory factor. Both GC-rich elements in oligo 1 appear to be important for high affinity binding, as two separate oligonucleotides containing either of the GC-rich elements showed very weak binding (data not shown). GCBF is predicted to have a broad tissue distribution as the 3.9-kb transcript is down-regulated in most somatic tissues. Consistent with this prediction, a preliminary survey has established that this factor is also present in liver, lung, and kidney (data not shown).

Proteins Binding to the FP-2 and FP-4 Sites Are Members of...
The Sp1 Family—The protected region FP-2 contains an inverted GT box (CCCAACC), which is similar to the inverted GC box (CCCCGCC) recognized by Sp1. Recently, several novel Sp1-related factors, that also bind GC and GT boxes, have been described (17–19). Therefore, the strategy we used to identify the protein interacting with the FP-2 site included experiments to assess the involvement of Sp1 or a related family member. Oligo 2 which spans FP-2 (−47 to −10, Table I), and oligo Sp1 were analyzed by EMSA using nuclear extracts from L-cells, brain tissue, and LMG. Two protein-DNA complexes (I and II)3 were seen when nuclear extract from either L-cells (Fig. 4A, lane 2) or LMG (lane 4) was incubated with oligo 2. This binding activity was very low in brain and only a weak band, corresponding to complex I, was observed (lane 3). When purified Sp1 protein was incubated with oligo 2, an intense upper band that comigrated with complex I was observed (Fig. 4B, lane 1), or with nuclear extract from L-cells (L, lane 2), brain (Br, lane 3), and lactating mammary gland (LMG, lane 4) or Sp1 protein (Sp1, lane 5), and analyzed by EMSA. The position of the two protein-DNA complexes (I and II) is shown by the solid arrows (see footnote 3). The open arrow indicates the position of the free probe. B, identical to A except labeled oligo Sp1 was used. C, unlabeled oligo 2 (lanes 2 and 3) or oligo Sp1 (lanes 4 and 5) at the indicated molar excess was used as competitor for the formation of specific protein-DNA complexes between labeled oligo 2 and L-cell nuclear extract. D, L-cell nuclear extract and labeled oligo 2 were incubated without (−, lane 1) or with irrelevant serum (IS, lane 2) or anti-Sp1 antibodies (Sp1, lane 3).

3 The formation of two specific protein-DNA complexes upon incubation of oligonucleotide probes containing single Sp1 binding sites, with nuclear extracts from various tissues, has been reported (20, 21). The two complexes are attributed either to differentially phosphorylated forms of Sp1 with apparent molecular masses of 95 and 105 kDa (22), or to different Sp1 family members.

Fig. 4. Sp1 or Sp1-related nuclear factor(s) binds to the FP-2 site. A, labeled oligo 2 (−47 to −10) spanning the FP-2 site was incubated without (−, lane 1) or with nuclear extract from L-cells (L, lane 2), brain (Br, lane 3), and lactating mammary gland (LMG, lane 4) or Sp1 protein (Sp1, lane 5), and analyzed by EMSA. The position of the two protein-DNA complexes (I and II) is shown by the solid arrows (see footnote 3). The open arrow indicates the position of the free probe. B, identical to A except labeled oligo Sp1 was used. C, unlabeled oligo 2 (lanes 2 and 3) or oligo Sp1 (lanes 4 and 5) at the indicated molar excess was used as competitor for the formation of specific protein-DNA complexes between labeled oligo 2 and L-cell nuclear extract. D, L-cell nuclear extract and labeled oligo 2 were incubated without (−, lane 1) or with irrelevant serum (IS, lane 2) or anti-Sp1 antibodies (Sp1, lane 3).
bands corresponding to complexes I and II (lane 4).

To demonstrate if the formation of complexes I–IV was specific and corresponded to the four putative protein binding motifs identified in this region (Sp1, AP2, GC-rich element, and CTF/NF1 half-site; see Fig. 2), competition assays, using labeled oligo 3 and unlabeled oligo Sp1, oligo AP2, and oligo C/N containing the respective consensus binding site as competitors, were performed. The results of one such assay, using the LMG extract, showed that addition of a 250-fold molar excess of unlabeled oligo 3 inhibited complexes I–III; complex IV was partially diminished (Fig. 5B, lane 2), suggesting that the formation of all four complexes is specific.

Complex I formation was abolished in the presence of unlabeled oligo 3 (Fig. 5B, lane 3) and anti-Sp1 antibodies (Fig. 5C, lane 2), confirming that Sp1 or an Sp1-like protein binds to the perfect Sp1 motif (GGCCGGG) at the FP-3 site. Similar results were obtained with the nuclear extract from L-cells (data not shown). It was surprising that Sp1 binding to this site was weak since both L-cells and the LMG contain relatively high levels of Sp1 (see Fig. 4B). This may be due to competition between multiple factors binding to overlapping sequence elements at the FP-3 site.

Complex II formation was abolished in the presence of a 250-fold molar excess of unlabeled oligo AP2 (Fig. 5B, lane 4) and anti-AP2 antibodies (Fig. 5C, lane 3), confirming that nuclear factor AP2 binds to the AP2 motif (GCTTCGCGG) at the FP-3 site.

A number of observations led to the conclusion that complex III, which was seen with all three nuclear extracts, may result from the binding of GCBF or a GCBF-like factor to the GC-rich sequence (GGCCGGG): (i) The GC-rich motif at the FP-3 site is identical to the GCBF binding site (+24 to +30) upstream of FP-1. (ii) The mobility of complex III is similar to that of the complex seen with oligo 1. (iii) Complex III formation is highest in brain and GCBF levels are also highest in this tissue. (iv) The formation of complex III is not inhibited by an excess of unlabeled oligo Sp1 (Fig. 5B, lane 3), nor by anti-Sp1 antibodies (Fig. 5C, lane 2), as noted for GCBF.

Complex IV formation, which is unique to the LMG, was abolished in the presence of a 250-fold molar excess of unlabeled oligo C/N (Fig. 5B, lane 5) and greatly diminished by anti-CTF/NF1 antibodies (Fig. 5C, lane 4), indicating that CTF/NF1 or a CTF/NF1-like factor binds to the CTF/NF1 half-site (TGCC). This nuclear factor has a greater affinity for the palindromic consensus CTF/NF1 site than the half-site, as oligo C/N competed more effectively for the formation of complex IV than oligo 3 (compare complex IV in lanes 5 and 2, Fig. 5B). It is noteworthy that inhibition of complex IV (either by oligo C/N or anti-CTF/NF1 antibodies) enhanced the formation of complex II, suggesting that there is competition between binding of CTF/NF1 and AP2 to their respective site. However, CTF/NF1 appears to preferentially bind to the FP-3 site, as complex IV is the major band seen with the LMG nuclear extract.

Analogous experiments using L-cell nuclear extract showed that complex II formation was not competed by unlabeled oligo AP2 but it was competed by unlabeled oligo C/N (data not shown). These results are consistent with the fact that AP2 is a tissue-restricted transcription factor that is present in LMG but not in L-cells or brain (23) (see Fig. 6A). Therefore, in the LMG, complex II formation is due to AP2, whereas in L-cells it is due to CTF/NF1. These results suggest that two different forms of CTF/NF1, with varying mobilities, exist in the LMG and L-cells. To test this directly, labeled oligo C/N was incubated with each nuclear extract (Fig. 5D). An intense, heterogeneous band with mobility comparable to complex II was observed with all three extracts (lanes 1–3), consistent with the widespread distribution of CTF/NF1 (24). A higher mobility band, similar to complex IV, was seen only with the LMG extract (lane 3), suggesting the presence of a mammary gland-specific form of CTF/NF1 which we term, MG-C/N. Although CTF/NF1 is abundant in all three tissues, its binding to oligo 3 is reduced in L-cells and absent in brain. This may be attributed to the fact that the ubiquitous form of CTF/NF1 has a greater affinity for the full palindromic binding motif (TGCC/ C/A)(N)2/GCCA) than the half-site (TGCC) present in oligo 3 (24, 25). Alternatively, competition may occur between multiple factors binding to overlapping sites in this region. As noted earlier, MG-C/N also has a greater affinity for the full site than the half-site (Fig. 5B), but it appears to bind to the half-site (in...
the context of the FP-3 site) better than the ubiquitous form of CTF/NF1.

In summary, the results from the EMSA are in agreement with the DNase I footprinting analysis and confirm that the major interaction at the FP-3 site is mammary gland-specific and is the result of binding a mammary gland-specific form of CTF/NF1 (MG-C/N) to the CTF/NF1 half-site.

The Tissue-restricted Transcription Factor AP2 Binds to Both the FP-5 and FP-6 Sites—Footprints FP-5 and FP-6 were only seen with the nuclear extract from the LMG and therefore it was expected that the factor binding to these sites would be restricted in its tissue distribution. Consistent with this, a recognition motif for the tissue-restricted transactivator, AP2, was identified in each of the footprinted regions. To determine if AP2 binds to the FP-5 site, oligo 5 (-162 to -127, Table I) was analyzed by EMSA. As seen in Fig. 6A, only the extract from the LMG exhibited a prominent protein-DNA complex (lane 4). Complex formation was abolished by the addition of a 50- or a 250-fold molar excess of unlabeled oligo 5 (Fig. 6B, lanes 2 and 3). Since only partial inhibition was observed using the same molar excess of oligo AP2 (lanes 4 and 5), the binding protein appears to have a greater affinity for the AP2 motif within the FP-5 site than the consensus AP2 binding sequence. However, this nuclear protein was unequivocally shown to be AP2 (or related to AP2) as incubation with anti-AP2 antibodies caused a supershift of the specific complex (Fig. 6C, lane 3).

Analogous experiments using an oligonucleotide spanning the FP-6 site gave similar results, although the intensity of the complex was reduced (data not shown). Since the regions represented by FP-5 and FP-6 were equally well protected in the DNase I protection assay (Fig. 1C), these data suggest cooperative binding to the two AP2 sites. For example, binding of AP2 at the FP-5 site may stabilize binding at the FP-6 site. More importantly, these results show that the mammary gland-specific interactions at the FP-5 and the FP-6 sites are due to the binding of AP2 or an AP2-like protein that is absent in L-cells and brain.

The Ubiquitous Transcription Factor Sp1 Binds to Multiple GC Boxes Upstream of the 4.1-kb Start Site—We have previously shown that the 4.1-kb β1,4-GT transcript is ubiquitously expressed at similar levels in all somatic tissues, except brain tissue, where the levels are 10-fold lower. The sequence upstream of the 4.1-kb start site (at -190) shares features in common with other housekeeping genes, including the lack of a consensus TATA box, multiple start sites, high GC content, and multiple putative binding sites for the transcription factor, Sp1 (26). Transfection studies in L-cells and Drosophila SL2 cells show that a 287-bp (-474 to -187) DNA fragment immediately upstream of the 4.1-kb start site has promoter activity and that some or all of the Sp1 binding sites within this region are functional (6).

Six protected regions (FP-7 to FP-12), demarcated by hypersensitive sites, were seen when the DNA fragment (-474 to +55) was analyzed by the DNase I footprinting assay (Fig. 7; FP-7 is better visualized in the bottom half of Fig. 1C). FP-7, FP-8, and FP-9 were observed with nuclear extracts from all three tissues and each footprint contains an inverted GC or GT box (Fig. 2). FP-10, FP-11, and FP-12 were seen with the L-cells and LMG extracts but not with the brain extract. FP-11 was more pronounced with the LMG extract compared to the L-cells extract on the noncoding strand, however, on the coding strand both extracts showed equivalent protection (data not shown). FP-10 and FP-11 contain an imperfect, inverted GC box and an inverted GA box, respectively (Fig. 2). The protection of the FP-12 region was qualitatively different between the L-cells and LMG extracts; the L-cell extract showed better protection at the top (3')-half of FP-12, whereas the LMG extract protected the bottom (5')-half better. The reason for this became apparent when an inspection of this protected sequence revealed overlapping binding sites for AP2 (absent in L-cells) and Sp1.

**Fig. 6. Identification of AP2 as the nuclear factor that binds to the FP-5 site.** A, oligo 5 (-162 to -127), spanning the FP-5 site, was labeled and incubated without (−, lane 1) or with nuclear extract from L-cells (L, lane 2), brain (Br, lane 3), and lactating mammary gland (LMG, lane 4) and analyzed by EMSA. The solid arrow designates the position of the protein-DNA complex obtained with the LMG extract. The open arrow indicates the position of the free probe. B, a 50- or a 250-fold molar excess of both oligo 5 (lanes 2 and 3) or oligo AP2 (lanes 4 and 5) was used as a competitor for the formation of the specific protein-DNA complex between labeled oligo 5 and LMG nuclear extract. C, a binding reaction with labeled oligo 5 and LMG nuclear extract was performed without (−, lane 1) or with irrelevant serum (IS, lane 2) or anti-AP2 antibodies (AP2, lane 3). The position of the supershifted band in lane 3 is shown.
Oligonucleotides corresponding to FP-7 to FP-12 were then analyzed by EMSA and protection at each site was shown to be the result of binding by Sp1 or a related family member (data not shown). As expected, the oligonucleotide corresponding to the FP-12 site also showed weak binding by AP2 with the LMG extract. These data confirm that Sp1, or a family member, interacts at multiple sites in the immediate vicinity of the 4.1-kb start site and that the region upstream of this start site may well function as a housekeeping promoter. Consistent with this conclusion is the correlation between the levels of Sp1 binding activity and the 4.1-kb mRNA in the three different tissues tested. Brain which has 10-fold lower steady state levels of the 4.1-kb mRNA compared to L-cells and LMG, also shows the lowest level of Sp1 binding activity, whereas L-cells and LMG which have comparable amounts of the 4.1-kb transcript, have similar levels of Sp1 binding activity (Figs. 4 and 6). The relative Sp1 binding activity most likely reflects the amount of Sp1 protein present in each tissue, as the study by Saffer et al. (27) shows that Sp1 protein levels are very low in the brain tissue.

Analysis of the Region between −474 to −805 for Nuclear Factor Binding—Even though the above analyses show that regulatory elements necessary for expression of the 3.9- and 4.1-kb β4-GT transcripts reside between −474 and +55, we examined additional upstream sequence for nuclear factor binding, since DNA sequence analysis identified putative AP2 binding sites upstream of −2474. When the sequence from −828 to −449 was subjected to footprinting analysis, three footprinting results were obtained (Figs. 8, FP-13, FP-14, and FP-15) were observed (Fig. 8). FP-13 was seen with nuclear extracts from all three tissues and contained a full palindromic CTF/NF1 recognition sequence (TGGCGGAGCGCCA; Fig. 2). As expected, the factor binding to this site was identified as CTF/NF1 by EMSA (data not shown). It should be noted that both the ubiquitous and the mammary gland-specific forms of CTF/NF1 were found to bind to the FP-13 site with the LMG extract, as seen earlier with oligo C/N (Fig. 5D, lane 3). Footprints FP-14 and FP-15 were specific to the mammary gland and were shown to bind AP2. Therefore, CTF/NF1 and AP2 binding sites, found in the proximal promoter region and implicated in high level expression of the 3.9-kb transcript, are present further upstream and may function in an enhancer-like capacity.

**DISCUSSION**

**β4-GT: One Gene, Three Transcriptional Start Sites, and Three Promoters**—The organization of the 5′-end of the murine β4-GT gene is unusual in that three transcriptional start sites are embedded within an −800-bp contiguous piece of DNA (Figs. 2 and 9). The most distal start site (relative to the translation initiation codon) is male germ cell-specific (Gc) and...
it is used exclusively in late pachytene spermatocytes and round spermatids (28). The “middle” start site (4.1 kb) is predominantly used in all somatic cell types examined (6) as well as spermatogonia (29). The proximal start site (3.9 kb) is preferentially used in the mid- to late pregnant and lactating mammary gland (6). The differential use of the three start sites suggests the presence of both tissue-specific and housekeeping promoters, each regulating the expression of the respective mRNA species.

We have recently shown that a 798-bp genomic fragment spanning the male germ cell start site is sufficient to target expression of the reporter gene, β-galactosidase, exclusively to male germ cells (28). The “middle” start site (4.1 kb) is predominantly used in all somatic cell types examined (6) as well as spermatogonia (29). The proximal start site (3.9 kb) is preferentially used in the mid- to late pregnant and lactating mammary gland (6). The differential use of the three start sites suggests the presence of both tissue-specific and housekeeping promoters, each regulating the expression of the respective mRNA species.

We have recently shown that a 798-bp genomic fragment spanning the male germ cell start site is sufficient to target expression of the reporter gene, β-galactosidase, exclusively to male germ cells (28). The “middle” start site (4.1 kb) is predominantly used in all somatic cell types examined (6) as well as spermatogonia (29). The proximal start site (3.9 kb) is preferentially used in the mid- to late pregnant and lactating mammary gland (6). The differential use of the three start sites suggests the presence of both tissue-specific and housekeeping promoters, each regulating the expression of the respective mRNA species.

We have recently shown that a 798-bp genomic fragment spanning the male germ cell start site is sufficient to target expression of the reporter gene, β-galactosidase, exclusively to male germ cells (28). The “middle” start site (4.1 kb) is predominantly used in all somatic cell types examined (6) as well as spermatogonia (29). The proximal start site (3.9 kb) is preferentially used in the mid- to late pregnant and lactating mammary gland (6). The differential use of the three start sites suggests the presence of both tissue-specific and housekeeping promoters, each regulating the expression of the respective mRNA species.

We have recently shown that a 798-bp genomic fragment spanning the male germ cell start site is sufficient to target expression of the reporter gene, β-galactosidase, exclusively to male germ cells (28). The “middle” start site (4.1 kb) is predominantly used in all somatic cell types examined (6) as well as spermatogonia (29). The proximal start site (3.9 kb) is preferentially used in the mid- to late pregnant and lactating mammary gland (6). The differential use of the three start sites suggests the presence of both tissue-specific and housekeeping promoters, each regulating the expression of the respective mRNA species.

We have recently shown that a 798-bp genomic fragment spanning the male germ cell start site is sufficient to target expression of the reporter gene, β-galactosidase, exclusively to male germ cells (28). The “middle” start site (4.1 kb) is predominantly used in all somatic cell types examined (6) as well as spermatogonia (29). The proximal start site (3.9 kb) is preferentially used in the mid- to late pregnant and lactating mammary gland (6). The differential use of the three start sites suggests the presence of both tissue-specific and housekeeping promoters, each regulating the expression of the respective mRNA species.

We have recently shown that a 798-bp genomic fragment spanning the male germ cell start site is sufficient to target expression of the reporter gene, β-galactosidase, exclusively to male germ cells (28). The “middle” start site (4.1 kb) is predominantly used in all somatic cell types examined (6) as well as spermatogonia (29). The proximal start site (3.9 kb) is preferentially used in the mid- to late pregnant and lactating mammary gland (6). The differential use of the three start sites suggests the presence of both tissue-specific and housekeeping promoters, each regulating the expression of the respective mRNA species.
suggest that 3.9-kb steady state mRNA levels are determined by the balance between this negative factor and the positive factors, MG-C/N, AP2, and Sp1. For example, brain tissue which lacks the 3.9-kb mRNA, has high levels of GCBF and low levels of Sp1. L-cells make low levels of this transcript and contain moderate amounts of both GCBF and Sp1. The LMG which synthesizes high levels of the 3.9-kb mRNA, contains moderate amounts of GCBF and Sp1, but high levels of the mammary gland-enriched factors, MG-C/N and AP2. These findings suggest that the binding of positive factors to sites adjacent to the GCBF binding site displaces GCBF, thus allowing transcription from the 3.9-kb start site.

Although GCBF or a GCBF-like protein appears to bind to the GC-rich motif at the FP-3 site (Figs. 2 and 5), this binding does not seem to have a negative effect, as a reduction in CAT activity was not observed when the FP-3 region was included in one of the β4-GT/CAT constructs (172 to 13) previously analyzed (6). Therefore, the sequence context of the GC-rich element may determine whether GCBF acts as a negative or positive regulator. Examples of transcription factors exhibiting dual function are YY1 (40), Egr-1 (41), and WT-1 (42).

Role of CTF/NF1—CTF/NF1 constitutes a family of proteins which bind to the palindromic sequence, TGG(C/A)(N)GCCAA, or with lower affinity to the half-site, TGG(C/A) (24, 25, 43, 44). Although generally considered to be a ubiquitous factor, CTF/NF1 has also been implicated in the mammary gland-specific expression of MMTV and several milk protein genes including α-lactalbumin (32, 49–52). Moreover, a mammary gland-specific form of CTF/NF1, similar to the one we observed (MG-C/N), has been described in rat (51) and bovine (53) LMG. However, it is not known if this size variant represents a unique gene product, a spliced variant, or a partially degraded form.

Our data show that both the ubiquitous form of CTF/NF1 and MG-C/N bind with higher affinity to the palindromic sequence than to the half-site, but the half-site in FP-3 is notable in that it binds MG-C/N with higher affinity than the ubiquitous form. This may result from cooperative interaction with AP2, which also binds at the FP-3 site. It has been proposed that CTF/NF1 binding may be stabilized by interactions with factors bound to adjacent sites (25).

The FP-13 site, containing the palindromic CTF/NF1 sequence, binds both forms and shows equivalent protection using nuclear extract from L-cells, brain, or LMG. While this might suggest that this site is involved in 4.1-kb mRNA expression, we think it unlikely as β4-GT/CAT constructs, which contained or lacked this sequence, exhibited similar CAT activities (6). However, binding at this site may be important for 3.9-kb mRNA expression in the LMG, as it is juxtaposed between two AP2 sites (Figs. 2 and 9).

Role of AP2—AP2 was first identified in Hela cells as a transcription factor that binds to GC-rich motifs in the enhancer regions of SV40 and the human metallothionein genes (23). It was shown to be tissue-restricted (23, 54) and has been implicated in the control of gene expression in the neural crest (54) and epidermal cell (55, 56) lineages. Recently, AP2 was found to be involved in MMTV expression in the mammary gland (57). The 5′ enhancer of the MMTV long terminal repeat contains four elements, AP2, CTF/NF1, "F12," and "mp4." While mutation of any one motif decreases enhancer activity, the most significant reduction results from mutation of the AP2 site.

AP2 also appears to be involved in 3.9-kb mRNA expression as it is found only in the LMG and not in L-cells or brain. The dose proximity of the three AP2 sites to the CTF/NF1 half-site just upstream of the 3.9-kb start site suggests that these factors may function cooperatively, as proposed for MMTV, to increase transcription from the 3.9-kb start site. The three additional AP2 sites and the palindromic CTF/NF1 site, located upstream of the 4.1-kb start site, may function in an enhancer-like capacity. A redundancy of cis-acting elements involved in tissue-specific expression has been noted in other genes. For example, multiple binding sites for factors (CTF/NF1 and mammary gland factor (MGF)) critical for mammary gland-specific expression of the whey acidic protein gene are found in the promoter proximal and distal regions, and it has been suggested that interaction at both sites is necessary for high level expression (52).

Transcription Factors Involved in the Expression of Genes in the Mammary Gland—The primary function of the mammary gland is to synthesize and secrete a group of milk specific proteins which include caseins, whey acidic protein, β-lactoglobulin, and α-lactalbumin, a variety of lipids and carbohydrates (e.g. lactose) required by the newborn. While the milk proteins are abundantly expressed exclusively in the mammary gland, different members of this group are expressed asynchronously, beginning in mid-pregnancy and continuing throughout lactation.

As discussed, the 3.9-kb β4-GT transcript is predominantly expressed in the mid-to late pregnant and lactating mammary gland, therefore, it was of interest to compare the regulatory elements involved in its expression with those of the milk protein genes. CTF/NF1 has been implicated in the expression of α-lactalbumin (49) and β-lactoglobulin (50) and has been shown to be functionally important for the expression of whey acidic protein (52). MMTV, which is expressed primarily in the late pregnant and lactating mammary gland, also contains a functional CTF/NF1 site (32) in addition to a functional AP2 site (57). Binding sites for the mammary gland-enriched factor, MGF, are found in all milk protein genes (reviewed in Groenen and van der Poel (58)) and have been shown to be functionally involved in the expression of β-casein (59), whey acidic protein (52), and β-lactoglobulin (60). However, this site is not present in the β4-GT gene sequence analyzed.

Recruitment of β4-GT for Lactose Biosynthesis in Mammals—The evolutionary route, resulting in the formation of lactose synthetase (a heterodimer between α-lactalbumin and β4-GT) in mammals, is both remarkable and unique. α-Lactalbumin and lysozyme are homologous proteins that have evolved from a common ancestral gene. It has been estimated that the α-lactalbumin gene line diverged from the lysozyme gene line about 400 million years ago, prior to the divergence of tetrapods and fishes (61) to emerge in mammals as a milk protein gene. In contrast, the β4-GT gene has been recruited from the non-mammalian vertebrate pool of constitutively expressed genes. This is evidenced by the fact that β4-GT from non-mammalian vertebrate species, such as chicken, can functionally interact with α-lactalbumin in vitro, indicating that the α-lactalbumin binding domain in β4-GT predates the rise of mammals.

With the recruitment of β4-GT for lactose biosynthesis, the problem arose as to how to increase the levels of this enzyme in the LMG, while maintaining the relatively low levels of constitutively expressed enzyme in all somatic tissues. Based on our analysis of the structure and regulation of the murine β4-GT gene, we would argue that this was achieved by the generation

1. J. A. Meurer, N. L. Shaper, D. H. Joziasse, R. L. Schnaar, and J. H. Shaper, manuscript submitted for publication.
of the 3.9-kb start site and its accompanying tissue-restricted regulatory elements. It is interesting to note in this regard that both AP2 and GCBF, two of the transcription factors implicated in the regulation of transcription from the 3.9-kb start site, bind to GC-rich sequence motifs, which could have been generated by mutations in the GC-rich regions flanking the 4.1-kb start site.

In summary, the results presented in this study support the conclusion that the presence of the 3.9-kb start site in the mammalian β4-GT gene is a direct consequence of the recruitment of β4-GT for the mammary gland-specific biosynthesis of the uniquely mammalian disaccharide, lactose.

Acknowledgements—We thank Michael Collector for assistance in obtaining mouse organs, Dr. Naoko Tanese for the kind gift of anti-CTF/NF1 antibodies, Dr. Yoshikuni Nagamine for critical reading of this manuscript, and Ann Larocca for editorial assistance.

REFERENCES

1. Hill, R. L., Brew, K., Vanaman, T. C., Trayer, I. P., and Mattock, P. (1968) Brookhaven Symp. Biol. 21, 139–154
2. Brodbeck, U., Denton, W. L., Tanahashi, N., and Ebner, K. E. (1967) J. Biol. Chem. 242, 1391–1397
3. Turkington, R. W., Brew, K., Vanaman, T. C., and Hill, R. L. (1968) J. Biol. Chem. 243, 3382–3387
4. Russo, R. N., Shaper, N. L., and Shaper, J. H. (1990) J. Biol. Chem. 265, 5519–5525
5. Roy, R. J., Gosselin, P., and Guerin, S. L. (1991) Genes & Dev. 5, 1391–1397
6. Vayou, U., and Godin, C. (1994) Mol. Cell. Biol. 14, 3494–3503
7. Li, S., and Rosen, J. M. (1994) Mol. Cell. Biol. 14, 7963–7973
8. Shaper, N. L., Wright, W. W., and Shaper, J. H. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 791–795
9. Shaper, N. L., Harduin-Lepers, A., and Shaper, J. H. (1991) J. Biol. Chem. 266, 25165–25171
10. Leask, A., Byrne, C., and Fuchs, E. (1991) J. Biol. Chem. 266, 8963–8967
11. Grobler, J. A., Ramakrishna, R., Pervaiz, S., and Brew, K. (1994) Arch. Biochem. Biophys. 313, 360–366
Transcriptional Regulation of Murine 1,4-Galactosyltransferase in Somatic Cells: 
ANALYSIS OF A GENE THAT SERVES BOTH A HOUSEKEEPING AND A 
MAMMARY GLAND-SPECIFIC FUNCTION

Bhanu Rajput, Nancy L. Shaper and Joel H. Shaper

J. Biol. Chem. 1996, 271:5131-5142.
doi: 10.1074/jbc.271.9.5131

Access the most updated version of this article at http://www.jbc.org/content/271/9/5131

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 59 references, 30 of which can be accessed free at
http://www.jbc.org/content/271/9/5131.full.html#ref-list-1