Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2748–2751, March 1994
Cell Biology

A family of cellular proteins related to snake venom disintegrins

(cell adhesion/integrin ligand/metalloprotease)

GISELA WESKAMP AND CARL P. BLOBEL*

Cellular Biochemistry and Biophysics Program, Memorial Sloan–Kettering Cancer Center, New York, NY 10021

Communicated by James E. Rothman, December 9, 1993 (received for review December 6, 1993)

ABSTRACT **Disintegrins are short soluble integrin ligands that were initially identified in snake venom. A previously recognized cellular protein with a disintegrin domain was the guinea pig sperm protein PH-30, a protein implicated in sperm–egg membrane binding and fusion. Here we present peptide sequences that are characteristic for several cellular disintegrin-domain proteins. These peptide sequences were deduced from cDNA sequence tags that were generated by polymerase chain reaction from various mouse tissues and a mouse muscle cell line. Northern blot analysis with four sequence tags revealed distinct mRNA expression patterns. Evidently, cellular proteins containing a disintegrin domain define a superfamily of potential integrin ligands that are likely to function in important cell–cell and cell–matrix interactions.**

Cell–cell and cell–matrix interactions are of vital importance for the development and maintenance of an organism. During embryogenesis and morphogenesis, cells must move relative to one another, adhere, communicate, and in certain cases fuse. These events are mediated by cell surface proteins, many of which can be grouped into protein families such as integrins (1), cadherins (2), selectins (3, 4), and immunoglobulin superfamily proteins (4). The sequences of another potential type of integrin ligand, the heterodimeric guinea pig sperm protein PH-30 (5) have recently been reported. The β subunit of mature PH-30 contains a membrane-anchored disintegrin domain which is highly related to soluble snake venom integrin ligands called disintegrins (6–9). PH-30 is likely to interact with an integrin on the egg plasma membrane (5, 10, 11). The PH-30 α subunit is related to PH-30 β and contains a potential fusion peptide that may trigger membrane fusion (5, 12).

The high level of sequence conservation between the sperm protein PH-30 and snake venom disintegrins implied that a family of cellular proteins related to snake venom disintegrins should exist (10). To test this hypothesis, we systematically searched for cellular disintegrins by PCR in several mouse tissues and a mouse muscle cell line. We have identified eight sequence tags that are characteristic for the disintegrin protein family. Discovery of these sequence tags, together with other recently reported disintegrin proteins (13, 14), firmly establishes the existence of a cellular disintegrin protein superfamily with a variety of possible functions in cell–cell and cell–matrix binding and signaling and in cell–cell fusion.†

MATERIALS AND METHODS

Mouse Tissues and mRNA Isolation. BALB/c mice were obtained from the transgenic breeding facility at the Sloan–Kettering Institute, and were euthanized by cervical dislocation. Total cellular RNA was isolated from mouse lung, muscle, testis, and spleen as described (15).

Cells. C2 mouse muscle cells (16) were obtained from R. Benezra (Memorial Sloan–Kettering Cancer Center) and grown in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum and 0.2% chicken embryo extract (GIBCO/BRL) for 2 days. The cells were then switched to DMEM with 2% horse serum (GIBCO/BRL) to induce myogenic differentiation. RNA was isolated from cells that were ≈20% fused.

cDNA Preparation and PCR. cDNA was reverse transcribed from RNA isolated from mouse tissues or C2 mouse muscle cells (Superscript kit, BRL). PCR was performed with first-strand cDNA, the degenerate primers listed below, and reagents and *Taq* polymerase from Boehringer Mannheim. Primers were designed from conserved disintegrin protein domains (Fig. 1): sense primer for the metalloprotease region, primer A (5′-ATG-ACN-CA-Y-GAR-ATG-GGN-CAY-AA-3′); sense primers for the disintegrin domain; primers B1 (5′-GAR-GGN-GAR-GAY-TGY-GAY-TG-3′) and B2 (5′-GGN-GAR-GAY-TGY-GAY-TGY-GG-3′); antisense primers for the disintegrin domain; primers C1 (5′-CA-RTA-YTC-NGG-NAR-RTC-RCA-3′) and C2 (5′-TA-YTC-NGG-NAR-RTC-RCA-YTC-3′). Primary PCRs were performed with primer pairs A/C1 and B1/C1. Nested PCRs were performed on 1:300 diluted primary reaction mixtures with primers B2 and C2. Primers B2 and C2 were synthesized with an *Eco*RI site (GGAATTCC) at their 5′ end for subcloning of PCR fragments into the *Eco*RI site of a pBluescript vector (Stratagene). PCR products were electrophoresed in 5% polyacrylamide gels with Tris/borate/EDTA buffer, and bands of about 180–200 bp were excised, eluted overnight in water, preciptated, cut with *Eco*RI, and subcloned into a pBluescript vector. Individual subclones were sequenced by the dideoxy chain-termination method using Sequenase enzyme (United States Biochemical). Sequences were analyzed with MACVECTOR computer software (Kodak/IBI).

Northern Blot Analysis. RNA blots were performed as follows: 15 μg of total cellular RNA isolated as described above was run per lane in a 1% agarose gel in Mops/formaldehyde buffer, transferred to a Nytran membrane (Schleicher and Schuell) by capillary transfer, and crosslinked by UV irradiation (19). RNA blots were prehybridized at 60°C for 3–5 hr in 6× standard saline/citrate (SSC)/10% dextran sulfate/0.1% SDS/10× Denhardt's solution and hybridized at 60°C for 18 hr in 6× standard saline/phosphate/EDTA (SSPE)/10% dextran sulfate/40% formamide/1% SDS containing ^{32}P-labeled cDNA probe (2×10^6 cpm/ml). Probe synthesis was performed with random primers, and with gene-specific primers B2 and C2 (see above), in a labeling reaction on disintegrin cDNA. Washes were carried out at 60°C in 0.1× SSC/0.1% SDS. The blots were exposed to x-ray film (Kodak XAR) with an intensifying screen at −70°C.

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.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank database (accession nos. U06144–U06151).

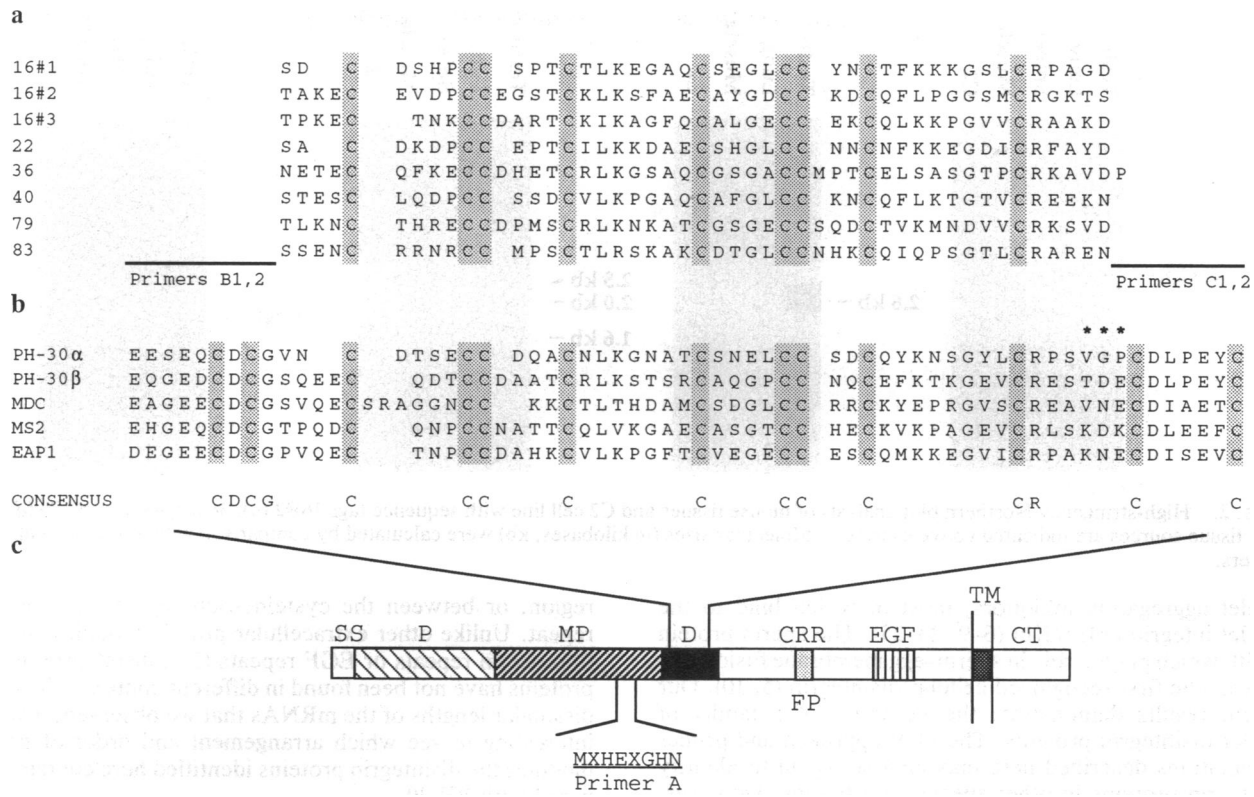

FIG. 1. Alignment of newly identified (*a*) and previously known (*b*) characteristic disintegrin peptide sequences, and a diagram of disintegrin protein domains (*c*). The sequence alignment in *b* shows most of the disintegrin domain of five known cellular disintegrin proteins. They are, from top to bottom, PH-30 α and β (5, 17), the candidate breast cancer tumor-suppressor MDC (13), the mouse monocyte disintegrin MS2 (18), and the rat epididymal epithelial protein EAP1 (14). The relative position of the conserved disintegrin domain in a membrane-anchored cellular disintegrin protein is indicated below the sequence alignment. The model domain structure (*c*) is that of PH-30 α, where a signal sequence (SS) is followed by a pro domain (P), a metalloprotease domain (MP), a disintegrin domain (D), a cysteine-rich region (CRR) with a fusion peptide (FP), an epidermal growth factor repeat (EGF), a transmembrane domain (TM), and a cytoplasmic tail (CT). The predicted binding site, corresponding to the position of the sequence Arg-Gly-Asp (RGD) in snake venom disintegrins (10), is indicated by stars (*b*). The metalloprotease domains of PH-30 β and the MDC protein are structurally related, but do not have the metalloprotease consensus sequence HEXXH (13, 17). The conserved disintegrin peptide regions that were used for degenerate primer design are indicated by lines above the known cellular disintegrins. The conserved sequence in the metalloprotease domain that was used for primer design is shown below the diagram of the metalloprotease region. The eight disintegrin sequence tags in *a* are shown aligned with the known disintegrin domains in *b*. The observed spacing of cysteines and some additional amino acids is characteristic for disintegrin-domain proteins.

RESULTS

We applied a nested PCR approach to search for disintegrins in various mouse tissues and in C2 cells (16), a mouse cell line capable of myogenic differentiation. Degenerate oligonucleotide primers were designed based on conserved sequences in known disintegrin proteins (Fig. 1*b*). As the overall degree of sequence identity among these proteins is low, few regions lend themselves to the design of conserved primers. The alignment shown in Fig. 1*b* represents one of the most conserved disintegrin protein domains. We used a sequence near the amino terminus of the disintegrin domain to design nested sense-orientation primers (B1 and B2; Fig. 1*b*), and a second, less conserved region near the carboxyl terminus of the disintegrin domain to design nested antisense primers (C1 and C2; Fig. 1*b*). We also synthesized a degenerate primer corresponding to the metalloprotease consensus sequence HEXGHN present in many disintegrin proteins (refs. 17, 20–22; Fig. 1*c*). First-strand cDNA was reverse transcribed from mRNA isolated from several mouse tissues (testis, lung, spleen, and skeletal muscle) and from C2 mouse muscle cells. We performed two separate sets of primary PCRs, (A/C1 or B1/C1) and reamplified the products with the nested primers B2 and C2. Resulting PCR fragments in the predicted size range (180–200 bp) were excised, subcloned, sequenced, and translated into peptide sequences. These were scored for features that are diagnostic for disintegrin sequences: cysteine residues in a characteristic position and spacing that is unique to disintegrins, as well as additional conserved amino acids (Fig. 1). We found eight previously unknown disintegrin sequence tags (Fig. 1*a*). All of these characteristic sequences were generated by nested PCR from a primary reaction with primers A and C1, implying the presence of a metalloprotease consensus sequence.

Next we analyzed the expression pattern of four sequence tags by high-stringency Northern blot analysis of mRNA from mouse tissues and from C2 cells (Fig. 2). Sequence tag 16#2 hybridized well to an mRNA of 4.2 kb in mouse lung and gave a weak signal at 4.2 kb with mRNA from C2 muscle cells (Fig. 2*a*). Sequence tags 36 and 79 both hybridized strongly to bands in testis mRNA of 2.6 kb and 1.7 kb, respectively (Fig. 2*b* and *d*). Sequence tag 83 hybridized to three mRNA bands from C2 cells, at 2.5, 2.0 and 1.6 kb, which may be spliced mRNA variants (Fig. 2*c*). The length and different tissue distribution of the detected mRNAs implies that each of the tested sequence tags hybridizes to a different mRNA species.

DISCUSSION

We have systematically searched for cellular disintegrin proteins and have identified eight sequence tags in mouse that are characteristic for members of this protein family. Disintegrins were first isolated as short, soluble snake venom

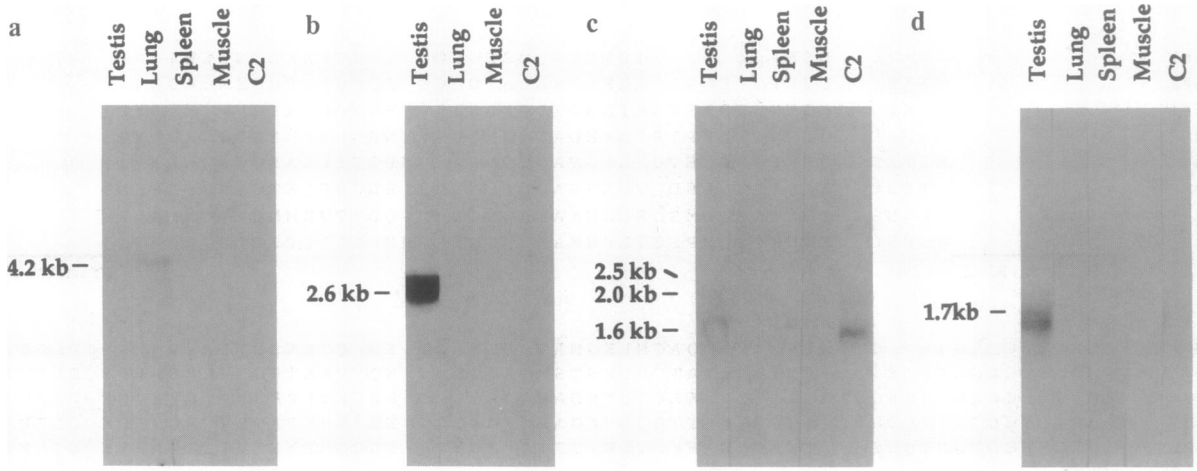

FIG. 2. High-stringency Northern blot analysis of mouse tissues and C2 cell line with sequence tags 16#2 (*a*), 36 (*b*), 83 (*c*), and 79 (*d*). The RNA tissue sources are indicated above each lane. Molecular sizes (in kilobases, kb) were calculated by comparison with RNA molecular size markers.

platelet aggregation inhibitors, most of which bind to the platelet integrin gpIIb/IIIa (6–9, 23–25). The sperm protein PH-30, which plays a role in sperm–egg membrane fusion (26, 27), was the first recognized cellular disintegrin (5, 10). Our present results demonstrate the existence of a family of cellular disintegrin proteins. The PCR approach and primer combinations described here may also be useful to identify disintegrin proteins in other species and tissues. However, due to the low degree of sequence conservation between disintegrin proteins, these particular primer combinations can only be designed to detect a subset of the few presently known disintegrin proteins. We therefore predict that many more members of this protein superfamily are likely to exist that are not identifiable with this primer combination.

Northern blot analysis of the expression pattern of four disintegrin sequences revealed that two were expressed in testis, one in C2 cells and in lung, and one in C2 cells only (Fig. 2). The relationship between PH-30 α or β and the two disintegrins expressed in testis is unclear, as their sequences are not highly related. Our finding that both sequence tags expressed in C2 muscle cells were not detected on a Northern blot of adult muscle may indicate a role of these two disintegrin proteins in myogenesis rather than in mature muscle. The expression pattern of sequence 16#2 in C2 cells and in lung implies that, like certain integrins, individual disintegrins may also function in a number of different tissues. Additional cellular disintegrin proteins have been reported in epididymal epithelial cells (14), in mouse monocytes (18) and in the nematode *Caenorhabditis elegans* (B. Podbilewicz, personal communication), thus further expanding this protein superfamily. Recently a gene encoding a metalloprotease/disintegrin, MDC, was identified as a candidate breast cancer tumor-supressor gene on chromosome 17q21.3 (13).

What can known disintegrin proteins imply about the domain structure of the cellular disintegrin proteins we have identified? All membrane-anchored disintegrin proteins known to date have a very similar arrangement of protein modules (Fig. 1*c*): along with a disintegrin domain, they contain at their amino terminus a metalloprotease domain related to snake venom metalloproteases, a pro domain, and a signal sequence. Carboxyl-terminal to the disintegrin domain is a cysteine-rich region, an epidermal growth factor (EGF) repeat, a transmembrane domain, and a cytoplasmic tail (10, 13, 17). The protein domains present in some larger snake venom proteins (20, 21, 28–30), and in the MDC protein (13) are found in the same order as in membrane-anchored disintegrins (10, 17). However, these sequences are truncated either between the disintegrin domain and the cysteine-rich region, or between the cysteine-rich region and the EGF repeat. Unlike other extracellular protein modules, such as fibronectin repeats or EGF repeats (31), disintegrin-domain proteins have not been found in different contexts. Given the dissimilar lengths of the mRNAs that we observed, it will be interesting to see which arrangement and order of protein modules the disintegrin proteins identified here contain compared with PH-30.

What are the possible functions of cellular disintegrin proteins? Snake venom disintegrins are proven integrin ligands (6–9, 23–25), so cellular proteins with a disintegrin domain are also likely to be integrin ligands (10). Depending on the presence or absence of a membrane anchor, the function of cellular disintegrin proteins may be very different. A membrane-anchored disintegrin domain implies a role in integrin-mediated cell–cell membrane binding (10). The sperm membrane protein PH-30 β is thus thought to bind to one of several integrins present on the egg plasma membrane (5, 11, 32). If a membrane-anchored disintegrin contains a hydrophobic fusion peptide or is associated with a protein containing a fusion peptide, it might play a role in cell–cell fusion (5, 12).

Cellular disintegrin proteins lacking a membrane anchor, such as the candidate breast cancer tumor-suppressor MDC (13), could have one of two functions in adhesion. In analogy to soluble snake venom disintegrins, they may competitively inhibit other integrin–ligand interactions, perhaps leading to de-adhesion of cells. However, soluble disintegrins may also function as ligands that are immobilized on the extracellular matrix or cell surface. Besides a role in adhesion, membrane-anchored, soluble, and extracellular matrix-bound disintegrin proteins may trigger a signal through their receptor (10, 11). Finally, because the conserved domain structure of known disintegrin proteins usually contains a metalloprotease domain as well as a disintegrin domain, and since the primary PCR was carried out with a primer that encodes the metalloprotease consensus sequence HEXXH (17, 20–22), the disintegrin sequence tags are likely to be part of a metalloprotease gene.

In conclusion, the disintegrin sequence tags presented here, together with other known disintegrin proteins, clearly define a protein superfamily. Identification of these characteristic sequence tags thus provides the critical first step toward understanding the role of these cellular disintegrins in cell–cell and cell–matrix interactions.

We thank Martha S. Reid for excellent technical assistance and Stephanie Woods for assistance in preparing the manuscript. This

research was supported by Cancer Center Support Grant P30-CA08748 from the National Cancer Institute.

1. Hynes, R. O. (1992) *Cell* **69,** 11–25.
2. Grunwald, G. B. (1993) *Curr. Opin. Cell Biol.* **5,** 797–805.
3. Bevilacqua, M., Butcher, E., Furie, B., Furie, B., Gallatin, M., Gimbrone, M., Harlan, J., Kishimoto, K., Lasky, L., McEver, R., Paulson, J., Rosen, S., Seed, B., Siegelman, M., Springer, T., Stoolman, L., Tedder, T., Varki, A., Wagner, D., Weissman, I. & Zimmerman, G. (1991) *Cell* **67,** 233.
4. Springer, T. A. (1990) *Nature (London)* **346,** 425–434.
5. Blobel, C. P., Wolfsberg, T. G., Turck, C. W., Myles, D. G., Primakoff, P. & White, J. M. (1992) *Nature (London)* **356,** 248–252.
6. Gould, R. J., Polokoff, M. A., Friedman, P. A., Huang, T. F., Holt, J. C., Cook, J. J. & Niewiarowski, S. (1990) *Proc. Soc. Exp. Biol. Med.* **195,** 168–171.
7. Musial, J., Niewiarowski, S., Rucinski, B., Stewart, G. J., Cook, J. J., Williams, J. A. & Edmunds, L. H. (1990) *Circulation* **82,** 261–273.
8. Dennis, M. S., Henzel, W. J., Pitti, R. M., Lipari, M. T., Napier, M. A., Deisher, T. A. & Bunting, S. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2471–2475.
9. Scarborough, R. M., Rose, J. W., Naughton, M. A., Phillips, D. R., Nannizzi, L., Arfsten, A., Campbell, A. M. & Charo, I. F. (1993) *J. Biol. Chem.* **268,** 1058–1065.
10. Blobel, C. P. & White, J. M. (1992) *Curr. Opin. Cell Biol.* **4,** 760–765.
11. Myles, D. G. (1993) *Dev. Biol.* **158,** 35–45.
12. White, J. M. (1992) *Science* **258,** 917–924.
13. Emi, M., Katagiri, T., Harada, Y., Saito, H., Inazawa, J., Ito, I., Kasumi, F. & Nakamura, Y. (1993) *Nat. Genet.* **5,** 151–157.
14. Perry, A. C. F., Jones, R., Barker, P. J. & Hall, L. (1992) *Biochem. J.* **286,** 671–675.
15. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
16. Yaffe, D. & Saxel, O. (1977) *Nature (London)* **270,** 725–727.
17. Wolfsberg, T. G., Bazan, J. F., Blobel, C. P., Myles, D. G., Primakoff, P. & White, J. M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1–5.
18. Yoshida, S., Setoguchi, M., Higuchi, Y., Akizuki, S. & Yamamoto, S. (1990) *Int. Immunol.* **2,** 586–591.
19. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
20. Hite, L. A., Shannon, J. D., Bjarnason, J. B. & Fox, J. W. (1992) *Biochemistry* **31,** 6203–6211.
21. Takeya, H., Oda, K., Miyata, T., Omori-Satoh, T. & Iwanaga, S. (1990) *J. Cell Biol.* **265,** 16068–16073.
22. Gomis-Rüth, F. X., Kress, L. F. & Bode, W. (1993) *EMBO J.* **12,** 4151–4157.
23. Huang, T. F., Holt, J. C., Lukasiewicz, H. & Niewiarowshi, S. (1987) *J. Biol. Chem.* **262,** 16157–16163.
24. Huang, T. F., Holt, J. C., Kirby, E. P. & Niewiarowshi, S. (1989) *Biochemistry* **28,** 661–666.
25. Shebuski, R. J., Ramjit, D. R., Gencen, G. H. & Polokoff, M. A. (1989) *J. Biol. Chem.* **264,** 21550–21556.
26. Blobel, C. P., Myles, D. G., Primakoff, P. & White, J. W. (1990) *J. Cell Biol.* **111,** 69–78.
27. Primakoff, P., Hyatt, H. & Tredick-Kline, J. (1987) *J. Cell Biol.* **104,** 141–149.
28. Au, L. C., Huang, Y. B., Huang, T. F., Teh, G. W., Lin, H. H. & Choo, K. B. (1991) *Biochem. Biophys. Res. Commun.* **181,** 585–593.
29. Kini, R. M. & Evans, H. J. (1992) *Toxicon* **30,** 1–29.
30. Takeya, H., Nishida, S., Miyata, T., Kawada, S., Saisaka, Y., Morita, T. & Iwanaga, S. (1992) *J. Biol. Chem.* **267,** 14109–14117.
31. Engel, J. (1991) *Curr. Opin. Cell Biol.* **3,** 779–785.
32. Tarone, G., Russo, M. A., Hirsch, E., Odorisio, T., Altruda, F., Silengo, L. & Siracusa, G. (1993) *Development* **117,** 1369–1375.