

Neurexin III α : Extensive alternative splicing generates membrane-bound and soluble forms

YURI A. USHKARYOV AND THOMAS C. SÜDHOF*

Howard Hughes Medical Institute and Department of Molecular Genetics, University of Texas Southwestern Medical School, Dallas, TX 75235

Communicated by Michael S. Brown, March 26, 1993

ABSTRACT The structure of neurexin III α was elucidated from overlapping cDNA clones. Neurexin III α is highly homologous to neurexins I α and II α and shares with them a distinctive domain structure that resembles a cell surface receptor. cDNA cloning and PCR experiments revealed alternative splicing at four positions in the mRNA for neurexin III α . Alternative splicing was previously observed at the same positions in either neurexin I α or neurexin II α or both, suggesting that the three neurexins are subject to extensive alternative splicing. This results in hundreds of different neurexins with variations in small sequences at similar positions in the proteins. The most extensive alternative splicing of neurexin III α was detected at its C-terminal site, which exhibits a minimum of 12 variants. Some of the alternatively spliced sequences at this position contain in-frame stop codons, suggesting the synthesis of secreted proteins. None of the sequences of the other splice sites in this or the other two neurexins include stop codons. RNA blot analysis demonstrate that neurexin III α is expressed in a brain-specific pattern. Our results suggest that the neurexins constitute a large family of polymorphic cell surface proteins that includes secreted variants, indicating a possible role as signaling molecules.

The development and function of the nervous system are dependent on an intricate network of cell signaling and cell–cell interaction mechanisms (1). We have identified (2) a family of cell surface receptors named neurexins with characteristics that suggest a role in such cell–cell interactions.

Two genes for neurexins were identified; each gene contains two independent promoters that drive transcription of two classes of mRNAs. The two classes of transcripts encode proteins with different N termini; the larger proteins are referred to as α -neurexins and the shorter are β -neurexins. The expression of the neurexins is brain-specific and concentrated in neurons. Neurexins are highly polymorphic due to extensive differential splicing with the potential to express hundreds of different proteins.

Neurexins have a distinctive domain structure. In the α -neurexins, a signal sequence is followed by three copies of a long repeat that make up >75% of the total protein. Each of the long repeats contains a central epidermal growth factor (EGF) domain flanked by right and left arms that are weakly homologous to each other. Data bank searches (GenBank Release 70 and Protein Identification Resource Release 30) demonstrated that the extracellular repeats in neurexins are related to sequences found in several extracellular matrix proteins that participate in cell–cell interactions during development, including laminin A (3), agrin (4, 5), and slit (6). After the three long repeats, α -neurexins contain an O-linked sugar domain, a single transmembrane region, and a short cytoplasmic tail. The N-terminal sequence of β -neurexins differs from that of the α -neurexins but is spliced into the α -neurexin sequences at the end of the third EGF domain. α -

and β -neurexins have identical C termini including the O-linked sugar region, transmembrane region, and cytoplasmic domain (2).

Structures of the neurexins are suggestive of cell surface receptors. Indeed, the neurexins were originally found because of the identity of the sequence of neurexin I α and sequences from the high molecular weight subunit of the receptor for a presynaptic neurotoxin, α -latrotoxin (7). Immunocytochemistry of rat brain frozen sections revealed that neurexin I is highly enriched in synapses in agreement with the localization of α -latrotoxin (2). These findings suggest that at least neurexins I α and I β are synaptic proteins that, based on their structure and homologies, may represent a class of neuronal cell-interaction molecules (2).

We now report the identification, structure, and expression of neurexin III α .[†] By extending the neurexin family of proteins, our results demonstrate the generality of the features of this protein family: their conserved alternative splicing patterns, brain-specific expression, and similar domain structure with distinct patterns of shared and divergent sequences. The expanded number of expressed neurexins lends further support to a putative role of the neurexins as polymorphic identifier molecules. The surprising finding, however, was that neurexin III α may be produced in secreted forms, suggesting a higher degree of complexity than indicated by the structures of neurexins I and II.

EXPERIMENTAL PROCEDURES

cDNA Cloning and Sequencing. cDNA cloning and mapping were performed as described using ³²P-labeled oligonucleotides as probes (8). DNA sequencing was performed on single-stranded DNA from M13 subclones using the dideoxynucleotide-termination method (9) with fluorescent primers and *Taq* polymerase (Promega). Reaction products were analyzed on an ABI 373A automatic DNA sequencer.

PCRs. First-strand cDNA synthesis was performed in 100 μ l by standard procedures (10) using 10 μ g of poly(A)⁺-enriched RNA from rat forebrain, cerebellum, or spinal cord. PCR mixtures contained 2–5 μ l of cDNA, 2.5 mM MgCl₂, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 20 pmol of each primer, all four dNTPs (each at 200 μ M), and AmpliTaq DNA polymerase (Perkin-Elmer) at 30 units/liter. PCRs were performed in 32 cycles, with the first two cycles at 95°C for 5 min, 60°C for 1 min, and 72°C for 5 min, followed by 30 cycles at the same temperatures but for 1 min, 1 min, and 2 min. Neurexin III was initially identified with two oligonucleotides corresponding to sequences that are highly conserved among neurexins [(nucleotides in parentheses denote redundant positions with multiple nucleotides) GGAGATCTAGT-TGCGGTA CTGTACATGGC(G/A)TAGAG(G/A)AGGAT and AGCAAGCTTGGCACCGTGCCCATCGCCAT-

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.

Abbreviation: EGF, epidermal growth factor.

*To whom reprint requests should be addressed.

[†]The sequences reported in this paper have been submitted to GenBank (accession nos. L14851).

CAA]. Subsequent PCRs to map the extent of alternative splicing were carried out with oligonucleotides I, II, and III in Fig. 3 (I, GCGAAGCTTGTTCATCTGCTGAATGCT; II, GCGGCATGCCACATTCTTGAAAGTTGC; III, GCCTCTAGAACCCGTCTGATTCTGGCT). Two types of PCR experiments were performed: (i) PCRs were performed in duplicate with one labeled primer and one unlabeled primer. The products of such reactions were then analyzed by denaturing PAGE and autoradiography to determine the minimal number of fragments generated that incorporate both primers. This avoids artifacts caused by single-oligonucleotide priming or heteroduplex formation (11). (ii) PCRs were performed with unlabeled primers, analyzed by nondenaturing PAGE, and stained by ethidium bromide. PCR products were then subjected to Southern blot analysis with internal oligonucleotides or excised and cloned into M13 vectors and sequenced. All of the PCR products of Fig. 3 were sequenced except for products C, E, and G, which were identified on the basis of their size in relation to the sequenced other products and of their hybridization with an insert-specific oligonucleotide (ACCAAGTACAGGTAG-GTC).

RNA Blot Analysis. This procedure was performed (2) using total RNA from various rat tissues and single-stranded ^{32}P -labeled DNA probes. Blots were rehybridized with a probe to cyclophilin to control for RNA loading.

RESULTS AND DISCUSSION

Cloning of Neurexin III α . Neurexin III α was initially identified by PCRs performed with oligonucleotides corresponding to conserved sequences in the C termini of neurexins I and II. Analysis of the PCR products by DNA sequencing demonstrated that in addition to characterized neurexins, another neurexin transcript was amplified (referred to hereafter as neurexin III α). An oligonucleotide specific for neurexin III α was then synthesized based on the PCR sequence and used to isolate cDNAs encoding neurexin III α from a rat brain library. To obtain the 5' end of the neurexin III α mRNA, further 5' clones were isolated using additional neurexin III α -specific oligonucleotides.

Twelve overlapping cDNA clones were selected for sequence analysis (Fig. 1). These clones show internal differences, suggesting that neurexin III α , like the other neurexins, is extensively alternatively spliced (see below). The full-length structure of neurexin III α was assembled from the sequences of the cDNA clones and is shown with the deduced amino acid sequence of neurexin III α in Fig. 2.

The mRNA for neurexin III α contains a long 5' untranslated region (>700 bp). The translation initiation codon was identified as the first methionine codon in the sequence that

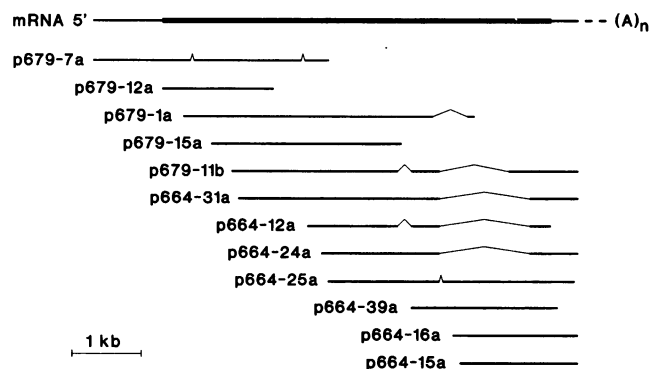


FIG. 1. Structures of the mRNA for neurexin III α and the cDNA clones encoding neurexin III α used in the current study. Sequence segments deleted by alternative splicing are indicated by thin arched lines. The thick line indicates the coding region.

is followed by a long open reading frame. This identification is supported by the fact that the putative initiator methionine is preceded by an in-frame stop codon (underlined in Fig. 2) and that its sequence context conforms well to the consensus site of initiator methionines (12).

Six of the cDNA clones had poly(A) tails at their 3' end. However, these poly(A) tails initiated at three positions in the nucleotide sequence (Fig. 2). None of the poly(A) tails was preceded by a classical polyadenylation consensus sequence. Therefore, it is doubtful that these poly(A) tracts represent the poly(A) tails of the neurexin III α mRNA. The 3' untranslated region of the neurexin III α mRNA is rich in adenosines, suggesting that the poly(A) tails in the cDNAs may represent oligo(dT)-priming sites that are not necessarily poly(A) tracts.

Alternative Splicing of Neurexin III α . The 12 cDNA clones that were sequenced contained four internal positions at which different clones exhibited sequence variations (Figs. 1 and 2). These differences consisted in the presence or absence of a sequence stretch, suggesting that they arose by alternative splicing. Comparison of the sequences of neurexins I α , II α , and III α demonstrates that, at each of these four positions, neurexin I α , neurexin II α , or both are also alternatively spliced (2). This result raises the possibility that the three neurexins are alternatively spliced at similar positions.

A simple pattern of alternative splicing with only two variants was observed in three of the four alternatively spliced sequences based on the cDNA clones. At these positions, a sequence is either included or deleted from the mRNA (Fig. 2). Remarkably, the third of these sequences, a conserved 90-bp sequence in the C-terminal half of the neurexins, is alternatively spliced in all neurexins.

A complex pattern of alternative splicing was observed at the fourth position (Fig. 2). Here, cDNA clones contained five sequences, some of which exhibited in-frame stop codons, which predict the synthesis of secreted proteins without a transmembrane region. To investigate whether these cDNA clones represented cloning artifacts arising from partially processed heterogeneous nuclear RNA, PCR experiments on single-stranded cDNA made from poly(A)⁺-enriched RNA from rat brain were carried out. Two sets of oligonucleotide primers (Fig. 3) were used to amplify the alternatively spliced sequences of the neurexin III α mRNA at this position. PCR products were analyzed on a Southern blot with internal oligonucleotides and by DNA sequencing.

The PCR experiments confirmed and extended the cDNA cloning results by identifying additional splice variants (see Fig. 3). The results from the PCR experiments agreed well with the cDNA cloning data. Thus, evidence for at least 12 splice variants at this site was obtained. PCR is a very sensitive technique that is prone to artifacts. However, the following findings suggest that the C-terminal splice variants containing stop codons correspond to physiological mRNAs: (i) These forms were independently isolated in multiple cDNA clones. In the course of studying neurexins, we have isolated >100 neurexin cDNAs. None of the sequenced splice sites of these cDNAs contained stop codons except for the C-terminal splice site of neurexin III (ref. 2 and Y.A.U. and T.C.S., unpublished observations). (ii) PCR experiments on different splice sites in all three neurexins led to >100 sequencing events of neurexin splice sites, none of which exhibited in-frame stop codons except for the C-terminal splice site of neurexin III α . (iii) PCRs with rat genomic DNA under the same conditions with positive controls failed to amplify a DNA fragment, suggesting that the primers are separated by a large intron. Thus these data suggest that contamination of the cDNA library and PCRs with partially processed heterogeneous nuclear RNA is not a source of artifacts and that contamination by genomic DNA also cannot explain our findings.

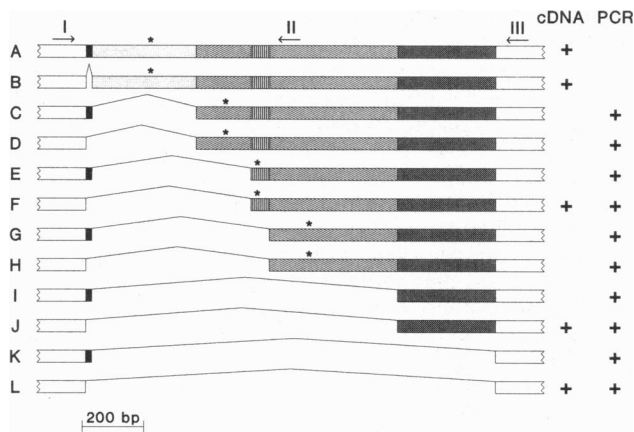


FIG. 3. Splice-site variants observed at the C-terminal splice site of neurexin III α (nt 4868–6199, boxed in Fig. 2). Variants were identified by cDNA cloning and PCR; the relative positions of the three oligonucleotides used for PCR are indicated above variant A (labeled I, II, and III). The method of identification of each variant is indicated on the right. The different spliced sequence segments are marked by different textures. Positions of the first in-frame stop codon in a transcript are shown by asterisks; these stop codons create secreted forms of neurexin III α .

These results suggest that the splice variants observed at the C-terminal site of neurexin III α are physiological. Since most of the alternatively spliced sequences in neurexin III α are widely separated and occur in multiple combinations, it is likely that they are independent of each other. Combinations of the different inserts at the four sites could generate at least 96 transcripts for neurexin III α , indicating that neurexin III α is also a very polymorphic protein. Several of the neurexin III α transcripts generated by alternative splicing at the C-terminal site contain stop codons before the transmembrane region, suggesting they encode secreted proteins. No splice variant of the other neurexins exhibits this feature.

Protein Structure of Neurexin III α . Neurexin III α is highly homologous to neurexins I α and II α . Fifty-eight percent of the residues are invariant in the three neurexins, and neurexin III α is approximately equally homologous to neurexins I α and II α . As expected from the high degree of homology between the neurexins, neurexin III α has the same overall domain structure as the other neurexins (Fig. 4). Analysis of the neurexin sequences for signal sequences and signal peptidase cleavage sites (13) predicts that all three neurexins contain cleaved signal sequences at their N termini. This analysis suggests that the conserved sequence Gly-Leu-Glu-Phe represents the probable cleavage site, with the cleavage occurring after the glycine (data not shown).

After the signal peptide, neurexin III α , as the other neurexins, contains three overall repeats (labeled A, B, and C in Fig. 4). These three repeats are composed of a central EGF domain flanked by right and left arms that are weakly homologous to each other. The repeats are followed by a threonine- and serine-rich sequence that probably represents an O-linked sugar domain. At the C terminus, neurexin III α contains a single transmembrane region and a short cytoplasmic tail.

Analysis of the sequences of the three neurexins demonstrates a patchy distribution of identical and divergent sequences. In the repeat region, the three α -neurexins show the least homology to each other in the left arm of the first overall repeat (33% identity) and the most homology in the right arm of the first overall repeat (83% identity). The remaining left and right arms are between 56% and 70% identical. In some regions, blocks of identical residues in all three neurexins flank completely dissimilar sequences. The C termini of the neurexins are among the most conserved, with the trans-

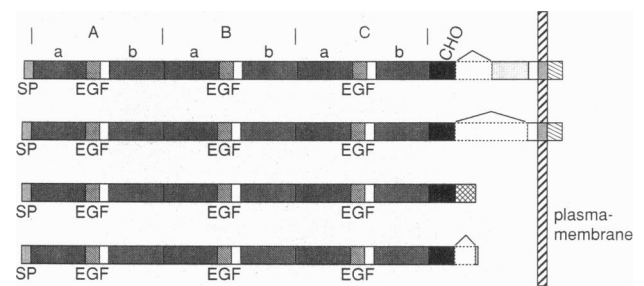


FIG. 4. Domain models of membrane-bound and secreted forms of neurexin III α as predicted from the sequences of the cDNAs and of the PCR products. Similar to neurexins I α and II α , neurexin III α contains an N-terminal signal peptide (SP) followed by three long repeats (labeled A, B, and C). The three repeats are composed of central EGF-like domains flanked by right and left arms (boxes labeled a and b). The right and left arms of each repeat are weakly homologous to each other and to the G-domain repeats of laminin A and the corresponding sequence repeats found in merosin, agrin, perlecan, and slit (2). After the repeats, neurexin III α contains a threonine- and serine-rich sequence typical of O-linked carbohydrate attachment sites (labeled CHO) followed by a transmembrane region and a cytoplasmic C terminus. The diagram shows the four principal protein forms resulting from the extensive C-terminal splice site variants (Fig. 3): two membrane-bound forms (Upper) corresponding to splice variants I to L and two secreted forms (Lower) corresponding to variants A to H.

membrane region being identical in all neurexins. Recent experiments have demonstrated that the α -latrotoxin receptor and the neurexins interact with synaptotagmin, a synaptic vesicle membrane protein, and that the cytoplasmic C-terminal domains of neurexins bind to synaptotagmin *in vitro* (14, 15). The high degree of sequence identity in the cytoplasmic domains of neurexins supports a functional role for these interactions.

Data bank searches indicate that the repeats of neurexin III α have a similar degree of homology to agrin (4, 5), slit (6), laminin A (3), merosin (16), and perlecan (17–19), as was observed for neurexins I α and II α (2). In particular, those residues tend to be conserved among the neurexins, agrin, laminin A, slit, and perlecan that are also conserved among repeats within a neurexin (data not shown). This finding suggests that these repeats serve as protein modules with related structures and functions. Since at least some of the domains corresponding to the neurexin repeats appear to be involved in cell signaling and cell attachment in agrin and laminin A, a possible function of this protein domain is that of a cell-interaction domain.

Most of the alternatively spliced sequences in the neurexins are located in conserved regions in the three repeats and insert or delete small sequences. The exception to this observation is the highly polymorphic splice site in neurexin III α that contains at least 12 variants. This site is located between the O-linked sugar domain and the transmembrane region. Several of the splice variants at this alternatively spliced position contain in-frame stop codons, suggesting the synthesis of secreted forms of neurexin III α (Fig. 2). These secreted forms contain all repeats of neurexin III α and the O-linked sugar region but end in hydrophilic C-terminal sequences of variable length (Fig. 4). Secreted forms of neurexins would have the potential to serve as soluble signaling molecules analogous to agrin, which has a role in synaptogenesis and is also alternatively spliced (4, 5, 20). The presence of secreted and membrane-bound forms of neurexins increases the number of potential signaling roles of these molecules.

Expression of Neurexin III α . RNA blots of various rat tissues were hybridized with probes from neurexin III α to determine its tissue distribution (Fig. 5). Like the other

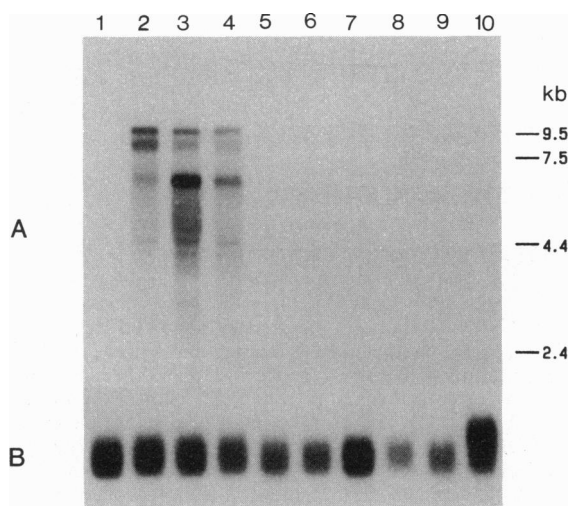


FIG. 5. Expression of neurexin III α in rat tissues by RNA blot analysis. Total RNA from various tissues (10 μ g per lane) was denatured, electrophoresed, and blotted. The blot was hybridized with a uniformly 32 P-labeled DNA probe from the 3' end of neurexin III α (A). After exposure (1 day at -70°C with screen), the blot was rehybridized with a cyclophilin probe to control for RNA loads (B). Numbers on the right indicate the positions of RNA size markers. Lanes: 1, PC12 cells; 2, brain; 3, cerebellum; 4, spinal cord; 5, adrenal; 6, intestine; 7, kidney; 8, liver; 9, spleen; 10, testis.

neurexins, neurexin III α was found to be expressed exclusively in brain. Even prolonged exposure failed to visualize a hybridization signal in nonneuronal tissues (data not shown). RNA from the adrenal gland, which expresses many otherwise neuron-specific genes, also showed no expression of neurexin III α . In the three brain regions tested (forebrain, cerebellum, and spinal cord), transcripts of different sizes were observed whose levels varied among the brain regions (Fig. 5). It is likely that the different RNA transcripts observed correspond to splice variants of neurexin III α and possibly to an as yet unidentified neurexin III β .

CONCLUSION

Neurexins I and II are cell surface proteins that were originally identified by the sequence identity of one of its members with the high molecular weight component of the α -latrotoxin receptor (2, 3). We have now determined the full-length structure, alternative splicing, and expression of neurexin III α , a third member of the neurexin gene family. Neurexin III α is highly homologous to neurexins I α and II α and is also expressed in a brain-specific manner. Neurexins I α , II α , and III α have the same overall domain structure and homologies to cell-attachment proteins (such as laminin A and agrin). Furthermore, neurexin III α is also subject to extensive alternative splicing with the potential to generate at least 96 forms. However, neurexin III α differs from other neurexins in that alternative splicing produces truncated and membrane-bound forms, suggesting that some forms of neurexin III α may be secreted.

Four positions of alternative splicing in the neurexin III α mRNA were observed by cDNA cloning and PCRs. Interestingly, neurexin I α , neurexin II α , or both are also alternatively spliced at each of these positions (2), suggesting that the alternative splicing is conserved among neurexins. Most sites of alternative splicing interrupt sequence regions that are highly conserved among neurexins, supporting the notion

that these alternative splicing events are functionally important. Except for the C-terminal splice site, all proposed alternative splicing events consist of in-frame deletions/insertions. Some of the inserts in the C-terminal splice site of neurexin III α contain in-frame stop codons, a property different from those of other neurexins. At least 12 splice variants were found at this position in neurexin III α , suggesting that the total number of neurexin transcripts could be much larger than those currently known.

Neurexins are neuronal cell surface proteins whose structures suggest that they serve as receptors. The homologies of neurexins to laminin A, agrin, perlecan, and slit raise the possibility that neurexin may participate in cell-cell interactions. One unusual feature of the neurexins is their high degree of polymorphism. The finding of a third gene for neurexins indicates that the neurexins may constitute a large and conserved gene family. The conservation of alternative splicing among all neurexins suggests that this alternative splicing may be functionally important and could represent a mechanism by which different neurons display different combinations of these molecules on their surface. The presence of potentially secreted forms of neurexin III α suggests that these molecules may be incorporated into the extracellular matrix or may themselves represent signaling molecules.

We thank I. Leznicki and A. Hopkins for excellent technical assistance, and Drs. M. S. Brown and J. L. Goldstein for advice. This work was supported by the Howard Hughes Medical Institute and the Perot Family Foundation.

1. Sheperd, G. M., ed. (1990) *The Synaptic Function of the Brain* (Oxford Univ. Press, New York), 3rd Ed.
2. Ushkaryov, Y. A., Petrenko, A. G., Geppert, M. & Südhof, T. C. (1992) *Science* **257**, 50–56.
3. Sasaki, M., Kleinman, H. K., Huber, H., Deutzmann, R. & Yamada, Y. (1988) *J. Biol. Chem.* **263**, 16536–16544.
4. Rupp, F., Payan, D. G., Magill-Solc, C., Cowan, D. M. & Scheller, R. H. (1991) *Neuron* **6**, 811–823.
5. Ruegg, M. A., Tsim, K. W. K., Horton, S. E., Kroger, S., Escher, G., Gensch, E. M. & McMahon, U. J. (1992) *Neuron* **8**, 691–699.
6. Rothberg, J. M., Jacobs, J., Goodman, C. S. & Artavanis-Tsakonas, S. (1990) *Genes Dev.* **4**, 2169–2187.
7. Petrenko, A. B., Lazaryeva, V. D., Geppert, M., Tarasyuk, T. A., Moomaw, C., Khokhlatchev, A. D., Ushkaryov, Y. A., Slaughter, C., Nasimov, I. V. & Südhof, T. C. (1993) *J. Biol. Chem.* **268**, 1860–1867.
8. Südhof, T. C., Lottspeich, F., Greengard, P., Mehl, E. & Jahn, R. (1987) *Science* **238**, 1142–1144.
9. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4509–4513.
10. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
11. Zorn, A. M. & Krieg, P. M. (1991) *Biotechniques* **11**, 181–183.
12. Kozak, M. (1991) *J. Cell Biol.* **115**, 887–903.
13. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
14. Petrenko, A. G., Perin, M. S., Davletov, B. A., Ushkaryov, Y. A., Geppert, M. & Südhof, T. C. (1991) *Nature (London)* **353**, 65–68.
15. Hata, Y., Davletov, B., Petrenko, A. G., Jahn, R. & Südhof, T. C. (1993) *Neuron* **10**, 307–315.
16. Ehrig, K., Leivo, I., Argraves, W. S., Ruoslahti, E. & Engvall, E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3264–3268.
17. Noonan, D. M., Fulle, A., Valente, P., Cai, S., Horgan, E., Sasaki, M., Yamada, Y. & Hassel, J. R. (1991) *J. Biol. Chem.* **266**, 22939–22947.
18. Kallunki, P. & Tryggvason, K. (1992) *J. Cell Biol.* **116**, 559–571.
19. Murdoch, A. D., Dodge, G. R., Cohen, I., Tuan, R. S. & Iozzo, R. V. (1992) *J. Biol. Chem.* **267**, 8544–8855.
20. Ferns, M. J. & Hall, Z. W. (1992) *Cell* **70**, 1–3.