[Frontiers in Bioscience 11, 2483-2495, September 1, 2006]

Identification of an octamer-binding site controlling the activity of the small breast epithelial mucin gene promoter

Florent Hubé [1,2], Shilpa Chooniedass-Kothari [2], Mohammad K. Hamedani [2], Richard J. Miksicek [3], Etienne Leygue [2] and Yvonne Myal [1]

[1] Department of Pathology and [2] Department of Biochemistry and Medical Genetics, University of Manitoba, 770 Bannatyne Avenue, Winnipeg R3E 0W3, Manitoba, Canada. [3] Department of Physiology, Michigan State University, East Lansing, MI 48824, USA.

TABLE OF CONTENTS

1. ABSTRACT

The human small breast epithelial mucin (SBEM) gene has been identified as being preferentially expressed in mammary epithelial cells and over-expressed in breast tumors. In this report, we have characterized the promoter of SBEM gene in order to identify sequences responsible for this strong mammary expression. A series of SBEM promoter/luciferase constructs were transiently transfected into both breast (MCF-7, BT-20) and non-breast (HeLa and HepG2) cell lines. In addition to the minimal promoter and to a repressor region, we have identified an 87-bp sequence (-357/-270) driving a strong breast-specific expression. Site-directed mutagenesis of a putative octamer-binding transcription factor binding site located within this latter region led to a strong decrease of the transcriptional activity of the SBEM promoter. Furthermore, transient over-expression of Oct1 and Oct2 not only increased SBEM promoter reporter activity, but also enhanced endogenous SBEM mRNA level. Overall, the data suggest that octamer-binding transcription factors participate in the strong expression of SBEM gene in breast tissues. Clarifying the SBEM gene regulation will help to dissect mechanisms underlying transcription of normal breast and breast cancer-associated genes.

2. INTRODUCTION

Breast cancer, with an estimated 1 million new cases every year, remains the most common type of cancer in women (23% of all cancers) in developed countries (1). Tumor progression in breast cancer is a multistep process beginning with a benign stage, progressing through hyperproliferation of the breast epithelium, and ultimately leading to invasive carcinoma. The progression is characterized by multiple changes in gene expression (2-5). It has been suggested that the identification of genes differentially expressed during the transition from a normal to a cancer cell, together with the elucidation of the mechanisms controlling their expression, will lead to the establishment of novel diagnostic and therapeutic strategies for the clinical management of this disease.

In this context, a gene called SBEM (small breast epithelial mucin, For review, see (6)), preferentially expressed in breast epithelial cells and over-expressed in breast tumor, has been identified (7-9). Dot blot analyses revealed a strong SBEM gene expression in breast tissues and salivary glands, whereas all other normal tissues such as brain, ovary, uterus, prostate, or lung were negative (8,9). This apparent breast-specific expression was further

Table 1. Oligonucleotide primers used in this study

Name	Sequence (5'-3')	Position	Prim
SBEM-F	GATCTTCAGGTCACCACCATG	-18/+2	RT
SBEM-R	GGGACACACTCTACCATTCG	+251/+270	RT
GAP-F	ACCCACTCCTCCACCTTTG	+868/+886	RT
GAP-R	CTCTTGTGCTCTTGCTGGG	+1027/+1045	RT
RACE-R1	CAGAAGACTCAAGCTGATTCC	+277/+297	R
RACE-L2	TCTTTACGAGCAGTGGTAGAA	+214/+234	R
Adap-F1	CTAATACGACTCACTATAGGGC	[1]	R
Adap-F2	AAGCAGTGGTATCAACGCAGAGT	[1]	R
P947-F	GCTCCCCATTTTCCATCTCGAGACTC	-964/-939	PC
P632-F	CAATGGTTGCTAACTCGAGTAAGGTC	-646/-621	PC
P531-F	AGGGAGGCCATGACTCGAGGAATG	-545/-522	PC
P357-F	CTCCAACTCGAGATAGGAGCTGG	-364/-342	PC
P270-F	GGAGCATATTTAACTCGAGAGACTCG	-284/-259	PC
P170-F	CTCCAGTGGCTCGAGTCCCAACGTT	-181/-157	PC
P132-F	TAACCTGGATCTCGAGTGACAGCTCC	-143/-118	PC
P106-F	CCTGATTGGTGCCTCGAGCATATATATTGTC	-119/-89	PC
P51-R	CTTCAAAGCCTAAGCTTAGGCAAGGCGC	-66/-39	PC
P270nde-F	CATATTTAACATATGAGACTCCAATTGAAACCTG	[2]	M
P357nde-R	GGGGCATCTCCAACATATGATAGGAGC	[2]	M
87kpn-F	$(AC)_5$-GGTACCTCATAGGAGCTGGTAATTATGG	[2]	M
87xho-R	GAAGTTGCCTGGAGCATATTTAACCTCGAG-$(GGTT)_3$	[2]	M
P4OM	GTTGCCTGGTCTAATGTAAACATGAGAGACTCG		

Position: Nucleotide position relative to the translation start site (+1). In the mRNA sequence for RT-PCR and RACE-PCR, and in the genomic sequence for plasmid constructs. [1] Position is on the adaptator sequence.[2] Position is on the vector sequence used as template for mutant constructs.[d] Prim: Primers used for amplification of RT-PCR products (RT), for 5'-RACE (R), for plasmid contructs (PC) and for mutant constructs (M). The restriction sites are underlined and bold letters indicate mutated nucleotides compared to the SBEM genomic sequence. '-F' and '-R' are for forward and reverse primers, respectively.

established by RT-PCR as most of the mammary cancer cell lines tested expressed SBEM transcript (7 out of 8), as opposed to none out of the 10 non-breast tumor cell lines examined (8); unpublished data). SBEM mRNA expression, widely detectable throughout breast tumorigenesis and tumor progression with more than 90% of primary and metastatic breast tumors expressing this transcript, is increased in tumor compared to normal breast tissues (7,9). Altogether, data accumulated suggest that SBEM, one of the most "breast-specific" genes described to date, may become an attractive candidate with obvious potential as breast tumor biomarker and breast-specific target for future therapeutic strategies.

In the present study, we characterized the promoter region of the SBEM gene in order to identify putative regulatory sequences and transcription factors driving its strong expression in breast tissue.

2. MATERIAL AND METHODS

2.1. Cell Culture

Cell lines were obtained from the American Type Culture Collection and were cultured in DMEM with 5% fetal bovine serum (Life Technologies, Inc., Burlington, ON, Canada) supplemented with 100 units/mL penicillin, 100 µg/mL streptomycin, 2 mM glutamine (Life Technologies), 15 mM sodium bicarbonate and 2 mM glucose. Cells were grown at 37 ºC in an atmosphere of 95% air and 5% CO2. Confluent cells, which viability was determined by the Trypan blue dye exclusion test, were detached by 0.05% trypsin-0.02% EDTA (Life Technologies), seeded in 24-well plates (Corning Inc., NY, USA) and cultured in complete medium until use.

2.2. RNA isolation and reverse transcription-polymerase chain reaction (RT-PCR)

Total RNA was extracted from cells using the perfect RNA Mini Kit (Eppendorf, Hambourg, Germany) according to the manufacturer's instructions and reverse transcribed as previously described (8). Briefly, 2 µg of total RNA were reverse-transcribed for 1 hour at 37°C in 1x incubation buffer containing 300 µM of each deoxynucleotide triphosphate, 50 ng random hexamers, 12 units of RNase Out and 300 units of MMLV Reverse Transcriptase (Life Technologies). The primers used for SBEM amplification consisted of SBEM-F and SBEM-R and are described in Table 1. PCR were performed as previously (8), i.e. 1 µL of each reverse transcription mixture was amplified in a final volume of 50 µL, in the presence of 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl2, 200 µM of each deoxynucleotide triphosphate, 200 ng of each SBEM primer, and 0.5 unit of Taq DNA polymerase. Each PCR consisted of 35 cycles (15 s at 94°C, 15 s at 56°C, and 15 s at 72°C). Primers for the ubiquitously expressed GAPDH gene (GAP-F and GAP-R) are shown in Table 1. To amplify cDNA corresponding to GAPDH, 30 cycles of PCR were used (15 s at 94°C, 15 s at 52°C, and 15 s at 72°C). PCR products were then separated on a 1.5% agarose gel containing 1 mg/ml ethidium bromide and visualized under UV irradiation using the GelDoc2000/ChemiDoc System (BioRad).

2.3. Rapid amplification of the 5'-cDNA end (RACE)

A 5'-rapid amplification of cDNA end (RACE) reaction was performed using the SMART RACE cDNA amplification kit (BD Biosciences Clontech, Palo Alto, CA,

USA) according to the manufacturer's instructions. Briefly, total RNA extracted from breast cancer cells (MCF-7, T5 and MDA-MB-468) by the method described above served as starting material. 5'-RACE PCR was successively performed using two nested reverse primers (RACE-L1 and RACE-L2, see table 1) specific for the SBEM gene. The amplified fragments were subcloned into pCR4 vector (Life Technologies) and sequenced on both strands.

2.4. Identification of transcription factors binding sites within the 87-bp enhancer region (ENH)

Identification of transcription factor binding sites was done with MatInspector v2.2 program based on TransFac database (10-12) available online (MatInspector, www.genomatix.de; TransFac, www.gene-regulation.com). All parameters were set to default except for the core similarity (1.00), matrix similarity (Optimized +0.04), and matrix group (Vertebrates).

2.5. Plasmids

To characterize and identify potential regulatory regions in the SBEM promoter, various deletion constructs of the 5'-flanking region were generated by PCR using modified primers that contained restriction sites (Table 1). The XhoI and HindIII (Invitrogen) restriction enzyme-digested fragments obtained were subsequently subcloned in the promoterless, enhancerless expression vector pGL3-Basic (Promega, Madison, WI, USA) upstream of the luciferase reporter gene. Mutated SBEM/luciferase plasmids (P4OM, P1plus, P1minus and P4plus) were constructed by directed mutagenesis and PCR strategy using the QuickChange Multi Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA). P4OM corresponded to the P4 plasmid in which the octamer-binding site GGAGCATATTTAA (-284/-272) was replaced by GGTCTAATGTAAA. When appropriate, PCR-mutated fragments were subsequently digested by NdeI (P1plus and P1minus) or KpnI and XhoI (P4plus) restriction enzymes (Life Technologies) and subcloned in parental vectors. Sequences of all promoter constructs were then confirmed by dideoxynucleotide chain-termination sequencing.

The expression plasmids pOct1, pOct2 and the corresponding empty vector peOct (13) were kindly provided by Dr. Winshop Herr (Cold Spring Harbor Laboratory).

2.6. Transient DNA transfections and luciferase assays

The human SBEM/luciferase reporter constructs were transiently transfected into the mammary and non-mammary cell lines (cells cultured in a 24-well plate until 70-80% confluent) using LipofectAMINE Plus® Reagent (Life Technologies) and following the manufacturer's instructions. Briefly, 1.33 nM of appropriate plasmid was mixed with 8 µg of LipofectAMINE for each cell lines in a final volume of 200 µL of complete media without FBS. MCF-7 and HeLa were supplemented with 4 µL of Plus® Reagent, and BT-20 and HepG2 with 10 µL of Plus® Reagent. For each condition, the renilla luciferase reporter vector (100 ng) was always co-transfected to normalize for transfection efficiency. Cells were then lysed 24 h after transfection. Luciferase and renilla luciferase activities were measured using Dual-LuciferaseÒ Reporter Assay System according to the manufacturer's protocol (Promega, Madison, WI, USA) and a Lmax Luminometer (Molecular Devices, Sunnyvale, CA, USA). Resulting luciferase activities were expressed in relative light units (RLU) and adjusted to the renilla luciferase activity. A positive control containing CMV promoter sequence (pGL3-Control) and a negative control (pGL3-Basic) were used for each experiment. Results, adjusted to the positive control, are representative of at least 3 independent experiments, expressed in fold of pGL3-Basic activity +/- standard error of the mean (SEM). When appropriate, the statistical significance of differences observed between luciferase activities was determined using the student t-test.

Co-transfection experiments using pOct1, pOct2 and peOct were performed as stated above except that 0.26 nM of the expression vector was co-transfected with 1.33 nM of the P4 construct.

3. RESULTS

3.1. Identification of the transcription initiation sites

The SBEM gene, located on chromosome 12q13, spans a 3.9-kb long region consisting of 4 exons and 3 introns (Figure 1A). The corresponding transcript, shown to be approximately 600-700 bp long, encodes a secreted protein of 90 amino acids (8,9). Sequence analysis of the 5'-flanking region of the SBEM gene revealed the presence of two putative overlapping TATA boxes in -100/-95 and -98/-93 (Figure 1B) upstream from the translation start site (ATG). To precisely locate the beginning of the promoter region and to identify the exact transcription initiation site(s) of the SBEM gene, we performed a rapid amplification of 5'-cDNA ends followed by PCR amplification (RACE-PCR) on total RNA from 3 different breast cancer cell lines (MCF-7, T5 and MDA-MB-468) as described in the "Material and Methods" section. The RACE-PCR products were subsequently cloned and sequenced (data not shown), allowing us to identify experimentally two distinct transcription initiation sites in -67 and -69 upstream of the ATG (Figure 1B). RNAs initiated at both transcription initiation sites were found in all the three cell lines (data not shown).

3.2. Endogenous SBEM promoter activity in mammary and non-mammary cancer cells

To further study the activity of the SBEM promoter in a mammary and a non-mammary context, we selected 2 breast (MCF-7, BT-20) and 2 non-breast (cervix: HeLa; liver: HepG2) cancer cell lines. We first assessed the activity of the endogenous SBEM promoter in these cell lines by investigating SBEM mRNA expression using reverse transcription followed by PCR amplification, as described in the "Material and Methods" section. As shown in Figure 2, a single band of 288 bp corresponding to SBEM mRNA was readily detected in both breast tumor but not in non-breast tumor cell lines. In contrast, the expression of the housekeeping gene GAPDH was uniform

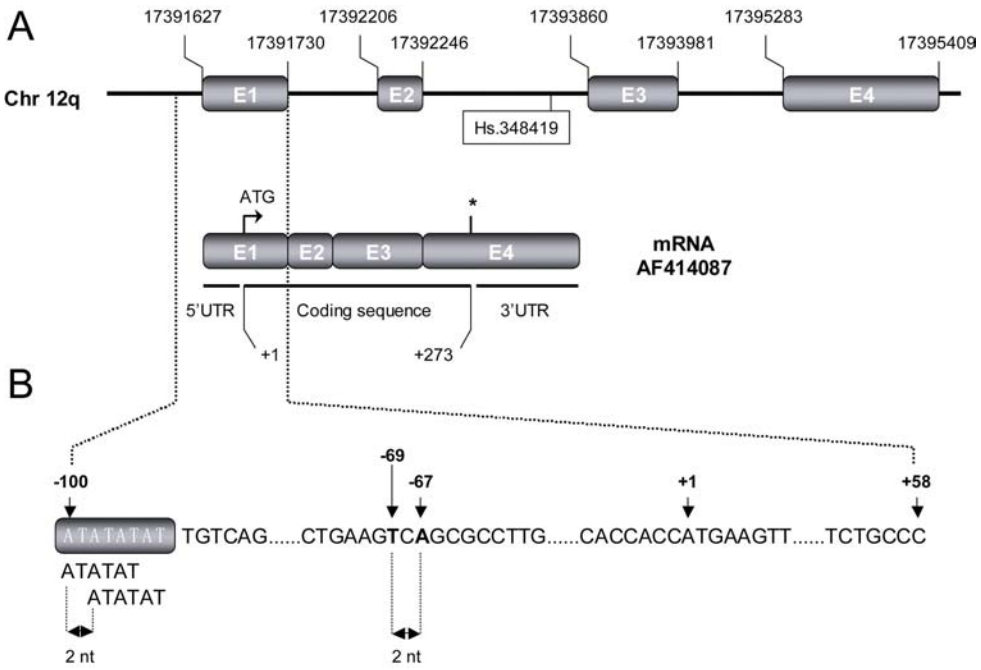

Figure 1. Schematic representation of the SBEM gene and transcription initiation sites. (A) Schematic representation of the SBEM genomic structure and the transcribed SBEM mRNA [GenBank Accession Number AF414087]. The number on the top refers to the chromosomal numbering. (B) Mapping of transcription initiation sites by rapid amplification of 5'-cDNA end analysis led to the identification of 2 transcription initiation sites, localized in -67 and -69 (bold) upstream from the translation initiation site (+1). The 2 nucleotides shift (2 nt) observed is the same as observed for the 2 putative TATA-boxes (-100/-95 and -98/-93).

in all the cell lines analyzed. Identities of all PCR products were subsequently confirmed by sequencing (data not shown) and shown to correspond to the previously published sequences [GenBank accession nos. AF414087 and NM_002046 for SBEM and GAPDH mRNA sequences, respectively].

3.3. Analysis of SBEM promoter activity in mammary and non-mammary cancer cells

A series of promoter fragments fused to the coding region of the luciferase gene was constructed through successive deletions of the ~900-bp 5'-region (-947/-51) as described in the "Material and Methods" section. The exact sequences used (P1-P8) are given in Figure 3 whereas a schematic representation is shown in Figure 4A. These constructs were then transfected into the mammary (MCF-7, BT-20) and non-mammary (HeLa and HepG2) cancer cells, and resulting luciferase activities were measured. As seen in Figure 4B, the longest promoter construct P1, containing regions from -947 to -51 upstream of the ATG, led to a strong luciferase activity in both mammary cell lines. This activity of P1 construct ranged from 28 and 38 fold over the empty vector (the baseline control) in MCF-7 and BT-20, respectively. In contrast, this construct resulted in an activity lower than 4 fold the control in the two non-breast cells, HeLa and HepG2 (Figure 4B), thus suggesting a breast-specific regulation of the SBEM promoter. Subsequent partial 5'-deletions (P2 and P3) did not change the promoter activity significantly in any of the cell lines analyzed. Interestingly, the removal of -531/-357 region (P4) led to a strong increase in the promoter activity in breast cells (124 and 95 fold for MCF-7 and BT-20, respectively), suggesting the presence of negative regulatory elements in the region encompassed between -531 and -357.

The P4 construct (-357/-51) had an 85-130 fold stronger promoter activity in mammary cells compared to non-mammary cells (Figure 4B). The deletion of the subsequent 87-bp sequence (-357/-270) in the P5 construct reduced the promoter activity dramatically in the mammary cells (6.2 and 2.7 fold less than the P4 activity for MCF-7 and BT-20, respectively), suggesting the existence of a putative breast-specific enhancer region located within this 87-bp fragment. Further deletion of the region between -132 and -106 bp completely abolished the reporter activity (less than 1.2 fold the activity driven by the promoterless vector for all cells).

In these experiments, independently of the constructs used, promoter activity was always higher in mammary cells compared to non-mammary cells. Four different SBEM promoter regions were identified: from -531 to -357, which contains an apparent repressive activity; the 87-bp region from -357 to -270, which possesses an enhancer activity (ENH region), the -132/-51 region, which represents the minimal promoter, and the region from -947 to -51 containing the basal breast-specific activity conserved in the full-length promoter.

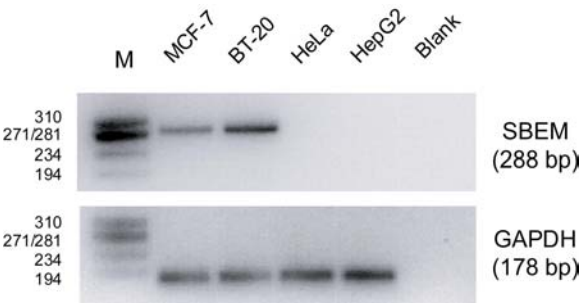

Figure 2. RT-PCR analysis of SBEM gene expression in 2 mammary and 2 non-mammary cancer cell lines. Total RNA was extracted from human breast cancer cells (MCF-7 and BT-20) and non-breast cancer cells (HeLa and HepG2), reverse-transcribed and PCR amplified with SBEM or GAPDH primers, as described in "Materials and Methods" and in Table 1. Lane M: PhiX174 RF DNA/Hae III DNA ladder. Data are representative of at least two independent RNA extractions for each cell lines. Size shown in bp. Figure was digitally inversed.

3.4. The ENH region (-357/-270) drives a strong breast-specific promoter activity

In order to address the possible role of the ENH region (87-bp region between -357 and -270 upstream from the ATG) in SBEM promoter activity, two mutants, P1plus and P1minus, were constructed. P1plus consisted of the basic P1 construct supplemented with an additional copy of the 87-bp region whereas P1minus lacked this region (Figure 5A). These constructs were then transfected in mammary and non-mammary cells, and luciferase activities were measured, as described previously. As shown in Figure 5B, addition of a second ENH region to the full-length promoter (P1plus) did not significantly modify the activity of the promoter in MCF-7 or BT-20 cells. However, deletion of this region (P1minus) decreased the luciferase activity in all the mammary cells (from 28 to 9.5 and 38 to 4.7 fold the empty vector activity in MCF-7 and BT-20, respectively). As expected, addition or deletion of the 87-bp region to the full-length promoter did not significantly modify the luciferase activity in non-mammary cells (Figure 5B).

It was surprising that while the removal of the 87-bp region strongly reduced the promoter activity, the addition of one extra copy of this region did not increase the transcriptional activity. As underlined earlier, our results suggested the existence of a repressor region (-531/-357), which might repress the enhancer activity of the 87-bp ENH region. To address this possibility, a P4plus mutant promoter was constructed, consisting of P4 supplemented with an additional copy of the ENH region (Figure 6A). The activity of P4plus mutant was then compared to the activity observed with P4 (only one copy of the 87-bp region) and P5 (without this particular 87-bp sequence). As shown in Figure 6B, the luciferase activity of the P4plus construct in all mammary cells was approximately twice the P4 activity, i.e. from 124 to 273 and 95 to 195 fold in MCF-7 and BT-20, respectively. This represented an average of a 10-fold increase compared to the P5 construct activity. Interestingly, this construct still remained free of any luciferase activity in the non-mammary cells.

3.5. Importance of octamer-binding transcription factors motif in the SBEM promoter activity

In order to further identify potential sequences involved in the strong enhancer effect of the 87-bp region on the reporter gene activity, we searched for putative transcription factor binding sites, using MatInspector software as described in the "Material and Methods". As shown in Figure 3, only 3 different motifs were identified within this region. The first two motifs overlapped in a region located between -361 and -335 and consisted of binding sites for AIRE (autoimmune regulator) and Nkx2-5 (cardiac-specific homeobox protein NK-2 homolog E). The third, located in -284/-272, corresponded to an octamer-binding transcription factor site (Oct motif).

To determine whether the octamer-binding site participated to the strong breast expression of the reporter gene, an Oct-mutated SBEM promoter construct (P4OM) was generated, by substituting the octamer-binding site GGAGCATATTTAA located in -284/-272 by GGTCTAATGTAAA. P4 and P4OM were transiently transfected in mammary and non-mammary cell lines, and luciferase activities were measured as stated above. As shown in Figure 7, mutation of the octamer-binding site in P4OM totally abolished the luciferase activity in MCF-7 and BT-20 (from 111 and 97 to less than 1 fold, respectively). Despite the fact that P4 activity was already extremely low in non-mammary cell lines, the mutation of the octamer-binding site nonetheless led to a further decrease of the luciferase activity.

3.6. Oct1 and Oct2 enhance both exogenous and endogenous SBEM promoter activities

In order to determine the potential role of the Oct transcription factors in the regulation of the SBEM promoter activity, the P4 construct was co-transfected with pOct1 and pOct2 expression vectors, or the empty vector as indicated in the "Material and Methods" section. Luciferase activity was then measured 24 hours following transfection as described earlier.

Co-transfection with the empty vector (peOct) did not alter the luciferase activity of the P4 construct (Figure 8A). However, transient over-expression of either Oct1 or Oct2 transcription factors in mammary cell lines led to an increase in the P4 luciferase activity. In MCF-7, an increase from 100 to 342 and 100 to 264 fold was observed following Oct1 and Oct2 over-expression, respectively. Similarly, the increase in BT-20 was from 93 to 191 and from 93 to 173 fold following Oct1 and Oct2 over-expression, respectively. This corresponded to an average of a 2.5 fold increase compared to the P4 activity co-transfected with the corresponding empty vector. Over-expression of Oct1 and Oct2 transcription factors did not modify the luciferase activity in the non-mammary cells.

To investigate a putative role of Oct1 and Oct2 on the regulation of the endogenous SBEM gene

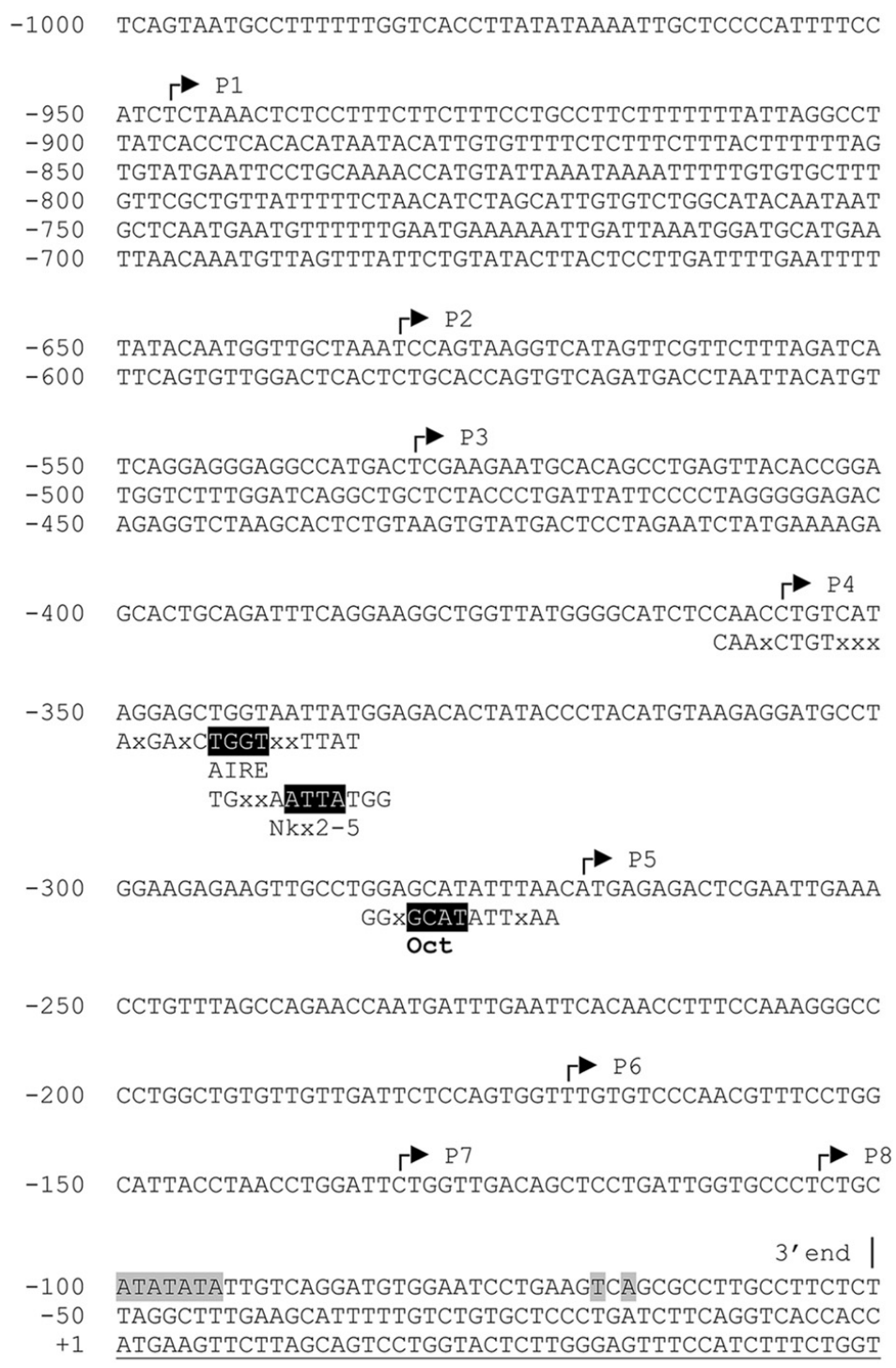

Figure 3. Nucleotide sequence of the SBEM promoter. The coding region (underlined) starts at the ATG (+1) and non-coding sequence is in uppercase non-underlined. The overlapping TATA-boxes and the two transcription initiation sites are boxed in gray. Promoter fragments used are shown (P1 to P8), starting in 5' by arrows and finishing in 3' by the vertical line (-51). Consensus binding sites for transcription factors in the 87-bp region are shown below the promoter sequence. Nucleotides indicated corresponded to the MatInspector matrix used, with the core sequences boxed in black. Nkx2-5, NK2 transcription factor related; AIRE, Autoimmune regulator; Oct, binding site for octamer-binding transcription factor.

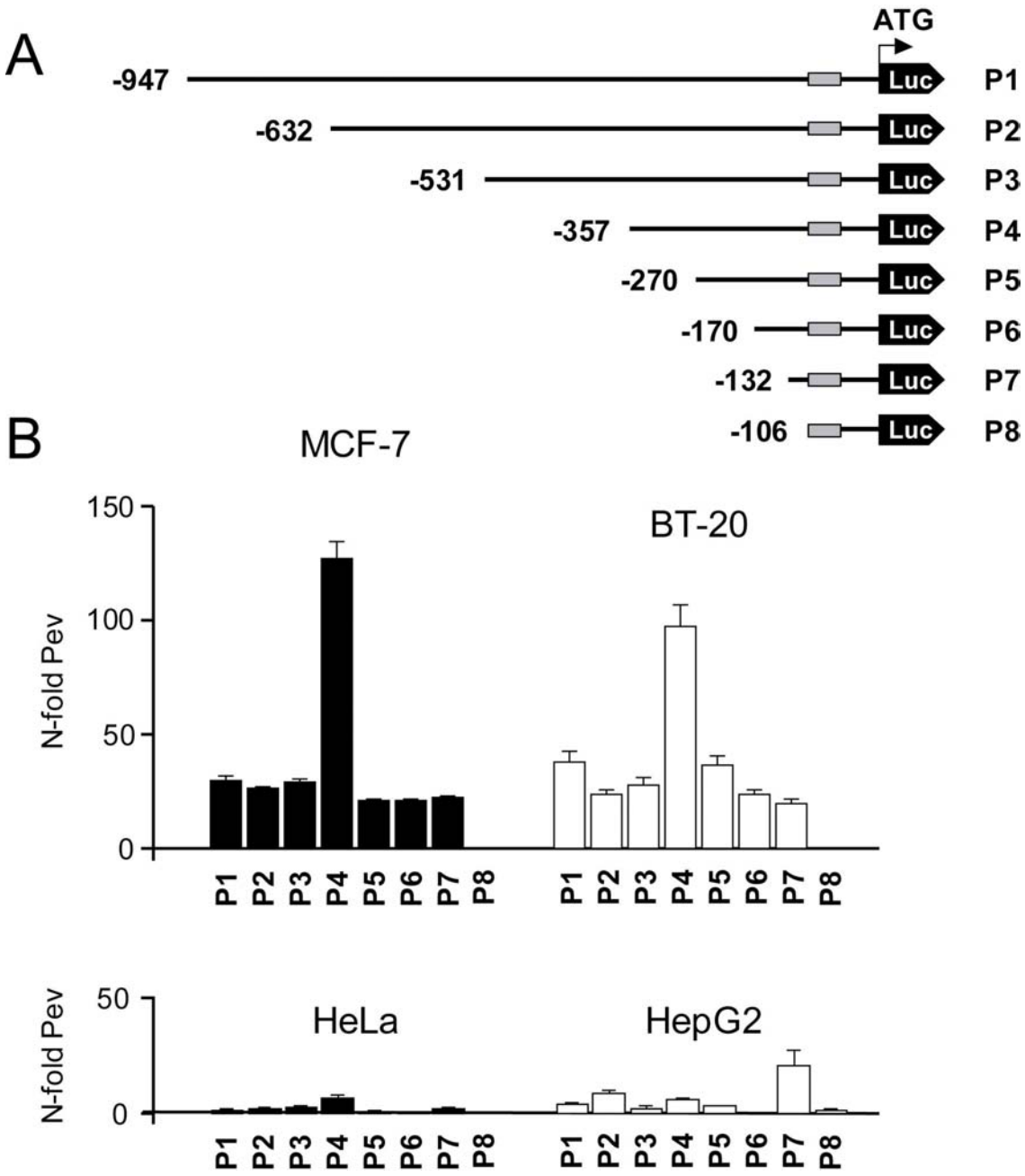

Figure 4. Functional analysis of the SBEM promoter. (A) Different constructs used in luciferase assay are numbered from P1 (longer construct) to P8 (smaller construct). Partial 5'-deleted SBEM promoter regions were cloned in pGL3-basic containing the luciferase reporter gene as described in "Material and Methods". Number on the left indicates positions in the promoter sequence relative to the translation start site (ATG). The gray box corresponds to the putative TATA-boxes. (B) Promoter activities of the SBEM promoter deletion constructs transiently transfected in MCF-7, BT-20, HeLa and HepG2 cells. Luciferase activity, shown as n-fold value compared to cells transfected with the promoterless pGL3-basic vector, was measured as indicated in the "Material and Methods". Data represent the means of at least three independent transfection experiments.

expression, total RNA was extracted from both mammary and non-mammary cells 24 hours following the transfection with pOct1, pOct2 or the empty expression vector. Transcripts were then reverse transcribed and PCR amplified using SBEM and GAPDH primers as described in "Material and Methods". As shown in Figure 8B, endogenous SBEM gene expression was up-regulated in mammary cell lines following Oct1 and Oct2 over-expression. In contrast, the GAPDH gene expression was uniform in all conditions, and no variation in the SBEM gene expression was observed in non-breast cancer cells.

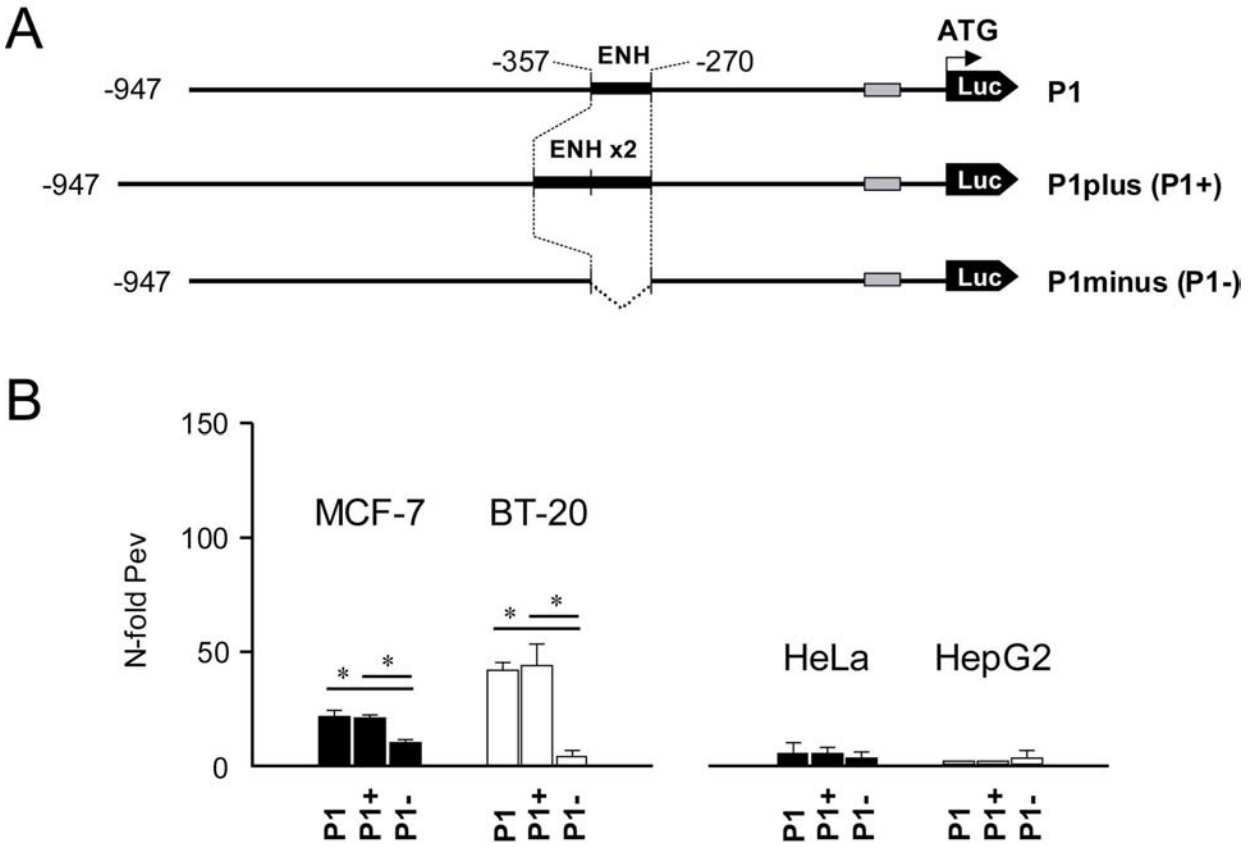

Figure 5. Activity of mutated P1 constructs used in luciferase assay. (A) P1 modified constructs were cloned in pGL3-basic containing the luciferase reporter gene as described in "Material and Methods". P1plus (P1+) contains two copies of the ENH region, and P1minus (P1-) does not contain any. The gray box corresponds to the putative TATA-boxes and ENH region is in black. (B) Promoter activities of the SBEM mutated P1 constructs transiently transfected in MCF-7, BT-20, HeLa and HepG2 cells. Luciferase activity, shown as n-fold value compared to cells transfected by the promoterless pGL3-basic vector, was measured as indicated in the "Material and Methods". Stars indicate activities that are statistically different (t-test; $p<0.05$) from activity of P1minus. Data represent the means of at least three independent transfection experiments.

4. DISCUSSION

In the present study, we have investigated the transcriptional activity of the foremost ~900 bp of the SBEM promoter using sequential 5'-deletion and transfection of both mammary and non-mammary cells. We have identified experimentally an 87-bp enhancer region, which increased the promoter activity in mammary but not in other cell tested.

Three particular motifs were identified within this short sequence. The first two overlapping motifs corresponded to binding sites for Nkx2-5 and the autoimmune regulator (AIRE). Nkx2-5, which belongs to the NK2 family of homeobox transcription factor, is essential for cardiac development , but its expression has never been detected in breast tissue (14). Similarly, the expression of the autoimmune regulator AIRE, a transcription factor previously shown to control the self-reactivity of the T cell repertoire, is mainly restricted to the thymus, with low detectable levels in lymph node, fetal liver and spleen (15), but not in breast tissue. It was therefore unlikely that these two overlapping sites were participating to the strong activity of the SBEM promoter in breast cells.

The identification of an octamer-binding transcription factors (Oct) motif within the 87-bp enhancer region was, however, of particular interest. Indeed, this octamer motif is recognized by Oct1 and Oct2 transcription factors (16). Both factors belong to the POU (Pit-1, Oct1/2 and Unc-86) domain family of transcription factors, which are strongly involved in embryogenesis, organ development and cell-type determination (17). Oct1, ubiquitously expressed, as well as Oct2, previously described as a B-lymphocyte-specific factor, are both expressed in breast cancer cell lines (18-20). Furthermore, Oct1 and Oct2 are involved not only in the transcriptional regulation of genes strongly expressed in mammary glands such as beta-casein, Prolactin and cyclin D1 genes (21-24), but also in the activity of the exogenous MMTV (mouse mammary tumor virus) promoter, predominantly active in breast tissue (25,26). The presence, within the strong enhancer region (87-bp ENH), of an octamer-binding site strongly

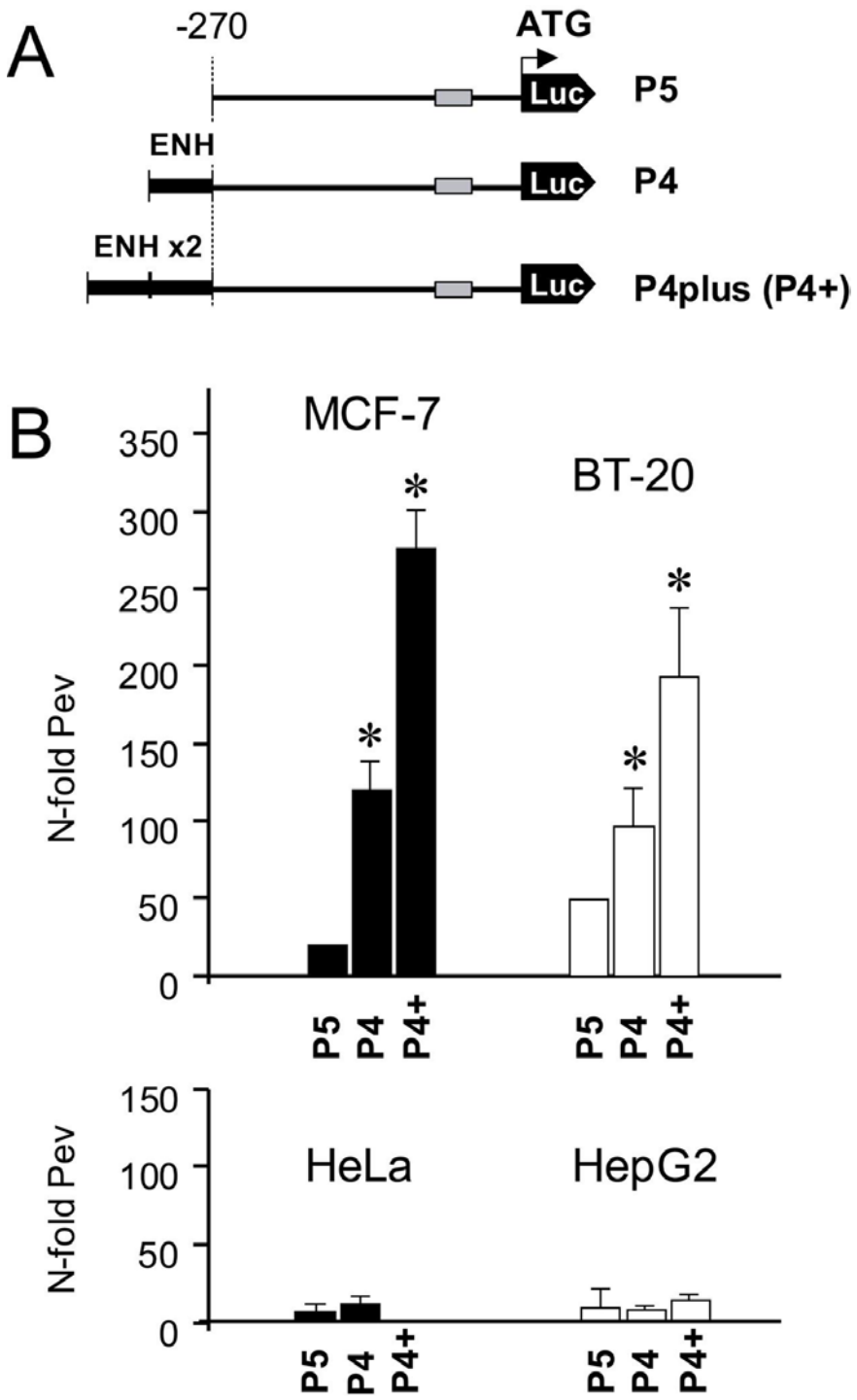

Figure 6. Activity of mutated P4 constructs used in luciferase assay. (A) P4 modified constructs were cloned in pGL3-basic containing the luciferase gene as described in "Material and Methods". P4plus contains two copies of the ENH region, and P5 does not contain any. The gray box corresponds to the putative TATA-boxes and ENH region is in black. (B) Promoter activities of the SBEM mutated P4 constructs transiently transfected in MCF-7, BT-20, HeLa and HepG2 cells. Luciferase activity, shown as n-fold value compared to cells transfected by the promoterless pGL3-basic vector, was measured as indicated in the "Material and Methods". Stars indicate activities that are statistically different (t-test; $p<0.05$) from activity of P5. Data represent the means of at least three independent experiments.

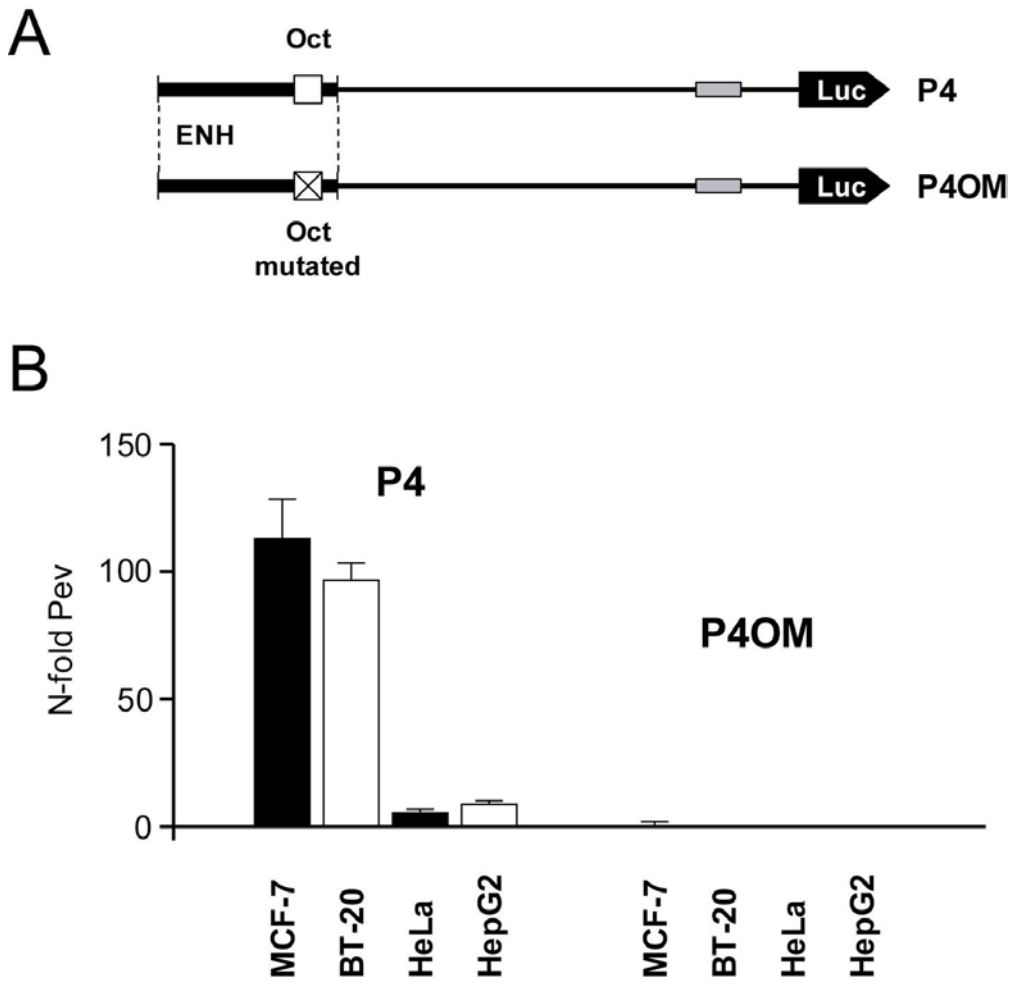

Figure 7. Activity of P4 and P4OM constructs used in luciferase assay. (A) P4OM was mutated in the octamer-binding site located in (-282/-274) as compared to the wild-type construct P4 (See "Material and Methods"). The gray box corresponds to the putative TATA-boxes and ENH region is in black. (B) Promoter activities of the SBEM P4 and P4OM constructs transiently transfected in MCF-7, BT-20, HeLa and HepG2 cells. Luciferase activity, shown as n-fold value compared to cells transfected by the promoterless pGL3-basic vector, was measured as indicated in the "Material and Methods". Data represent the means of three independent experiments.

suggested a possible involvement of these factors in the mechanisms underlying SBEM promoter activity in breast cancer cells. The suppression of the reporter activity following the octamer-binding site mutation supports the hypothesis that this motif, rather than Nkx2-5 or the AIRE motif, indeed regulates SBEM promoter activity. Furthermore, the subsequent increase in both exogenous reporter activity and endogenous SBEM mRNA level following over-expression of either Oct1 or Oct2 transcription factor in MCF-7 and BT-20 cells corroborates this assumption. We have recently confirmed the involvement of this octamer-binding motif using a newly developed method, the promoter competition assay (27).

It has been widely assumed in the past that the octamer-binding factor Oct2 mediated tissue-specific promoter activity, whereas the ubiquitously expressed Oct1 mediates general promoter activity (28). However, studies have more recently challenged this dogma, underlying a role of Oct1 in the tissue-specific expression of numerous genes (28-31). It has now been established that both Oct1 and Oct2 were involved in the regulation of tissue-specific genes such as immunoglobulin genes (18,29,31-33). As SBEM gene is one of the most "breast-specific" gene identified to date, it could be proposed that these two transcription factors, in addition to regulating the strong expression of the SBEM gene, could also be responsible for the its breast-specificity of its expression.

However, since the over-expression of either Oct1 or Oct2 alone is not sufficient to restore SBEM gene expression in non-mammary cell lines, it is reasonable to assume that a participation of octamer-binding transcription factors to the breast-specific expression of SBEM, if any, will involve other partners. The need for other molecules, beside octamer-binding transcription factors, to induce

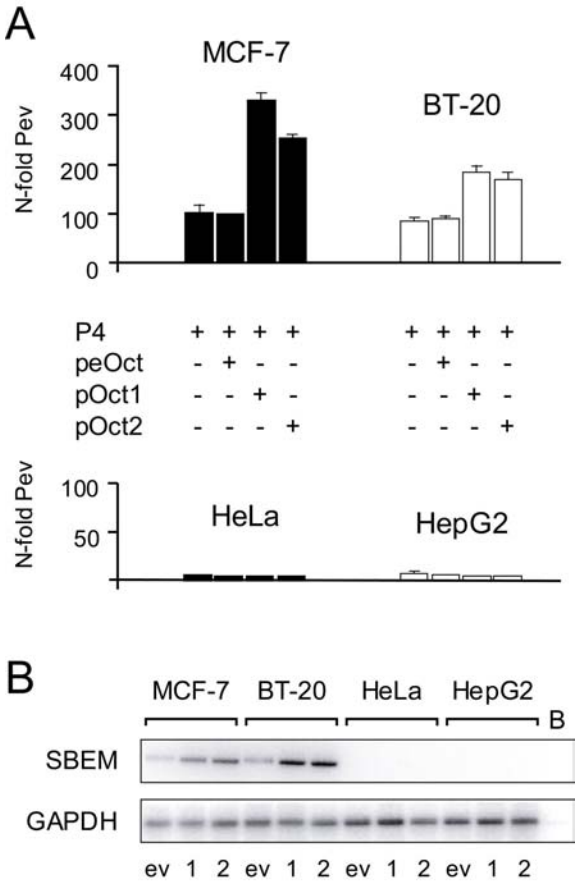

Figure 8. Up-regulation of exogenous and endogenous SBEM promoter activities. (A) Promoter activities of the SBEM P4 constructs co-transfected with expression vectors pOct1, pOct2 and the corresponding empty vector peOct in MCF-7, BT-20, HeLa and HepG2 cells. Luciferase activity, shown as n-fold value compared to cells transfected by the promoterless pGL3-basic vector, was measured as described earlier. Data represent the means of at least three independent transfection experiments. (B) RT-PCR analysis of RNA extracted from human breast cancer cells (MCF-7 and BT-20) and non-breast cancer cells (HeLa and HepG2) formerly transfected with peOct, pOct1, and pOct2. Total RNA was reverse-transcribed and PCR amplified with SBEM or GAPDH primers, as described in "Material and Methods" and in Table 1. Data are representative of two independent RNA extraction for each cell lines.

tissue specificity has been demonstrated in other models (34-38). For example, the ubiquitous Oct1 and a B-cell-specific co-activator are both required for the B-lymphocyte-specific expression of immunoglobulins (35). Similarly, Runx2, a bone-specific transcription factor belonging to the Runt family, participates to the Oct1 induced osteoblast differentiation and chondrocyte maturation (39). Interestingly, Runx2 has recently been identified in mammary epithelial cells (40) and was found to participate to the formation of a complex with Oct1 to subsequently contribute to the expression of the mammary gland-specific beta-casein gene in breast tissues (38). We propose that such factors, remaining to be identified, interact with octamer-binding transcription factors to regulate the breast specific expression of SBEM gene.

It should be stressed that two other members of the Oct transcription factors family, Oct3 and Oct11, have also been detected in breast cancer cells and tissues, but not in normal breast tissues (20). The over-expression of SBEM during breast tumorigenis (7,9) could therefore result from a change in expression of these alternative octamer-binding transcription factors. However, as of today, very few data have been accumulated regarding these two new members of the POU family and knowledge of DNA binding properties is still limited. It is therefore premature to speculate any involvement of the Oct3 and Oct11 in SBEM transcriptional regulation and further studies are needed to clarify this issue.

Our data clearly demonstrated that octamer-binding transcription factors are involved in the regulation of the SBEM promoter activity. Further studies are still needed to fully understand the mechanisms behind the specific expression of the SBEM gene in breast and to identify the molecules involved. We however strongly believe that the understanding of these mechanisms will be invaluable in designing novel diagnostic as well as therapeutic strategies for the clinical management of breast cancer.

5. ACKNOWLEDGEMENT

We are grateful to Winshop Herr (Cold Spring Harbor Laboratory, New York, USA) who provided the pOct1, pOct2 and peOct expression vectors. We particularly thank George Deng for his technical help. F.H. is a Manitoba Health Research Council Fellow and S.C-K. has a Manitoba Health Research Council Studentship. Y.M. has a Manitoba Medical Service Foundation Career Award. This work is supported by the Canadian Institute of Health Research (CIHR # MOP 62971) and Friends You Can Count On (USA).

6. REFERENCES

1. Parkin D.M, F Bray, J Ferlay & J Pisani : Global cancer statistics, 2002. CA Cancer *J Clin* 55, 74-108 (2005)
2. Wellings S.R, & H M. Jensen: On the origin and progression of ductal carcinoma in the human breast. *JNatl Cancer Inst* 50, 1111-1118 (1973)
3. Krishnamurthy S & N Sneige: Molecular and biologic markers of premalignant lesions of human breast. *Adv Anat Pathol* 9, 185-197 (2002)
4. Schmidt E.V: Genes involved in breast cancer progression: analysis of global changes in gene expression or retroviral tagging? *Am J Pathol* 161, 1973-1977 (2002)
5. Garnis C., T. P. Buys & W. L. Lam.: Genetic alteration and gene expression modulation during cancer progression. *Mol Cancer* 3, 9 (2004)
6. Hube F, M Mutawe, E Leygue & Y Myal: Human small breast epithelial mucin: the promise of a new breast tumor biomarker. *DNA Cell Biol* 23, 842-849 (2004)

7. Houghton R.L, D. C. Dillon, D. A. Molesh., B.K. Zehentner, J. Xu, J Jiang, C. Schmidt, A Frudakis, E Repasky, F. A. Maltez, M Nolasco, R Badaro, X Zhang, P. C. Roche, D. H. Persing & S.G Reed: Transcriptional complementarity in breast cancer: application to detection of circulating tumor cells. *Mol Diagn* 6, 79-91 (2001)
8. Miksicek R.J, Y. Myal, P.H.Watson, C Walker, L.C. Murphy & E Leygue: Identification of a novel breast- and salivary gland-specific, mucin-like gene strongly expressed in normal and tumor human mammary epithelium. *Cancer Res* 62, 2736-2740 (2002)
9. Colpitts T.L, P Billing, E Granados, M Hayden, S Hodges, L Roberts, J. Russell, P. Friedman, & S Stroupe: Identification and immunohistochemical characterization of a mucin-like glycoprotein expressed in early stage breast carcinoma. *Tumour Biol* 23, 263-278 (2002)
10. Quandt K, K. Frech, H. Karas, E Wingender & T. Werner: MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* 23, 4878-4884 (1995)
11. Wingender E, X. Chen X, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, Ohnhauser R., Pruss M., Schacherer F., S. Thiele & Urbach S.: The TRANSFAC system on gene expression regulation. *Nucleic Acids Res* 29, 281-283 (2001)
12. Hube F, P. Reverdiau, S. Iochmann, C. Cherpi-Antar, & Y. Gruel: Characterization and functional analysis of TFPI-2 gene promoter in a human choriocarcinoma cell line. *Thromb Res* 109, 207-215 (2003)
13. Tanaka M, & W. Herr: Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell* 60, 375-386 (1990)
14. Brown C.O, III, X. Chi, E. Garcia-Gras, M. Shirai, X.H. Feng & R. J. Schwartz: The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. *J Biol Chem* 279, 10659-10669 (2004)

15. Su M.A. & M.S. Anderson: Aire: an update. Curr. Opin. Immunol. 16, 746-752 (2004)
16. Gstaiger M, L. Knoepfel, O. Georgiev, W. Schaffner & C.M. Hovens: A B-cell coactivator of octamer-binding transcription factors. *Nature* 373, 360-362 (1995)
17. Chen H, & S. Sukumar: Role of homeobox genes in normal mammary gland development and breast tumorigenesis. *J Mammary Gland Biol Neoplasia* 8, 159-175 (2003)
18. Sturm R.A., G. Das & W. Herr: The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev* 2, 1582-1599 (1988)
19. D.S. Latchman: The Oct-2 transcription factor. Int. J. Biochem. *Cell Biol* 28, 1081-1083 (1996)
20. Jin T, D.R. Branch, X. Zhang, S. Qi, B. Youngson & P.E. Goss: Examination of POU homeobox gene expression in human breast cancer cells. *Int J Cancer* 81, 104-112 (1999)
21. Voss J.W, L. Wilson & M.G. Rosenfeld: POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and can cooperate to induce expression of the prolactin promoter. *Genes Dev* 5, 1309-1320 (1991)
22. Zhao F.Q, K. Adachi, & T. Oka: Involvement of Oct-1 in transcriptional regulation of beta-casein gene expression in mouse mammary gland. *Biochim Biophys Acta* 1577, 27-37 (2002)
23. Boulon S, J.C. Dantonel, V. Binet, A. Vie, J.M. Blanchard, R.A. Hipskind & A. Philips: Oct-1 potentiates CREB-driven cyclin D1 promoter activation via a phospho-CREB- and CREB binding protein-independent mechanism *Mol Cell Biol* 22, 7769-7779 (2002)
24. Cicatiello L, R. Addeo, A. Sasso, L. Altucci, V.B. Petrizzi, R. Borgo, M. Cancemi, S. Caporali, S. Caristi, C. Scafoglio, D. Teti, F. Bresciani, B. Perillo & A. Weisz: Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter. *Mol Cell Biol* 24, 7260-7274 (2004)
25. Kim M.H & D.O. Peterson: Stimulation of basal transcription from the mouse mammary tumor virus promoter by Oct proteins. *J Virol* 69, 4717-4726 (1995)
26. Kim M.K, L.A. Lesoon-Wood, B.D. Weintraub & J.H. Chung: A soluble transcription factor, Oct-1, is also found in the insoluble nuclear matrix and possesses silencing activity in its alanine-rich domain. *Mol Cell Biol* 16, 4366-4377 (1996)
27. Hube F, Y. Myal & E. Leygue: The promoter competition assay (PCA): a new approach to identify motifs involved in the transcriptional activity of reporter genes. *Front Biosci* 11, 1577-1584 (2006)
28. Dong K.W, H.M. Zheng, Z.Y. Wen & Z.G. Chen: The POU homeodomain protein Oct-1 binds Cis-regulatory element essential for the human GnRH upstream promoter activity in JEG-3 cells. *J Clin Endocrinol Metab* 86, 2838-2844 (2001)
29. Prefontaine G.G, R. Walther, W. Giffin, M.E. Lemieux, L. Pope & R.J. Hache: Selective binding of steroid hormone receptors to octamer transcription factors determines transcriptional synergism at the mouse mammary tumor virus promoter. *J Biol Chem* 274, 26713-26719 (1999)
30. Gonzalez M.M & C. Carlberg: Cross-repression, a functional consequence of the physical interaction of non-liganded nuclear receptors and POU domain transcription factors. *J Biol Chem* 277, 18501-18509 (2002)
31. Belikov S, C. Astrand, P.H. Holmqvist & O. Wrange: Chromatin-mediated restriction of nuclear factor 1/CTF binding in a repressed and hormone-activated promoter in vivo. *Mol Cell Biol* 24, 3036-3047 (2004)
32. Fletcher C, N. Heintz, & R.G. Roeder: Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. *Cell* 51, 773-781 (1987)
33. Malone C.S, L. Patrone, & R. Wall: An essential octamer motif in the mb-1 (Igalpha) promoter. *Mol Immunol* 37, 321-328 (2000)
34. Inamoto S, N. Segil, Z.Q. Pan, M. Kimura & R.G. Roeder: The cyclin-dependent kinase-activating kinase (CAK) assembly factor, MAT1, targets and enhances CAK activity on the POU domains of octamer transcription factors. *J Biol Chem* 272, 29852-29858 (1997)
35. Luo Y, H. Ge, S. Stevens, H. Xiao & R.G. Roeder: Coactivation by OCA-B: definition of critical regions and

synergism with general cofactors. *Mol Cell Biol* 18, 3803-3810 (1998)
36. Malone C.S & R Wall: Bob1 (OCA-B/OBF-1) differential transactivation of the B cell-specific B29 (Ig beta) and mb-1 (Ig alpha) promoters. *J Immunol* 168, 3369-3375 (2002)
37. Lins K, A. Remenyi, A. Tomilin, S. Massa, M. Wilmanns, P. Matthias & H.R. Scholer: OBF1 enhances transcriptional potential of Oct1. *EMBO J* 22, 2188-2198 (2003)
38. Inman C.K, N. Li, & P. Shore: Oct-1 counteracts autoinhibition of Runx2 DNA binding to form a novel Runx2/Oct-1 complex on the promoter of the mammary gland-specific gene beta-casein. Mol. Cell Biol. 25, 3182-3193 (2005)
39. Komori T: Runx2, a multifunctional transcription factor in skeletal development. *J Cell Biochem* 87, 1-8 (2002)
40. Barnes G.L, A. Javed, S.M. Waller, M.H. Kamal, K.E. Hebert, M.Q. Hassan, A. Bellahcene, A.J. Van Wijnen, M.F. Young, J.B. Lian, G.S. Stein & L.C. Gerstenfeld: Osteoblast-related transcription factors Runx2 (Cbfa1/AML3) and MSX2 mediate the expression of bone sialoprotein in human metastatic breast cancer cells. *Cancer Res* 63, 2631-2637 (2003)

Abbreviations: ENH, enhancer region; GAPDH, Glyceraldehyde-3-phosphate dehydrogenase; Oct, octamer-binding transcription factor; RACE, Rapid amplification of the 5'-cDNA end; RLU, relative luciferase unit; SBEM, small breast epithelial mucin; MMTV, mouse mammary tumor virus

Key Words: Breast cancer, Breast-specific, Enhancer, Octamer-binding transcription factor, Promoter, SBEM

Send correspondence to: Dr Etienne Leygue, Department of Biochemistry and Medical Genetics, University of Manitoba, 770 Bannatyne Avenue, Winnipeg, MB, R3E 0W3, Canada, Tel: 204-977-5608, Fax: 204-789-3900, E-mail: eleygue@cc.umanitoba.ca

http://www.bioscience.org/current/vol11.htm