Isolation and sequencing of a cDNA coding for the human DF3 breast carcinoma-associated antigen

J. Siddiqui, M. Abe, D. Hayes, E. Shani, E. Yunis, and D. Kufe*

Laboratory of Clinical Pharmacology and Division of Immunogenetics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115

Communicated by Emil R. Unanue, December 9, 1987

ABSTRACT  The murine monoclonal antibody (mAb) DF3 reacts with a high molecular weight glycoprotein detectable in human breast carcinomas. DF3 antigen expression correlates with human breast tumor differentiation, and the detection of a cross-reactive species in human milk has suggested that this antigen might be useful as a marker of differentiated mammary epithelium. To further characterize DF3 antigen expression, we have isolated a cDNA clone from an Agt11 library by screening with mAb DF3. The results demonstrate that this 309-base-pair cDNA, designated pDF9.3, codes for the DF3 epitope. Southern blot analyses of EcoRI-digested DNAs from six human tumor cell lines with 32P-labeled pDF9.3 have revealed a restriction fragment length polymorphism. Variations in size of the alleles detected by pDF9.3 were also identified in Pst I, but not in HindIII, DNA digests. Furthermore, hybridization of 32P-labeled pDF9.3 with total cellular RNA from each of these cell lines demonstrated either one or two transcripts that varied from 4.1 to 7.1 kilobases in size. The presence of differently sized transcripts detected by pDF9.3 was also found to correspond with the polymorphic expression of DF3 glycoproteins. Nucleotide sequence analysis of pDF9.3 has revealed a highly conserved (G+C)-rich 60-base-pair tandem repeat. These findings suggest that the variation in size of alleles coding for the polymorphic DF3 glycoprotein may represent different numbers of repeats.

A human breast carcinoma-associated antigen has been identified by using a murine monoclonal antibody (mAb), designated DF3. mAb DF3 was prepared against a membrane-enriched fraction of a human breast carcinoma metastatic to liver (1). DF3 antigen is expressed on the apical borders of secretory mammary epithelial cells and in the cytosol of less differentiated malignant cells (1). DF3 antigen expression also correlates with the degree of breast tumor differentiation and estrogen receptor status (2). These findings and the detection of DF3 antigen in human milk (3) have suggested that mAb DF3 reacts with a differentiation antigen expressed in breast carcinoma cells.

The DF3 antigen has been characterized as a high molecular weight mucin-like glycoprotein (4, 5). DF3 antigen in human MCF-7 breast carcinoma cells consists of two distinct glycoproteins with molecular weights of 330,000 and 450,000 (4, 5). Moreover, previous studies have demonstrated that DF3 antigen circulates at elevated levels in the plasma of patients with breast cancer (6). The circulating mAb DF3-reactive antigens also have molecular weights ranging from approximately 300,000 to 450,000. However, the electrophoretic mobility patterns for circulating DF3 antigen differ among individuals (6). Subsequent studies in family members have demonstrated that the electrophoretic heterogeneity of plasma DF3 antigen is determined by codominant expression of multiple alleles at a single locus (7).

The present studies were performed to further examine genetic mechanisms responsible for the heterogeneity of DF3 antigen expression. We describe the isolation of a partial cDNA clone,1 designated pDF9.3, coding for the DF3 antigen. This cDNA clone encompasses 309 nucleotides of DF3 mRNA and has tandemly repeated sequences. By using this clone, we also demonstrate that the heterogeneity of DF3 antigen is related to size variations in DF3 alleles and DF3 transcripts.

MATERIALS AND METHODS

Library Screening. An oligo(dT)-primed cDNA library prepared from human MCF-7 breast carcinoma cells in Agt11 was kindly provided by P. Chambon (Institut de Chimie Biologique, Strasbourg, France) (8). Immunologic screening of the Agt11 library was performed as described (9) by using affinity-purified mAb DF3 (0.25 μg/ml) and antimouse IgG conjugated with alkaline phosphatase (Promega Biotec, Madison, WI). Positive plaques were isolated and the phage was further purified to homogeneity by repeated antibody screening. DNA was isolated from mAb DF3-positive recombinant phage, digested with EcoRI, and electrophoresed in 1.2% agarose gels containing ethidium bromide to determine the size of the insert.

Analysis of Lysogens for Fusion Protein. Lysogenization of Escherichia coli Y1089 with phage and induction of fusion protein with isopropyl β-D-thiogalactoside (IPTG) were performed as described (9, 10). The lysate of IPTG-induced lysogen was subjected to electrophoresis in NaDodSO4/7.5% polyacrylamide gels (11) and transferred to nitrocellulose filters for immunoscreening (12).

Southern and RNA Transfer Blot Analyses. The human breast carcinoma cell lines (BT-20, T47D, MCF-7, ZR-75-1), an ovarian carcinoma cell line (OV-D), and the HL-60 promyelocytic leukemia cell line were maintained in exponential phase (13–15). High molecular weight DNA and total cellular RNA were isolated by the guanidine isothiocyanate/cesium chloride method (16). The DNA was digested with EcoRI, Pst I, or HindIII. The DNA fragments were separated by electrophoresis in 0.6% agarose gels and then transferred to nylon membranes. The prehybridization and hybridization conditions were as described in the Zeta-Probe manual (Bio-Rad). The purified RNA (20 μg) was analyzed by electrophoresis in 1.5% agarose/formaldehyde gels; this was followed by transfer to nitrocellulose paper. The hybridization conditions were as described (15). The pDF9.3 cDNA probe was labeled with [32P]dCTP (Amersham) by the random primer method (17) to a specific activity of ~1010 cpm/μg of DNA.

Abbreviation: mAb, monoclonal antibody(ies).
*To whom reprint requests should be addressed.
1This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03651).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
Immunoblot Analysis. Cells were suspended in phosphate-buffered saline (0.15 M NaCl/0.01 M sodium phosphate, pH 7.4 (PBS)), 0.2 mM phenylmethylsulfonyl fluoride, and aprotinin (0.015 trypsin inhibitor unit/ml). The suspensions were sonicated and protein concentration was determined by the Bio-Rad protein assay. The protein samples (100 μg) were analyzed by electrophoresis in NaDodSO4/7.5%-gradient polyacrylamide gels and transferred to nitrocellulose paper (18). The nitrocellulose filters were washed with 5% bovine serum albumin in PBS for 1 hr at room temperature and incubated with mAb DF3 (0.25 μg/ml) for 2 hr, rabbit anti-mouse immunoglobulin for 1 hr, and then 125I-labeled protein A for 2 hr. The filters were washed five times, dried, and exposed to x-ray film.

Nucleotide Sequence Analysis. The 309-base-pair (bp) pDF9.3 cDNA insert was subcloned into the EcoRI site of E. coli phage M13mp8 and M13mp9. The DNA sequence was determined by sequencing both strands by means of the dideoxy chain-termination method (19) using Klenow fragment DNA polymerase I (New England Biolabs) and [α-35S]dCTP (Amersham).

RESULTS

Isolation and Characterization of cDNA Clones Coding for DF3 Antigen. mAb DF3 was used to screen the Agt11 library prepared from MCF-7 cells. Screening of 800,000 plaques yielded three positive clones that were further purified by repeated antibody screenings. Physical mapping showed that each of these recombinant clones contained inserts of similar size and that they had similar restriction maps (data not shown). One clone, designated pDF9.3, was characterized further. A β-galactosidase fusion protein was prepared by infecting E. coli Y1089 with pDF9.3 and then analyzed by immunoblotting. The Agt11 lysogen produced a protein corresponding in molecular weight and antigenicity to β-galactosidase (Fig. 1, lane 1). mAb DF3 was unreactive with β-galactosidase and other antigens present in the bacterial lysate (data not shown). In contrast, the recombinant pDF9.3 lysogen produced a fusion protein with an estimated molecular weight of 126,000 that reacted with mAb DF3 (Fig. 1, lane 2) and the anti-β-galactosidase antibody (data not shown).

Competition assays were also performed to further confirm that the epitope expressed by pDF9.3 shares homology with that identified by mAb DF3 on the DF3 glycoprotein. Thus, mAb DF3 was preincubated with purified DF3 antigen (13) before immunoblot analysis of the pDF9.3 fusion protein. Preincubation of mAb DF3 with increasing amounts of purified DF3 antigen progressively inhibited reactivity of the antibody with the fusion protein (Fig. 1, lanes 3–5). This finding indicates that the epitope on the fusion protein originates from the same reading frame that codes for the DF3 epitope.

Southern Blot Analysis of Genomic DNA. Identification of the cDNA was further studied by Southern blot hybridizations using 32P-labeled pDF9.3 prepared by subcloning the 309-bp insert into the EcoRI site of pUC8. Southern blot analyses of genomic DNAs from the human tumor cell lines digested with EcoRI, Pst I, and HindIII are shown in Fig. 2. Hybridization of the 309-bp cDNA with the EcoRI and Pst I DNA digests revealed restriction fragment length polymorphisms. The EcoRI digest yielded two fragments ranging from 7 to 12 kilobases (kb) in size for DNAs from each of the cell lines except BT-20. Similar findings were obtained with the Pst I fragments, which ranged in size from 3.5 to 6 kb. The single EcoRI and Pst I restriction fragments obtained with BT-20 DNA indicate the presence of two alleles of identical size or only a single allele. In contrast to these results, digestion of each of the DNA preparations with HindIII revealed only a single fragment of 23 kb. This finding corresponds to the absence of a HindIII restriction site in the alleles identified by pDF9.3.

RNA Transfer and Immunoblot Analyses of DF3 Expression. Total cellular RNA was prepared from each of the human tumor cell lines and monitored by RNA transfer blot analysis for transcripts that hybridized to the pDF9.3 probe. A single 4.7-kb mRNA was detectable in BT-20 cells (Fig. 3A). In contrast, cell lines derived from the other breast and ovarian carcinomas expressed two transcripts that ranged in size from approximately 4.1 to 7.1 kb (Fig. 3A). Furthermore, no hybridization was detectable with RNA from HL-60 cells (Fig. 3A).

These findings by RNA transfer blot analysis were compared to those obtained by immunoblotting with mAb DF3 and extracts prepared from each of the cell lines. The results indicate concordance in patterns of expression at the RNA level.
glycoproteins with molecular weights ranging from 300,000 to >450,000 (4, 5). DF3 antigenicity was found to be sensitive to neuraminidase and proteases (4, 5). These results suggested that sialyl oligosaccharides on a peptide backbone are required for DF3 antigenicity. In the present study, mAb DF3-positive plaques were isolated by using a λgt11 cDNA library prepared from human MCF-7 breast carcinoma cells. The MCF-7 cells have been shown to express DF3 antigen (13). One of the positive λ clones (pDF9.3) was further purified and found to produce a β-galactosidase fusion protein that specifically reacted with mAb DF3. The reactivity of mAb DF3 with plaques from this expression library and the fusion protein suggests that this antibody reacts with the core protein of DF3 antigen. However, DF3 antigenicity has also been shown to be sensitive to neuraminidase (4, 5). Thus, mAb DF3 binding to the protein may be enhanced by the presence of glycosidic linkages.

Although patients with breast cancer and certain other carcinomas have higher levels of circulating DF3 antigen, the electrophoretic mobilities of the mAb DF3-reactive species are similar to those in normal subjects (6, 20). Indeed, more recent results have indicated that the variation in electrophoretic mobility of circulating DF3 antigen among family members is related to a genetically determined polymorphism (7). The present findings support this genetic polymorphism. Thus, considerable fragment size variation was observed after hybridization of the pDF9.3 probe to EcoRI and Pst I restriction digests of DNA from different cell lines. The EcoRI restriction fragments varied from 7 to 12 kb in size, and the different cells had only one or two bands. Furthermore, the Pst I fragments varied from 3.5 to 6 kb and each DNA preparation similarly yielded one or two bands. In contrast, hybridization of pDF9.3 probe to HindIII DNA digests revealed only one 23-kb band and indicated that this restriction enzyme has digestion sites outside the region identified by this probe.

The variation in allele size identified with pDF9.3 corresponded with the presence of differently sized transcripts. Thus, cells with two restriction fragments in the EcoRI or Pst I DNA digests had two differently sized mRNAs. In contrast, BT-20 cells had only one detectable restriction fragment in these DNA digests and expressed only one transcript. This relationship also extended to the variation in electrophoretic mobilities of DF3 antigen. BT-20 cells expressed a single mAb DF3-reactive species, whereas the other epithelial tumor cells expressed two DF3 antigens. Moreover, HL-60 cells had no detectable transcripts and no detectable DF3 antigen. Taken together, these findings support our previous findings that the heterogeneity of DF3 antigen production is controlled by multiple alleles at a single locus expressed in an autosomal codominant fashion (7).

The nucleotide sequence analysis of pDF9.3 provides a possible explanation for the variability in restriction fragment size and the polymorphic patterns of DF3 expression. In this regard, we have identified a 309-bp cDNA clone that consists of multiple tandem repeats. These repeats are (G+C)-rich and encompass 60 bp. Variation in the size of the DF3 alleles could thus be due to differences in the

DISCUSSION

We have previously demonstrated that DF3 antigen in human breast tumors and milk is comprised of mucin-like

```
CGACCGGGCTG GGGGCGGCGG GTGAGGCGGG GGGGCGGCGG CTCGGGGGCC CGAGGGCGA  60
  CGTG........CG........CGGTGGGGCGG........CGGCGGCGG  120
  ..G.....G........G...G........G  180
  ........................................G  240
  ........................................G  300
```

![FIG. 4. Nucleotide sequence of the pDF9.3 cDNA insert.](image-url)
number of these repeats and occur as a result of unequal crossing-over events. The presence of closely related repeats may also explain the finding that mAb DF3 binds to two or more epitopes on the same DF3 molecule (6). The total number of these repeats in the full-length cDNA, however, requires further investigation.

Similar variable tandem repeats have been reported for other genes, including those coding for carcinoembryonic antigen (21), insulin (22), α- and β-globulin (23, 24), Epstein–Barr virus (25), c-Ha-ras (26), and a hypervariable minisatellite family (27). Furthermore, the human complement receptor (CR1) gene consists of homologous repeats ~1.6 kb in size (28). Allelic variants of CR1 differ by 1.6 kb and also correlate with variations in size of the CR1 transcripts and products (28). The lengths of most internal repeats, however, range between 120 and 300 bp (29). Moreover, homology of the internal repeats for many vertebrate proteins ranges between only 20% and 50% (29). In contrast, the internal repeats identified in the present study exhibit a particularly high degree of homology. This finding could suggest that the DF3 gene evolved more recently by duplication of a primordial gene or by exon shuffling.

We have recently demonstrated that the DF3 antigen is a member of a family of related, but not identical, high molecular weight tumor-associated antigens (13). The DF3 glycoprotein is also detected by another mAb, designated Ca1, that reacts with a wide range of human tumors (30, 31). The binding site for mAb Ca1 on the DF3 glycoprotein, however, is distinct from that defined by mAb DF3 (13). Recent studies have described a genetic polymorphism of the Ca1 antigen in human urine (32). Moreover, the locus coding for these urinary mucins (PUMs, from peanut lectin binding urinary mucins) is a hypervariable minisatellite of human DNA (33). After completion of the present work, a report appeared that describes the isolation of a partial cDNA clone (pMUC10) that codes in part for PUM (33). The EcoRI restriction fragments of genomic DNA identified by the pMUC10 probe are similar to those found in the present study (34). However, the nucleotide sequence of the pMUC10 cDNA has not been published and therefore it is not possible to directly compare that clone with the pDF9.3 cDNA isolated in the present study.

This investigation was supported by Public Health Service Grants CA-38869 and CA-20531 awarded by the National Institutes of Health, Department of Health and Human Services, and by a Burroughs Wellcome Clinical Pharmacology Scholar Award (D.K.).