Molecular Cloning and Characterization of a Human Adenocarcinoma/Epithelial Cell Surface Antigen Complementary DNA


Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana 46285

ABSTRACT

A human adenocarcinoma-associated antigen (KSA) defined by the monoclonal antibody KS1/4 has become the focus of several site-directed strategies for tumor therapy. KSA, a 40,000 Da cell surface glycoprotein antigen, is found at a high density in all adenocarcinomas examined to date and in corresponding normal epithelial tissues. Here we describe the cloning and sequencing of overlapping complementary DNA clones which encode the entire KSA as expressed in UCLA-P3, a human lung adenocarcinoma cell line. We have deduced the 314-amino acid sequence and have compared it to the N-terminal amino acid sequence data of the affinity-purified antigen. The KSA is synthesized as a 314-residue-long preproprotein that is then processed to a 232-residue-long antigen. KSA appears to have a single transmembrane domain of 23 residues that separates the highly charged 26-residue cytoplasmic domain from the extracellular domain. The N-terminal region of the propeptide is rich in cysteines and contains three potential N-glycosylation sites. Computer-assisted analyses at both the DNA and protein levels have found no significant similarities of this protein to known sequences, but a GC-rich 5' terminus is evident. Northern blot analysis shows that transcription of KSA can be detected in RNA isolated from normal colon but not in RNA isolated from normal lung, prostate, or liver.

INTRODUCTION

The development of hybridoma technology has provided numerous monoclonal antibody probes which appear to define tumor-associated gene products. MoAbs to human tumor-associated antigens are currently being widely studied as possible diagnostic and therapeutic reagents (1–8). To date, however, relatively little information is available on the molecular nature of the targeted antigens. However, in some cases, structural information of human tumor-associated antigens has been deduced from cDNA analysis (9–11). Such structural information could help the basic understanding of tumor cell biology and may aid in the design of novel antigen directed monoclonal antibodies.

Adenocarcinomas of the lung, colon, prostate, and breast encompass some of the major forms of solid tumors. With this in mind, the adenocarcinoma-associated antigen recognized by KS1/4 (12) has been under investigation as a target for site-directed therapy approaches with covalent monoclonal antibody oncolytic drug conjugates (13–16). To date this antigen has been found on all adenocarcinoma cells tested and in various corresponding normal epithelial cells. In this study, we report on the molecular cloning of this unique epithelial malignancy/epithelial tissue glycoprotein from the human lung adenocarcinoma cell line, UCLA-P3, and discuss initial studies on its expression.

MATERIALS AND METHODS

Construction and Screening of a λgt11 cDNA Library. Total RNA was prepared from UCLA-P3 cells grown in culture. The RNA was extracted with guanidinium thiocyanate (17) and polyadenylated RNA was isolated using oligo(dT) deoxythymidylate cellulose column chromatography. Double stranded cDNA was synthesized using the RNase H procedure (18) and size selected on an agarose gel. DNA over 700 base pairs long was methylated and ligated to synthetic EcoRI linkers and then inserted into EcoRI-digested λgt11 DNA.

Antibody and Radiolabeled Probe Screening. The λgt11 library was screened at approximately 60,000 plaques/100-mm plate by the method of Young and Davis (19). KS1/4 monoclonal antibody or a primary rabbit polyclonal antibody to KSA was used for initial screening. The secondary biotinylated horse antibody to mouse IgG or a goat antibody to rabbit IgG and avidin-conjugated horseradish peroxidase was purchased from Vector Laboratories. The positive plaques were cycled through four rounds of screening. Plaque hybridization with [32P]RNA probes generated from an SP6 vector were carried out essentially as described by Melton et al. (20).

Blot Hybridization. The RNAs were electrophoresed through a 1.5% agarose/6% formaldehyde gel, stained with acridine orange, and transferred to a Genescreen nylon membrane. RNAs were prehybridized according to the Genescreen manual in 10 ml of 50% formamide, 10 × Denhardt's, 0.05 M Tris (pH 7.5), 1 M NaCl, 0.1% sodium pyrophosphate, 0.1% SDS, 10% dextran sulfate, and 100 µg/ml Escherichia coli DNA for 6 h at 42°C with agitation. Nick-translated StuI-BglII cDNA fragment of Ag1338 (3 × 106 cpm/ml; specific activity, 3.9 × 109 cpm/µg) was then added and the RNAs were hybridized overnight at 42°C with agitation. Washes were carried out with constant agitation as follows: 2 times for 5 min each in 2 × SSC at room temperature; 2 times for 30 min each in 2 × SSC and 0.5% SDS at 65°C; and 1 time for 5 min in 0.1% SSC at room temperature.

Recombinant Antigen Synthesis and Detection. Cell pellets of Asso = 1 were prepared from cultures grown at 32°C for uninduced pl and at 42°C for induced pl. Lysates were prepared by incubation on ice for 1 h in lysis mix [1 mM EDTA (pH 8.0), 6 µg/ml DNase, and 170 µg/ml lysozyme] and subsequently carried through three cycles of freeze/thaw. Cell debris was removed by centrifugation, lysates were dot blotted onto nitrocellulose, and the blot was probed with a rabbit polyclonal antibody raised against purified KSA (21).

RESULTS

Isolation and Characterization of Antigen cDNA Clones. KSA cDNAs were isolated from a UCLA-P3 λgt11 library of approximately 291,000 individually derived clones. No positive plaques were obtained when this library was screened with the KS1/4 MoAb, but rescreening with a rabbit polyclonal antibody raised against affinity-purified antigen (21) did yield a single positive plaque which was plaque purified and found to contain a 249-base pair insert (Ag1). Ag1 was radiolabeled and used to rescreen the library for a larger insert clone. Five positive plaques were found, the largest of which was 1338 base pairs long (Ag1338). The 750-base pair 5' region of Ag1338 was then radiolabeled and used to probe a RNA blot (22, 23) of UCLA-P3 and normal human liver mRNAs. The 750 base pair probe hybridized to a single UCLA-P3 mRNA species of approximately 1700 bases but did not hybridize to normal human...
liver mRNA (data not shown). It, therefore, appeared unlikely that Ag1338 represented a full length mRNA sequence. This same probe was then used to screen the library a third time and detected a positive plaque containing a 933-base pair insert (Ag933) that overlapped Ag1338.

Fig. 1 shows the complete nucleotide sequence of the initial clone, Ag1, and the two overlapping clones, Ag1338 and Ag 933. Together these clones code for an antigen sequence of 1516 base pairs. If one assumes a typical mRNA to have a 200- to 250-nucleotide polyadenylated tail (24), this composite sequence should closely represent full length KSA mRNA. The sequence contains a 405-nucleotide 3' noncoding region, excluding the polyadenylated tail, which contains the AATAAA polyadenylation signal (25) (1486 to 1491) 13 nucleotides upstream from the polyadenylated tract. The 5'-untranslated region is GC rich similar to the transcription initiation regions of many genes (26). This region contains an untranslated sequence that is 78.9% GC and a proposed signal peptide sequence that is 71.4% GC. The remainder of the coding region is 41.5% GC (Fig. 2).

There is a single long open reading frame that codes for a deduced peptide of 314 residues. Only one AUG translation initiation codon is present at the 5' end of the cDNA. This methionine codon is in frame with the remainder of the peptide and lies in a nucleotide sequence that closely corresponds to a consensus translation initiation site (27). It is followed by a hydrophobic signal-like sequence. Inspection of the deduced amino acid sequence suggests a number of potential sites for cleavage of the signal peptide (28); the most likely site is after residue 23 (alanine).

The 23-residue hydrophobic carboxy-terminal domain (residues 266 through 288) represents a potential transmembrane region that separates a highly charged 26-residue cytoplasmic region that separates a highly charged 26-residue cytoplasmic (residues 266 through 288) represents a potential transmembrane domain (21).

The extracellular domain contains a region moderately rich in cysteine residues (12 of 109). The number of disulfide bonds is not known although some must be present since the reduced antigen has a molecular mass of 36 kDa by SDS-polyacrylamide gel electrophoresis, in contrast to the 40–42-kDa mass of the unreduced form. This apparent disulfide-dependent conformation must also provide the epitope recognized by MoAb KS1/4, since the reduced molecule is unreactive with this antibody. The extracellular domain also contains three possible sites for asparagine-linked glycosylation.

KSA appears to be synthesized as a preproantigen peptide. N-terminal sequence analysis of affinity purified KSA, prepared from UCLA-P3 cells grown either as a nude mouse xenograft tumor or in tissue culture, starts with residue 82 (alanine) of the cDNA-deduced peptide sequence (21).

Expression of the KSA. In an attempt to show that the cDNA codes for KSA as recognized by the KS1/4 MoAb, an E. coli expression vector was constructed that contains the predicted mature KSA sequence (residues 82 through 265). This protein was synthesized under the control of the λP promoter and immunologically detected as dot blots using the previously mentioned rabbit polyclonal antibody. As seen in Fig. 3, the polyclonal antibody recognizes only the KSA construction in which the pL promoter was induced. It does not recognize those constructions in which the pL promoter was uninduced or those containing an unrelated gene in an identical vector that was induced. The KS1/4 MoAb failed to recognize the E. coli-synthesized antigen (data not shown) which probably represents an inability of E. coli to properly fold and/or posttranslationally modify KSA.

Blot Hybridization of RNA from Different Tissues. The KSA cDNA was also used to survey various tissue for the presence of this antigen. A 1112-nucleotide SstII to BglII fragment which encompasses most of the coding region of KSA was nick-translated and used to probe an RNA blot of normal human colon, lung, and prostate; CaCO2 (a human colon carcinoma); and UCLA-P3 RNAs. As shown in Fig. 4, a positive signal was

---

Fig. 1. A, sequencing strategy for cDNA clones. Arrows, direction and extent of sequencing. Restriction enzyme cleavage sites refer to fragments that were subcloned into M13 vectors for dideoxy sequencing and subsequent manipulation. B, composite nucleotide sequence obtained from Ag1338 and Ag933 overlapping clones and the deduced amino acid sequence. Cysteinyl residues are boxed and potential N-glycosylation sites are overlined. Arrow, cleavage site that gives rise to mature KSA; heavy line, potential transmembrane region; dashed line, polyadenylation signal sequence.
ADENOCARCINOMA/EPITHELIAL CELL SURFACE ANTIGEN cDNA

Fig. 2. Analysis of KSA cDNA and protein structure. A, percentage of GC content averaged over 100 nucleotides along the 1516-base pair DNA. B, Kyte-Doolittle hydrophobicity averaged over nine amino acids along the predicted 314-amino acid protein sequence (38).

Fig. 3. Dot blot showing antibody recognition of E. coli-synthesized KSA (14). Each column represents serial 10-fold dilutions from top to bottom. A, lysate of uninduced pl. expression plasmid coding for a protein unrelated to KSA; B, induced pl. expression of A; C, uninduced pl. expression of KSA; D, induced pl. expression of KSA.

detected with normal human colon and CaCO2 but not with lung or prostate RNAs. These results are consistent with immunoperoxidase studies (29) that have shown the KS1/4 antigen to be highly expressed in normal colon epithelial cells but expressed only at low levels in lung or prostate which contain only a fraction of epithelial cells.

DISCUSSION

We present here the molecular cloning of KSA from the human adenocarcinoma cell line UCLA-P3. This work represents one of only several reports characterizing monoclonal antibody-defined cell surface antigens and the first epithelial malignancy integral membrane glycoprotein. KSA shows structural features that are common to a number of cell membrane proteins such as a cysteine-rich region, N-glycosylation sites, a hydrophobic transmembrane domain, and a highly charged cytoplasmic anchorage domain (30-32). Computer-assisted analyses at both the nucleotide and amino acid levels found no significant homology of KSA with known sequences. Thus, the KSA sequence provides one of the first detailed descriptions of the primary structure of a membrane-bound tumor-associated antigen.

What emerges from this study is that KSA is a cell surface protein that is synthesized as a 314-residue preproprotein that is processed to a 232-residue antigen. The alanine 82-processing site is proceeded by two arginine residues which could serve as a trypsin-like proteolytic cleavage signal. At this time there is no biochemical information on KSA processing in normal cells or other adenocarcinoma-derived cell lines. Processing removes 7 of the 12 cysteine residues as well as one of the three potential N-glycosylation sites found in the proantigen sequence. The mature 26,340-Da antigen peptide appears to be highly glycosylated since based on polyacrylamide gel electrophoresis the molecular mass is approximately 40,000 Da (12). Presumably both the 183-residue extracellular domain and the sugar moiety contain potential antibody-binding epitopes.

Fig. 4. Northern blot hybridization (15) of KSA cDNA to 10 µg of total RNA isolated from CaCO2 cells (Lane 1), normal human colon (Lane 2), prostate (Lane 3), lung (Lane 4), and UCLA-P3 cells (Lane 5). Autoradiograph shown is from a 3.5-h exposure. RNA marker sizes are shown at right in kilobases.

The proposed cytoplasmic domain has nine positively charged and five negatively charged residues giving it a net positive charge of 4 which is similar to the net positive charge of 6 found in the abbreviated cytoplasmic domain of the 55,000-Da interleukin 2 receptor protein (33). There is a single serine residue in the cytoplasmic domain that may serve as a phosphorylation site. None of the above-described characteristics, however, has given us any insight into the biological role of this protein. In our initial tissue survey it appears that only colon, a tissue rich in epithelial cells, has detectable levels of KSA RNA transcription. In situ hybridization techniques will be attempted to give a more accurate picture of transcriptionally active tissue.

With the cloning of KSA, structural and functional studies of this glycoprotein are now possible. The cloning of KSA will also facilitate the generation of MoAbs directed at epitopes other than that recognized by KS1/4 and facilitate the possibility of expressing sufficient quantities of KSA in mammalian cells for studies of the molecular interactions between KSA and the KS1/4 MoAb. In addition, the relationship of the KSA glycoprotein to other tumor associated antigens with similar molecular characteristics can now be explored (34-37). In this regard, preliminary evidence suggests that antibodies 17-1A (34) and AUAl (35) do react with the affinity-purified KSA from tumor xenografts which was proven to be homogenous by
N-terminal amino acid sequence. These data may unify both preclinical and clinical data of various antibodies based on the utility of KSA as a target for immunotherapy.

ACKNOWLEDGMENTS

We thank Stan Burgett for sequencing help, Dr. Craig A. Halverson for supplying tissue samples, and Dr. Bernie J. Abbott for his critical comments. We also thank Dr. Irving S. Johnson for his support and encouragement.

REFERENCES


Molecular Cloning and Characterization of a Human Adenocarcinoma/Epithelial Cell Surface Antigen Complementary DNA
