Identification of a Proviral Structure in Human Breast Cancer

Bingren Liu, Yue Wang, Stella M. Melana, Isabelle Pelisson, Vesna Najfeld, James F. Holland, and Beatriz G-T. Pogo

Departments of Medicine [B. L., Y. W., S. M. M., I. P., V. N., J. F. H.] and Microbiology [B. G-T. P.], Mount Sinai School of Medicine, New York, New York 10029

ABSTRACT

Involvement of a virus similar to mouse mammary tumor virus (MMTV) in human breast cancer has long been postulated but never demonstrated. We have detected by PCR a 660-bp sequence similar to the env gene of MMTV but not to the known endogenous viruses, in 38% of human breast cancers examined (Wang et al., Cancer Res., 55: 5173–5179, 1995). This sequence was expressed in 66% of the env-positive tumors as detected by reverse transcription-PCR (Wang et al., Clin. Cancer Res., 4: 2565–2568, 1998).

In this article we report the amplification of a whole proviral structure from each of two human breast carcinomas that were env positive. Using nested extra-long PCR and primers from specific MMTV sequences, overlapping env-long terminal repeat (LTR), LTR-gag, gag-pol, and pol-env segments were successfully amplified. The 9.9-kb provirus is 95% homologous to MMTV but only 57% to human endogenous retrovirus K10 in 3.5 kb of the gag and pol genes. The provirus displays typical features of a replication competent virus, plus the open reading frame for the superantigen and the glucocorticoid responsive element. Fluorescence in situ hybridization with a 2.7-kb env-LTR sequence of an env-positive breast cancer cell line revealed that the sequence is inserted in several chromosomes but not in chromosomes from normal breast cells.

The origin of the MMTV-like sequences is uncertain. Because they are undetectable in normal tissues, because the similarity between the two isolates is high (96%), and because they maintain open reading frames, they appear to be exogenous.

INTRODUCTION

The participation of a virus similar to MMTV in human breast cancer has long been postulated but never demonstrated (1, 2). Despite a large body of positive evidence, the presence of endogenous retroviral sequences with 58% homology to MMTV (3, 4) and the failure to observe virus particles in tumors have confounded the evidence.

In an effort to exclude confusion with known endogenous retroviruses, we selected a 660-bp region of the MMTV env gene that has low homology to HERV-K10, the prototype human retrovirus most similar to MMTV (4), designed primers to this env gene region, and used the PCR to detect the 660-bp sequence. We found this specific env gene sequence in 38% of unselected breast cancers of American women. The sequence was rarely present (2%) in normal breasts but was not found in other tumors or tissues (5). The sequence was detected by Southern blot hybridization under stringent conditions. Similar results have been reported recently by Etkind et al. (6). We have also shown that this sequence is expressed in 66% of the sequence-positive tumors and in some breast cancer cell lines (7). These observations also suggested that the sequence was present in a very low copy number. When the phenotypic characterization of tumors containing the sequence was studied, a correlation was found between env-positive tumors and expression of laminin receptor (8), a characteristic parameter of breast cancer aggressiveness (9). In addition to the env sequences, we have reported the detection of MMTV-like LTR sequences in human breast cancers that were env sequence positive. A 1.6-kb env-LTR sequence was also found that provided evidence that there was continuity between the two genes.4

We set out to establish if other retroviral genes homologous to MMTV are present in env gene-positive human breast cancers and, if so, whether they are distinct from the known endogenous proviruses, which are defective. We have now amplified, cloned, and sequenced a series of overlapping fragments that correspond to a 9.9-kb proviral sequence from each of two human breast cancers. Characterization of this entire sequence indicated that it is highly homologous to MMTV but not to HERV-K10 and that it has potentiality for replication.
described previously (5, 7). The same strategy was used to amplify and sequence the 2.7-kb env-LTR fragment from tumors MSSM 1 and MSSM 2.

**Amplification and Cloning of 1.3-kb LTR-gag Sequence.** Primers LTR5 and GAG4 were used to amplify a 1.58-kb LTR-gag sequence. Nested PCR with primers LTR5 and GAG3 was then performed, using AmpliTaq polymerase following conditions recommended by the supplier (Perkin-Elmer) to amplify 1.3 kb. Perfect-match PCR Enhancer (Stratagene) was used to increase product specificity. Thermocycling conditions were as follows: 1 cycle of 94°C for 2 min, followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, and 68°C for 10 min, and 1 cycle at 72°C for 10 min. The 3.09-kb PCR product was electrophoresed in 1% agarose gel, followed by Southern hybridization with labeled POL5. The same 380-bp sequence between pol and env genes was amplified with primers POL1R and POL3'R, hybridized with labeled POL9, and sequenced directly from the PCR product.

**Amplification of Sequences from Tumor MSSM 2.** A different strategy was used to amplify and clone some of the sequences from the second tumor. The same primers and PCR conditions were used to clone the 2.7-kb env-LTR, 1.3-kb LTR-gag, and 0.8-kb pol-env fragments. However, the pol-env region was amplified with primers GAG2 and POL8R for the first-round PCR and with primers GAG3R and POL8R for the next-round PCR, and Southern blot hybridization was performed with labeled probe POL4. The pol-env region was amplified with primers POL8 and 3L for the first-round PCR and then with primers POL8 and ENV5R for the second-round PCR. The same PCR conditions were used as described for amplification of pol-env sequences. The amplified PCR products were separated by 2% agarose gel electrophoresis, and Southern hybridization with labeled POL5 was carried out as above.

**Cloning and Sequencing of the Amplified Sequences.** The PCR products were ligated into the PCR II vector from Original TA Cloning Kit (Invitrogen, San Diego, CA) and grown in INV<sup>®</sup>F<sup>™</sup> One shot competent cells (Invitrogen). White colonies were selected, and isolation and purification of the plasmids were accomplished using Plasmid Maxi Kit (Qiagen, Chatsworth, CA). EcoRI digestion and 1% agarose gel electrophoresis and Southern blot hybridization with labeled probes were performed to test for the presence of cloned inserts, and finally, the purified DNA was sequenced by automated DNA sequencing. The sequences were then compared with known sequences of MMTV, HERV-K10, and others through the GenBank program.

**FISH.** The 2.7-kb env-LTR fragment was used as a probe after nick-translation with dioxygenin (Boehringer Mannheim). The conditions described by Najfeld et al. (11) were used for hybridization and detection.

### Table 1

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5′-3′)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTR5</td>
<td>GGTGGACACACGAGCTAAAT</td>
<td>735-754 of MMTV (9308-9327)</td>
</tr>
<tr>
<td>LTR3</td>
<td>CGAACAGAAGACACACACAGCA</td>
<td>1345-1364 of MMTV (9918-9937)</td>
</tr>
<tr>
<td>GAG2</td>
<td>GCACGACTCCACGGAGAGGA</td>
<td>1565-1582</td>
</tr>
<tr>
<td>GAG3</td>
<td>GCACGCTTGGATACCTGCTGT</td>
<td>2053-2072</td>
</tr>
<tr>
<td>GAG4</td>
<td>GATGATGAGCTAATCCTTCC</td>
<td>2499-2518</td>
</tr>
<tr>
<td>GAG5</td>
<td>TTGGATTGTGATTTCGG</td>
<td>2643-2660</td>
</tr>
<tr>
<td>GAG7</td>
<td>AGATTTATCATTTTGGAAAACG</td>
<td>3143-3163</td>
</tr>
<tr>
<td>POL4</td>
<td>TCTAAAACACAGCCAGGACA</td>
<td>3393-3412</td>
</tr>
<tr>
<td>POL7</td>
<td>AAAGGAGAAAGAAGAGGACA</td>
<td>3756-3775</td>
</tr>
<tr>
<td>POL8</td>
<td>CCTTGGTTGGAGAGACATT</td>
<td>4135-4154</td>
</tr>
<tr>
<td>POL5</td>
<td>TGATGAGCTCCCTAAAGGAT</td>
<td>4501-4520</td>
</tr>
<tr>
<td>POL6</td>
<td>AAATACCTACGGGAGATTT</td>
<td>5021-5040</td>
</tr>
<tr>
<td>POL3</td>
<td>GGCTGTACATAYAATATTAATT</td>
<td>5345-5364</td>
</tr>
<tr>
<td>POL2</td>
<td>GTGGATTACCCCTTGTGAAT</td>
<td>5811-5830</td>
</tr>
<tr>
<td>POL10</td>
<td>GCTCCCTGATGACGACAACT</td>
<td>5851-5870</td>
</tr>
<tr>
<td>POL1</td>
<td>GCCGGCCCTGCTGGCGCC</td>
<td>6214-6231</td>
</tr>
<tr>
<td>POL9</td>
<td>TTGGAGGATGCGCCTATAG</td>
<td>6269-6289</td>
</tr>
<tr>
<td>POL3'NR</td>
<td>CCGGACAAAAAACCTATAGTGC</td>
<td>6576-6595</td>
</tr>
<tr>
<td>ENV5SR</td>
<td>AGACGAGCTTGTTGGAGGGCA</td>
<td>6685-6696</td>
</tr>
<tr>
<td>ENV1</td>
<td>AATTACCAAACATTTATGGG</td>
<td>7061-7080</td>
</tr>
<tr>
<td>5L</td>
<td>CGGATCCACGAGTTTAAAG</td>
<td>7284-7295</td>
</tr>
<tr>
<td>EPR</td>
<td>AAACCTAGACCCGCCCTTGTACG</td>
<td>7453-7471</td>
</tr>
<tr>
<td>3L</td>
<td>TACGAGTGAGACGACCTATG</td>
<td>7951-7970</td>
</tr>
</tbody>
</table>

Reverse primers were made based on the specific sequences and indicated by R.

---

**Fig. 1. Amplification of a 2.7-kb env-LTR sequence.** A, position of the primers in the MMTV genome and results of sequencing and comparison with MMTV and HERV-K10. B (left), agarose gel electrophoresis of the PCR product using primers 5L and LTR3. Arrow, position of 2.7 kb; right, Southern blot hybridization with 5′-32P-end-labeled LTR5. M, molecular weight marker; N, negative control; 1, amplified tumor DNA.
RESULTS

The primers and probes used for amplification and sequencing of the proviral sequences are shown in Table 1 together with their location in the MMTV genome (12). All PCRs were performed with tumor and normal breast DNAs. Only the amplified fragments for tumor DNA are shown, because normal DNA failed to be amplified with these primers.

Amplification of a 2.7-kb env-LTR Sequence. In Fig. 1A, the position of the primers is indicated, and in Fig. 1B, the product of two rounds of XL-PCR using primers 5L and LTR3 is shown as well as Southern blot hybridization with probe LTR5. The results of sequencing and comparison to MMTV strain BR6 and to endogenous HERV-K10 are summarized in Fig. 1A. The 2.7-kb sequence was 94 and 95% homologous to MMTV in 2.59 kb of the two breast cancers but only 55 and 64% homologous to HERV-K10 in 116 bp (GenBank accession number AF243039).

Amplification of a 1.3-kb LTR-gag Sequence. In Fig. 2A, the position of the primers is indicated, and in Fig. 2B, the product of nested PCR with primers LTR5 and GAG4, followed by PCR with primers LTR5 and GAG3, is shown as well as Southern blot hybridization with probe LTR3. The 1.3-kb sequence from the two tumors showed 94% homology to MMTV in 1.3 kb and 66 to 67% homology to HERV-K10 in 179 bp (Fig. 2A; GenBank Accession number AF248271).

Amplification of a 2.2-kb 5’ gag-pol Sequence. In Fig. 3A, the position of the primers is indicated, and in Fig. 3B, the product of nested XL-PCR with primers GAG4R and EPR, followed by PCR with primers GAG4R and POL5R, is shown as well as Southern blot hybridization with probe POL4. The 2.2-kb sequence from the two tumors was 94 to 95% homologous to MMTV over its entire length but only 59% homologous to HERV-K10 in 118 bp (Fig. 3A). The 2.2-kb sequence contains the ORFs for the dUTPase and the protease (prt) genes (GenBank accession number AF248272).

Amplification of a 3.09-kb gag-pol Sequence. In Fig. 4A, the position of the primers is indicated, and in Fig. 4B, the product of nested XL-PCR with primers GAG4R and EPR, followed by PCR with primers GAG4R and POL5R, is shown as well as Southern blot hybridization with probe POL4. The 3.09-kb sequence from the two tumors was 94 to 95% homologous to MMTV over its entire length but only 59% homologous to HERV-K10 in 118 bp (Fig. 4A). The 3.09-kb sequence contains the ORFs for the dUTPase and the protease (prt) genes (GenBank accession number AF248272).
with primers GAG7 and POL1, is shown as well as Southern blot hybridization with POL5. The 3.09-kb sequence from the two tumors was 95% homologous to MMTV in 3.02 kb and 69 to 70% homologous to HERV-K10 in 160 bp (Fig. 4A; GenBank accession number AF248269).

**Amplification of a 0.8-kb pol-env Sequence.** In Fig. 5A, the position of the primers is indicated, and in Fig. 5B, the product of the XL-PCR with GAG4R and EPR, followed by PCR with primers POL3’R and 5LR, is shown as well as Southern blot hybridization with probe ENV1. The 0.8-kb sequence from the two tumors was 94 and 95% homologous to MMTV in 745 bp and 60% to HERV-K10 in 25 bp (Fig. 5A; GenBank Accession number AF248270).

The whole proviral structure is shown in Fig. 6. Total homology from the two carcinomas was 95% to MMTV in 9.9 kb and 57% to HERV-K10 in 3.5 kb. Homology to the endogenous retrovirus was seen primarily in the pol gene, which is known to be conserved among different retroviruses and partially in the gag gene. Similarity between the two tumors was 96%, indicating that they were closely related but not identical. The differences were randomly distributed across the provirus genome.
To learn more about the genomic localization of the proviral sequences, FISH analysis was performed using the 2.7-kb env LTR sequence as probe in EK-2 cells, a primary cell line derived from the pleural fluid of a patient with env-positive breast cancer. These cells have been karyotyped and shown to be human. They contain env, LTR, gag and pol sequences, as well as env-LTR and gag-pol junctional sequences, as detected by amplification with PCR. An average of 3.5 signals per cell were detected in 80% of the 100 interphase EK-2 cells counted, whereas only 2 cells of the 100 normal breast MTSV1–7 cells counted and 2 of the 100 MCF-10 normal breast cells counted showed signals. Between 1 and 14 signals were observed in different chromosomes in nine metaphases in EK-2 cells, but the labeled chromosomes could not be identified with certainty. Fig. 7 shows one metaphase with hybridization signals produced by the 2.7-kb env-LTR sequence on four chromosomes.

DISCUSSION

The results reported here indicate that the whole proviral structure, homologous to MMTV, is present in the genome of two breast cancers that were positive for the env gene sequence. Although primary tumors are extremely heterogeneous, the preponderance of evidence from positive results in cell lines and negative results in other tissues (5) supports the concept that these sequences are related to the mammary cancer cells.

The strategy followed for detecting the proviral sequences was the use of primers with low homology to known genes and viral endogenous sequences, followed by amplification of sequences of contiguous genes using the putative primers from the known sequences (chromosome walking).

Analysis of the proviral sequences demonstrates that they have potential for expression and replication. The LTRs contain all of the enhancer and promoter elements characteristic of a replication competent MMTV-like virus as well as the glucocorticoid responsive element and the superantigen sequences. The other genes did not show evidences of deletions or early stop codons as do replication-incompetent endogenous retroviruses (13). The gag gene contains the dUT-Pase and prt ORFs (Fig. 6).

The putative HMTV overall homology to MMTV is 95% in both isolates. However, when 14 cloned MMTV-like env human sequences were compared with MMTV strains, the homology varied from 84 to 99%, suggesting that they are homologous to but not identical with MMTV (14). Furthermore, sequencing of 10 LTR human isolates revealed that there are certain specific deletions and insertions at the COOH-terminal of the superantigen sequences that clearly differentiate them from the MMTV strains. Comparison of the two isolates with the HERV-K10, the endogenous retrovirus closely related to MMTV (4, 15), revealed 57% homology in 3.5 kb of the pol and prt genes. Endogenous retroviruses contain deletions and early stop codons in different genes that make them replication incompetent. This is not the case of the two HMTV isolates.

The HERV-K family, which accounts for 1% of the human genome, is expressed to different degrees among individuals and cell types (16). HERV-K10 is expressed in some teratocarcinoma cell lines (13). One of these defective proviral elements, the HML6 BC1, was found to be expressed in only 1 breast carcinoma of 60 examined (17). Yin et al. (17) concluded that the role of their element in breast cancer remains unclear.

The origin of the proviral sequences that we have detected only in breast carcinomas is not yet established. Their absence from normal breast samples (2%) and from other tissues of the same patients whose tumors contain the sequences (5) suggests an exogenous origin. The possibility that they are endogenous and detectable only in tumors because of aneuploidy is improbable, because many aneuploid tumors analyzed were negative and many euploid tumors were positive (8). The likelihood of a hybrid virus, resulting from recombination between endogenous and exogenous viruses has been proposed (2), but...
analysis of the sequences described here does not support this postu-
late. The sequences represent a complete competent virus, closely
related to MMTV.

To ascertain if the sequences reported here play a role in breast
cancer pathogenesis, it is necessary to localize the sites of insertion
and to show evidence for viral expression and infectivity. MMTV
promotes cell growth through insertional mutagenesis, with activation
or inactivation of nearby genes that are related to growth control.
Thus, it is important to establish where these sequences are localized
in the human genome and whether they promote expression of nearby
genes. Among the multiple random insertion sites, only one may be of
importance.

With regard to viral expression, it has been shown by immunoflu-
orescence that envelope glycoprotein (gp52) expression occurred in
39% of the breast cancers studied (18). Viral particles have been
reported in 37% of breast cancer biopsies (19), in milk (20), and in
breast cancer cell lines in culture (15). Antibodies against gp52 have
been detected in the sera of 25% of American women with breast
cancer (21), indicating immunoreactivity. By analogy to other virus-
induced tumors, we can speculate that viral infection-promoting cell
growth could have taken place early, but by the time breast cancer
develops, only certain viral genes necessary to maintain the malignant
state are expressed. We have observed viral particles in env-gene-
positive specimens and are seeking specific gene activation patterns
in tumors.

The presence of an env gene sequence highly homologous to
MMTV in 38% of American women’s breast cancers (5), the expres-
sion of these sequences in breast cancers (7), the results from FISH
analysis indicating presence of the env-LTR sequence only in breast
cancer cells, and the identification of an entire proviral sequence in
two breast carcinomas taken together support the interpretation that a
substantial proportion of breast cancer is associated with retroviral
sequences that may represent an HMTV.

Recently, provocative epidemiological inference for the participa-
tion of MMTV in human breast cancer has been published (22).

REFERENCES

3. Westley, B., and May, F. E. B. The human genome contains multiple sequences of
4. Oso, M., Yasunaga, T., Miyata, T., and Ushikubo, H. Nucleotide sequence of human
endogenous retrovirus genome related to the mouse mammary tumor virus genome.
5. Wang, Y., Holland, J. F., Brieuweiss, I. K., Melana, S., Liu, X., Pelisson, I., Cantarella,
A., Stiellech, K., Mani, S., and Pogo, B. G-T. Detection of mammary tumor virus env
virus-like env gene sequences in human breast tumors and in a lymphoma of a breast
Ménard, S. Sequences homologous to the MMTV env gene in human breast carci-
9. Ménard, S., Casalini, P., Tomasic, G., Pilotti, S., Cascalini, N., Bufalino, R., Perrone,
F., Rilke, F., and Colnaghi, I. Pathobiologic identification of two distinct breast
carcinoma syndromes with diverging clinical behaviors. Breast Cancer Res. Treat.,
10. Berdichevsky, F., and Taylor-Papadimitriou, J. Morphological differentiation of hy-
brids of human mammary epithelial cell lines is dominant and correlates with the
R. S., Nagel, R. L., and Rybicki, A. C. The gene for human erythrocyte protein 4.2
sequence of a milk transmitted mouse mammary tumor virus: two frameshift sup-
pression events are required for translation of gag and pol. J. Virol., 61: 480–490,
1987.
13. Tonjes, R., Czudzerna, F., and Kurth, R. Genome-wide screening, cloning, chromo-
somal assignment and expression of full-length human endogenous retrovirus type K.
15. Seifarth, W., Baust, C., Murr, A., Skladny, H., Krieg-Schneider, F., Blusch, J.,
Werner, T., Hehlmann, R., and Leib-Mösch, C. Provirial structure, chromosomal
location and expression of HERV-K-T47D, a novel human endogenous retrovirus
16. Medstrand, P., Andersson, M-L, Borg, A., Olsson, H., and Blomberg, J. Transcrip-
tion of human endogenous retroviral sequences related to mouse mammary
tumor virus in human breast and placenta: similar pattern in most malignant and
17. Yin, H., Medstrand, P., Kristoffersen, A., Dietrich, U., Aman, P., and Blomberg,
J. Characterization of human MMTV-like (HML) elements similar to a sequence that
was highly expressed in a human breast cancer: further definition of the HML-6
18. Mesa-Tejada, R., Keydar, I., Ramanarayanan, M., Othono, T., Fenoglio, C., and
Spiegelman, S. Detection in human breast carcinomas of an antigen immunologically
19. Feller, W. F., and Chopra, H. C. A small virus-like particle observed in human breast
mouse mammary tumor virus in sera of breast cancer patients. Int. J. Cancer, 23:
incidence highest in the range of one species of house mouse, Mus domesticus.
Identification of a Proviral Structure in Human Breast Cancer

Bingren Liu, Yue Wang, Stella M. Melana, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/61/4/1754

Cited articles
This article cites 19 articles, 9 of which you can access for free at:
http://cancerres.aacrjournals.org/content/61/4/1754.full#ref-list-1

Citing articles
This article has been cited by 21 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/61/4/1754.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.