

# Sequence and expression of caveolin, a protein component of caveolae plasma membrane domains phosphorylated on tyrosine in Rous sarcoma virus-transformed fibroblasts

(cell signaling/membranes/tyrosine kinase/transformation)

JOHN R. GLENNEY, JR.\* AND DANIEL SOPPET\*†

Lucille P. Markey Cancer Center and Department of Surgery, University of Kentucky College of Medicine, Lexington, KY 40536-0093

Communicated by Hidesaburo Hanafusa, August 10, 1992 (received for review July 21, 1992)

**ABSTRACT** Caveolae are flask-shaped plasma membrane invaginations abundant in endothelium and muscle but may be present in all cells. They contain a filamentous coat material thought to be important in their structure and function. Recent studies have demonstrated that a 22-kDa protein (caveolin) phosphorylated on tyrosine in Rous sarcoma virus-transformed chicken fibroblasts is a component of the caveolae coat on the inner aspect of the membrane. We now report the deduced protein sequence of chicken caveolin derived from cDNA PCR products and genomic DNA clones. Caveolin is a unique protein of 178 amino acids and displays little sequence similarity to other proteins in the GenBank data base. Hydrophobicity predictions indicate an unusual 40-amino acid hydrophobic region near the C terminus that may be used to anchor the protein to the membrane. When chicken caveolin was expressed in mouse 3T3 cells and detected by immunofluorescence microscopy, the typical caveolae pattern was observed. This includes brightly fluorescent membrane patches in many cases concentrated at the margin of cells and in arrays. Caveolae may be distinct from other membrane domains due at least in part to caveolin.

Tyrosine phosphorylation is a common signal leading to cellular proliferation or differentiation. Although many tyrosine kinases are known to generate this signal, the substrates that mediate the effects of the tyrosine kinases are less well understood. The tyrosine phosphorylation of several cellular proteins is now thought to provide crucial regulation of the growth factor receptor or tyrosine kinase oncogene product response (see ref. 1 for a recent review). In the analysis of tyrosine-phosphorylated proteins most of the substrates were characterized in other systems first (as the unmodified enzyme or protein) and the regulatory tyrosine phosphorylation was revealed subsequently. Given that the identity or function of the vast majority of mammalian proteins is unknown it seems likely that many of the functionally important tyrosine phosphorylations have been overlooked.

At least two other approaches have been used to identify and characterize other potential substrates of the tyrosine kinases. Substrates that have a sequence motif termed the SH2 domain (2) can be directly cloned from an expression cDNA library, thereby allowing the identification of new proteins in this class (3). Alternatively, proteins are isolated using an anti-phosphotyrosine antibody affinity column and monoclonal antibodies (mAbs) are raised to the protein backbone of these components (4, 5). The mAbs then can be used to study individual substrates in isolation.

In our laboratory Rous sarcoma virus-transformed chicken embryo fibroblasts were used to generate antibodies to several tyrosine phosphorylated proteins of 215 kDa, 76 kDa,

and 22 kDa (6-8). The 22-kDa protein was found to be tyrosine phosphorylated in chicken embryo fibroblasts expressing transforming *v-src* but not in cells expressing non-myrystoylated *v-src*, cells that are nontransformed (9).

The subcellular location of the 22-kDa substrate has provided a clue to its function. Immunolocalization reveals that the 22-kDa protein is present in caveolae, small plasma membrane specializations present in most cells, and the 22-kDa protein was thus termed caveolin. Caveolae are known to have a unique filamentous coat and caveolin was assigned to this coat material by immunoelectron microscopy. To begin addressing these questions about the structure and function of caveolin, cDNA encoding caveolin was cloned and sequenced.‡ Caveolin is 178 amino acids in length with an unusual hydrophobic membrane attachment domain. No significant sequence homology with proteins in the GenBank data base was observed. Cloned chicken caveolin when expressed in 3T3 cells targeted to caveolae.

## MATERIALS AND METHODS

**Isolation of Caveolin and Peptides.** Chicken gizzards were extracted with 5 volumes of 1% SDS, boiled for 5 min, and clarified by centrifugation for 1 hr at 100,000 × *g*. The supernatant was adjusted to 1% Triton X-100/0.2% SDS/0.15 M NaCl/20 mM Tris, pH 7.4 (buffer A), and passed through an affinity column containing 5 mg of mAb 20b (4) per ml of agarose. The column was washed with buffer A and then washed with 40 mM NaCl; elution from the column was with 0.1 M diethylamine (pH 11.5). Protein was run on a 4-15% SDS/acrylamide gel and the 22-kDa band was visualized by precipitation with ice-cold 0.4 M KCl. Caveolin was excised from the gel, electrophoretically eluted, dialyzed, lyophilized, and digested with either trypsin, chymotrypsin, or V8 protease and the peptides were resolved by reverse-phase HPLC. Peptides were sequenced by the Macromolecular Core Facility, University of Kentucky.

**Amplification of Caveolin cDNA.** A degenerate oligonucleotide based on the amino acid sequence of peptide 1 with the sequence GTGCGAATTC(C,T)TT(A,C,G,T)GG(A,G)TC-(A,C,G,T)C(G,T)(A,G)TT(A,C,G,T)AC was used in a reverse transcriptase reaction followed by PCR using the above oligonucleotide together with the oligonucleotide GCCGGG-AATTCTA(C,T)GT(A,C,G,T)GA(C,T)(A,T)(C,G)-(A,C,G,T)GA(A,G)GG (based on peptide 2). The reaction was carried out for 40 cycles at 95°C, 60°C, and 72°C for 0.75, 1.5, and 2 min, respectively. The PCR mix was digested with

Abbreviation: mAb, monoclonal antibody.

\*To whom reprint requests should be addressed.

†Present address: Laboratory of Mammalian Genetics, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, MD 21702.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. L01582).

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.

*EcoRI* and cloned into the R1 site of  $\lambda$ gt11 (Stratagene). The resulting phage was screened with a mixture of mAbs to caveolin. Positive phage were plaque purified twice and the insert was amplified with the oligonucleotide primers TTGACACCAGACCAACTGGTAATG and GGTGGCGACGACTCCTGGAGCCCG ( $\lambda$  sequences) by PCR as above. A longer [500 base pairs (bp)] unique caveolin PCR product was then made by amplification with a unique primer (CGCCTGGAATTCCTGTACGCGGCCCGTCAGGGAA) at the 5' end of the insert in the  $\lambda$  clone (see *Results*) together with a degenerate oligonucleotide [CCGATGAATTC(C,T)TC(A,C,G,T)-C(G,T)(A,C,G,T)AC(A,C,G,T)GT(A,C,G,T)GC-(A,C,G,T)C] corresponding to the C terminus of peptide 7.

**Genomic DNA Clones.** An EMBL3 chicken genomic DNA library (Clontech) was screened with a  $^{32}$ P-labeled probe from chicken caveolin cDNA by hybridization at 42°C in 5× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate), 40% formamide, 0.5% SDS, 1× Denhardt's solution, and 100  $\mu$ g of salmon sperm DNA per ml. Filters were washed with 0.2× SSC at 60°C for 1 hr. Cloned DNA was subjected to restriction enzyme digestion and Southern blotting using hybridization conditions described above. Restriction fragments were subcloned into the Bluescript plasmid and selected by colony hybridization. The sequence manipulation and analysis were performed with the MACVECTOR program (Kodak).

**Expression of Chicken Caveolin in Mouse 3T3 Cells.** The coding region of chicken caveolin was amplified from gizzard RNA with the oligonucleotide primers CCGCGCTAGC-CATATGTCCGGCACCAAATACGTG and CGGCCTC-GAGAATCTCTTCCGTAAGTGGC incorporating restriction sites for cloning into the pMAMneo vector (Clontech). Transfection was performed using the Lipofectin reagent (BRL) and cells were selected with G418 (400  $\mu$ g/ml). Induction of chicken caveolin protein was accomplished with the addition of dexamethasone to 4  $\mu$ M and incubation for 12 hr. Immunofluorescence microscopy was performed as described (4). For Western blotting, cells were lysed in SDS sample buffer, run on a 4–15% SDS/acrylamide gradient gel, transferred to Immobilon, and treated with the anti-caveolin mAb 2222 antibody followed by  $^{125}$ I-labeled goat anti-mouse second antibody and autoradiography.

## RESULTS

Since caveolin is known to be expressed at high levels in smooth muscle (9), we employed a two-step method to purify the protein from chicken gizzard using an anti-caveolin mAb (4) followed by preparative SDS gel electrophoresis. No sequence was observed for intact caveolin or several derived peptides, pointing to a blocked N terminus. Eight other peptides yielded the sequences shown in Table 1. A search of the GenBank data base with the longest of these peptide

sequences did not reveal any obvious sequence homology. We therefore decided to clone the cDNA encoding caveolin using oligonucleotides predicted from peptide sequences.

To facilitate cloning of the cDNA encoding caveolin, degenerate oligonucleotides corresponding to the 5' end of one peptide and the 3' end of a second was used in a PCR with RNA derived from chicken gizzard. The resulting cDNA was cloned into  $\lambda$ gt11 and immunoscreened with a mixture of mAbs to caveolin. When this method was used with oligonucleotides corresponding to the N terminus of peptide 2 together with the C terminus of peptide 1, immunopositive  $\lambda$  clones were observed, whereas the reverse experiment (N terminus of peptide 1 with C terminus of peptide 2) was immunonegative. Multiple ethidium bromide-stained bands were observed after the PCR, one of which corresponded to the 156-bp insert from the positive clones (Fig. 1). cDNA from this clone was sequenced and found to encode additional peptide sequences derived from gizzard caveolin. A longer cDNA fragment was then generated using a unique oligonucleotide based on the 5' end of this new sequence together with a degenerate oligonucleotide predicted from the C terminus of peptide 7 in a PCR reaction. A single band of 500 bp (not shown) was sequenced and used to screen cDNA and genomic DNA libraries.

A chicken caveolin  $^{32}$ P-labeled DNA probe was used to screen several chicken cDNA libraries without success. Since the 500-bp PCR product from chicken gizzard was predicted to encode the majority of a 22-kDa protein, we decided to ascertain the polypeptide ends from the sequence of genomic DNA. We screened an EMBL3 library that was constructed from chicken genomic DNA. Several positive clones were selected that fell into one of two groups by restriction analysis and Southern blotting. One set of clones was found to encode an exon containing the entire 3' two-thirds of the PCR product (Fig. 2b). In addition, we were able to extend the sequence to include the putative termination codon. The second set of clones encoded the N terminus divided into two exons. Again we were able to extend the available PCR-derived sequence in this case to include the initiator ATG. The sequence surrounding the predicted protein start site is consistent with the features identified by Kozak (10) for the initiator AUG region. All of the peptides derived from chicken gizzard caveolin (Table 1) were found in the deduced protein thereby confirming its identity. To confirm this organization unique PCR primers corresponding to the predicted 5' and 3' ends of the coding region were used to amplify cDNA from chicken gizzard RNA. As expected, a single 534-bp product was generated. This fragment was sequenced and exactly corresponded to the sequence shown in Fig. 2.

Hydrophilicity analysis of the deduced caveolin protein by the Kyte–Doolittle method (11) revealed an extensive hydrophobic domain (Fig. 3). This extremely hydrophobic region of

Table 1. Sequences of peptides derived from chicken caveolin

Peptide	Amino acid sequence	Protease	Amino acid positions*
1	EIDLVNRPDK	Trypsin	48–57
2	YVDSEGFYAXPVR	Trypsin	6–19
3	VFSSIR	Trypsin	166–171
4	MMADELSEK	Trypsin	31–39
5	AAPVREQNIY	Chymotrypsin	15–25
6	KASFTTF	Chymotrypsin	86–92
7	SSIRATVR	V8	168–176
8	GFLYAAPVRE	V8	11–20

All of the peptide sequences from three separate protease digestions of caveolin isolated from chicken gizzard are listed. The positions of these peptides in the deduced amino acid sequence can be found in Fig. 2.

\*Amino acid positions in chicken cDNA sequence.

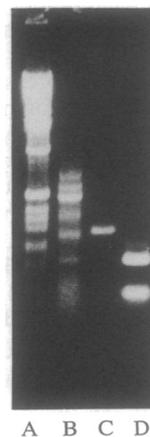


FIG. 1. Amplification of caveolin cDNA sequences. Degenerate oligonucleotides deduced from the C terminus of peptide 1 together with the N terminus of peptide 2 (see Table 1) were used in a PCR, a sample of which was run on an agarose gel stained with ethidium bromide (lane B). The DNA from this reaction was cloned into the *EcoRI* site of  $\lambda$ gt11 and screened with antibodies to caveolin. The insert of an antibody-positive clone was amplified with  $\lambda$  primers outside the R1 site (lane C), which was then digested with *EcoRI* (lane D) to remove the  $\lambda$  sequences. The 1-kilobase (kb) ladder was run in lane A for size determination.

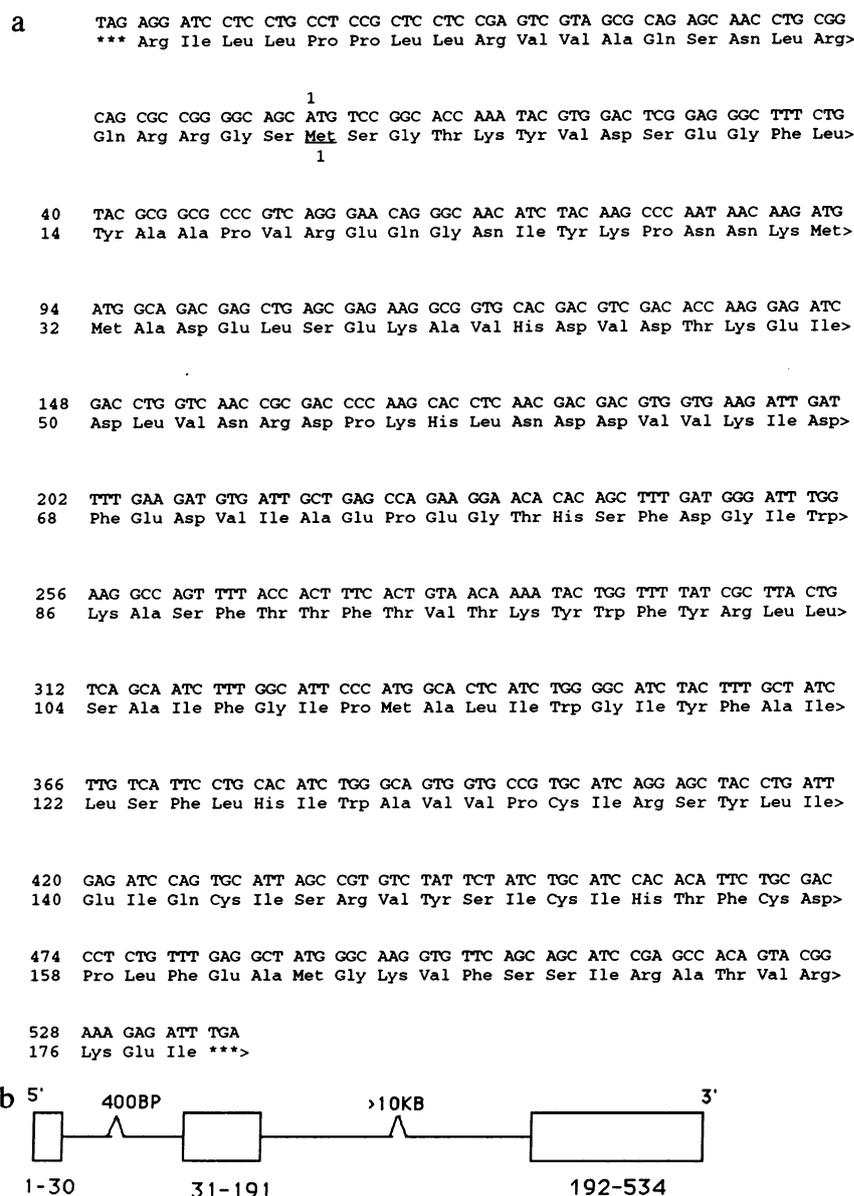


FIG. 2. (a) DNA sequence and deduced protein sequence of chicken caveolin. Numbers on the left indicate the nucleotide positions within the predicted protein coding sequence. The deduced amino acid sequence (three-letter code) is given under the corresponding nucleotide sequence. The termination codon is indicated by (\*\*\*). The coding sequence was derived from sequencing PCR products and parts of genomic clones. Regions outside the coding sequence were from genomic clones. (b) The intron-exon organization of the caveolin gene was derived from the sequence of corresponding EMBL3 clones. Boxes are the exon coding sequences and numbers represent the positions in the cDNA sequences above.

caveolin is approximately twice the 20-amino acid length of a simple  $\alpha$ -helical membrane-spanning region. In addition, secondary structure predictions would suggest more of a  $\beta$ -sheet conformation in this region (Fig. 3). It seems likely that this represents the domain for membrane attachment and explains the recent finding that caveolin is not extractable with agents that generally remove peripheral membrane proteins (12).

The entire chicken cDNA coding region was cloned into a mammalian expression vector (pMAMneo) where protein expression can be controlled by the dexamethasone-inducible mouse mammary tumor virus long terminal repeat. The plasmid that contained the caveolin cDNA was transfected into mouse 3T3 fibroblasts and cell lines were selected by growth in G418. As shown in Fig. 4, the chicken caveolin protein band was only detected in transfected cells, where it was visualized as a single species corresponding to a molecular mass of 22 kDa by SDS/PAGE and Western blotting. Some caveolin protein was seen without dexamethasone

induction, but clearly a significantly higher level was seen 24 hr after addition of dexamethasone.

Expression of chicken caveolin in 3T3 cells was also analyzed by immunofluorescence microscopy (Fig. 5). When cells expressed high levels of caveolin the distribution was generally similar to the pattern previously seen in fibroblasts (4, 12). A punctate pattern was observed, in many instances with large accumulations of caveolin concentrated in patches and at the margin of cells. In a few instances, large intracellular inclusions were observed by phase-contrast optics and these were surrounded with caveolin. It is not known what these inclusions represent or whether caveolin may be responsible for their existence.

## DISCUSSION

Plasma membrane specializations termed caveolae are widespread and routinely seen in the electron microscope. Although a filamentous coat material has been clearly demon-

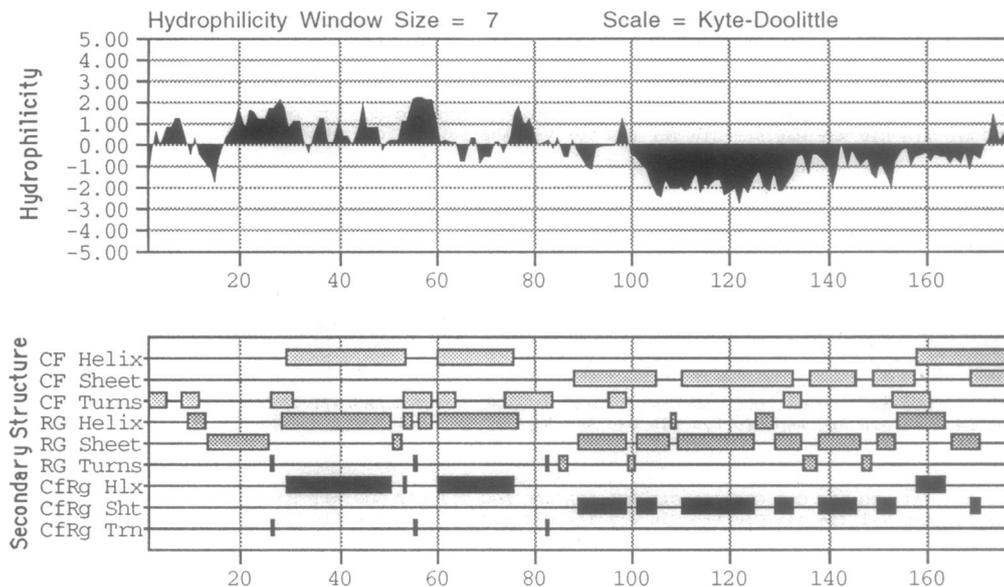


FIG. 3. Hydrophilicity plot (*Upper*) and secondary structure predictions (*Lower*) of the amino acid sequence of chicken caveolin. (*Upper*) Shaded areas above or below the zero line indicate relatively hydrophilic or hydrophobic regions, respectively. (*Lower*)  $\alpha$ -Helix,  $\beta$ -sheet, or turns were predicted by three methods using the MACVECTOR computer program.

strated, until recently no protein marker for the cytoplasmic surface of these structures was known. The localization of a 22-kDa protein that is phosphorylated on tyrosine in transformed chicken fibroblasts to this caveolae coat provided such a marker. For this reason the protein was named caveolin and may serve a structural role in the same way that clathrin makes up coated pits and vesicles (13). The deduced protein sequence of caveolin, however, suggests that if it is a major structural element of caveolae it would behave quite differently than clathrin. The most obvious differences reside in the hydrophobic nature of caveolin and the smaller polypeptide size compared to clathrin (14).

Caveolin is clearly present on the inner surface of caveolae (12) and cloned caveolin targets to caveolae (see above), yet we cannot conclude that caveolin makes up the caveolae coat or is even a structural element in caveolae. It is possible, for instance, that caveolin provides a membrane anchor for the attachment of the filament coat. Other components may well be more important to the structure of the caveolae coat with caveolin associated with these structural elements. Although

this question can be answered by expressing the cloned caveolin into cells that lack caveolae, virtually all cells are thought to contain these structures. The 22-kDa caveolin has been shown to be present at high levels in muscle and lung tissue (9), but most tissues or cells have a detectable amount.

The unique hydrophobic region in caveolin could provide the stability of caveolae or targeting of the protein to caveolae in the plasma membrane. The extensive 40-amino acid  $\beta$ -sheet conformation predicted for caveolin is unlike a typical  $\alpha$ -helical transmembrane domain that is usually 20 amino acids in length. A hint of a possible role of the caveolin hydrophobic domain comes from previous studies in which cells depleted of cholesterol lack caveolae (15) and treatment of cells with the cholesterol-binding drug filipin or nystatin results in the progressive dissolution and fragmentation of the caveolin coat (12). This may reflect the self-association of caveolin within the plane of the membrane that requires sterols.

Previous studies have noted that sedimentation of caveolin through a sucrose gradient in detergent solution is consistent with a protein having a molecular mass of  $\approx 150$  kDa (9). We can interpret this now to at least partially reflect the association of caveolin with detergent micelles. In addition, however, we have detected a high molecular mass form of caveolin that may be due to self-association. Alternatively, caveolin targeting and associations in caveolae may have specific lipid requirements that are found in caveolae, and the caveolin hydrophobic region may serve as a targeting signal. A hydrophobic targeting sequence is not without precedent. A positive membrane targeting sequence, for instance, has been demonstrated for the hydrophobic region of the protein Gal T that resides in Golgi membranes (16).

Transformation of chicken embryo fibroblasts by Rous sarcoma virus leads to the phosphorylation of caveolin on one or more tyrosine residues (9). Many proteins have been shown to be tyrosine phosphorylated in Rous sarcoma virus-transformed chicken embryo fibroblasts, yet most of those proteins are phosphorylated equally well in cells expressing nonmyristoylated *src*, cells that are not transformed (17, 18). Only two proteins, a 120-kDa protein (19) and caveolin (9), display this selective phosphorylation. In the case of caveolin, this lack of phosphorylation could be due to the fact that caveolin is an integral membrane protein and the N-terminal myristate of pp60<sup>v-src</sup> provides a hydrophobic membrane anchor to the membrane. Thus, even though *src* is found in focal adhesions (20), some may reside in caveolae, at least transiently, to phosphorylate caveolin. Although caveolin



FIG. 4. Expression of chicken caveolin in mouse 3T3 cells detected by Western blot. Control (lane A), transfected (lane B), or transfected and induced (lane C) 3T3 cell extracts were subjected to Western blotting with the 2222 mAb, which recognizes only the chicken protein (and not the endogenous mouse caveolin). The 22-kDa caveolin position is indicated.

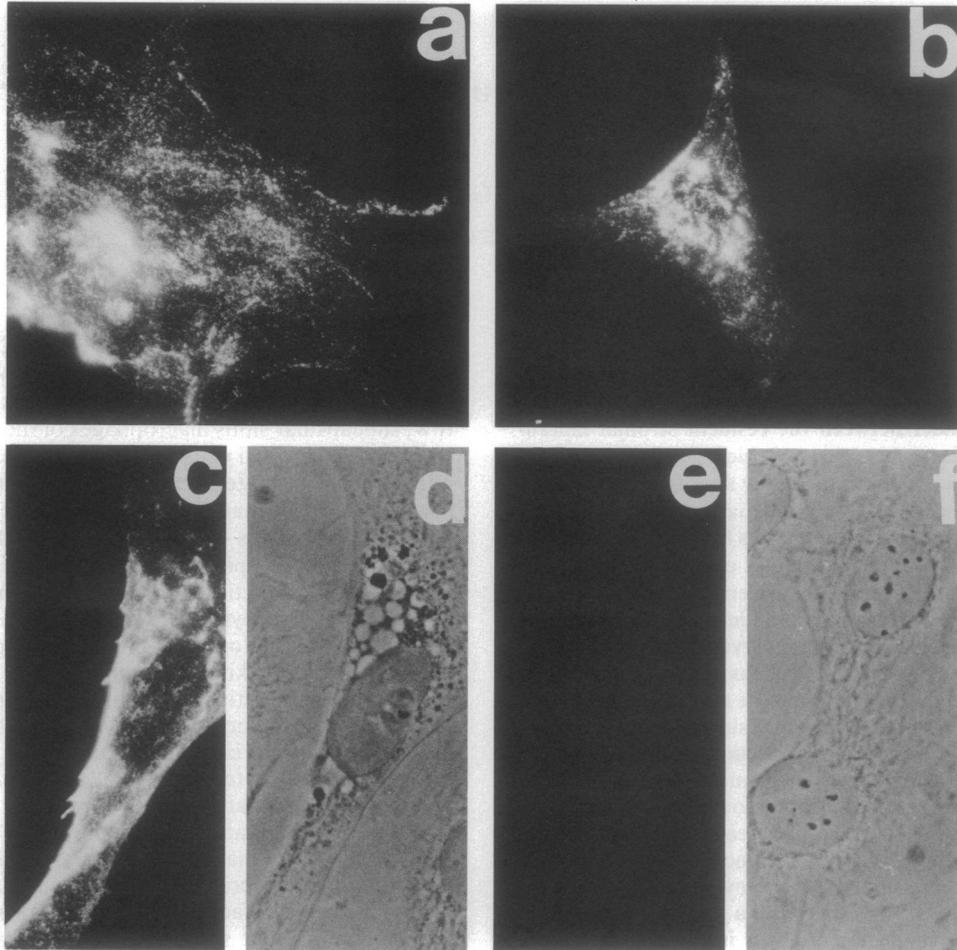


FIG. 5. Chicken caveolin in mouse 3T3 cells detected by immunofluorescence microscopy. Transfected (*a–d*) and control 3T3 cells (*e* and *f*) were treated with dexamethasone and stained with the 2222 mAb; this was followed by fluorescein isothiocyanate-labeled anti-mouse antibody and visualization by fluorescence (*a–c* and *e*) or phase-contrast (*d* and *f*) optics. ( $\times 200$ .)

appears to have a somewhat different distribution in Rous sarcoma virus-transformed chicken embryo fibroblasts compared to their normal counterpart, this may simply reflect a difference in cell morphology in these transformed cells. Future studies must address the regulation of caveolin and how this may influence the structure or function of caveolae.

We thank Mike Mendenhall, Roberto Campos, and Sue Jamison for encouragement and suggestions. This work was supported by Public Health Services Grant CA49298.

- Glenney, J. R. (1992) *Biochim. Biophys. Acta* **1134**, 113–127.
- Koch, C. A., Anderson, D., Moran, D. F., Ellis, C. & Pawson, T. (1991) *Science* **252**, 668–674.
- Skolnik, E. Y., Margolis, B., Mohammadi, M., Lowenstein, E., Fischer, R., Drepps, A., Ullrich, A. & Schlessinger, J. (1991) *Cell* **65**, 83–90.
- Glenney, J. R. & Zokas, L. (1989) *J. Cell Biol.* **108**, 2401–2408.
- Kanner, S. B., Reynolds, A. B., Vines, R. R. & Parsons, J. T. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3328–3332.
- Davis, S., Lu, M. L., Lo, H. L., Lin, S., Butler, J. A., Druker, B. J., Roberts, T. M., An, Q. & Chen, L. B. (1991) *Science* **252**, 712–715.
- Turner, C., Glenney, J. R. & Burrige, K. (1990) *J. Cell Biol.* **111**, 1059–1068.
- Turner, C. (1992) *J. Cell Biol.* **115**, 201–207.
- Glenney, J. R. (1989) *J. Biol. Chem.* **264**, 20163–20166.
- Kozak, M. (1986) *Cell* **44**, 283–292.
- Kyte, J. & Dolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
- Rothberg, K. G., Heuser, J. E., Donzell, W. C., Ying, Y.-S., Glenney, J. R. & Anderson, R. G. W. (1992) *Cell* **68**, 673–682.
- Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russel, D. W. & Schneider, W. J. (1985) *Annu. Rev. Cell Biol.* **1**, 1–39.
- Pierce, B. M. F. (1978) *J. Mol. Biol.* **126**, 803–812.
- Rothberg, K. G., Ying, Y.-S., Kamen, B. A. & Anderson, R. G. W. (1990) *J. Cell Biol.* **111**, 2931–2938.
- Teasdale, R. D., D'Agostaro, G. & Gleeson, P. A. (1992) *J. Biol. Chem.* **267**, 4084–4096.
- Kamps, M. P., Buss, J. E. & Sefton, B. (1986) *Cell* **45**, 105–112.
- Hamaguchi, M., Grandori, C. & Hanafusa, H. (1988) *Mol. Cell. Biol.* **8**, 3035–3042.
- Downing, J. R. & Reynolds, A. B. (1991) *Oncogene* **6**, 607–613.
- Rohrschneider, L. R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3514–3518.