*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 10425–10429, November 1992
Immunology

# Cloning and sequencing of a trophoblast-endothelial-activated lymphocyte surface protein: cDNA sequence and genomic structure

**(leucine zipper/lymphocyte activation antigen/*myc*/nuclear protooncogenes/syncytiotrophoblast antigen)**

JOSEPH R. VOLAND, RICHARD J. WYZYKOWSKI, MARK HUANG, AND RICHARD W. DUTTON

Department of Biology and Cancer Center, University of California at San Diego, La Jolla, CA 92093-0063

*Communicated by Russell F. Doolittle, July 7, 1992 (received for review April 16, 1992)*

ABSTRACT **We have previously described the distribution of a surface glycoprotein recognized by monoclonal antibody B721. We now report the molecular characterization of this molecule at the protein, cDNA, and genomic level. A 75-kDa glycoprotein can be immunoprecipitated from B721$^+$ tissues. We have isolated a near full-length cDNA containing the complete coding region and a full-length genomic clone. We present evidence that this protein has similarity to several classes of nuclear transcription factors, particularly the myc family of proteins. The 721P protein was found to have a leucine zipper-like structure, a possible basic region immediately upstream from the leucine zipper, and a protein kinase A phosphorylation site. 721P protein is encoded by a gene distinct from any deposited in existing data bases, and displays several features associated with proteins involved in signal transduction and gene regulation.**

Many nuclear proteins involved in DNA binding and transcriptional regulation share common structural motifs. Some of these proteins possess a dimerization domain used for protein–protein interactions and an adjacent basic region necessary for DNA-binding activity (1–5); these proteins have been divided into two large classes on the basis of their dimerization domains and the nature of their basic DNA-binding regions. The first group of proteins, typified by c-fos, have been termed B-zip proteins because they have a basic region directly adjacent to a leucine zipper and they use the leucine zipper as a dimerization domain (1, 6). The second group of proteins has been termed helix–loop–helix proteins on the basis of two amphipathic $\alpha$-helices in their dimerization domain (3, 4). In all cases only a dimer with an intact basic region can bind to DNA (1, 3, 4). Although leucine zippers have been reported in several membrane-anchored proteins, these proteins do not have adjacent basic regions (7, 8).

In this report we describe the cloning and sequencing of a molecule, 721P, expressed on the plasma membrane of activated lymphocytes, most vascular endothelium, and syncytiotrophoblast (9, 10). Our results demonstrate a protein that, although expressed at the cell surface, has many structural motifs in common with nuclear transcription factors. The predicted amino acid sequence indicates a leucine zipper-like motif and an adjacent polar region. These regions resemble the leucine zipper and basic region of the myc family of proteins. We also present the intron/exon structure for the gene encoding this protein and show that the putative leucine zipper and the polar region are encoded by separate exons. These data show that 721P protein is encoded by a previously unreported gene* and that this protein is structurally similar to proteins involved in gene regulation.

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.

## MATERIALS AND METHODS

**Cell Lines and Monoclonal Antibody (mAb) Purification.** All cell lines were maintained as described (9). mAb B721 was purified from spent culture supernatant on either protein A or protein G columns (Pharmacia). The purified antibody was coupled to CNBr-activated Sepharose 4B (Pharmacia), according to the method supplied by the manufacturer.

**Isolation of Syncytiotrophoblast Microvilli and Endothelium.** Chorionic villi membranes were isolated according to the methods of Webb *et al.* (11), except that 6 M urea was used for 1 M ammonium isothiocyanate. Endothelium was isolated from umbilical veins by collagenase treatment according to the methods of Corkey *et al.* (12).

**Isolation and Activation of Peripheral Blood Lymphocytes.** Normal peripheral blood lymphocytes were obtained by venipuncture from healthy volunteers. They were purified on a Ficoll/Hypaque gradient and activated by phytohemagglutinin (Sigma) as described (10). All experiments were done with phytohemagglutinin-activated 5-day blasts.

**Iodination and Immunoprecipitation of 721P Protein.** All cell populations were labeled by surface iodination with the lactoperoxidase method (13). A minimum of $10^7$ cells were used per iodination. For trophoblast, membrane vesicles obtained from 10 g of placenta were used per iodination. Cells were lysed in RIPA buffer (0.15 mM NaCl/0.05 mM Tris·HCl, pH 7.2/1% Triton X-100/1% deoxycholate/0.1% SDS) (13), immunoprecipitated using a mAb B721-affinity column, and analyzed by SDS/PAGE and autoradiography.

For N-glycosidase F (PNGase F) digestion, the immunoprecipitated 721P protein from iodinated placental membranes was removed from the column by boiling in 0.1 M $Na_2PO_4$, pH 6.0/5 mM EDTA. Five units of PNGase F (Boehringer Mannheim) were added; the samples were incubated at 37°C for 24 hr and analyzed by SDS/PAGE and autoradiography.

**Generation of Polyclonal Antiserum to 721P Protein.** Because mAb B721 does not recognize denatured 721P protein, a polyclonal serum was generated. The mAb B721-affinity column was used to immunoprecipitate 721P protein. The protein was purified by preparative SDS/PAGE, cut from the gel, homogenized in phosphate-buffered saline, and used to immunize New Zealand White rabbits. Approximately 50 $\mu$g of protein was used per immunization. The animals were boosted three times at 1-mo intervals. Their serum contained antibodies that could precipitate denatured 721P protein and could recognize 721P protein in immunoblots (data not shown).

**Screening of Placental λgt11 Expression Library.** A placental cDNA library was screened according to the methods of Huynh *et al.* (14). The antiserum was used as a 1:100 dilution

Abbreviations: PNGase F, N-glycosidase F; mAb, monoclonal antibody.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M99578).

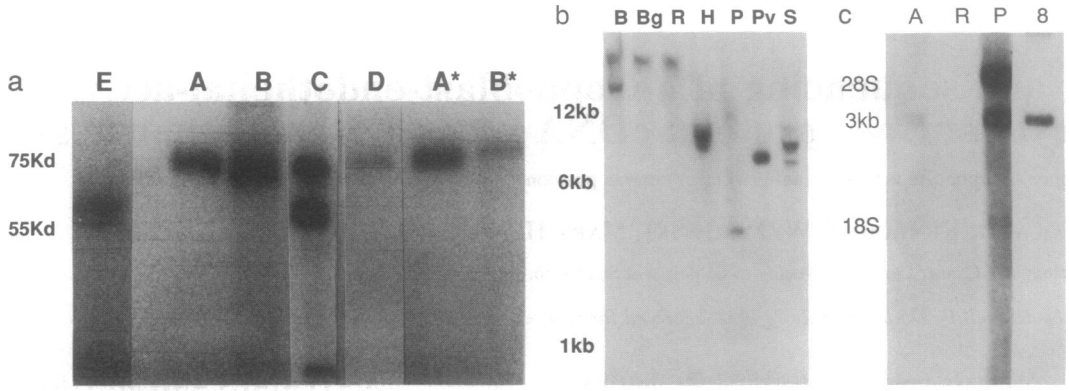


FIG. 1. (*a*) Immunoprecipitation of 721P protein. The molecule from all tissues is ≈75 kDa. Treatment with PGNase reduces the molecular size to 60 kDa. Lanes: A, placenta; B, phytohemagglutin-activated day 5 lymphocytes; C, endothelium; D, 8392; A*, placenta (nonreduced); B*, activated lymphocytes (nonreduced); E, PGNase treated. (*b*) Hybridization of Plac-3 to human genomic DNA. The single band seen with most restriction enzymes is consistent with a single-copy gene. B, *Bam*HI; Bg, *Bgl* II; R, *Eco*RI; H, *Hin*dIII; P, *Pst* I; Pv, *Pvu* II; S, *Sst* I. (*c*) Expression of 721P mRNA. Ten micrograms of poly(A) RNA was loaded per well. A 3-kb message was seen in activated, but not resting, lymphocytes, placenta, and the 8392 cell line. The weaker intensity of the activated lymphocyte band is consistent with the fact that only ≈40% of total activated peripheral blood lymphocytes are mAb B721 positive (10). Lanes: A, activated peripheral blood lymphocytes; R, resting peripheral blood lymphocytes; P, placenta; 8, 8392 cell line. The band at 28S in the placenta presumably represents residual ribosomal RNA; however, a second message cannot be ruled out.

of the crude serum with 20% fetal calf serum as a blocking agent. The antibody was visualized by immunohistochemical methods with a rabbit ABC horseradish peroxidase kit (Vector Laboratories) and 3-amino-9-ethylcarbazole as substrate as described (9).

**Characterization and Subcloning of Plac-3 Sequence.** Plac-3 sequence was subcloned into the *Eco*RI site of pGem3 (Promega), and this clone was used for all further experiments. Restriction mapping was done by standard protocols, and appropriate restriction fragments were subcloned into M13MP18 and M13MP19 vectors (15).

**RNA and Southern Blotting.** Genomic DNA was extracted from peripheral blood lymphocytes as described (16). Digested genomic DNA was resolved on 1% agarose gels, transferred to nylon membranes, and probed with full-length Plac-3 sequence, which was labeled by the random-primer method.

RNA was extracted from relevant tissues as described (17), poly(A) enriched over an oligo(dT)-cellulose column (Pharmacia), resolved on 1.25% agarose/formaldehyde gel, transferred to nylon membranes, and probed as above.

All blots were washed sequentially with 2× standard saline/citrate (SSC)/0.5% SDS at room temperature, 2× SSC/0.5% SDS at 55°C, and 0.1% SSC/0.5% SDS at 55°C. Each wash cycle was repeated twice for 30 min. Exposure was with Cronex Lightning Plus intensifying screens at −70°C.

**Transfection of Plac-3 Sequence and Flow Cytometry Analysis.** Plac-3 sequence was subcloned into the *Eco*RI site of pCK3 and transfected into COS cells by the lipofection method. Cells were cultured for an additional 48 hr. The cells were stained with fluorescein isothiocyanate-conjugated mAb B721 as described (10). Stained cells were analyzed on a Becton Dickinson FACscan.

**Sequencing of Plac-3.** All sequencing was done by the Sanger dideoxynucleotide chain-termination method (18) with Sequenase (United States Biochemical). Both M13 universal primer and oligonucleotide primers derived from internal cDNA sequences were used. 7-Deaza-GTP and ITP were used to solve compressions. Double-stranded sequencing was done on the original Plac-3 clone around the restriction sites used for subcloning to insure that multiple restriction sites were not present. Both strands were sequenced, and all subclones were sequenced at least three times.

**Isolation of Genomic Clone and Determination of Intron/Exon Junctions.** A full-length genomic clone coding for 721P protein was isolated and mapped by standard procedures (Stratagene). Intron/exon junctions were determined by double-stranded sequencing of the genomic clone. Primers for sequencing were generated from the sequence of the cDNA at intervals of 100 base pairs (bp). Intron/exon junctions were identified by divergence from the cDNA sequence combined with the appropriate consensus sequences for the 5′ and 3′ ends of introns. Approximate intron lengths were estimated from restriction-map distances.

## RESULTS

**Molecular Characterization of 721P Protein.** Immunoprecipitation studies demonstrated a 75-kDa protein from four tissue types (Fig. 1*a*). The additional bands seen in lane C (endothelium) were determined by immunoblotting to be immunoglobulin light and heavy chains (data not shown). Comparison of gels run under reducing and nonreducing conditions indicated that 721P was a single-chain polypeptide. PNGase F treatment reduced the molecular mass of 721P from 75 kDa to ≈60 kDa, indicating that 20% of the

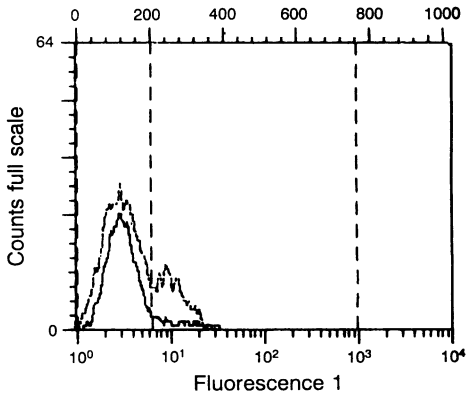


FIG. 2. Flow cytometric analysis of transfected COS cells. COS cells were transfected with Plac-3 and stained with fluorescein isothiocyanate-conjugated mAb B721 48 hr later. Nontransfected cells were used as control. The transfected cells showed ≈20% positive cells.

molecular mass of the glycoprotein was from N-linked carbohydrate.

**Cloning of cDNA Encoding 721P Protein.** Screening of $1 \times 10^6$ λgt11 plaques with polyclonal serum produced three positive clones; only one clone remained positive on subsequent screenings. This clone was designated Plac-3 and was further analyzed.

**Characterization of Plac-3 Clone.** The Plac-3 clone was found to be 2.5 kilobases (kb) in length. Plac-3 generally hybridized to a single band on a genomic blot (Fig. 1*b*), suggesting that it is encoded by a single-copy gene.

Hybridization of Plac-3 to RNA blots revealed a 3-kb band in placenta, the 8392 cell line, and activated but not resting peripheral blood lymphocytes (Fig. 1*c*). This result indicated that Plac-3 mRNA expression was consistent with the pattern of mAb B721 reactivity.

To determine whether Plac-3 actually encoded 721P protein, the cDNA was transfected into COS cells and assayed for transient expression of protein 48 hr later. Approximately 20% of cells transfected with the cDNA reacted with mAb B721 (Fig. 2), indicating that Plac-3 did, indeed, code for 721P.

**Sequencing of Plac-3 and Predicted Protein Sequence.** Plac-3 is 2626 bp in length and contains a long open reading frame of 1650 bp beginning with an ATG codon at nucleotide 172 (Fig. 3). There is an upstream, in-frame stop codon, suggesting that the identified start codon is the correct one. In addition, the start codon is surrounded by the consensus sequence for initiation codons (19). The 5′-untranslated sequence is 84% G+C.

The predicted protein sequence of 721P is 550 amino acids long, and the predicted pI is 10.38. There is a leader sequence and a consensus leader cleavage site at residue 21 (20). The protein has an estimated molecular mass of ≈61 kDa. There are two N-glycosylation sites toward the amino terminus of the protein, and a tentative transmembrane domain begins at residue 203. The tentative cytoplasmic domain contains a sequence at residue 505–508 (Arg-Arg-Arg-Ser), which is the consensus sequence for serine phosphorylation by protein

```
   GGCGACGGCGGTGGCGGCGTCGGAGGCGCCTCCGGGGGACGGTGGCGGCTCCCGGCGGTGAGGCCGCGCCTGTCCGGGG   79
   ATCGTCGAGGGACGGCGGGAGCTTGGGCCAGCGGCGGCGGCGGCCTGGGACGCAGGCGGAGCCCCGCGCAGGCCCAAGG  158
 1                 M  A  A  A  T  I  V  H  D  T  S  E  A  V  E  L  C
   TCCCGGAGGCT ATG GCA GCG GCT ACC ATC GTG CAC GAC ACG TCT GAG GCC GTG GAG CTC TGC  220
18  P  A  Y  G  L  Y  L  K  P  I  T  K  M  T  I  S  V  A  L  P
   CCT GCT TAC GGC TTG TAC CTG AAG CCC ATC ACC AAG ATG ACC ATC AGC GTG GCA CTC CCG  280
38  Q  L  K  Q  P  G  K  S  I  S  N  W  E  V  M  E  R  L  K  G
   CAG CTG AAG CAG CCG GGG AAG TCC ATC TCC AAC TGG GAG GTG ATG GAG AGG CTG AAG GGC  340
58  M  V  Q  N  H  Q  F  S  T  L  R  I  S  K  S  T  M  D  F  I
   ATG GTG CAG AAC CAC CAG TTC TCC ACG CTG CGT ATT TCC AAG AGC ACC ATG GAC TTC ATC  400
78  R  F  E  G  E  V  E  N  K  S  L  V  K  S  F  L  A  C  L  D
   CGC TTC GAG GGG GAG GTG GAG AAC AAG AGC CTG GTC AAG TCT TTT CTG GCC TGC CTG GAC  460
98  G  K  T  I  K  L  S  G  F  S  D  I  L  K  V  R  A  A  E  F
   GGC AAG ACC ATC AAG CTC AGC GGC TTC TCC GAC ATC CTG AAG GTG CGC GCG GCC GAG TTC  520
118  K  I  D  F  P  T  R  H  D  W  D  S  F  F  R  D  A  K  D  M
   AAG ATC GAC TTC CCC ACC CGC CAC GAC TGG GAC TCC TTC TTC CGC GAC GCC AAG GAC ATG  580
138  N  E  T  L  P  G  E  R  P  D  T  I  H  L  E  G  L  P  C  K
   AAC GAG ACC CTG CCG GGG GAG CGG CCG GAC ACC ATC CAC CTG GAG GGG CTG CCC TGC AAG  640
158  W  F  A  L  K  E  S  G  S  E  K  P  S  E  D  V  L  V  K  V
   TGG TTC GCC CTG AAG GAG TCG GGC TCC GAG AAG CCC AGC GAG GAC GTC CTG GTC AAG GTG  700
178  F  E  K  F  G  E  I  R  N  V  D  I  P  M  L  D  P  Y  R  E
   TTT GAG AAG TTC GGG GAG ATC CGG AAT GTG GAC ATC CCC ATG CTG GAC CCC TAC CGG GAG  760
198  E  M  T  G  R  N  F  H  T  F  S  F  G  G  H  L  N  F  E  A
   GAG ATG ACG GGC CGC AAC TTC CAC ACC TTC AGT TTC GGG GGG CAC TTG AAC TTC GAG GCC  820
218  Y  V  Q  Y  R  E  Y  M  G  F  I  Q  A  M  S  A  L  R  G  M
   TAT GTG CAG TAC CGT GAG TAC ATG GGC TTC ATC CAG GCC ATG AGC GCC CTG CGC GGG ATG  880
238  K  L  M  Y  K  G  E  D  G  K  A  V  A  C  N  I  K  V  S  F
   AAA CTC ATG TAC AAG GGC GAG GAC GGC AAG GCC GTG GCC TGC AAC ATC AAG GTT TCT TTT  940
258  D  S  T  K  H  L  S  D  A  S  I  K  K  R  Q  L  E  R  Q  K
   GAT TCG ACC AAA CAC CTG AGT GAT GCC TCA ATT AAG AAG CGG CAG CTG GAG AGG CAG AAG 1000
278  L  Q  E  L  E  Q  Q  R  E  E  Q  K  R  R  E  K  E  A  E  E
   CTT CAG GAA CTG GAG CAG CAA AGA GAA GAA CAA AAG CGC AGA GAG AAG GAA GCG GAG GAG 1060
298  R  Q  R  A  E  E  R  K  Q  K  E  L  E  E  L  E  R  E  R  K
   AGG CAG CGA GCG GAG GAA AGG AAA CAA AAG GAG CTG GAA GAG CTG GAG CGA GAG AGG AAA 1120
318  R  E  E  K  L  R  K  R  E  Q  K  Q  R  D  R  E  L  R  R  N
   AGA GAA GAG AAG CTT CGC AAG AGG GAG CAG AAG CAG AGG GAC CGT GAG CTG CGC CGG AAT 1180
338  Q  K  K  L  E  K  [L]  Q  A  E  E  Q  K  [Q]  L  Q  E  K  I  K
   CAG AAG AAG CTG GAG AAG CTG CAG GCG GAG GAG CAG AAG CAG CTG CAG GAG AAG ATC AAG 1240
358  [L]  E  E  R  K  L  L  [L]  A  Q  R  N  L  Q  S  I  R  L  I  A
   CTG GAG GAG CGC AAG CTG CTG GCC CAG AGG AAC CTG CAG TCC ATC CGG CTC ATC GCC 1300
378  E  L  L  S  R  A  K  A  V  K  L  R  E  Q  E  Q  K  E  E  K
   GAG CTG CTC AGC AGA GCC AAG GCT GTG AAG CTA CGG GAA CAG GAG CAG AAG GAG GAG AAG 1360
398  L  R  L  Q  Q  Q  E  E  R  R  R  L  Q  E  A  E  L  R  R  V
   CTG AGG CTC CAG CAG CAG GAG GAG CGG CGG CGG CTG CAG GAG GCC GAG CTG CGG CGC GTG 1420
418  E  E  E  K  E  R  A  L  G  L  Q  R  K  E  R  E  L  R  E  R
   GAG GAG GAG AAG GAG CGC GCG CTG GGC CTG CAG CGG AAA GAG CGG GAG CTG CGC GAG CGG 1480
438  L  L  S  I  L  Q  S  K  K  P  D  D  S  H  T  H  D  E  L  G
   CTG CTG AGC ATC CTG CAG AGC AAG AAG CCG GAC GAC AGC CAC ACA CAC GAC GAG CTG GGC 1540
458  V  A  H  G  P  A  A  A  R  P  G  H  P  A  D  R  V  V  R  L
   GTG GCA CAC GGA CCT GCT GCA GCC CGT CCT GGA CAT CCT GCA GAC CGT GTC GTC CGG CTG 1600
478  C  E  R  H  H  A  A  P  P  R  G  P  A  P  G  R  C  P  Q  G
   TGT GAG CGC CAC CAC GCT GCA CCC CCT CGG GGG CCA GCC CCC GGC CGG TGC CCC CAA GGA 1660
498  E  P  G  P  P  R  G  [R  R  R  S]  Q  K  R  E  R  E  R  G  R
   GAG CCC GGC CCC CCC AGA GGC CGA CGG CGC TCC CAA AAG CGT GAA CGG GAG CGT GGC CGA 1720
518  G  G  P  M  Q  G  G  S  E  L  L  S  C  G  P  R  G  W  L  S
   GGA GGC CCC ATG CAA GGA GGT TCA GAG CTC CTG TCG TGT GGT CCC CGA GGA TGG CTC TCC 1780
538  R  E  E  V  P  G  R  R  P  L  L  H  S  -
   AGA GAA GAG GTG CCC GGG CGG CGT CCT CTC CTG CAT TCC TGA CAACAACCAACAGCCCAAGGGCA 1845
   TCCCTGCCTGCGAGCAGAATGTCTCCAGAAAGGACACCCGGTCAGAACAGGACAAGTGCAACCGGGAGCCCAGCAAGGG 1924
   CCGGGGCCGGGCCACCGGAGACGGGCTTGCTGACCGGCACAAGCGGGAGAGGAGCCGGGCCAGGCGGGCCAGCAGCAGG 2003
   GAGGACGGGAGGCCACGCAAGGAGCGGCGGCCCCACAAGAAGCACGCCTACAAGGATGACAGCCCCCGCCGGCGCAGCA 2082
   CGAGCCCGGACCACACCCGGTCCCGGAGGTCCCACAGCAAAGACAGGCACCGGAGGGAGCGGAGCCGGGAGCGGAGGGG 2161
   CAGCGCCAGCAGGAAGCACAGCCGCCACCGCCGCCGAAGCGAGCGGTCGCGCTCCCGGTCCCCGAGCAGGCACCGCAGT 2240
   ACCTGGAACAGGTAATGACGGGCACGGCCTCCCCACGGCCTGTCCGGGAAAGACCAGGACCTGCTCGAGCCTCCTGGCC 2319
   GCTCCTTGGCCGCTCTCCGTCCACCCCTGCAAAGCCAAGACCCTTCTGCAGCCACGAATGTCCACGGAGCCCGCCGGCA 2398
   GGAAGGAAGACACCATGCTTTAGAGATCCATCTTTCTCCACTCACCGCAGCGTACTTGGCACTTCAGTTTCAAACACGT 2477
   AGTCCTTTAAAACTTGATCCGATAGCTTTAATGCGGCCGGTCCTCTCTCAGTCAGGAAAATTGCACAGACCGACAGTCG 2556
   TGAGGATGGCAGAGCTGCTGCATTCCCCCACACGGGGATTTCTGTGTCTGCTTGGCGACCTCCTAC              2622
```

FIG. 3. Nucleotide sequence and predicted amino acid sequence of 721P. The start ATG is surrounded by the Kozak consensus initiation site (dotted underlining), and the in-frame stop codon at position 61 is marked. The leader sequence and potential transmembrane domain are underlined, and the two N-glycosylation sites are double-underlined. The leucines forming the leucine zipper motif and the consensus phosphorylation site are bracketed.

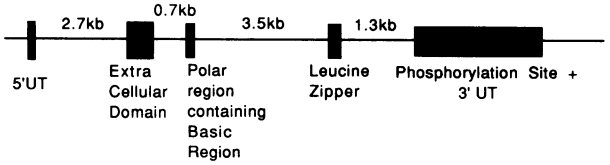


FIG. 4. Genomic organization of 721P. The gene encoding 721P protein was found to be ≈10.5 kb in length, and there are five exons. The first exon codes the G+C-rich 5'-untranslated region (5'UT); the second exon encodes the extracellular and putative transmembrane domain; the third exon encodes the tentative basic region; the fourth exon encodes the leucine zipper-like region; and the fifth exon encodes the phosphorylation site and 3'-untranslated region (3'UT). The intron/exon junctions bracketing exons 3 and 4 (basic region and leucine zipper) fall between codons, suggesting that these exons could be spliced in or out as a unit without disturbing the reading frame of the rest of the sequence.

kinase A (21). The predicted protein contains six cysteine residues. Two cysteines bracket the second N-glycosylation site (residues 95 and 156), and two cysteines bracket the serine phosphorylation site (residue 494 and residue 530).

**Cloning and Intron/Exon Structure of the Gene for 721P.** A 15-kb genomic clone was isolated that contained the entire coding region for 721P protein. Restriction mapping and internal sequencing revealed that the gene was composed of five exons (Fig. 4, Table 1). The entire gene was ≈10.5 kb in length. The exon sequences from the genomic clone were an identical match of the sequences obtained for the cDNA.

**Sequence Homology of 721P Protein.** Plac-3 sequence was compared with sequences in GenBank and the European Molecular Biology Library gene bank. No significant homology was found at the DNA level, suggesting that 721P protein encoded an undeposited gene. A search of the Protein Identification Resource data base (22) revealed a surprising similarity, at the amino acid level, between 721P and human N-myc protein. There were also similarities to c-myc from a variety of species. Weaker homologies were also seen to the gag proteins of the Moloney and AKR murine leukemia viruses.

The homologous area of 721P protein begins at residue 280 and extends to residue 400. The homologous area in N-myc was located in the carboxyl-terminal region of the molecule. The carboxyl-terminal region of N-myc contains the "leucine zipper" and the amphipathic sequence thought important in homodimer and heterodimer formation (3, 4). We projected the 721P sequence onto an idealized α-helix and compared

Table 1. Intron/exon junctions of gene encoding 721P

| Exon | Intron | Exon |
|---|---|---|
| 152 CCGCGCAG | ≈2.7 kb gtgagt....gtttcag | 153 GCCCAAG |
| 933 ACATCAAG | ≈0.7 kb gtgtgtc....tttatag | 934 GTTTCTTT |
| 1082 GAGGAAAG | ≈3.5 kb gtacct....acacag | 1083 GAAACAA |
| 1323 GAGCCAGG | ≈1.3 kb gtaccc....ccgcag | 1324 GCTGTGA |

Adjacent exon (uppercase) and intron (lowercase) sequences are given for each junction. Also shown are nucleotide positions of exon/intron boundaries in the 721P cDNA sequence and intron sizes (superscript numbers). Intron sizes are approximate.

this with a projection on N-myc (Fig. 5*a*). 721P can be projected onto the helix to produce a leucine zipper, with 40% of the amino acids showing identity with the leucine zipper region of N-myc. Moreover, along the surface of the helix facing the leucine zipper region, percent identity rises to 66%. Similar degrees of similarity were seen between the leucine zipper of 721P and the leucine zippers of c-myc and L-myc.

The fos/jun class of regulatory proteins have a basic–spacer–basic structure immediately upstream from the leucine zipper, with the spacer region being composed of two to three alanine residues (1). Examination of 721P did not reveal any basic–spacer–basic motif in this region. However the basic region of the helix–loop–helix proteins and, in particular, the myc family of proteins has been located further from the leucine zipper region (1). 721P protein has a cluster of basic residues very similar to the basic region in myc proteins responsible for DNA binding (4, 23). Spacing between the basic region and the leucine zipper was also similar for the two proteins (Fig. 5*b*).

## DISCUSSION

We have characterized a trophoblast/lymphocyte/endothelial cross-reactive antigen at both the protein and nucleic acid levels and presented evidence that mAb B721 recognizes a specific plasma membrane-associated glycoprotein encoded by a previously unreported gene containing five exons. That we have cloned the correct molecule is supported by the fact that (*i*) the size of the predicted protein agrees with the observed molecular mass of the deglycosylated native protein, (*ii*) there are appropriate N-glycosylation sites, (*iii*) the cDNA reacts with an mRNA in appropriate tissues, and

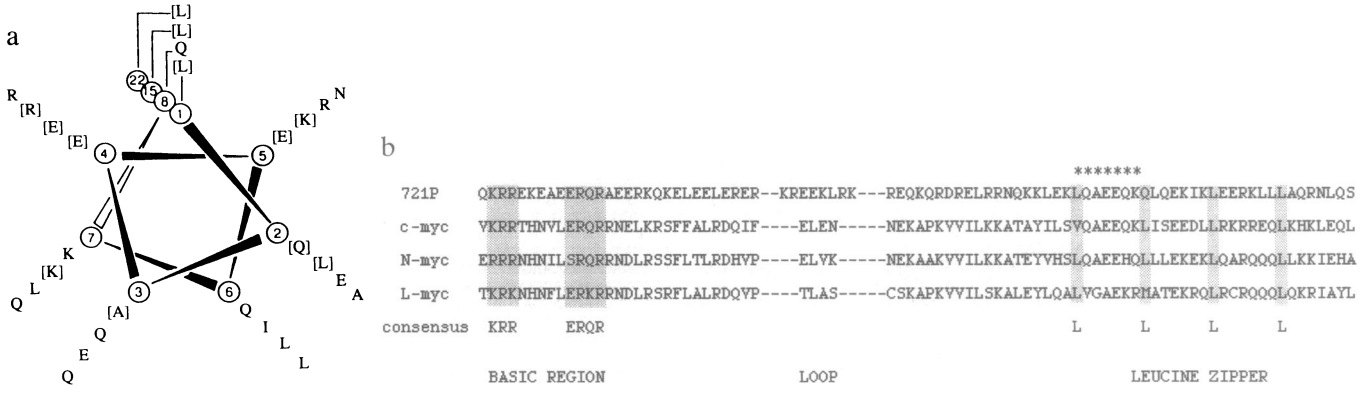


FIG. 5. (*a*) Leucine zipper motif of 721P. Beginning with Leu-344, 721P protein was projected onto an idealized α-helix after the method of Landshultz *et al.* (2). Bracketed amino acids are identical residues with N-myc projected on a similar wheel. (*b*) Comparison of 721P to the basic region and leucine zipper of the myc family of oncogenes. The region of identity from the protein data base search, LQAAEQK (marked with asterisks), was used to align 721P (stating at Glu-288) to the corresponding myc proteins. The leucine zipper region and basic region are shaded and compared with the myc consensus sequences (3, 4). There is 55–60% identity between amino acids in the leucine zipper region of 721P compared with the leucine zipper of the myc proteins.

finally (*iv*) COS cells transfected with the cDNA react with mAb B721. We recognize that the mAb could cross-react with some other protein expressed on the COS cell; however, we do not think this likely in that COS cells are not of human origin and do not normally stain with this mAb. The possibility that we accidentally cloned a protein with a cross-reactive epitope or that we cloned a regulatory protein that happens to up-regulate 721P protein on COS cells is remote.

That 721P protein is on the plasma membrane is clear both from the iodination data and from the flow cytometry analysis. Whether 721P is a transmembrane protein is somewhat more tentative. Generally, transmembrane proteins have a membrane-spanning region that is strongly hydrophobic and that usually contains several leucine, isoleucine, valine, or alanine residues and no positive charges (24). The only possible transmembrane region of 721P is a stretch of amino acids beginning at Asn-203, which contains no positive charges and only one acidic amino acid and is the longest stretch of hydrophobic or uncharged amino acids in the protein. The N-glycosylation sites are all to the amino-terminal side of this region. Although these characteristics favor a transmembrane domain, the region also contains a significant number of uncharged polar residues and has only one leucine and one valine residue. However, other transmembrane proteins with polar amino acids in their transmembrane domains have been reported (25), so this region of 721P could act as a transmembrane spanner.

We do not yet know the function of 721P protein; however, the predicted sequence suggests some intriguing possibilities. The most striking feature was the strong similarity to the leucine zipper region of the myc family of proteins. The only anomaly in 721P protein was the substitution of a glutamine for a leucine in the second position of the zipper. Other investigators have shown that, at least in the fos/jun heterodimer system, an individual leucine in the zipper can be replaced without abolishing heterodimer formation (26), and as long as the first leucine in the zipper is unaltered, DNA binding is also unaffected (27). Moreover, it has been proposed that the major requirement for an amino acid in the zipper region is an amino acid with a long side chain (no substitutions on the $\beta$ carbon) and a relatively bulky end group (2). Glutamine with a three-carbon side chain may fit this model. Finally, the recently described myc-binding protein, max, has been shown to have a histidine in the second position of its leucine zipper, indicating that a leucine is not essential (4). We do not imply that the similarity between 721P protein and the myc family means that 721P is a member of the myc family of proteins or that there is necessarily an evolutionary relationship between the two proteins. Rather, we believe that the presence in 721P protein of structures involved in dimerization and DNA binding suggest that this protein could enter into similar interactions.

That 721P protein is a membrane-associated protein may not necessarily exclude it from interacting with either nuclear proteins or DNA. The interleukin 1 receptor, upon binding interleukin 1, has been shown to be translocated intact to the nucleus (28). Conceivably, either the entire 721P molecule or, at least, the carboxyl-terminal half of the molecule could also translocate to the nucleus. Although the leucine zipper and basic region could be a site of protein–protein interaction at the membrane, no such combined motif has been reported in a transmembrane protein.

Cloning and sequencing of a trophoblast/lymphocyte/endothelial cross-reactive surface protein has revealed that 721P protein is a discrete plasma-membrane-associated glycoprotein that appears to have interesting similarities to the myc family of nuclear proteins. The function of the molecule is presently unknown, and inferences from its predicted structure are, of course, speculative. However, the presence of a leucine zipper region and a cluster of basic residues with strong similarity to the myc family of proteins, as well as a terminal phosphorylation site, are all consistent with a protein involved in signal transduction and in gene regulation.

The authors thank Drs. Russell Doolittle, Stephen Hedrick, and Cornelius Murre for critically reviewing the manuscript. We also thank Drs. Andrew Weinberg, Gerald Siu, and Alex Duncan for helpful discussions. Dr. Sarka Southern initially subcloned Plac-3, and Dr. Peter Scully performed the COS transfections. We thank Dr. Jose Millan for providing us with the placental cDNA library, Dr. Douglas Hanahan for providing us with the pCK3 plasmid, and Dr. Wen-Hwa Lee for the OS2 human genomic library. Expert technical assistance was provided by Mr. David Levi and Mr. Homero Sepulvada. This work was supported, in part, by National Institutes of Health Grants AI 23287 and AI 08795 (R.W.D.) and by The State of California Universitywide Taskforce on AIDS and National Institutes of Health Grant AI 31870 (J.R.V.).

1. Busch, S. J. & Sassone-Corsi, P. (1990) *Trends Genet.* **6**, 36–40.
2. Landshultz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
3. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56**, 777–783.
4. Blackwood, E. & Eisenman, R. (1991) *Science* **251**, 1211–1217.
5. Nakabeppu, Y. & Nathans, D. (1989) *EMBO J.* **8**, 3833–3841.
6. Maekawa, T., Sakura, H., Kanei-Ishii, C., Sudo, T., Yoshimura, T., Fujisawa, J.-I., Yoshida, M. & Ishii, S. (1989) *EMBO J.* **8**, 2023–2028.
7. Nakayama, E., von Hoegen, I. & Parnes, J. R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1352–1356.
8. Ikuta, K., Takami, M., Kim, C. W., Honjo, T., Miyoshi, T., Tagaya, Y., Kawabe, T. & Yodoi, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 819–823.
9. Voland, J. R., Frisman, D. F. & Baird, S. M. (198) *Am. J. Reprod. Immunol. Microbiol.* **11**, 24–30.
10. Voland, J. R., Frisman, D. F., Fan, Q. & Dutton, R. W. (1987) *Cell. Immunol.* **110**, 197–208.
11. Webb, P. D., Evans, P. W., Molloy, C. M. & Johnson, P. M. (1985) *Am. J. Reprod. Immunol. Microbiol.* **8**, 113–119.
12. Corkey, R. F., Corkey, B. E. & Gimbrone, M. A. (1981) *J. Cell. Physiol.* **106**, 425–434.
13. Harlow, E. & Lane, D. (1988) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
14. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) *DNA Cloning: A Practical Approach* (IRL, Oxford), Vol. 1, 49–78.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
16. Weinberg, A. D., English, M. & Swain, S. L. (1990) *J. Immunol.* **144**, 1800–1807.
17. Weinberg, A. D. & Swain, S. L. (1990) *J. Immunol.* **144**, 4712–4720.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 1825–1848.
20. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
21. Spiess, M., Handschin, C. & Baker, K. P. (1989) *J. Biol. Chem.* **264**, 19117–19124.
22. Doolittle, R. F. (1987) *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences* (University Science Books, Mill Valley, CA).
23. Prendergast, G. C. & Ziff, E. B. (1991) *Science* **251**, 186–188.
24. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
25. Hedrick, S. M. (1989) in *Fundamental Immunology*, ed. Paul, W. E. (Raven, New York), pp. 291–314.
26. Ransone, L. J., Visvader, J., Sassone-Corsi, P. & Verma, I. M. (1989) *Genes Dev.* **3**, 770–781.
27. Kouzarides, T. & Ziff, E. (1988) *Nature (London)* **336**, 646–651.
28. Curtis, B. M., Widmer, M. B., deRoos, P. & Qwarnstrom, E. E. (1990) *J. Immunol.* **144**, 1295–1303.