Identification of a Novel Breast- and Salivary Gland-specific, Mucin-like Gene Strongly Expressed in Normal and Tumor Human Mammary Epithelium

Richard J. Miksicek, Yvonne Myal, Peter H. Watson, Christina Walker, Leigh C. Murphy, and Etienne Leygue

Department of Physiology, Michigan State University, East Lansing, Michigan 48824 [R. J. M.]; Department of Pathology, University of Manitoba, Winnipeg, Manitoba, R3E 0W3 Canada [Y. M., P. H. W.]; Department of Biochemistry and Medical Genetics and Manitoba Institute of Cell Biology, University of Manitoba, Winnipeg, Manitoba, R3E 0W3 Canada [P. H. W., C. W., L. C. M., E. L.]

Abstract

Expression profiling using the public expressed sequence tag (EST) and serial analysis of gene expression (SAGE) databases resulted in the identification of a putative breast-specific mRNA that we have termed small breast epithelial mucin (SBEM). Hybridization analysis performed on 43 normal human tissues revealed that the SBEM gene was only expressed in mammary and salivary glands. Further reverse-transcription PCR analyses confirmed SBEM expression in most of established human breast epithelial cell lines analyzed (7 of 8) but not in cell lines of non-breast origin (0 of 6). SBEM mRNA expression was detected in >90% of invasive ductal carcinomas and correlated with the expression of a previously characterized breast-specific gene, mammaglobin-1 (n = 54; Spearman r = 0.34, P = 0.011). Interestingly, a higher SBEM:mammaglobin-1 ratio was observed in primary tumors with axillary lymph node metastasis than in node-negative tumors (n = 46; Mann-Whitney, P = 0.04). In a subset of 20 primary breast tumors and their matched axillary lymph nodes, a high concordance (Fisher’s exact test, P < 0.0001) was seen between PCR detection of SBEM mRNA in lymph node tissue and their histopathological status, indicating that SBEM mRNA expression is conserved in nodal metastasis. The SBEM gene is predicted to code for a putative low molecular weight, secreted sialoglycoprotein, potentially useful for the diagnosis of metastatic breast cancer.

Introduction

Early detection remains a central goal in breast cancer treatment to enable intervention at a localized and potentially curable stage and to maximize the opportunity for breast conservation. The 5-year survival rate for women with breast cancer increases dramatically when it can be diagnosed at an early stage, from >95% in patients with a localized tumor to ~75% with regional disease and <25% in women with disseminated cancer (1). Nevertheless, only 60% of all breast cancers are diagnosed at a local stage, and any improvement in early detection would have a significant impact on reducing overall breast cancer mortality.

Improving the diagnosis and clinical management of breast cancer requires access to a wider range of biomarkers able to reflect the molecular phenotype of breast tissue. A special need exists to identify novel genes whose expression is restricted to the mammary epithelium, because these genes have the greatest potential to enhance detection of micrometastatic disease and the potential to report on proliferative changes in the breast, analogous to the ability of elevated serum prostate-specific antigen levels to indicate the presence of hyperplasia or cancer of the prostate gland (2).

The identification of new tissue-specific markers has benefited especially from expansion of public and private databases for ESTs (3, 4) and by large-scale efforts to profile patterns of gene expression using techniques such as serial analysis of gene expression (5). Using sequence analysis software and web-based tools developed for molecular profiling, we have identified a novel putative breast-specific gene, belonging to a recently regrouped cluster (UniGene identifier Hs.348419), which represents an attractive candidate for a breast tumor marker with obvious potential for cancer diagnostics.

Materials and Methods

Database and Sequence Analysis. The cdna xProfiler tool (6) was used to search for novel breast-specific ESTs. Protein sequence analysis used the NetOGlyc algorithm (7) to search for the presence of a signal sequence (6) and the NetOGlyc algorithm (8) to predict sites of potential glycosylation (7).

RNA Hybridization Analysis for Tissue Specificity. A 32P-labeled SBEM probe, generated using the cloned SBEM PCR product (396 bp) and the RadPrime DNA labeling system (Life Technologies, Inc., Burlington, Ontario, Canada), was hybridized to a commercially available RNA Master Blot (Clontech, Palo Alto, CA), containing poly(A)+ RNA (100–500 ng) isolated from a variety of adult and fetal human tissues, according to the manufacturer’s instructions.

Cell Culture and RNA Preparation. Cell lines were obtained from the American Type Culture Collection or other sources and were cultured as follows: DMEM with 10% fetal bovine serum (MCF7, MCF10AT1, MCF10AT3c, SK-UT-1B, and HepG2); DMEM with 10% calf serum (MDA MB-231, HeC 1A, and HeLa); DMEM/Ham’s F12 (1:1) with 10% fetal bovine serum (ZR-75-1 and RL95-2); RPMI 1640 with 10% fetal bovine serum (T-47D and LNCaP); or MSU-1 medium (8) with 5% fetal bovine serum (M13SV-1). All media were supplemented with penicillin (50 units/ml), streptomycin (50 µg/ml), HEPE (pH 7.4; 5 µM), and glutamine (2 mM). MCF7, T-47D, and ZR-75-1 cells also received bovine insulin (10 µg/ml). Media and sera were obtained from Life Technologies, Inc. (Life Technologies, Inc., Grand Island, NY). RNA was extracted from cultured cells using guanidinium isothiocyanate, followed by centrifugation through a 5.7 M cesium chloride cushion as described (9). RNA from cultured primary HMECs obtained by reduction mammaryplasty was a kind gift from P. Ervin (Biotherapeutics, Inc., Ann Arbor, MI).

Breast Tumors and Axillary Lymph Nodes. Fifty-four invasive ductal carcinomas were selected from the Manitoba Breast Tumor Bank (Winnipeg, Manitoba, Canada). Cases spanned many ER (0–298 fmol/mg protein) and PR (0–1199 fmol/mg protein) levels, as determined by ligand binding assay. Tumors also spanned many grades (Nottingham grade scores from 5 to 9). For

4 The abbreviations used are: EST, expressed sequence tag; SBEM, small breast epithelial mucin; HMEC, human mammary epithelial cell; ER, estrogen receptor-α; PR, progesterone receptor; RT-PCR, reverse transcription-PCR; GAPDH, glyceraldehyde phosphate dehydrogenase; MUC1, mucin 1.
5 Internet address: http://www.ncbi.nlm.nih.gov/UniGene/. UniGene is a system for automatically partitioning GenBank sequences, including ESTs, into a nondisjoint set of gene-oriented clusters.
6 Internet address: http://cgap.nci.nih.gov/CGAP/Tissues/xProfiler.
7 Internet address: http://www.cbs.dtu.dk/services/SignalP/.
8 Internet address: http://www.cbs.dtu.dk/services/NetOGlyc/.
and their matched frozen lymph nodes containing (n = 20), frozen primary human breast tumor samples and their matched frozen lymph nodes containing (n = 14) or not (n = 6) histologically detectable metastatic cancer cells were available from the Mani-
toba Breast Tumor Bank. For the primary tumor samples, the ER levels, determined by ligand binding assays, ranged from 2.3 fmol/mg protein to 298 fmol/mg protein, whereas PR levels ranged from 10.1 fmol/mg protein to 112 fmol/mg protein.

RNA Analysis by RT-PCR. Total RNA was extracted from 20-μm frozen tissue sections (five sections/tumor) and reverse transcribed as described previously (10). The primers used for SBEM amplification consisted of SBEM-U (5′-CTTTGAAAGCTTATTTTGCTGTG-3′; sense) and SBEM-L (5′-AAGGTAATGTTTGAGGAAT-3′; antisense). PCR amplifications were performed, and PCR products were analyzed as described previously (11), with minor modifications. Briefly, aliquots of each reverse transcription mixture (2 μl for Fig. 2 and 0.8 μl for Fig. 3) were amplified in a final volume of 20 μl, in the presence of 20 μM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 200 μM of each deoxynucleotide triphosphate, 5 ng/μl of each SBEM primer, and 0.5 unit of Taq DNA polymerase. Each PCR consisted of 35 cycles (15 s at 94°C, 30 s at 58°C, and 60 s at 74°C).

Primers used for mammaglobin-1 were: Mam-1 (5′-CCAGACAGCACG- CAGCCAAC-3′; sense strand) and Mam-2 (5′-CCGGTGTTTGTTTCTCT -CAC-3′; antisense strand). Primers for the ubiquitously expressed GAPDH gene were GAP-1 (5′-ACACCACCTCCTCCACCTTTTG-3′; sense strand) and GAP-2 (5′-CCTGCTTGCTGGTCTGGG-3′; antisense strand). To amplify cDNA corresponding to mammaglobin-1 and GAPDH, 30 cycles of PCR were used (30 s at 94°C, 30 s at 55°C, and 30 s at 72°C). All buffers were the same as for SBEM PCR, except that 2.5 mM MgCl₂ was used when amplifying mammaglobin-1 cDNA. PCR products were then separated on a 1.5% agarose gel. After electrophoresis, the gels were stained with ethidium bromide (0.5 μg/ml).

Quantification and Statistical Analysis. Three independent PCRs were performed for tumor specimens using SBEM, mammaglobin-1, and GAPDH primers and signals, visualized with UV irradiation on a GelDoc2000/Chem-Doc System (Bio-Rad). Quantitation was determined by densitometry using the Quantity One software (version 4.2; Bio-Rad). SBEM and mammaglobin-1 expression was normalized to GAPDH expression as described previously (11). Correlation between SBEM expression and tumor characteristics or mammaglobin-1 expression was tested by calculation of the Spearman coefficient r. Differences between tumor subgroups were tested using the Mann-Whitney two-tailed test or Fisher’s exact test.

In Situ Analysis of SBEM mRNA Expression. In situ hybridization was performed on adjacent paraffin-embedded breast tumor tissue sections corresponding to a case shown to express high levels of SBEM mRNA by RT-PCR, using SBEM 32P-labeled sense and antisense probes, as described previously (12).

Results

Identification of a Putative Novel Breast-specific Gene. The cDNA xProfiler tool (see “Materials and Methods”) was used to identify tissue-restricted cDNAs with preferential representation in libraries prepared from normal breast tissue and breast tumors. This search identified a new cluster of ESTs now grouped under the UniGene identifier number Hs.346419. Of the 30 ESTs found in this cluster, 15 are ascribed to breast cDNA libraries, 9 were isolated from random activation of gene expression or pooled tissues, 5 were isolated from fetal sources (fetal heart and fetal skin), and 1 came from a head and neck tumor cDNA library. Alignment of these ESTs led to the construction of a 500-bp consensus cDNA sequence containing a 90-amino acid open reading frame in which the initiating methionine is framed by a nearly perfect consensus motif for translation initiation (5′-CCACCATGAG-3′; Ref. 13). Further database analysis showed that this sequence, interrupted by three introns, is present on chromosome 12q13.2. Primers were designed to span the open reading frame, and we cloned a 396-bp fragment from both MCF-7 cells and breast tissue, which we called SBEM (GenBank accession number AF414087). The presence of a hydrophobic signal peptide (residues 1–19; Ref. 6) within the protein sequence (GenBank accession number AAL02119) suggests that SBEM is a secreted protein subject to proteolytic processing. The NetOGlyc glycosylation algorithm (7) further predicts this protein to be O-glycosylated on most of its 16 threonine residues. The SBEM protein contains three tandem copies of a neutral octapeptide core repeat (ThrThrAlaAlaXxxThrThrAla, where Xxx corresponds to Ala, Pro, or Ser). The NH₂ and COOH termini of the processed polypeptide are otherwise charged and fairly polar. These features suggest strong similarity to many sialomucins, although this protein lacks a transmembrane domain and is substantially shorter than most other known epithelial mucins (14, 15).

Expression of SBEM mRNA Is Restricted to the Mammary and Salivary Glands. Database searches suggested that SBEM expression was mainly restricted to breast tissue. To confirm this prediction, we performed hybridization analysis with an RNA MasterBlot containing highly purified polyadenylated RNA from 43 adult and 7 fetal human tissue arrayed on a nylon membrane. A SBEM cDNA probe hybridized exclusively to mRNA from the mammary and salivary glands (Fig. 1). Of note, no expression was observed in colon, lung, uterus, ovary, liver, pancreas, kidney, or prostate, all of which represent common primary tumor sites. Additionally, no hybridization to any of the fetal RNAs was observed.
SBEM mRNA Is Expressed in Breast Cancer Cell Lines but not in Cell Lines of Non-Breast Origin. The profile of SBEM mRNA expression was further assessed using RT-PCR, followed by PCR amplification, in a panel of human breast and non-breast cell lines. A SBEM PCR product of the expected size (396 bp) was readily detected in MCF7 and ZR-75-1 breast tumor cells (data not shown). Lower but reproducible expression was also observed in primary HMECs and in several established breast epithelial cell lines including T-47D, M135V-1 (8), MCF10AT1, and MCF10AT3c (16). MDA MB-231 breast tumor cells were negative for SBEM expression, as were six tumor cell lines of non-breast origin (uterus: RL95-2, SK-UT-1B, Hec 1A; cervix: HeLa; prostate: LNCaP; and liver: HepG2). As controls, we also examined the expression of a housekeeping gene (GAPDH) and mammaglobin-1, an established mammary-specific gene that is being independently investigated as a promising marker for breast tumor diagnosis and nodal metastasis (12, 17). Of the cell lines tested, only HMEC and ZR-75-1 cells expressed mammaglobin-1, consistent with published reports.

Analysis of SBEM mRNA in Human Breast Tumors. Northern blot analyses performed on a small series of 10 cases revealed that SBEM mRNA was 600 bp long and differentially expressed from one sample to another (data not shown). To determine whether SBEM mRNA was widely expressed in human breast tumor tissue, 54 human breast tumors, spanning many ER and PR levels as well as tumor grade and nodal status, were selected from the Manitoba Breast Tumor Bank. Total RNA was extracted from frozen tissue sections and reverse transcribed. PCR amplification of GAPDH (control), mammaglobin-1, and SBEM cDNA was then performed. A PCR product, 396-bp long was detected in all but three tumors (data not shown) when using SBEM-specific primers. After cloning and sequencing, this product was shown to correspond to SBEM cDNA. Quantification of the SBEM signal relative to the GAPDH signal was performed as described in “Materials and Methods.” No correlation was found between SBEM expression and tumor characteristics such as ER (n = 54; Spearman r = 0.01, P = 0.89) and PR (n = 54; Spearman r = 0.03, P = 0.77) levels or tumor grade (n = 44; Spearman r = 0.06, P = 0.68). Interestingly, however, the SBEM signal correlated positively with mammaglobin-1 expression (n = 54; Spearman r = 0.340, P = 0.011). Subgroup comparison of SBEM and mammaglobin-1 expression confirmed our previous observation⁹ that mammaglobin-1 expression is higher in ER-positive and low-grade tumors (Table 1). Interestingly, although not statistically significant (P = 0.09), higher SBEM expression was found in lymph node-positive compared with node-negative tumors. Also of interest is the fact that the SBEM:mammaglobin-1 ratio is significantly (n = 46; Mann-Whitney, P = 0.04) higher in these lymph node-positive tumors.

SBEM mRNA Expression in Primary Breast Tumors and Their Corresponding Axillary Nodes. We next investigated the possibility that SBEM mRNA could be a tissue marker of axillary lymph node metastasis. Twenty independent cases were selected, including 14 tumors that were axillary lymph node positive and 6 that were node negative. Total RNA was extracted from frozen primary tumor sections and frozen node sections of corresponding axillary lymph nodes. The histological status of all tissues was confirmed in paraffin sections cut from adjacent mirror image paraffin tissue blocks that had been processed in parallel to the frozen blocks. These RNAs were reverse-transcribed and analyzed by RT-PCR using SBEM-specific primers. PCR was performed three times, giving the same result. A representative experiment is shown in Fig. 2. A signal corresponding to SBEM was detected in all lymph nodes containing metastatic cells by histopathological assessment (14 of 14 cases). In contrast, no signal was detectable in lymph nodes from cases without histologically detectable tumor cells (0 of 6 cases). RT-PCR detection of SBEM mRNA in axillary lymph nodes is therefore strongly associated (Fisher’s exact test, P < 0.001) with the histopathological detection of lymph node metastases. The higher sensitivity afforded by RT-PCR detection therefore indicates that SBEM, perhaps together with mammaglobin-1, represents an excellent marker for the detection or confirmation of occult breast tumor metastasis, where histopathology may not be definitive.

SBEM mRNA Is Expressed in Mammary Epithelial Tumor Cells in Vivo. To further establish whether SBEM was expressed by mammary epithelial cells in vivo, paraffin breast tumor tissue sections corresponding to a case shown to strongly express SBEM mRNA by RT-PCR were studied by in situ hybridization. No signal was detectable when using a sense probe (Fig. 3A). In contrast, a signal was observed in epithelial tumor cells when using an antisense probe (Fig. 3B). SBEM mRNA was not detected in stromal or inflammatory cells in any of the sections studied.

Discussion

This article reports the identification, cloning, and preliminary characterization of a cDNA encoding a novel mucin-like protein that displays an unusually narrow pattern of expression. Hybridization analysis revealed that SBEM mRNA was only detectable in two normal tissues, breast and salivary gland. Interestingly, the tissue-

---

Table 1: Medium values (arbitrary units) of SBEM expression, mammaglobin-1 (Mam) expression, and SBEM:mammaglobin-1 ratio in different tumor subgroups

<table>
<thead>
<tr>
<th>Tumors</th>
<th>n</th>
<th>SBEM</th>
<th>Mam</th>
<th>SBEM:mammaglobin-1 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER +</td>
<td>34</td>
<td>2.29</td>
<td>0.66</td>
<td>0.04</td>
</tr>
<tr>
<td>ER −</td>
<td>20</td>
<td>2.15</td>
<td>0.34</td>
<td>0.06</td>
</tr>
<tr>
<td>PR +</td>
<td>34</td>
<td>2.30</td>
<td>0.63</td>
<td>0.03</td>
</tr>
<tr>
<td>PR −</td>
<td>20</td>
<td>1.97</td>
<td>0.38</td>
<td>0.01</td>
</tr>
<tr>
<td>Node +</td>
<td>36</td>
<td>2.30</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>Node −</td>
<td>10</td>
<td>1.96</td>
<td>0.70</td>
<td>0.03</td>
</tr>
<tr>
<td>Grades 5-7</td>
<td>24</td>
<td>2.05</td>
<td>0.84</td>
<td>0.03</td>
</tr>
<tr>
<td>Grades 8-9</td>
<td>20</td>
<td>2.24</td>
<td>0.34</td>
<td>0.03</td>
</tr>
</tbody>
</table>

ER +, >3 fmol/mg of protein; ER −, ≤3 fmol/mg of protein; PR +, >10 fmol/mg of protein; PR −, ≤10 fmol/mg of protein, as determined by ligand binding assay. Node +, confirmed metastasis in axillary lymph nodes; Node −, absence of metastasis in analyzed axillary lymph nodes. Grade, Nottingham grading system. Subgroup comparison was performed using the Mann-Whitney two-tailed test.

---

⁹ E. Leygue, L. C. Murphy, and P. H. Watson, unpublished results.
specific expression that we observed experimentally directly reflects the distribution of ESTs within the Hs.348419 cluster. Indeed, as mentioned above, only two adult tissues (breast and head/neck tumors) have been shown to express SBEM-related ESTs. The fact that SBEM is also expressed in salivary tissue does not undermine the possible use of SBEM as a marker of breast cancer, because tumors of the salivary gland are less common and can readily be distinguished clinically.

Among the primary breast tumors examined in this study (representing mostly invasive ductal carcinoma), SBEM mRNA was observed by RT-PCR analysis in the majority (>90%) of cases. Despite a significant overall correlation between the expression of SBEM and mammaglobin-1 mRNA, a significantly higher SBEM:mammaglobin-1 ratio was observed in primary tumors associated with positive axillary lymph nodes as compared with node-negative tumors. This was mostly attributable to a trend toward higher SBEM expression in node-positive tumors. Although further analysis of a larger number of tumors will be required to confirm these observations, this may suggest differences in the biology of these tumors and also a possible role of SBEM and mammaglobin-1 in the mechanisms involved in tumor metastasis. Our findings indicate, however, that SBEM expression is a common feature of breast cancer and can furthermore serve as a useful marker for breast nodal metastasis, both for detection of micrometastatic cells within lymph nodes as well as in the differential diagnosis of the primary origin of an unknown metastasis. This potential is enhanced by the conserved SBEM expression in high grade and ER/PR-negative tumors that are most likely to metastasize.

The potential diagnostic relevance of SBEM is also increased by its predicted biochemical structure. The SBEM cDNA sequence codes for a 90-amino acid polypeptide that contains a distinctive tandem repeat, rich in alanine and threonine residues, that represents a probable target for O-glycosylation. Consistent with such posttranslational modification is the presence of a well-defined signal peptide, leading us to predict that SBEM is likely to be processed at the apical surface of luminal epithelial cells and to be secreted into the alveolar or ductal lumen. Further study is needed to ascertain whether higher SBEM expression occurs in association with tumors.

Secreted (or transmembrane) proteins that contain internally repeated, densely glycosylated neutral core motifs such as this are characteristic of mucins, which are typically expressed by the surface epithelium of secretory mucosae and by exocrine glands (14, 15). The role of mucins is primarily one of hydrating and lubricating epithelial linings, although several mucins have been implicated in modulating both cell adhesion and growth factor signaling (18, 19). Furthermore, mucins have a well-established link to cancer, best illustrated by the product of the MUC1 gene. MUC1 is overexpressed in a variety of epithelial tumors including breast cancer and gives rise to several well-characterized tumor antigens including CA15.3 and CA27.29 (20, 21). Combined with the loss of cell polarity and changes in glycosylation patterns observed in transformed epithelial cells, overexpression of MUC1 results in the appearance of mucin-derived tumor antigens in the sera of cancer patients that are not seen in normal controls (14, 22). We hypothesize that a similar situation may hold for SBEM expression in the human mammary epithelium and in human breast tumors. However, MUC1 displays relatively broad expression among epithelial tissues including the colon, breast, pancreas, ovary, prostate, tracheobronchial tree, stomach, and uterus. For this reason, MUC1-derived tumor antigens have relatively poor specificity for individual tumor types, and their clinical utility is limited to monitoring the efficacy of cancer therapy and warning of tumor relapse or malignant spread (21–23).

Parallels between SBEM and known epithelial mucins such as MUC1, together with its more narrowly restricted pattern of expression, suggest that this novel gene represents an attractive candidate for a breast biomarker with potential for cancer diagnostics, as well as being a possible future target for the development of a breast tumor vaccine. Moreover, the absence of SBEM expression in normal lymph node tissue suggests that this gene could also be used to detect breast micrometastases in axillary lymph nodes.

References

Identification of a Novel Breast- and Salivary Gland-specific, Mucin-like Gene Strongly Expressed in Normal and Tumor Human Mammary Epithelium

Richard J. Miksicek, Yvonne Myal, Peter H. Watson, et al.


Updated version Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/10/2736

Cited articles This article cites 20 articles, 5 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/10/2736.full#ref-list-1

Citing articles This article has been cited by 5 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/62/10/2736.full#related-urls

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.