Mammaglobin, a Mammary-specific Member of the Uteroglobin Gene Family, Is Overexpressed in Human Breast Cancer

Mark A. Watson and Timothy P. Fleming

ABSTRACT

In this report, we describe a novel cDNA isolated from a primary human breast adenocarcinoma and differentially expressed in several breast carcinoma cell lines. The protein encoded by this cDNA, which we have named mammaglobin, is homologous to a family of secreted proteins that includes rat prostatic steroid-binding protein subunit C3, human Clara cell 10-kilodalton protein, and rabbit uteroglobin. Expression of the mammaglobin gene is restricted to the adult mammary gland. More significantly, in a study of 35 breast tumor biopsies, mammaglobin mRNA levels were increased at least 10-fold relative to normal breast tissue in 23% of cases. The breast-specific expression of this potentially secreted protein and its frequent overexpression in primary human breast tumors suggest that mammaglobin may be a novel marker for the management of breast cancer.

INTRODUCTION

The evolution of breast cancer is accompanied by multiple genetic changes that result in qualitative and quantitative alterations in individual gene expression (1). Our hypothesis is that many of these quantitative genetic changes manifest themselves as alterations in the cellular complement of novel transcribed mRNAs. We believe that the identification of these mRNAs, if sufficiently characterized, could provide clinically useful information for patient management and prognosis while enhancing our understanding of breast cancer pathogenesis. Using a "shotgun" approach that uses modifications of the differential display PCR technique (2), we have previously identified several PCR fragments or DESTs (3) from human breast tumor biopsies. These PCR fragments correspond to mRNA transcripts that are under- or overexpressed in breast tumors relative to patient-matched normal breast tissue.

In this report, we describe the isolation of a novel full-length cDNA corresponding to one of these previously isolated DEST sequences. The protein encoded by this cDNA, which we have named mammaglobin, is homologous to several secreted epithelial proteins including hCC10 protein (4), rUg (5), and rPSC3 (6). We demonstrate that expression of the mammaglobin gene is restricted to the adult mammary gland, and that compared to normal breast tissue, 23% (8 of 35) of primary human breast tumors overexpress mammaglobin mRNA. The breast-specific expression of this potentially secreted protein and its frequent dysregulation in primary mammary tumors suggest that mammaglobin may be an important new marker for the management of breast cancer.

MATERIALS AND METHODS

Cell Culture and Tissues. Cell lines were obtained from American Type Culture Collection and grown in DMEM supplemented with 10% FCS. Tissue biopsy specimens and accompanying pathological reports were obtained from the Human Cooperative Tissue Network (7).

Isolation of Mammaglobin cDNA. Total cellular RNA from the cell line MDA-MB415 was isolated using the standard guanidinium isothiocyanate method. A sample of this RNA was used in the RACE PCR procedure using the 5'-Amplifier Race kit (Clonetech) and following the manufacturer's protocol. Amplification steps were performed with Vent DNA polymerase (New England Biolabs) in a volume of 50 μl containing 1× polymerase buffer [10 mM KCl, 20 mM Tris·HCl (pH 8.8), 10 mM (NH4)2SO4, 2 mM MgSO4, 0.1% Triton X-100], 400 μM dNTPs, 2 units Vent polymerase, and 0.2 μM each appropriate primer. Amplification was performed at 94°C for 45 s, 55°C for 1 min, and 72°C for 1 min over 40 cycles. A single, 400-bp amplified PCR product was obtained, digested with EcoRI, and ligated into the EcoRI and Smal sites of the plasmid vector pGEM7Z (Promega). All sequencing was performed on both strands using a Taq DNA polymerase thermal cycle sequencing kit according to the manufacturer's protocol (Promega) and mammaglobin sequence-specific primers.

RNA Isolation, RT-PCR, and Northern Analysis. All indicated total cellular RNA samples were isolated using the standard guanidinium isothiocyanate method and treated with RNase-free DNase (Promega). For RT-PCR analysis, 1 μg of total isolated RNA was reverse transcribed with oligo(dT)12,18 and Superscript II reverse transcriptase ( Gibco-BRL) in a volume of 20 μl containing 1× reverse transcriptase buffer [50 mM KCl, 75 mM KCl, 3 mM MgCl2] provided by the manufacturer, 10 mM DTT, 200 μM dNTPs, 100 ng oligo(dT)12,18, and 200 units reverse transcriptase. Reactions were performed at 45°C for 60 min. One-tenth of each RT reaction was subject to PCR analysis using a mammaglobin specific primers D2R (5'-ATA AGA AGA AGG TGT GG-3') and D2102 (5'-CAG CGG CU CCT TGA TGG GTG GC-3'). Amplification steps were performed in a 50-μl volume containing 1× polymerase buffer provided by the manufacturer, 2 mM MgCl2, 200 μM dNTPs, 0.5 μM each primer, and 2 units Taq DNA polymerase (Perkin Elmer Cetus). Amplification occurred for 40 cycles at 94°C for 30 s, 55°C for 1 min, and 72°C for 1 min. For Northern analysis, 20 μg of total RNA were analyzed as described previously (3) using the full-length mammaglobin cDNA probe. Integrity, equal loading, and uniform transfer of each RNA sample were assessed by ethidium bromide staining and visualization of 28S and 18S ribosomal bands. In no case did the intensity of ribosomal bands differ by more than 50% between tumors, tumor cell lines, or tumor/normal tissue pairs.

RESULTS

Isolation of the Full-length Mammaglobin cDNA. We previously isolated a DEST PCR fragment, the corresponding mRNA of which was abundant in the cell line MDA-MB415 (3). We therefore elected to use the RACE PCR technique (8) with this cell line to isolate the corresponding full-length cDNA. Sequence from the 403-bp fragment isolated by this technique was combined with a sequence obtained previously from the corresponding DEST sequence.

860
<table>
<thead>
<tr>
<th>Position</th>
<th>Nucleotide</th>
<th>Corresponding Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-18</td>
<td>TAC AGC GAC GGC TTC CTT GAT CCT TGC CAC CCG CGA CTG GAT GTC TCT GGC GCG GCC CTC TCC CAG CAC TGC</td>
<td>Met Lys Leu Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys</td>
</tr>
<tr>
<td>19-27</td>
<td>TAC GTA GCA GGC TGC TCG CAT TTA TTG GAG AAT GTG ATT TCC AAG ACA ATC AAT</td>
<td>Tyr Ala Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr Ile Asn</td>
</tr>
<tr>
<td>28-36</td>
<td>CCA CAA GTG TCT AAG ACT GAA TAC AAA GAA CTT CTT CAA GAG TTC ATA GAC GAC</td>
<td>Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Gln Glu Phe Ile Asp Asp</td>
</tr>
<tr>
<td>37-45</td>
<td>AAT GCC ACT ACA AAT GCC ATA GAT GAA TTG AAG GAA TCT TTT CTT AAG CAA ACG</td>
<td>Asn Ala Thr Thr Ala Ile Asp Glu Leu Lys Glu Cys Phe Leu Asn Gln Thr</td>
</tr>
<tr>
<td>46-54</td>
<td>GAT GAA ACT CTG AGC AAT GTT GAG GTT TTT ATG CAA TTA ATA TAT GAC AGC AGT</td>
<td>Asp Glu Thr Leu Ser Asn Val Glu Val Phe Met Gln Leu Ile Tyr Asp Ser Ser</td>
</tr>
<tr>
<td>55-63</td>
<td>CTT TGT GAT TTA TTT TAA CTT TCT GCA AGA CCT TTG GCT CAC AGA ACT GCA GGG</td>
<td>Leu Cys Asp Leu Phe ***</td>
</tr>
<tr>
<td>64-72</td>
<td>333 342 351 360 369 378</td>
<td></td>
</tr>
<tr>
<td>73-81</td>
<td>TAT GGT GAG AAA CCA ACT ACG GAT TGC TGC AAA CCA CAC CTT CTC TTT CTT ATG</td>
<td></td>
</tr>
<tr>
<td>82-90</td>
<td>441 450 459 468 477 486</td>
<td></td>
</tr>
<tr>
<td>91-99</td>
<td>TCT TTT TAC AAA CTA CAA GAC AAT TGT TGA AAC CTG CTA TAC ATG TTT ATT</td>
<td></td>
</tr>
<tr>
<td>100-108</td>
<td>495</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Human mammaglobin sequence. Nucleotide sequence of the cDNA and encoded protein are numbered respectively. Open bars, extent of the previously isolated DEST PCR sequence. Closed bars, sequence obtained from the RACE PCR clone. The putative signal peptide sequence of the mammaglobin protein and poly(A) addition site at the 3' end of the cDNA are shown in bold.
to deduce the full-length 503-bp cDNA sequence now designated as mammaglobin (Fig. 1). Because these sequence data were obtained from a tumor cell line by PCR amplification, it was conceivable that mutations may have been introduced into this cDNA clone. To exclude this possibility, two specific primers based on the MDA-MB415 mammaglobin cDNA sequence were used to amplify the mammaglobin cDNA from normal human breast tissue. Cloning and sequencing of the mammaglobin cDNA obtained from human breast tissue revealed no sequence discrepancies with the first MDA-MB415 clone. This demonstrated that the cDNA is not mutated in this cell line, nor were mutations introduced during the RACE procedure. The full-length mammaglobin sequence is displayed in Fig. 1. Within this 503-bp cDNA is a 279-bp open reading frame that encodes a polypeptide of 93 amino acids and a predicted molecular mass of 10.5 kDa (Fig. 2). The first 19 residues of this sequence also predict a hydrophobic peptide signal sequence. The initial methionine of the open reading frame contains a near-perfect Kozak consensus sequence, and the 60 bp upstream of this sequence contain no other in-frame methionines or translational stops. The 3'-untranslated sequence of the cDNA constitutes 163 bp and contains a polyadenylation signal, AATAAA, 12 bp upstream of the t1,AC oligonucleotide priming site of the original DEST sequence. On the basis of these data, we conclude that we have isolated the full-length mammaglobin cDNA.

Mammaglobin Is a Novel Member of the Uteroglobin Gene Family. Using the mammaglobin cDNA sequence to query Genbank with the BLAST search algorithm (9, 10), no obvious DNA sequence homologies were identified. However, using the amino acid sequence of the putative translation product, mammaglobin exhibited 42% amino acid identity (58% homology including conservative substitutions) with rPSC3 (Fig. 2). This protein is one of three subunits that forms a tetrameric structure, constituting the major secretory protein of the mammary gland. rPSC3 is expressed in rat ventral prostate (6), whereas expression of hCC10 and rUG has been demonstrated in numerous tissues including lung, uterus, prostate, and breast (5, 13). Because of the sequence homology between mammaglobin and these proteins, we sought to compare their patterns of tissue-specific expression. As shown in Fig. 3A, an ~500-bp mammaglobin message was easily detected in tumor specimen 2410 (the tissue from which the DEST corresponding to the mammaglobin cDNA was isolated) and to a much less extent in normal human breast tissue. However, expression of mammaglobin was undetectable in human uterus and lung, two sites of uteroglobin expression, as well as ovary and placenta. To broaden the scope and increase the sensitivity of our study, we used RT-PCR analysis in several other adult human tissues to assay for mammaglobin gene expression. As shown in Fig. 3B, RT-PCR analysis detected mammaglobin mRNA in both tumor 2410 and normal breast tissue. Of 15 other tissues surveyed, however, no mammaglobin expression was seen. This included tissues that normally express rPSC3 and uteroglobin (lung, uterus, prostate), hormonally responsive and steroidogenic tissues (ovary, tests, placenta), and other epithelial organs (colon, bladder). Although a band similar in size to the mammaglobin PCR product was detected in testis, this band originated from the GAPDH RT-PCR reaction. Therefore, it most likely represents a nonspecific product resulting from GAPDH oligonucleotide priming and not an authentic mammaglobin PCR product. Detection of GAPDH message (Fig. 3B) and epidermal growth factor receptor message (data not shown) in all reactions demonstrated that absence of expression was not due to degraded RNA or other trivial explanations. Therefore, to the extent of tissues examined and the sensitivity of the RT-PCR assay used, mammaglobin is a mammary-specific gene.
BT474, MDA-MB175, and MDA-MB468. In cell lines (and tumor tissues, see below) that demonstrated elevated mammaglobin expression, a second ~3-kb transcript was detected in addition to the 0.5-kb mammaglobin message. Because low-stringency Southern Blot analysis with a mammaglobin genomic probe has failed to reveal other human mammaglobin-related genes and sequence analysis of the mammaglobin gene predicts a 3-kb unspliced mammaglobin message, we believe that this 3-kb transcript represents unprocessed nuclear mRNA that is detectable in tissues and cell lines with a high

---

4 Unpublished data.
MAMMAGLOBIN EXPRESSION IN BREAST CANCER

Fig. 5. Mammaglobin expression in primary human tumors. A, Northern blot analysis of mammaglobin gene expression in normal breast and indicated tumor biopsy specimens. B, Northern blot analysis of mammaglobin expression in patient-matched pairs of breast tumor biopsies (CA) and surrounding normal breast tissue. Note that exposure times differ between A and B so that mRNA levels cannot be compared directly between these two experiments.

level of mammaglobin transcription. Overall, these results demonstrate that mammaglobin gene expression defines a unique phenotype to a subset of breast carcinoma cell lines and that, in culture, expression is absent in many of the normal cellular constituents that compose the mammary gland in vivo.

Overexpression of Mammaglobin in Primary Breast Carcinoma. To determine the significance of mammaglobin expression in vivo and its frequency of dysregulation in primary breast tumors, we next examined a panel of 35 primary breast carcinomas of differing histological types by Northern blot hybridization with the mammaglobin cDNA probe. Because of potential variability in expression due to environment influences (e.g., patient hormonal status), we also sought to compare tumor specimens directly with patient-matched normal breast tissues samples, although this was not possible in many cases. As shown in Fig. 5, the 500-bp mammaglobin mRNA was again detected in normal breast tissue and tumor 2410. Mammaglobin message was also abundant in seven other tumors. In four of these, expression was at least 10-fold higher than patient-matched normal breast tissue. In all, 23% (8 of 35) of tumors examined overexpressed mammaglobin mRNA. These data suggest that overexpression of mammaglobin is not unique to a single tumor specimen and is, in fact, relatively frequent among primary breast tumors. Review of clinical and pathological data available for each tumor (Table 1) revealed no clear relation between mammaglobin expression and tumor grade, stage, histological classification, or hormone receptor status.

DISCUSSION

We have described previously the use of a modified differential display PCR technique to identify sequences (DESTs) that are differentially expressed between breast carcinoma and normal breast tissue in vivo (3). Using this technique, we are interested in isolating genes that contribute to novel pathways in breast cancer progression and that may serve as useful prognostic markers for patient management. In this report, we have identified a novel cDNA that encodes a mammary-specific protein with homologies to uteroglobin and a prostatic steroid-binding protein. This protein, mammaglobin, is overexpressed in 23% of primary tumors examined thus far.

The amino acid sequence homology between mammaglobin, rPSC3, and uteroglobin suggests that these proteins constitute a gene family. Rat prostatic steroid binding protein (prostatin) is the major secretory protein in the rat ventral prostate (11). It binds steroid hormones, and its synthesis is stimulated by testosterone. The mature protein is a tetramer composed of two different dimeric subunits, C3/C1 and C3/C2, each linked by disulfide bonding. The C1, C2, and C3 genes all encode 8–12-kD peptides and are thought to have arisen from gene duplication, but whereas the C1 and C2 genes show strong homology to each other, they are much less similar to the C3 gene (16). Correspondingly, mammaglobin shows no sequence homology with the C1 or C2 peptides. The transcriptional regulation of the rPSC3 gene has been studied extensively (17). It is of interest that recent reports have demonstrated the rPSC3 promoter fused to SV40 T antigen produces both prostatic and mammary carcinomas in transgenic mice (18). However, the true biological function of prostatin and its individual subunits is unknown, and no corresponding human genes have been identified.

The hCC10 gene is the human homologue of the rUg gene (4, 5, 13). Uteroglobin was originally characterized as a secretory protein in rabbit uterus but has since been found in other epithelial organs including lung, breast, and prostate. The precise pattern of tissue expression, however, is species specific and related to a diverse set of regulatory elements in the rabbit and human genes. Unlike rat prostatin, uteroglobin is a homodimeric protein coupled by two disulfide linkages at the conserved residues Cys-2 and Cys-69. Uteroglobin gene transcription is regulated by tissue-specific responses to steroid hormones, and transcriptional control of this gene has been studied extensively both in vitro and in transgenic animal models (19). rUg is capable of binding progesterone and structurally related steroids and may interfere with the immune and inflammatory activity of several cell types. The hCC10 protein and corresponding mouse and rat

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Patient age</th>
<th>Histology</th>
<th>Grade</th>
<th>Stage</th>
<th>ER/PR*</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2410TU</td>
<td>73</td>
<td>Infiltrating lobular</td>
<td>II</td>
<td>I (±LN)</td>
<td>−/+</td>
<td>No data available for this patient</td>
</tr>
<tr>
<td>B016a</td>
<td>76</td>
<td>Mucinous</td>
<td>II</td>
<td>I (±LN)</td>
<td>+/+</td>
<td>&lt;8% cell proliferation/diploid DNA content</td>
</tr>
<tr>
<td>B032a</td>
<td>61</td>
<td>DCIS, comedo type</td>
<td>II</td>
<td>I (±LN)</td>
<td>−/+</td>
<td>Cath. D+/EGFR−/DNA index 1.3 (aneuploid)</td>
</tr>
<tr>
<td>75Aa</td>
<td>90</td>
<td>Infiltrating ductal</td>
<td>II</td>
<td>I (±LN)</td>
<td>+/+</td>
<td>Ki-67+/p53</td>
</tr>
<tr>
<td>T012</td>
<td>86</td>
<td>Mucinous</td>
<td>III</td>
<td>I (±LN)</td>
<td>−/+</td>
<td>Contralateral breast carcinoma, 12 yr earlier</td>
</tr>
<tr>
<td>87R</td>
<td>68</td>
<td>Infiltrating lobular</td>
<td>III</td>
<td>I (±LN)</td>
<td>−/+</td>
<td></td>
</tr>
<tr>
<td>B014</td>
<td>77</td>
<td>Infiltrating ductal</td>
<td>III</td>
<td>I (±LN)</td>
<td>+/+</td>
<td></td>
</tr>
</tbody>
</table>

*ER, estrogen receptor; PR, progesterone receptor; LN, lymph node; Cath. D, cathepsin D; DCIS, ductal carcinoma in situ; EGFR, epidermal growth factor receptor.

aTumors analyzed with patient-matched normal breast tissue.
homologues may have similar functional properties as rUg but are less well characterized (13). In all cases, the true biological function of these proteins is also unknown.

Unlike hCC10 or rPSC3, mammaglobin has been isolated based on its relative abundance in breast tumor tissue. As an initial attempt to determine its relevance to human breast cancer, we examined the expression of mammaglobin in several breast tumor biopsies. Among 35 tumor specimens examined, 23% overexpressed mammaglobin mRNA relative to normal breast tissue specimens. Although we examined a wide variety of tumor histologies in this study, the small number of tumors examined here prohibits any definitive correlation between tumor cell type and mammaglobin overexpression. We have identified both estrogen-receptor-positive and estrogen-receptor-negative tumors that express mammaglobin. It is possible, therefore, that mammaglobin expression independently defines a subclass of tumor cell phenotypes. Mammaglobin is expressed in early, noninvasive ductal carcinoma in situ, as well as in late stage invasive disease. We have not yet investigated whether mammaglobin is overexpressed in benign breast disease, but this information should assist in defining the relationship between breast cancer progression and mammaglobin expression. It will also aid in evaluating the clinical utility of measuring mammaglobin expression in breast cancer patients.

On the basis of the studies in this report, mammaglobin already demonstrates several properties of a clinically useful breast tumor marker. Unlike erb-B (20) or cyclin D (21), two other genes that are overexpressed in breast carcinoma, mammaglobin is a breast-specific protein, the overexpression of which may reflect a more cell-specific alteration of the mammary epithelium rather than representing a general increased growth potential or mitotic rate. As such, appearance of mammaglobin gene dysregulation may have more specific import for the therapeutic vulnerability or clinical course of a tumor, as is currently the case for the estrogen and progesterone receptors (22). Mammaglobin expression could not be detected in normal lymph nodes or peripheral leukocytes at the level of sensitivity afforded by a single-step RT-PCR assay. This suggests that analysis of mammaglobin transcripts in peripheral lymph nodes may also be useful for detecting occult breast cancer metastases, as has been suggested for other epithelial specific genes (23). Finally, because mammaglobin is a potentially secreted protein, its presence may be detectable in serum of patients whose tumor overexpresses this gene product. As such, mammaglobin may be equally or more clinically useful than other solid tumor markers used for managing patients with breast cancer (22).

REFERENCES


Mammaglobin, a Mammary-specific Member of the Uteroglobin Gene Family, Is Overexpressed in Human Breast Cancer

Mark A. Watson and Timothy P. Fleming


Updated version Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/56/4/860

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.