Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 8535–8539, November 1988
Cell Biology

cDNA-derived sequence of chicken embryo vinculin

(cytoskeleton/adherens junctions)

MICHELLE D. COUTU AND SUSAN W. CRAIG

Department of Biological Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205

Communicated by M. Daniel Lane, August 22, 1988 (received for review June 27, 1988)

ABSTRACT **We report the complete primary structure of chicken embryo vinculin. The amino acid sequence was derived from the nucleotide sequence of five overlapping cDNA clones isolated from a λgt11 phage library. Chicken embryo vinculin contains 1066 amino acids, has a calculated M_r of 116,990, a calculated pI of 5.9, and a hydropathy index of −4.22. A search of the National Biomedical Research Foundation protein sequence data base found no proteins with significant homology to vinculin. A striking feature of the linear sequence is a proline-rich region extending between residues 837 and 879. This region contains 45% proline and 19% aspartic plus glutamic acids; it is also the longest hydrophilic stretch in the molecule. The proline-rich region separates an amino-terminal domain with a calculated pI of 5.4 from a carboxyl-terminal domain with a calculated pI of 9.7. This feature suggests a structural basis for the specific interaction of vinculin with acidic phospholipids and a mechanism for the shuttling of vinculin between cytoplasm and membrane-associated junctional plaque.**

Vinculin, a 130-kDa protein, is one of the few identified components of microfilament-associated cell junctions (1). Because these cell–cell and cell–substrate junctions play key roles in growth and morphogenesis, definition of the events regulating the assembly of vinculin into the junctional plaque is an important goal. The distribution of vinculin between cytoplasm and junctional plaque varies depending on cell contact (2, 3), developmental stage (4), viral transformation (5), and stimulation with certain growth hormones (6, 7). A variety of biochemical observations have been made that may lead to a molecular understanding of the dynamics of vinculin assembly at adherens junctions. Among these are the *in vitro* interaction of vinculin with talin (8), α-actinin (9, 10), and acidic phospholipids (11, 12); myristoylation (13) and transformation-sensitive palmitoylation of vinculin (14); phosphorylation of vinculin both *in vitro* (11, 15) and *in vivo* (16); and the expression of both isoelectric (17–19) and molecular weight (20, 21) variants of vinculin. However, progress in understanding the structural basis for these observations and their relationship to the cellular properties of vinculin has been limited.

To overcome this limitation, we (22) and others (23, 24) have isolated cDNA clones coding for parts of the vinculin sequence. Price *et al.* (23) published the sequence of a 2.9-kilobase (kb) clone containing 80% of the vinculin coding sequence, and Bendori *et al.* (24) reported the isolation and restriction analysis of several clones encompassing 5.3 kb of the estimated 6.2-kb message. Both groups obtained their clones from a chicken embryo fibroblast library. In this paper we present the complete cDNA and derived amino acid sequence for vinculin from a 10-day chicken embryo.*

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

cDNA Library Screening. A chicken embryo cDNA library prepared by Clontech was screened with two synthetic oligonucleotides. Oligonucleotide V3 was complementary to the 24 most-3′ nucleotides of BCE68V (22) and V5 was complementary to the 19 most-5′ nucleotides of the vinculin coding region (23). The oligonucleotides were end-labeled (25) by using polynucleotide kinase from Boehringer Mannheim Biochemicals and [γ-^{32}P]ATP from New England Nuclear. Plaque screening was done by the method of Wood *et al.* (26), with the final wash done at 52°C for the 3′ probe and 45°C for the 5′ probe.

Sequencing. Dideoxy sequencing (27) was done on alkali-denatured supercoiled plasmids (28) with modified T7 DNA polymerase (Sequenase) from US Biochemical and [α-[^{35}S]thio]dATP from New England Nuclear. A series of deletions was made from each Bluescript plasmid to enable complete sequencing of the cDNA inserts. Deletions were made by following the protocol in the Stratagene manual (29) with two modifications. Exonuclease III (Bethesda Research Laboratories) reactions were done in 66 mM Tris, pH 8/6 mM $MgCl_2$/10 mM 2-mercaptoethanol. S1 nuclease (Bethesda Research Laboratories) was used in place of mung bean nuclease and the reaction buffer was 50 mM NaOAc, pH 4.5/50 mM NaCl/6 $ZnSO_4$. The DNA deletion clones were transformed into *Escherichia coli* JM109 or DH5α. Sequencing templates were prepared by a miniprep method (30). Clones pCE5V, BCE45V, and BCE68V, 256 base pairs (bp) from the 5′ end of BCE5′6V including 90 bp of overlap with BCE45V, and 569 bp from the 3′ end of BCE3′8V-1 including 190 bp of overlap with BCE68V were sequenced on one strand. Successive deletions overlapped by at least 17 bp. Each clone was sequenced at least twice. Ambiguous regions and regions of difference with previously published sequence (23) were confirmed either by sequence from the other strand or by their presence in two independently isolated clones.

RESULTS AND DISCUSSION

Cloning and Sequencing. Previously, we reported the isolation and characterization of three vinculin cDNAs from a λgt11 library made from a single 10-day chicken embryo (22). These cDNAs contained sequences accounting for 80% of the vinculin coding region. Overlapping clones that contain the rest of the vinculin coding sequence have now been isolated from the same library by using oligonucleotides as probes (Fig. 1). The screen with V3 resulted in three new cDNA fragments; one, λCE3′8V, contained the rest of the 3′ coding region. The screen with V5 identified one clone, λCE5′6V. DNA sequence analysis showed that λCE5′6V contained the same amino-terminal coding sequence reported by Price *et al.* (23), as well as some 5′ noncoding sequence (Fig. 1). The cDNA-derived amino-terminal sequence of vinculin was confirmed by direct protein sequencing of

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04126).

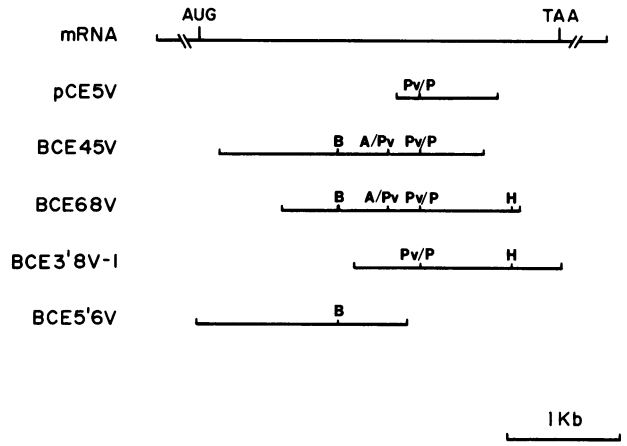

FIG. 1. cDNA clones used to obtain the sequence of vinculin mapped with respect to each other and the 6.2-kb mRNA (A, *Acc* I; B, *Bam*HI; H, *Hae* II; P, *Pst* I; Pv, *Pvu* II). Approximately 60,000 plaque-forming units were screened to obtain BCE5'6V and BCE3'8V-1.

vinculin isolated from a $NaDodSO_4$/polyacrylamide gel. Residues 2–20 of the cDNA-derived amino acid sequence were obtained in 10–20% yield (data not shown).

The amino acid sequence for chicken embryo vinculin was derived from the cDNA sequence (Fig. 2). The protein has a calculated molecular mass of 116,990 Da, only slightly different from the 130 kDa estimated from $NaDodSO_4$/polyacrylamide gel electrophoresis (1). A search of the National Biomedical Research Foundation (NBRF) protein sequence data base (April 14, 1988) with the program FAST-P (31), which aligns sequences optimally by inserting gaps, revealed no significant similarities to proteins in that data bank. A variety of protein sequences were retrieved that contained a proline-rich region but were otherwise not obviously related to each other or to the vinculin sequence. Possibly, proline-rich regions in proteins serve some general structural function that has not yet been defined. Our search also found that the amino acid sequence of vinculin at residues 139–969 showed 14.5% identity to the rod portion of nematode myosin (32). This most likely reflects the high content of α-helical structure in the two proteins (33).

The cDNA sequence and derived amino acid sequence presented in Fig. 2 are almost identical to the partial sequence (amino-terminal 80%) of a vinculin cDNA from chicken embryo fibroblasts (23). In the 2.6 kb of overlap in the coding region, there are only 11 nucleotide differences, 7 of which do not result in amino acid differences. Two of the nucleotide differences result in 2 amino acid differences (Gln-701 → His; Glu-880 → Lys). The two other nucleotide differences are 20 bp apart and result in a frameshift that gives rise to a stretch of 6 different amino acids (Fig. 2, nucleotides 1322–1341). To confirm the differences between our sequence and that of Price *et al.* (23), we sequenced two independent cDNA isolates that span the nucleotide differences. The sequences reported here were obtained in both isolates. Some of these nucleotide differences may be the result of mistakes made by reverse transcriptase, which occur at a frequency of 1/6000 (Bethesda Research Laboratories) during the production of the libraries; but most of the differences probably represent sequence polymorphisms between chickens or amino acid differences between vinculin isoforms.

Analysis of Vinculin Primary Structure. To gain insight into possible structure–function relationships, we used a variety of computer programs to analyze the primary structure of vinculin. Only those observations that are unique to our analysis or that extend those reported by Price *et al.* (23) are discussed.

The composition of the derived amino acid sequence is similar to that obtained by Evans *et al.* (34) from analysis of vinculin isolated from chicken gizzard. In particular, those amino acids that are found at significantly different levels in vinculin, as compared to the average for vertebrate proteins, are in excellent agreement (Table 1). This agreement between theoretical and experimentally determined amino acid composition is evidence, in addition to that already presented (22), that these cDNAs code for authentic vinculin.

Two provocative observations were made by examining the linear distribution of individual amino acids or their properties along the linear sequence. The most obvious feature is the nonrandom distribution of proline. Although the percent proline in vinculin is 4.9, which is the average for a vertebrate protein, there is a stretch of 42 amino acids (residues 837–878) of which 45% are proline. Of the other 23 amino acids in this region, 5 are glutamic acid and 3 are aspartic acid. The proline-rich region is also relatively hydrophilic (Fig. 3, hatched region), indicating that it is exposed to the solvent and not buried in the interior of the protein. These two characteristics of the proline-rich region make it a good candidate for the cleavage site of *Staphylococcus aureus* V8 protease, which cuts vinculin into two fragments. The calculated molecular masses of the two fragments created by cleavage within the proline-rich region are ≈94 kDa and ≈23 kDa, which are not very different from the observed sizes of 81–100 kDa and 27–30 kDa (36, 37).

A second intriguing feature was found by using the program CHARGEPRO from PC GENE to calculate apparent isoelectric points (pI) for the primary sequence and different subsegments (Table 2). The apparent pI for the entire sequence is 5.9; the observed pI for vinculin is 6.2–6.5 (34). If the amino acid sequence is cleaved at a potential V8 protease site in the proline-rich region, Glu-857, the calculated pI values for the two fragments are strikingly different (Table 1). The amino-terminal domain is more acidic (pI 5.4) than the intact molecule (pI 5.9), whereas the carboxyl-terminal domain is much more basic (pI 9.7). Although CHARGEPRO assumes that all ionizable groups are solvent-accessible, this is a reasonable assumption for the extended tail of vinculin. The basic carboxyl-terminal domain of the vinculin molecule may explain how vinculin interacts specifically and apparently electrostatically with acidic phospholipids (11, 12).

The analysis of the primary structure suggests the hypothesis that the proline-rich region separates the two domains of vinculin observed by rotary shadowing and partial proteolytic digestion with *S. aureus* V8 protease (36). The acidic, amino-terminal domain of ≈94 kDa may represent the globular, 81- to 100-kDa V8 fragment that retains talin-binding activity (8), and the basic, carboxyl-terminal domain of ≈23 kDa may represent the extended 20-nm tail and the 27- to 30-kDa fragment of V8 protease digestion. Preliminary isoelectric focusing analysis of the V8 proteolytic fragments supports this initial assignment (Candy DeBerry and S.W.C., unpublished data).

A functional role for the carboxyl-terminal domain is suggested by its basicity compared to the remainder of the molecule and by complete conservation of the carboxyl-terminal 96 amino acids between chicken and human vinculin (V. Koteliansky, personal communication). Perhaps a change in the basicity of this domain by phosphorylation alters the conformation of vinculin and thereby regulates the accessibility of binding sites for other molecules. With the primary structure of vinculin and corresponding cDNA clones available, it is now feasible to locate the sites of phosphorylation (11, 15, 16) and to determine the functional and structural consequences of their modification.

Analogous to other proteins recently classified as amphitropic (38), vinculin is not a typical membrane protein; it is

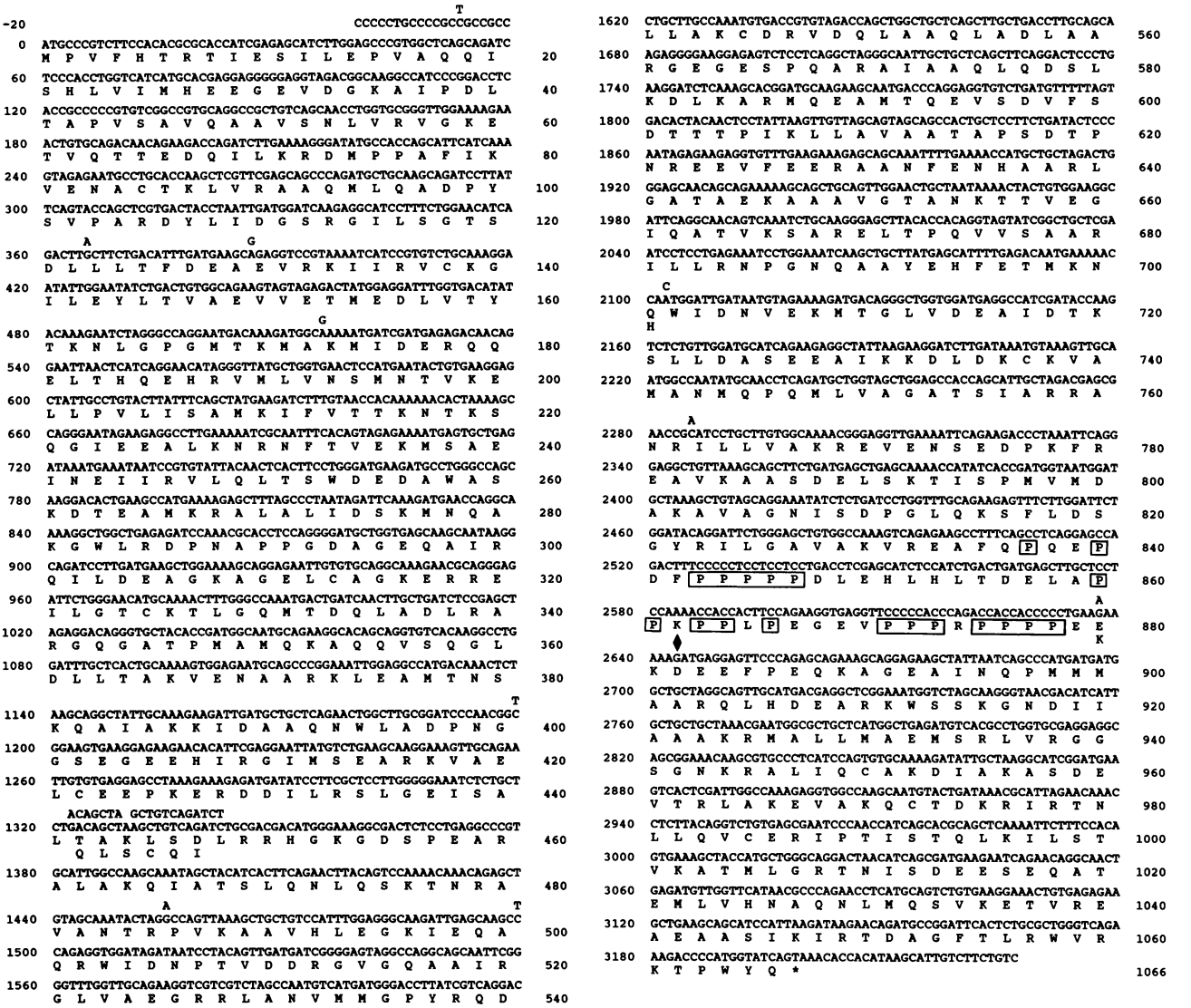

FIG. 2. Nucleotide and derived amino acid sequence of a chicken embryonic vinculin. Nucleotides are numbered on the left and amino acids (standard one-letter symbols) on the right. Differences from the partial sequence of chicken embryo fibroblast vinculin (23) are indicated above and below the main sequence. The end of the published sequence for fibroblast vinculin is marked (♦). Prolines in the proline-rich region are boxed.

solubilized in low-salt buffers in the absence of detergent. As anticipated, analysis of the hydropathicity index by the Kyte and Doolittle algorithm (35) using windows of 9 or 15 amino acids gives no evidence of a segment that is sufficiently hydrophobic over a long enough stretch to be a transmembrane segment (35). However, if vinculin associates reversibly with junctional membrane by interaction with the lipid bilayer, there are two possible mechanisms for this to occur in addition to the interaction with acidic phospholipids discussed earlier. First, the hydropathy plot (Fig. 3) shows two regions in the amino-terminal domain that, if exposed, are hydropathic enough and long enough to insert partially into the lipid bilayer (35). Second, the reported myristoylation (13) and palmitoylation (14) of vinculin could be the basis of amphitropic behavior. With regard to myristoylation it is disappointing that neither cultured chicken fibroblast vinculin (23) nor chicken embryonic vinculin contains amino-terminal glycine, an apparent requirement for myristoylation (reviewed in ref. 39). Identification of the site of myristoylation in vinculin may yield additional information on the requirements for this modification. Alternatively, there may be other isoforms of vinculin that do have amino-terminal glycine.

Palmitoylation occurs most frequently on cysteines and often on those followed by two aliphatic amino acids (reviewed in ref. 39). There are 10 cysteines in chicken embryonic vinculin, all of which are followed by two aliphatic amino acids; 3 cysteines are in the carboxyl-terminal domain and are probably accessible. Palmitoylation of vinculin is modulated upon transformation of chick fibroblasts with Rous sarcoma virus (14), an event that also alters vinculin distribution (5).

Analysis of Secondary Structure. The NOVOTNY and GARNIER programs from PC GENE were used to display the secondary-structure potential of the amino acid sequence. The success of these programs in predicting the secondary structural features of soluble, globular proteins that have been crystallized is only 55% (40). Therefore the programs cannot be used to discover the true secondary structure of a protein. However, the programs have been useful in defining repeating motifs of secondary structure, or domains of proteins with distinctive secondary structural properties. For vinculin, however, the programs predicted a homogenous profile of 18- to 30-residue α-helices separated by β-turns and short regions of extended or random-coil structure. There was no predicted secondary structural difference between the

Table 1. Comparison of the amino acid composition of vinculin with the average composition of all vertebrate proteins

Amino acid	mol % Chicken embryo vinculin*	Vertebrate proteins†	Chicken gizzard vinculin‡
Ala	12.7	7.2	12.1 ± 0.2
Leu	8.8	8.7	8.7 ± 0.1
Glu	8.6	6.8	ND
Lys	7.0	6.5	7.1 ± 0.2
Arg	6.3	4.9	6.5 ± 0.1
Val	6.2	6.8	6.3 ± 0.1
Thr	6.0	5.6	6.1 ± 0.2
Asp	5.7	5.4	ND
Gln	5.4	3.9	ND
Ile	5.3	5.0	5.1 ± 0.1
Gly	5.0	7.6	5.5 ± 0.1
Pro	4.9	4.9	5.4 ± 0.1
Ser	4.9	7.0	5.4 ± 0.2
Met	3.6	2.3	3.7 ± 0.1
Asn	3.6	4.1	ND
Phe	1.4	4.1	1.6 ± 0.1
His	1.3	2.2	1.4 ± 0.1
Cys	0.9	2.5	1.3 ± 0.2
Trp	0.8	1.3	0.7
Tyr	0.7	3.2	0.9 ± 0.1

Underlined vinculin values differ significantly from the average for all vertebrate proteins.
*Calculated from cDNA-derived amino acid sequence.
†Data from ref. 33.
‡Data from ref. 34.

regions amino-terminal and carboxyl-terminal to the proline-rich domain.

Attempts to relate the primary and predicted secondary structure to the shape of vinculin highlight the need for additional structural studies. In rotary-shadowed images, vinculin is shaped like a balloon on a string, with the head ≈80 Å in diameter and the tail ≈200 Å in length (36). If the electron microscopic image is formed by a monomer of vinculin as is thought currently, then the head cannot be a solid sphere. The calculated mass of an 80 Å sphere is 225 kDa, whereas a monomer of vinculin has a calculated mass of only 117 kDa and part of this must be used for the 200-Å tail. If the rotary shadowed image is produced by a single 117-kDa chain, then the head must be either hollow or disc-shaped.

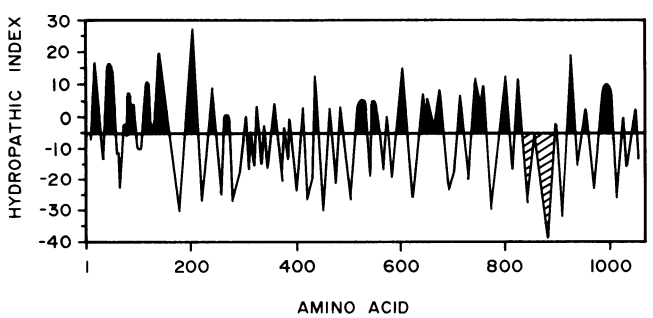

FIG. 3. Kyte and Doolittle (35) hydropathy plot of chicken embryo vinculin. The plot was generated with the SOAP program from PC GENE using a window of 15 residues. The horizontal line at −4.22 indicates the average hydropathic index of the entire sequence. Full scale on the y axis extends from −50 for most hydrophilic to +50 for the most hydrophobic sequence. With a 15-amino acid window, a potential membrane-spanning region would have a hydrophobic value >20. Hatched region represents the proline-rich domain between amino acids 837 and 878; black areas are the portions of vinculin having higher than average hydrophobicity; open areas are the regions of vinculin that are more hydrophilic than average.

Table 2. Theoretical domains of acidic and basic charge in vinculin

Segment	Residues	pI (calc.)*
Vinculin	1–1066	5.9
V8 protease head	1–857	5.4
V8 protease tail	858–1066	9.7

*Calculated values. The observed pI for chicken gizzard vinculin is 6.2–6.5 (34).

Another problem is to account for the 200-Å extended tail. If this tail is formed by the 23-kDa piece that is removed by V8 protease digestion as published data suggests (36), then there are about 200 amino acids in the tail. If we assume that 200 amino acids of a monomer form the 200-Å tail, the structure would have to be mostly that of a linearly extended α-helix. There have been only two examples described of helices that exist as fully solvent-exposed structures: a 28-amino acid stretch of calmodulin (41), and a 31-amino acid stretch of troponin C (42). These helices, which are rich in charged amino acids, are thought to be stabilized by the formation of salt bridges between adjacent residues.

In general, helices in proteins are stabilized by interaction with other helices. Therefore, it is more likely that the tail is a double chain, with the adjacent helices stabilized by their interaction with each other. Theoretically, the double chain could arise from a turn in the single 117-kDa polypeptide, although it is not obvious where this fold might occur. In this structure, even less mass would then be available for the head.

A simple alternative to the above models is that the rotary-shadowed image reflects a dimer of vinculin. If this were the case, parallel helices in the tail, each of which has a significant predicted amphiphilic character, could stabilize each other by interaction at their hydrophobic faces.

We thank Dr. Mario Amzel for consultation on protein structure predictions, Dr. Sam Casella for use of his computer and programs, Yolanda Kaczworoski for help in sequencing, Peggy Ford for typing, and Barry Byrne and Candy DeBerry for reading and improving this manuscript. The amino-terminal protein sequencing was done under contract by Craig Miles, National Jewish Hospital and Research Center. This work was supported by Grant GM41605 from the National Institutes of Health to S.W.C.

1. Geiger, B. (1979) *Cell* **18**, 193–205.
2. Volberg, T., Geiger, B., Kartenbeck, J. & Franke, W. W. (1986) *J. Cell Biol.* **102**, 1832–1842.
3. O'Keefe, E. J., Briggaman, R. A. & Herman, B. (1987) *J. Cell Biol.* **105**, 807–817.
4. Volberg, T., Sabanay, H. & Geiger, B. (1986) *Differentiation* **32**, 34–43.
5. David-Pfeuty, T. & Singer, S. J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6687–6691.
6. Herman, B. & Pledger, W. J. (1985) *J. Cell Biol.* **100**, 1031–1044.
7. Halegoua, S. (1987) *Dev. Biol.* **121**, 97–104.
8. Burridge, K. & Mangeat, P. (1984) *Nature (London)* **308**, 744–746.
9. Belkin, A. M. & Koteliansky, V. E. (1987) *FEBS Lett.* **220**, 291–294.
10. Wacksstock, D. H., Wilkins, J. A. & Lin, S. (1987) *Biochem. Biophys. Res. Commun.* **146**, 554–560.
11. Ito, S., Werth, D. K., Richert, N. D. & Pastan, I. (1983) *J. Biol. Chem.* **258**, 14626–14631.
12. Niggli, V., Dimitrov, D. P., Brunner, J. & Burger, M. M. (1986) *J. Biol. Chem.* **261**, 6912–6918.
13. Kellie, S. & Wigglesworth, N. M. (1987) *FEBS Lett.* **213**, 428–432.
14. Burn, P. & Burger, M. M. (1987) *Science* **235**, 476–479.
15. Werth, D. K., Niedel, J. E. & Pastan, I. (1983) *J. Biol. Chem.* **258**, 11423–11426.
16. Sefton, B. M., Hunter, T., Ball, E. H. & Singer, S. J. (1981) *Cell* **24**, 165–174.

17. Geiger, B. (1982) *J. Mol. Biol.* **159,** 685–701.
18. Belkin, A. M., Ornatsky, O. I., Kabakov, A. E., Glukhova, M. A. & Koteliansky, V. E. (1988) *J. Biol. Chem.* **263,** 6631–6635.
19. Endo, T. & Masaki, T. (1982) *Biochem. Biophys. Res. Commun.* **107,** 1467–1474.
20. Siliciano, J. D. & Craig, S. W. (1982) *Nature (London)* **300,** 533–535.
21. Feramisco, J. R., Smart, J. E., Burridge, K., Helfman, D. M. & Thomas, G. P. (1982) *J. Biol. Chem.* **257,** 11024–11031.
22. Coutu, M. D., Simon, D. J., Brown, A. E. & Craig, S. W. (1987) *Biochem. Biophys. Res. Commun.* **147,** 637–643.
23. Price, G. J., Jones, P., Davison, M. D., Patel, B., Eperon, I. C. & Critchley, D. R. (1987) *Biochem. J.* **245,** 595–603.
24. Bendori, R., Salomon, D. & Geiger, B. (1987) *EMBO J.* **6,** 2897–2905.
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
26. Wood, W. I., Gitschier, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1585–1588.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
28. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
29. Stratagene (1986) *Bluescript Exo/Mung DNA Sequencing System: Instruction Manual* (Stratagene, San Diego, CA), pp. 12–15.
30. Promega (1986) *GemSeq K/RT Sequencing System: Technical Manual* (Promega, Madison, WI) pp. 13–14.
31. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
32. Karn, J., Brenner, S. & Barnett, L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4253–4257.
33. Doolittle, R. F. (1986) *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences* (University Science Books, Mill Valley, CA).
34. Evans, R. R., Robson, R. M. & Stromer, M. H. (1984) *J. Biol. Chem.* **259,** 3916–3924.
35. Doolittle, R. F. & Kyte, J. (1982) *J. Mol. Biol.* **157,** 105–132.
36. Milam, L. (1985) *J. Mol. Biol.* **184,** 543–545.
37. Otto, J. J., Groesch, M. & Beltzer, D. (1987) *J. Cell Biol.* **105,** 131a (abstr.).
38. Burn, P. (1988) *Trends Biochem. Sci.* **13,** 79–83.
39. Sefton, B. M. & Buss, J. E. (1987) *J. Cell Biol.* **104,** 1449–1453.
40. Nishikawa, K. (1983) *Biochim. Biophys. Acta* **748,** 285–299.
41. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. & Cook, W. J. (1985) *Nature (London)* **315,** 37–40.
42. Herzberg, O. & James, M. N. G. (1985) *Nature (London)* **313,** 653–659.