Neuron, Vol. 14, 497–507, March, 1995, Copyright © 1995 by Cell Press

Cartography of Neurexins: More Than 1000 Isoforms Generated by Alternative Splicing and Expressed in Distinct Subsets of Neurons

Beate Ullrich, Yuri A. Ushkaryov,* and Thomas C. Südhof
Department of Molecular Genetics
Howard Hughes Medical Institute
University of Texas Southwestern Medical School
Dallas, Texas 75235

Summary

Neurexins, a family of cell surface proteins specific to brain, are transcribed from two promoters in three genes, resulting in three α- and three β-neurexins. In situ hybridization revealed differential but overlapping distributions of neurexin isoforms in different classes of neurons. PCRs demonstrated that α-neurexins are alternatively spliced at five canonical positions, and β-neurexins at two. Characterization of many independent bovine neurexin Iα cDNAs suggests that different splice sites are used independently. This creates the potential to express more than 1000 distinct neurexin proteins in brain. The splicing pattern is conserved in rat and cow. Thus, in addition to somatic gene rearrangements (immunoglobulins and T cell receptors) and large gene families (odorant receptors), alternative splicing potentially represents a third mechanism for creating a large number of cell surface receptors that are expressed by specific subsets of cells.

Introduction

Neurexins represent a family of cell surface proteins that are exclusively expressed in brain (Ushkaryov et al., 1992; Ushkaryov and Südhof, 1993; Geppert et al., 1992). There are at least six principal neurexins that are transcribed from three genes (designated I, II, and III). Each neurexin gene contains two independent promoters. The proteins encoded by the longer mRNAs are referred to as α-neurexins, and the ones encoded by shorter mRNAs as β-neurexins. α- and β-neurexins have distinct domain structures (Figure 1). α-Neurexins contain a classic N-terminal signal sequence followed by three copies of a long repeat. Each repeat is comprised of a central epidermal growth factor (EGF)–like domain flanked by right and left arms (A_A, A_B, B_A, B_B, C_A, and C_B in Figure 1). These arms are weakly homologous to each other and to sequences found in a variety of extracellular matrix proteins such as laminin A and agrin. After the three overall repeats, α-neurexins contain an O-linked sugar domain, a single transmembrane region, and a short cytoplasmic tail. The β-neurexins exhibit an N-terminal sequence that contains an atypical cleaved signal peptide followed by a short sequence unique to β-neurexins (Ushkaryov et al., 1994). The β-neurexins then splice into the α-neurexin sequences at the end of the EGF domain in the middle of the third repeat and have C-terminal halves identical to those of the α-neurexins (Figure 1). β-Neurexins can be considered as N-terminally truncated α-neurexins that contain only 35–43 amino acids unique to β-neurexins as compared with α-neurexins.

The domain structure of neurexins resembles that of a cell surface receptor. The first neurexin (Iα) was originally identified as the high molecular weight component of the receptor for a presynaptic neurotoxin, α-latrotoxin (Petrenko et al., 1993). The localization of the α-latrotoxin receptor to the presynaptic plasma membrane as well as immunocytochemical localization of neurexin I to synapses suggests that neurexins are synaptic proteins (Ushkaryov et al., 1992). The receptor-like structure of the neurexins and their homologies with extracellular matrix proteins involved in the development of the nervous system have raised the possibility that the neurexins function in mediating cell–cell interactions in the nervous system.

One of the most striking characteristics of the neurexins is their extensive alternative splicing. All three neurexin transcripts are alternatively spliced at multiple positions (Ushkaryov et al., 1992). Alignment of the sequences of the neurexins and their splice site positions reveals five alternative splice sites that are observed in at least two neurexins (labeled 1–5 in Figure 1). Extensive alternative splicing could theoretically generate hundreds and even thousands of receptor proteins that might impart different properties on the cells displaying these receptors. Receptor diversity is a mechanism to generate specificity in the immune system and in odorant perception (Tonegawa, 1983; Buck and Axel, 1991). However, in these cases diversity is not generated by alternative splicing but by gene rearrangements and the presence of large gene families, respectively. Although extensive alternative splicing has been observed for a number of proteins, most of these are intracellular proteins with fewer than 100 variants.

To determine whether neurexins could constitute a protein family in which alternative splicing generates a degree of diversity similar to that in other recognition processes (e.g., more than 1000 forms), we have now studied the distribution of neurexin isoforms and the pattern of alternative splicing. Previous studies demonstrated that neurexins are brain specific and extensively alternatively spliced but raised several questions. Are the different alternatively spliced sequences independent of each other? Is alternative splicing evolutionarily conserved? Is alternative splicing stochastic or regulated; e.g., is there a region-specific expression of splice variants in brain? Are different neurexins subject to alternative splicing at the same position, and do they have similar sequence variants? Are different neurexins and their splice variants homogeneously or differentially expressed in brain? Most importantly, how many different splice variants are there for each site, and how many different neurexins are there? In the current study, we have tried to answer these questions in a systematic study of the distribution and alternative splicing

*Present address: Biochemistry Department, Imperial College of Science, Technology, and Medicine, London SW7 2AY, England.

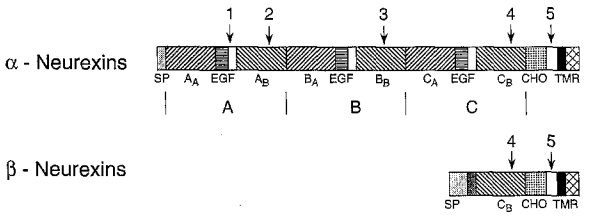

Figure 1. Generic Domain Model of Neurexins and Locations of Differentially Spliced Sequences

α-Neurexins contain a signal peptide (SP) followed by three overall repeats (A, B, and C), each of which consists of a central EGF domain (EGF) flanked by left and right arms (A_A, A_B, B_A, B_B, C_A, and C_B). β-Neurexins also start with a signal peptide (SP) followed by a short β-specific sequence, after which they splice into the α-neurexins at the end of the third EGF domain so that they only contain the right arm of the third overall repeat (C_B). The repeats are followed by a serine- and threonine-rich sequence that presumably represents an O-linked sugar domain (CHO). All neurexins except for some splice variants of neurexin III are anchored in the membrane by a transmembrane region (TMR) and end in a short cytoplasmic tail. The locations at which at least two neurexins are alternatively spliced are indicated by an arrow above the diagram and referred to as alternatively spliced sequences 1–5.

of the neurexins. Our data demonstrate that neurexins are among the most polymorphic proteins known, supporting their potential function as neural recognition molecules.

Results

Differential Expression of Neurexins

There are six principal neurexin mRNAs transcribed from two promoters in three genes. The domain structures of the neurexins are diagrammed in Figure 1. Expression of multiple isoforms of a protein often serves as an evolutionary mechanism of diversification. Thus, the question arises whether different neurexins are expressed in a complementary manner (i.e., cells only express one or the other neurexin isoform, similar to the exclusive expression of odorant receptors in olfactory neurons [Chess et al., 1994]), or whether different neurexins are expressed in a combinatorial manner (i.e., neurons express distinct combinations of neurexins characteristic for a given type). In the first case, neurons could be classified into six categories according to which neurexin they express, and the six classes would be coordinately regulated via their promoters. In the second case, functionally distinct isoforms could be coexpressed in variable patterns in neurons to confer specific properties onto expressing cells, and neurons could have a large number of potential expression patterns dependent on the ratio of expression of different neurexins. To determine whether neurexin isoforms are coexpressed in neurons or whether a given neuron expresses only a single form, we studied the localization of their mRNAs by in situ hybridization.

Rat brain sections were hybridized with oligonucleotides specific for the six individual neurexins and for synapsins as a neuronal marker (Figure 2). Control hybridizations were performed for all probes in the presence of a 50-fold excess of unlabeled oligonucleotide (shown for neurexin Iα in Figure 2). In addition, hybridization patterns with sense oligonucleotides or with multiple independent antisense probes for some neurexins were examined (data not shown). Film autoradiograms of horizontal sections reveal a striking differential expression of neurexins in brain with an exclusively neuronal expression pattern (Figure 2).

Neurexin Iα is expressed in all brain regions, with the highest levels in the claustrum and thalamic nuclei. In contrast, neurexin Iβ is restricted to a few brain areas, such as selected outer layers of the cerebral cortex (layers 2 and 3), the thalamus, and parts of the hippocampal formation (Figure 2; see also Figure 3 and Figure 4, below). Neurexin IIα expression also exhibits an interesting cortical distribution. Neurons in three cortical layers are preferentially labeled that were identified by emulsion autoradiography as neuronal subpopulations in layers 2, 4, and 6 (Figure 2 and data not shown). Significant levels of neurexin IIα mRNAs were also detected in the thalamus and cerebellum. Neurexin IIβ was more uniformly expressed than IIα, with particularly high levels in the outer layers of the cerebral cortex and the cerebellum. Neurexin IIIα probes revealed a generally low signal in most brain areas. A thin outer layer and the inner layers of the cerebral cortex were labeled more strongly than the middle layers, and the striatum, septal nuclei, and reticular thalamic nucleus (RN) were more prominent than the remaining areas of the central gray matter. Neurexin IIIβ transcripts were found abundantly in all brain structures without any marked regional preference. Studies on other brain regions such as the peripheral nervous system and the spinal cord also revealed a differential expression of neurexin subtypes in these structures (data not shown).

Expression of Neurexins in the Hippocampal Formation

Two brain regions, the hippocampal formation and the olfactory bulb, were chosen for in-depth studies to probe the combinatorial multiplicity of neurexins in neurons. The hippocampus, one of the best studied structures of the brain, consists of a limited number of types of neurons that are organized in a distinctive manner. This organization allows an unequivocal assignment of hybridization patterns at the cellular level (Figure 3). Hybridization patterns obtained with oligonucleotides specific for synapsins and for all neurexins ("pan-neurexin") reveal an identical uniform labeling of all cell types, suggesting that neurexins are coexpressed with synapsins in all neurons. Hybridizations with oligonucleotides specific for individual neurexins demonstrate slightly to dramatically different expression patterns for the individual neurexins. As observed for total brain in film autoradiograms (Figure 2), neurexins Iα, IIα, IIβ, and IIIβ are present in all types of neurons but exhibit distinct differences in expression intensity (e.g., note the relatively higher levels of IIβ in CA1 or of IIIβ in interneurons). In contrast, neurexins Iβ and IIIα show dramatic differences in expression. Pyramidal cells and interneurons of CA1 contain no detectable neurexin Iβ (Figure 3, arrowheads). CA1 pyramidal cells and dentate gyrus granule cells lack neurexin IIIα, but the interneurons

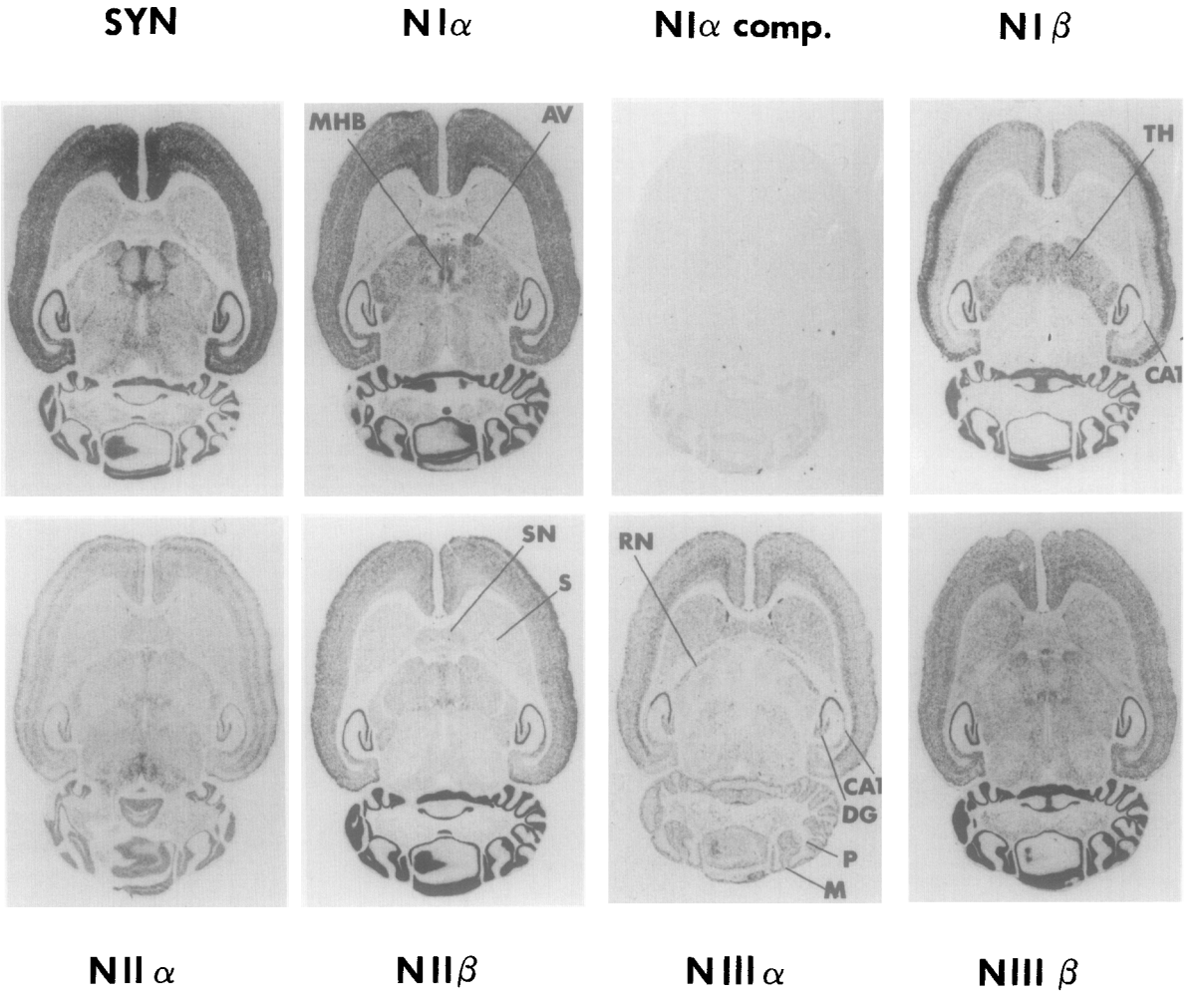

Figure 2. Differential Expression of Neurexins in Rat Brain

Oligonucleotides specific for the indicated neurexins (NIα to NIIIβ) and for synapsins as a neuronal marker (SYN) were hybridized to horizontal sections from adult rat brain. The distribution of neurexin Iα is similar to that of synapsins, except for an enhanced signal in the claustrum, anterior thalamic nuclei (MHB, medial habenular thalamic nuclei; AV, anteroventral thalamic nucleus), and deep cerebellar nuclei. As a control for specificity, competition experiments were performed with a 50-fold excess of unlabeled oligonucleotide, which eliminates the signal (shown for neurexin Iα in NIα comp.). Neurexin Iβ shows a prominent differential distribution with intense labeling of cortical layers 2 and 3, the thalamus, and the cerebellar granule cell layer. Neurexin IIα is differentially distributed among cortical layers with intense signal in layers 2, 4, and 6. Signal is also enhanced in the septal nuclei (SN), the reticular thalamic nucleus (RN), and some nuclei of the midbrain. For neurexin IIβ, an almost even hybridization is evident within the cortex, with slight enhancement in layer 2 and in the septal nuclei. Neurexin IIIα shows a prominent differential distribution: in the cerebral cortex, the signal is most intense in layer 6 and also detectable in layer 2. Relative to their cell density, the striatum, septal nuclei, and reticular thalamic nucleus are strongly labeled. In the cerebellum, an "inverted" hybridization pattern is evident compared with the other neurexins: intense labeling of the molecular and Purkinje cell layer contrasts with very low signal in the granule cell layer. In contrast, neurexin IIIβ is evenly expressed in most brain structures and only slightly more pronounced in the anterior and medial thalamic nuclei. In competition controls, the slight binding of this particular probe to white matter was shown to be nonspecific. Film autoradiographs have not been digitized. Exposure times were 3 days to 5 weeks.

in the same regions are strongly labeled. Thus, pyramidal cells of CA3 coexpress all six neurexin isoforms, whereas dentate gyrus granule cells selectively lack neurexin IIIα, and CA1 pyramidal cells lack neurexins Iβ and IIIα. Interneurons express variable ratios of neurexins in different regions, with neurexin IIIα being the most prominent in all interneurons (Figure 3).

Neurexins in Olfactory Bulb Neurons

Neurons are also organized in distinct, easily recognizable layers in the olfactory bulb, allowing an unequivocal identification of cell types (Figure 4a) (Shepherd, 1990). Synapsins are uniformly expressed in all neurons in the olfactory bulb (b). Hybridization of olfactory bulb sections with oligonucleotides that react with all neurexin mRNAs reveals the presence of neurexins in all neurons (as in all other brain regions examined) but very high expression levels in mitral and tufted cells (c). This suggests that neurexins are expressed particularly strongly by mitral cells, probably because of the high levels of neurexin Iβ expression in these cells (see below).

Examination of the distribution of individual neurexins

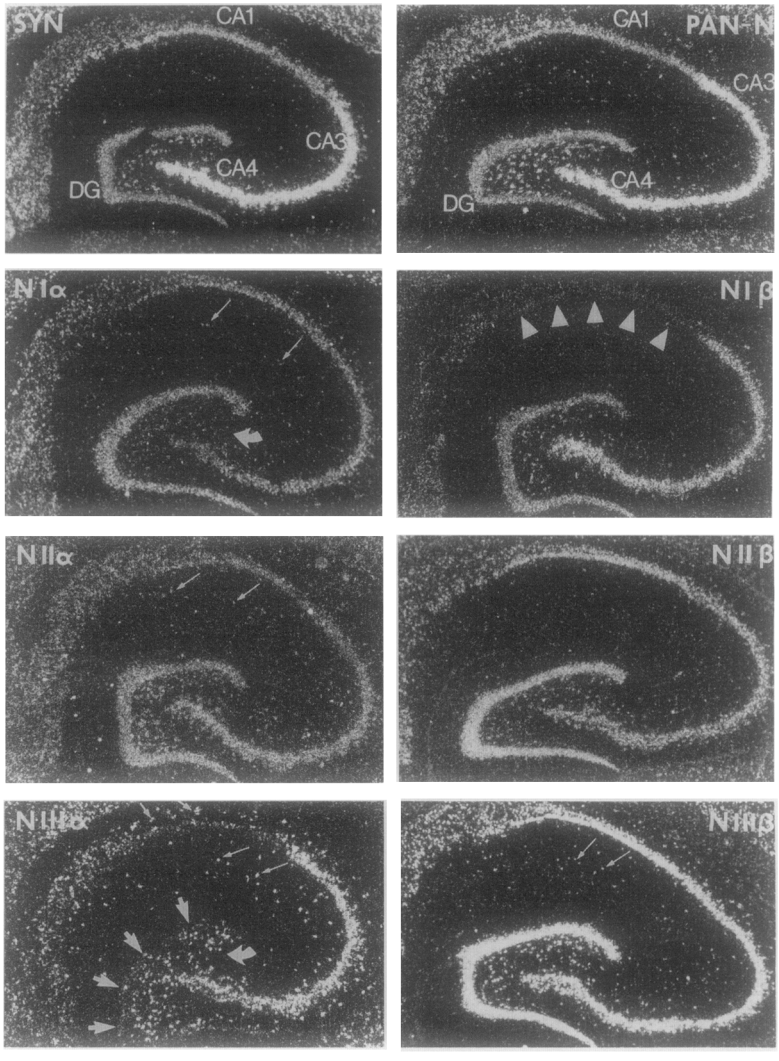

Figure 3. Distribution of Neurexins in the Hippocampal Formation

Emulsion autoradiography of rat brain sections hybridized with oligonucleotides that are specific for mRNAs encoding synapsin (SYN), all neurexins (PAN-N), or specific neurexin isoforms (NIα to NIIIβ). Note the differential distribution of individual neurexins in different segments of the hippocampal formation, e.g., the lack of neurexin Iβ and IIIα expression in CA1 (arrowheads in NIβ) and the presence of all neurexins in the dentate gyrus (DG) except for neurexin IIIα (thick arrows in NIIIα). Dentate gyrus interneurons express high levels of neurexin IIIα (curved arrow in NIIIα). The thin arrows in NIα, NIIα, NIIIα, and NIIIβ point to interneurons in CA1. Panels show dark-field photographs of sections coated with photographic emulsion and developed after 1–3 months.

in olfactory bulb neurons demonstrates that each neurexin isoform is expressed in a different pattern. Neurexin Iα is present at highest levels in periglomerular cells, but almost absent from tufted and mitral cells (Figure 4d). In contrast, neurexin Iβ is highly expressed in mitral and tufted cells (e); this pattern is specific since it can be competed off by a 50-fold excess of oligonucleotide (f). Neurexin IIα is only very weakly expressed in olfactory bulb neurons (g), whereas neurexin IIβ is highly expressed, in particular in the inhibitory periglomerular and granule cells (h). Neurexins IIIα and IIIβ are both present in granule cells, but only neurexin IIIβ exhibits significant levels in periglomerular cells (j and k, respectively). Thus, for the four types of neurons of the olfactory bulb, each type expresses a different combination of neurexins and is identifiable by the combination it expresses.

The results of the studies on the distributions of neurexins allow several conclusions, the major being that neurexins are expressed in a differential pattern specific for each form. It seems likely that all neurons coexpress multiple neurexins, but in different combinations and ratios. α- and β-neurexins are expressed independently of each other, although they are transcribed from the same gene, suggesting a rationale for two independent promoters.

Neurexin Iα: Splicing Patterns in Bovine Brain cDNAs

In addition to the multiplicity of neurexins due to the presence of multiple genes with independent promoters, neurexins are subject to extensive alternative splicing that could potentially generate a large number of cell surface proteins (Ushkaryov et al., 1992). It is unclear as to how many proteins are specified because the number of alternatively spliced sequences in the different neurexins are unknown, as are the number of insert variants at each site and the independence of alternative splicing events at different sites. To address these questions, two strategies were pursued. First, a large number of cDNA clones encoding bovine neurexin Iα were examined to assess the independence of alternative splicing events and the relative frequency of different types of variants. Bovine cDNAs were examined instead of rat cDNAs to evaluate the evolutionary conservation of alternative splicing. Second, polymerase chain reaction (PCR) analyses were performed on

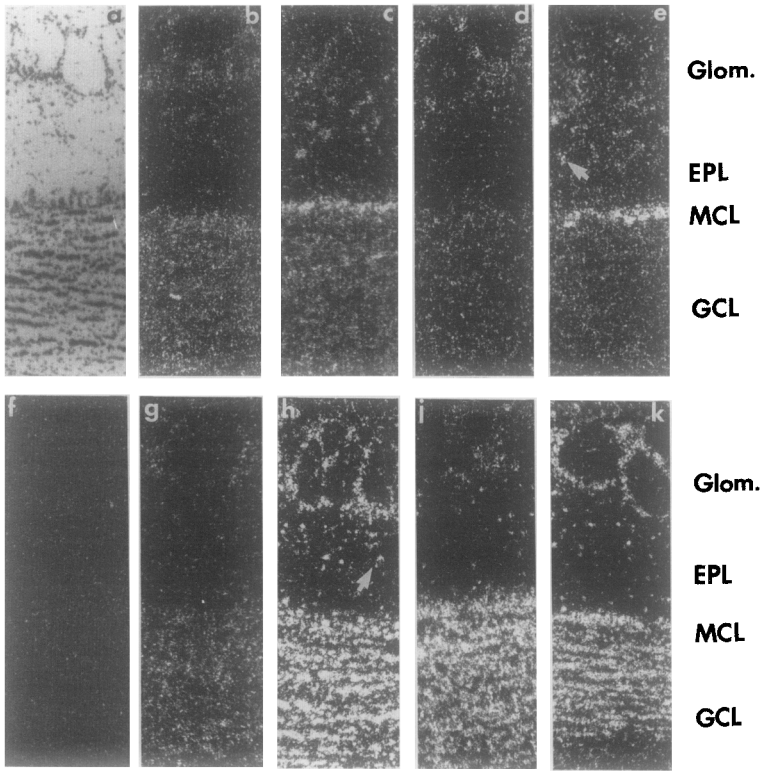

Figure 4. Distribution of Neurexins in Olfactory Bulb

(a) shows a bright-field micrograph of a thionin-stained section, and (b)–(k) show dark-field micrographs of corresponding sections from adult rat olfactory bulbs hybridized with different oligonucleotide probes. Probes used were synapsin (b), all neurexins (c), neurexin Iα (d), neurexin Iβ (e), neurexin Iβ with a 100-fold excess of unlabeled probe (f), neurexin IIα (g), neurexin IIβ (h), neurexin IIIα (j), and neurexin IIIβ (k). The layers of the olfactory bulb are identified on the right (Glom., glomerular layer; EPL, external plexiform layer; MCL, mitral cell layer; GCL, granular cell layer). Note that each neurexin exhibits a different labeling pattern between neurons in the olfactory bulb such that all types of neurons in the bulb express a different combination of neurexins.

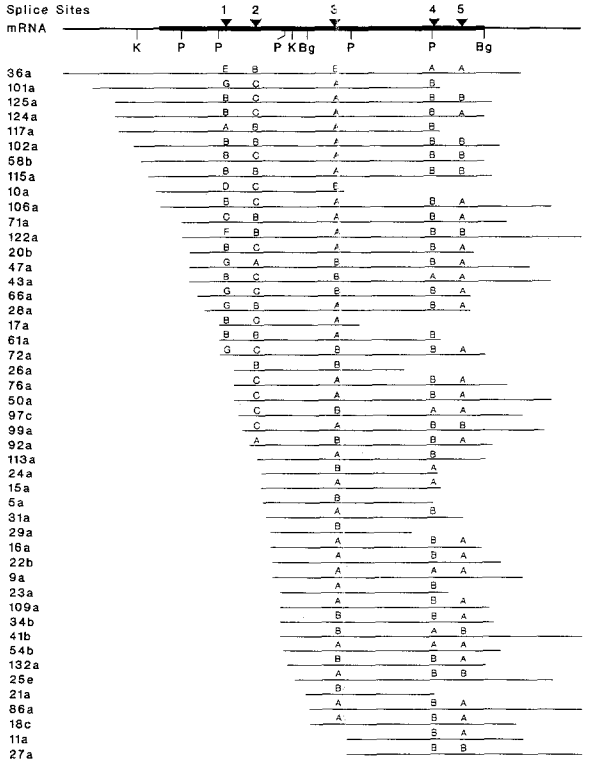

Figure 5. Analysis of the Alternative Splicing of Bovine Neurexin Iα by cDNA Cloning

The diagram shows the structures and alternative splicing patterns of 47 independent cDNA clones. The structure of the bovine neurexin Iα mRNA is shown on top of the graph together with the locations of the five alternatively spliced sequences (1–5) and of restriction enzyme sites (Bg, BglII; K, KpnI; P, PstI). For each alternatively spliced sequence in a given cDNA clone, the insert variant is indicated by letters. The terminology for alternatively spliced sequences 1–3 is explained in Table 1; for spliced sequences 4 and 5, variant A contains the insert and variant B lacks it. Clones 36a, 47a, and 124a were sequenced entirely to determine the primary structure of bovine neurexin Iα (GenBank accession #L14855).

rat brain RNA for all potential splice sites, to determine whether all neurexins are alternatively spliced at equivalent positions and to estimate the number of insert variants.

A bovine cortex cDNA library was screened with a randomly labeled probe encompassing the coding region of neurexin Iα. Sequencing and restriction enzyme mapping of 102 cDNA clones revealed that 87 encoded neurexin Iα, comprising 47 independent isolates. The remaining 15 clones either encoded neurexin Iβ (4 clones) or were not identified (11 clones). The splice site inserts of all cDNA clones were determined by restriction enzyme mapping, sequencing, and PCR (Figure 5). Together, the characterization of the bovine cDNAs allowed three conclusions.

First, the primary sequence of neurexin Iα is highly conserved evolutionarily. There are only 21 amino acid changes in the primary sequence between rat and cow; 3 of these are in the signal peptide and only 1 (at position 193) is nonconservative. This translates into a 99% conservation for this 1514 amino acid protein. The sequence conservation also includes the sequences of splice site inserts (see below).

Second, all of the five sites of alternative splicing pre-

Table 1. Splice Variants of Splice Sites 1–3 of Neurexins

Variant	Sequence	cDNA	PCR
Splice site 1			
Iα	CSQEDNNVEGLAHLMMGDQGKSKEDNNVEGLAHLMMGDQ----GKE		
A	CSQEDNNVEGLAHLMMGDQGKSKEDNNVEGLAHLMMGDQ----GKE	+	
B	CSQEDNNVEGLAHLMMGDQGKSK----------GKE	+	+
C	CSQ-DNNVEGLAHLMMGDQGKSK----------GKE	+	
D	CSQEDNNVEGLAHLMMGDQ----------GKE	+	+
E	CSQ-DNNVEGLAHLMMGDQ----------GKE	+	
F	CSQEDNNVE----------GKSK----------GKE	+	
G	CSQ----------GKE	+	+
IIα	CSEEEHPMEGPAHLTLNSEVGSLLFSEGGAGRGGAGNVHQPTKGKE		
A	CSEEEHPMEGPAHLTLNSEVGSLLFSEGGAGRGGAGNVHQPTKGKE	+	+
B	CSEEEHPME----------VGSLLFSEGGAGRGGAGDVHQPTKGKE		+
C	CSEEEHPME----------GKE		+
IIIα	CSEDVSQGPGLSHLMMSEQGRSK----------ARE		
A	CSEDVSQGPGLSHLMMSEQGRSK----------ARE	+	+
B	CSEDVSQGPGLSHLMMSEQ----------ARE	+	+
C	CSE------GLSHLMMSEQGRSK----------ARE		+
D	CSE------GLSHLMMSEQ----------ARE		+
E	CSEDVSQGP----------ARE		+
F	CSE----------ARE		+
Splice site 2			
Iα	WHDVKVTRNLRQHSGIGHAMVNKLHCSVTISVLTTTGYT		
A	WHDVKVTRNLRQHSGIGHAMVNKLHCSVTISVLTTTGYT	+	+
B	WHDVKVTRNLRQHSGIGHAM-------VTISVLTTTGYT	+	+
C	WHDVKVTRNLRQ----------VTISVLTTTGYT	+	+
IIα	WHDVRVTRNLRQHAGIGHAMVNKLHYLVTISVLTTTGYT		
A	WHDVRVTRNLRQHAGIGHAMVNKLHYLVTISVLTTTGYT	+	+
B	WHDVRVTRNLRQHAGIGHAM-------VTISVLTTTGYT		+
C	WHDVRVTRNLRQ----------VTISVLTTTGYT	+	+
IIIα	WHDVKVTRNLRQ----------VTISVLTTTGYT	+	+
Splice site 3			
Iα	ELDAGRVKLTVNLDCIRINCNSS---KGPETLFAGYNLN		
A	ELDAGRVKLTVNLDCIRINCNSS---KGPETLFAGYNLN	+	+
B	ELDAGRVKLTVNL---------G---KGPETLFAGYNLN	+	+
IIα	ELDGGQMKLTVNLDCLRVGCAPSAAAKGPETLFAGHKLN		
A	ELDGGQMKLTVNLDCLRVGCAPSAAAKGPETLFAGHKLN	+	+
B	ELDGGQMKLTVNLDCLRVGCAPT---KGPETLFAGHKLN	+	+
C	ELDGGQMKLTVNL---------G---KGPETLFAGHKLN		+
IIIα	ELDGGRVKLMVNLDCIRINCNSS---KGPETLYAGQKLN		
A	ELDGGRVKLMVNLDCIRINCNSS---KGPETLYAGQKLN	+	+
B	ELDGGRVKLMVNL---------G---KGPETLYAGQKLN	+	+

Sequences are named on the left, and their method of identification is indicated on the right. Sequences are shown in single-letter amino acid code, with gaps as hyphens. All sequences were from rat brain cDNAs and PCRs except for the site 1 variants for neurexin Iα, which were from bovine cDNAs (all variants), and rat cDNAs and PCRs (variants B, C, D, and G). At each splice site, comparison of the sequences of different insert variants suggests the presence of exchangeable modules that are either present or absent (e.g., E/DNNVE/GLAHLMMGDQ/GKSK in splice site 1 of neurexin Iα). Since it seems likely that these different modules occur in more combinations than observed in the limited sample analyzed, the total number of variants is probably much larger than is shown here.

viously observed for rat neurexin Iα (Ushkaryov et al., 1992) are also present in bovine neurexin Iα, with an almost complete conservation of the alternatively spliced coding sequences. Sequencing of the splice sites of the bovine cDNAs revealed that the number of alternatively spliced insert variants at splice site 1 is much larger than was previously known (listed together with the sequences obtained by PCR in Table 1). Some of these inserts have an unusual structure. For example, insert variant A contains a duplication of the same amino acid sequence interrupted by 4 amino acids. A total of seven different inserts were sequenced for splice site 1. Comparison of these inserts suggests that they consist of independently included or excluded sequence modules (e.g., the E at the beginning or the GKSK in the middle of the insert; Table 1). If these sequence modules were truly independent of each other, at least 63 different insert combinations could be generated for splice site 1 of neurexin Iα.

Third, the different splice sites seem to be inserted or deleted independently of each other. Of the 47 cDNA clones, 15 clones contain all five splice sites and exhibit ten different splice site patterns over the five sites. Only three splice site patterns were observed twice, and one splice site pattern was found in 3 independent cDNA clones (Figure 5). In this sample, some variants occur more frequently than others. For example, at splice site 1, 5 variants (A, C, D, E, and F) were observed only once, whereas variant G was observed 5 times and variant B was observed 10 times (see Table 1 for definition of these variants). Similarly, at splice site 3, variant A was observed

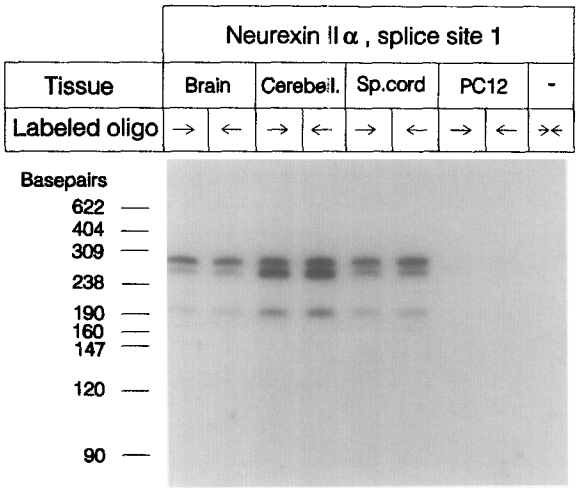

Figure 6. Analysis of Alternative Splicing by PCR: Alternatively Spliced Sequence 1 of Neurexin IIα as a Representative Example

PCRs were performed in duplicates on first-strand rat cDNAs with either one of the two oligonucleotide primers labeled. Products were analyzed on denaturing SDS gels and visualized by autoradiography. No detectable neurexin IIα mRNA is present in PC12 cells, in agreement with previous RNA blotting data (Ushkaryov et al., 1992).

29 times and variant B was observed 16 times (Figure 5). Together, these observations suggest that certain inserts are preferred but that, on the whole, the different splice sites are used independently of each other and occur in all combinations.

PCR Analysis of Neurexin Splicing in Rat Brain

Five sites of alternative splicing are observed in different neurexins, of which alternatively spliced sequences 1–3 are specific for α-neurexins, and sequences 4 and 5 are present in α- and β-neurexins (Figure 1). Previous cloning and PCR experiments demonstrated that sites 4 and 5 are present in neurexins I, II, and III (Ushkaryov et al., 1992, 1994; Ushkaryov and Südhof, 1993). However, these experiments did not reveal whether sites of alternative splicing 1, 2, and 3 are present in all neurexins and how many insert variants can be used at each site. Therefore, we used PCR to determine how many alternatively spliced sequences are conserved in all neurexins, how complex the alternative splicing is at each position, and how similar the sequence inserts are between different neurexins.

PCR reactions were performed on first-strand rat brain cDNA with primers specific for the respective neurexin and putative splice site studied. Three procedures were used to evaluate the PCR products and to guard against potential artifacts. First, PCRs were performed in duplicate in which one or the other of the two primers was labeled with ^{32}P (Figure 6). PCR products were separated on denaturing polyacrylamide gels and visualized by autoradiography. The separate labeling of each oligonucleotide allows direct visualization of specific products, and the denaturing electrophoresis avoids artifacts due to heteroduplex formation (Zorn and Krieg, 1991). Second, PCRs were performed with unlabeled oligonucleotides and analyzed by denaturing polyacrylamide gel electrophoresis and Southern blotting using internal oligonucleotide probes, thereby ensuring that the products observed contain the correct insert sequence. Third, PCR products were directly analyzed by sequencing. Together, these procedures allow a definitive identification of alternative splicing events as opposed to aberrant PCR reactions.

An example of an experiment demonstrating alternative splicing at site 1 of neurexin IIα is shown in Figure 6. In separate reactions, either one of the two oligonucleotide primers was ^{32}P-labeled. Three reaction products were observed which upon sequencing were found to correspond to the 3 insert variants defined by cDNA cloning (Table 1). Three rat brain regions and PC12 cells were analyzed separately. The utilization of alternatively spliced sequences differed between them (see also below). Using a similar analysis, we investigated the alternatively spliced sequences 1, 2, and 3 of all three α-neurexins. These experiments showed that all neurexins are alternatively spliced at sites 1 and 3, and that neurexins I and II but not III are also alternatively spliced at site 2. Sequencing of the PCR products confirmed and extended the prior cDNA cloning results, allowing us to postulate a list of insert sequences for splice sites 1–3 for the different neurexins (Table 1). Although we attempted to sequence as many PCR products as possible, the larger and less abundant products proved elusive. Accordingly, the insert variants whose sequences have been determined represent the minimal number of variants. Based on the PCR pattern, it seems probable that there are many additional variants. The sequences of the splice site variants also revealed that the insert sequences are conserved between different neurexins for most sites (Table 1).

Although PCRs across most sites of alternative splicing revealed a limited number of sequence variants, analysis of the alternatively spliced sequence 1 of neurexin Iα gave a surprising result: more than 30 different reaction products were observed (Figure 7). Southern blotting of the PCR products with multiple internal oligonucleotides confirmed that the observed products are specific. The modular design of the insert sequences for splice site 1 of neurexin Iα suggests 63 possible combinations, many of which would have identical sizes. Southern blotting with specific oligonucleotides that only hybridize to certain insert combinations showed that a subset of the more than 30 total products observed for splice site 1 of neurexin Iα contains each of the different junctional sequences, as would be expected if these products were generated by modular combinations of different insert sequences (Figure 7).

The data presented indicate that neurexins are subject to extensive alternative splicing at similar positions in their mRNAs. The question arises whether this splicing represents a baroque accident of nature or is physiologically relevant. To address this question, we investigated whether alternative splicing of neurexins is regulated. This was studied by probing the expression of splice variants in different brain regions. The data shown in Figure 6 indicate that the splice site pattern is not uniform between different

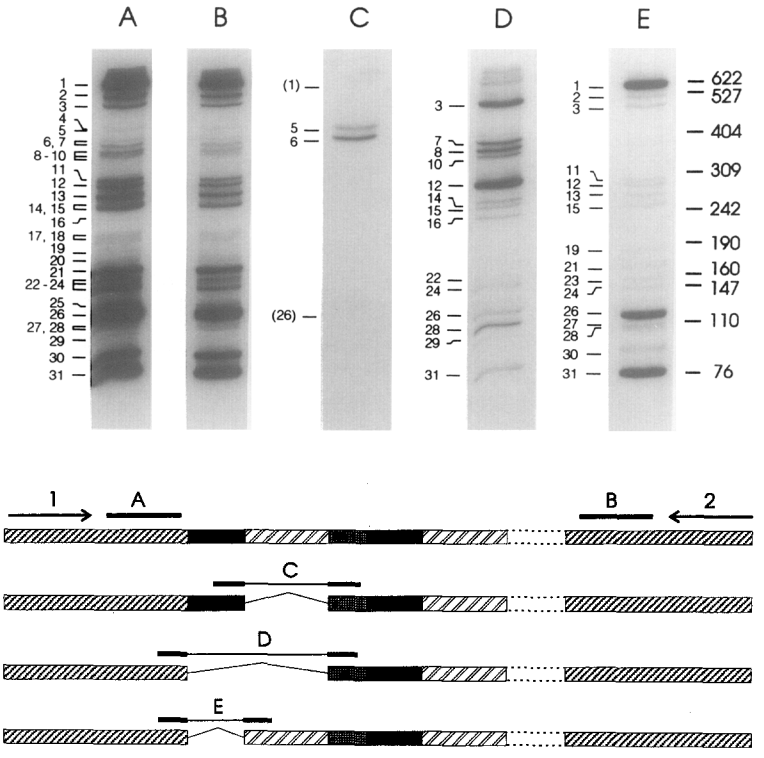

Figure 7. PCR Analysis of the Variants of Splice Site 1 of Neurexin Iα

PCR was performed with the oligonucleotide primers identified as 1 and 2 in the bottom diagram. Reaction products were analyzed by denaturing gel electrophoresis and Southern blotting with the oligonucleotides A to E that do not overlap with the PCR oligonucleotides. The top five lanes show autoradiograms of the Southern blot analysis of the PCR reactions with the different oligonucleotide probes whose relative positions are diagrammed below. In the bottom diagram, the top bar depicts the largest insert found in a cDNA clone in splice sequence 1 (variant A in Table 1), with the insert divided into five modular sequence blocks. Oligonucleotides C, D, and E were designed to test the 3 possible combinations of the modular sequences in the insert shown in the bottom 3 bars. On the autoradiograms on top, the bands visualized by the different oligonucleotide probes are identified by numbers on the left of each gel strip. Note that the number of specific PCR products observed far exceeds those predicted from a modular use of the sequence of the known splice site variants, suggesting the presence of additional as yet unidentified alternatively spliced sequences at this position. Molecular weight markers are indicated on the right (numbers are in base pairs).

brain regions. In this case, cerebellum contained much higher levels of the splice variants lacking part or all of the insert than did the rest of the brain. Similar observations were made for some but not all of the splice sites. Representative examples are shown for splice sites 1, 3, and 5 of neurexin III (Figure 8). This figure illustrates that splice site patterns can differ dramatically between different brain regions, revealing a selective expression of certain types of inserts in restricted types of neurons. These results indicate that not only neurexin isoform expression but also alternative splicing of neurexins is regulated.

Genomic Basis of Alternative Splicing

To investigate the genomic basis of the alternative splicing of the neurexins, part of the neurexin IIα gene was studied (Figure 9). Sequencing of the region of the gene corresponding to splice site 3 revealed the presence of a mini-exon that was separated from the preceding exon by a

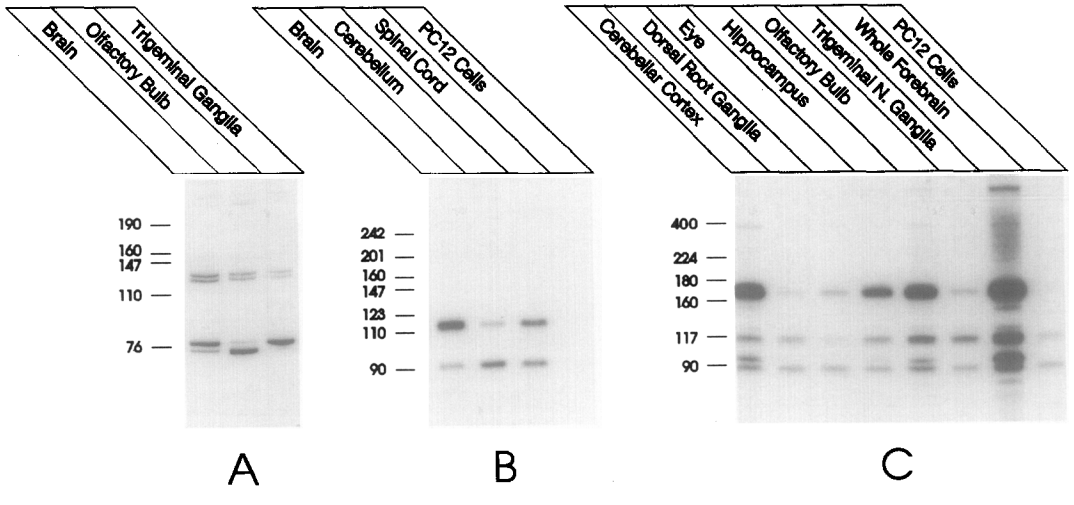

Figure 8. Differential Expression of Alternatively Spliced Sequences in Rat Brain

PCRs were performed on first-strand cDNA from the indicated brain regions with ^{32}P-labeled oligonucleotide primers specific for alternatively spliced sequences 1 (A), 3 (B), and 5 (C) of neurexin IIIα. Samples were analyzed by denaturing electrophoresis in a 5% polyacrylamide gel followed by autoradiography. Numbers on the left of each panel show the positions and sizes of DNA markers (in base pairs).

Figure 9. Structure of Neurexin IIα Gene Encoding Splice Site 3

On top, the intron–exon organization of the part of the neurexin II gene from mouse that encodes the sequences surrounding splice site 3 is shown (see Table 1). Exonic sequences are shown in boldface capital letters and are translated below the sequence. Splice site donor and acceptor sequences are shown in lower-case letters. The approximate sizes of the two introns in this region (6.6 and 1.2 kb) are indicated above the introns. Alternative splicing at these positions consists of a combination of two mechanisms: first, alternative use of the mini-exon encoding the sequence DCLRVGCAP, and second, alternative use of two overlapping splice acceptor sites. The part of the second splice acceptor site that is already in the exon if the first site is used is underlined on the right. The combination of the two different mechanisms could potentially result in 4 different sequence variants at this site, 3 of which were observed experimentally (Table 1).

rather large intron (6.6 kb). The PCR and cDNA cloning results suggest that this mini-exon is alternatively spliced in and out of the mRNA (Table 1). For the exon following the mini-exon, two possible splice acceptor sites could be detected, one with an imperfect consensus sequence (....ccccctg) followed by an overlapping perfect consensus splice acceptor site (.....ccccctggtgctgcag). Comparison of the translated sequences starting from these boundaries reveals that they match exactly the variants observed experimentally for splice site 3 (Table 1). Thus, at this position alternative splicing involves a combination of two mechanisms at the same site, the use of alternative exons and of differential splice acceptor sites. This could potentially result in 4 variants (Figure 9), 3 of which were observed in our clones and PCRs.

Discussion

Neurexins are highly polymorphic, brain-specific proteins that resemble cell surface receptors. We earlier hypothesized that the diversity of neurexins imparts specific cell surface properties upon neurons, one of which may be differential cell–cell recognition (Ushkaryov et al., 1992). In the current study, we have attempted to test two corollaries of this hypothesis. First, neurexins and their splice variants must not be uniformly distributed for this hypothesis to be valid, because a random distribution of diverse neurexins would have no information content. Second, in view of the complexity of neurons and synapses, a molecule that specifies neuronal interactions should occur in many independent forms. We have now shown that the six principal neurexins are differentially distributed in brain, suggesting that they occur in specific combinations in different neurons. This complexity is amplified by the large number of splice forms for each isoform, which themselves are at least partially differentially distributed in brain. Thus, neurexins have the requisite diversity to serve as identification markers on neurons.

Two approaches were used to investigate the distribution of neurexin isoforms in brain. First, an extensive in situ hybridization study determined the regions and cells in which the six principal neurexins are expressed. The results demonstrate that neurexin isoforms are specific for neurons and that all classes of neurons express multiple neurexins. The expression of the six neurexin isoforms is independently regulated, and each exhibits a differential distribution. Neurexins Iα, IIβ, and IIIβ were found to be more widely distributed than neurexins Iβ, IIα, and IIIα, which tended to be highly expressed only in selected cell types. In all brain regions studied, each type of neuron had a distinct pattern of neurexin coexpression, supporting the notion that the combinatorial expression of neurexins is a regulated property of neurons.

In a second approach, PCR experiments were performed for alternatively spliced neurexin sequences with RNA from different brain regions. In such experiments, the ratios of insert variants at each splice site exhibited distinct changes between different, rather large brain regions. This suggests that not only the expression of the six principal neurexins but also their alternative splicing varies between different brain regions. Thus, neurexins exhibit a differential, nonrandom expression in brain that at least partially extends to their alternative splicing.

The number of neurexins was also investigated by two methods. First, by cloning a large random collection of bovine cDNAs encoding neurexin Iα, we established the evolutionary conservation of neurexin Iα and its alternative splicing, and we demonstrated that the different alternative splicing events are independent of each other. Second, using PCR on rat brain cDNA, we showed that, with one exception (splice site 2 in neurexin III), all splice sites are uniformly present in all α-neurexins. Accordingly, there is not only an evolutionary conservation of alternative splicing between species but also a conservation of splice sites between neurexins, suggesting that alternative splicing is a property intrinsic to the functions of neurexins. Furthermore, the sequences of the different splice inserts in different neurexins at any splice site are highly homologous to each other. In summary, although the neurexins exhibit significant sequence differences that are evolutionarily conserved, they display a marked uniformity with regard to generating multiple forms: all neurexin genes appear to have dual promoters, and with one exception all neurexin transcripts are extensively alternatively spliced at identical positions.

How many different neurexins are there? If one assumes

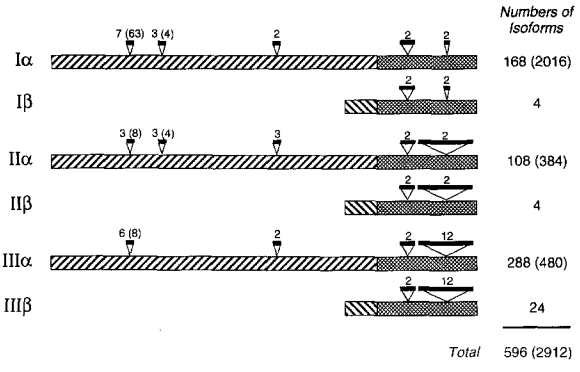

Figure 10. Overview of the Alternative Splicing of Neurexins

The six principal forms of neurexins are schematically represented with the regions shared between α- and β-neurexins shown in a cross-hatched pattern. For each alternatively spliced sequence, the number of sequenced variants is indicated above its sequence position, and the likely number of total variants is given in parentheses. On the right, the total number of distinct isoforms are indicated, calculated on the basis of the assumption that splice sites are used independently. The numbers without parentheses give the total number of isoforms for a neurexin based on the observed sequence variants, and the numbers in parentheses give the numbers of isoforms based on the potential number of variants at each spliced sequence.

that all splice sites are independent of each other, as suggested by the cDNA cloning experiments, and if one counts only splice variants that have actually been sequenced, the minimum number of different proteins would be close to 600 (Figure 10). On the other hand, if the modular elements of each splice site insert sequence allowed further variations, as suggested by the presence of uncharacterized specific products observed in the PCR experiments (Figure 7), there could be more than 3000 distinct neurexins generated by the splice sites in the currently known genes. Even the lesser of these two estimates suggests an unprecedented diversity of cell surface variants generated by alternative splicing. No other known protein family except for the odorant receptors in the nasal mucosa and antigen recognition molecules in the immune system has a similar diversity.

The structures of neurexins suggest functions as cell surface receptors. If the baroque alternative splicing of the neurexins has a functional role, this would most likely involve the specification of ligand interactions. Three of the alternative splice sites (sites 2, 3, and 4; Figure 1) interrupt repeats homologous to the laminin A and agrin repeats, and two of them are between domains (splice sites 1 and 5). If the laminin A–like repeats in neurexins are involved in binding to cell surface receptors, as has been suggested for the corresponding domains of laminin A (Gehlsen et al., 1992), their interruption would be expected to have an effect on this binding. The presence in brain of hundreds of different neurexins would give ample opportunity to confer specificity upon neurons by the expression of selected forms. Specificity would require a large number of specific ligands for the neurexins either belonging to separate protein families or themselves regulated by alternative splicing. A search for such ligands is currently under way.

Experimental Procedures

Bovine cDNA Cloning and Characterization

Approximately 2.5 million plaques of a size-selected amplified cDNA library from bovine brain cortex (gift of Dr. J. Dixon, Merck, Sharp, and Dohme) were screened with a randomly ^{32}P-labeled DNA probe encompassing the entire coding region of rat neurexin Iα (Ushkaryov et al., 1992). Of 150 hybridization-positive plaques identified, 102 were plaque purified, excised as plasmids, and evaluated by restriction enzyme mapping, Southern blotting, and DNA sequencing. Of the 102 original isolates, 87 encoded neurexin Iα, 4 encoded neurexin Iβ, and the remaining clones encoded either other neurexins or nonhomologous unidentified proteins. The 87 neurexin Iα clones contained several identical isolates, resulting in a total of 49 unique clones (Figure 5). Inserts in splice site 1 (Figure 1) were sequenced in all clones; inserts in site 2 were either sequenced or determined by PCR; inserts in sites 3 and 4 were assessed by restriction enzyme digests with enzymes whose ability to cut the DNA depended on the presence or absence of the insert. In addition, in all clones site 5 (and in some clones site 4 as well) was investigated by Southern blotting using splice variant–specific oligonucleotides. Only 3 cDNA clones (36a, 47a, 124a) were fully sequenced. DNA sequencing was performed with an ABI370A DNA sequencer using fluorescently labeled primers and *Taq* polymerase.

Genomic Cloning

Genomic clones containing parts of the neurexin II gene were isolated from murine λ and cosmid libraries (obtained from Stratagene) using standard procedures (Südhof et al., 1985).

Polymerase Chain Reaction

PCRs were performed on single-stranded rat brain cDNA synthesized on total RNA from different brain regions and PC12 cells (Ushkaryov and Südhof, 1993). Single-stranded cDNA was synthesized on 10–20 μg of total RNA in a standard reverse transcriptase reaction primed by oligo(dT). PCRs were performed using ^{32}P-labeled and/or unlabeled primers. A typical PCR mixture contained in 25 μl: 1–3 μl of cDNA template, 1 × PCR buffer (Promega), 2 mM $MgCl_2$, 1 μl of each primer (diluted to 5 OD_{260}), and 0.3 μl of *Taq* polymerase (Cetus). PCRs were run for 25–40 cycles of 30 s at 95°C, 30 s at 60°C, and 1–2 min at 72°C. For analysis of PCR products by electrophoresis, 1 μl of PCR reaction mix was boiled for 5 min with 15 μl of formamide dye (95% formamide, 10 mM EDTA, 0.02% bromophenol blue, 0.02% Xylene Cyanol FF) and immediately loaded on a 5% polyacrylamide minigel containing 8 M urea. Gels were fixed, dried, and exposed to film. For nonradioactive PCR primers, gels were electroblotted onto positively charged nylon membranes, and bands were visualized by hybridization. For sequencing, PCR products were isolated on 5% gels and subcloned into the appropriate M13 vectors.

In Situ Hybridizations

Adult Fisher rats (2–4 months) were anesthetized with halothane and killed. Brains were frozen on a flat block of dry ice covered with aluminum foil, wrapped in parafilm, and stored air-tight at −80°C until use. Frozen sections were cut at a thickness of 15 μm for film autoradiography and 10 μm for emulsion autoradiography and thaw-mounted on RNase-free poly-L-lysine-coated slides. Immediately after sectioning, slides were fixed in 4% paraformaldehyde in PBS (pH 7.4) for 10 min and acetylated (Gerfen et al., 1992), followed by two 5 min delipidizations in chloroform. Air-dried slides were stored desiccated at −25°C until processing. Oligonucleotides were 3′ end–labeled with [α-^{35}S]dATP (DuPont) using calf thymus terminal transferase (Boehringer) and purified by ethanol precipitation. Probes were stabilized with dithiothreitol, diluted to a final concentration of 1 μg/ml (106 cpm/l) in hybridization buffer (40% or 50% formamide, 5 × SSC, 25 mM sodium phosphate buffer [pH 7.0], 1 mM sodium pyrophosphate, 5 × Denhardt's solution, 0.2 mg/ml alkali-cleaved salmon sperm DNA [Sigma], 0.1 mg/ml polyadenylic acid [Boehringer]), and applied to dry slides. Controls con-

taining a 50- or 100-fold excess of unlabeled oligonucleotides to compete out specific binding sites were included in all hybridizations. Hybridizations were performed in a chamber with 5 × SSC, 50% formamide as a humidifying buffer for 12–16 hr at 42°C. After hybridization, slides were washed (Gerfen et al., 1992) in 1 × SSC at 60°C, dehydrated in ethanol, air dried, and exposed to Amersham Hyperfilm beta-max for 3 days to 8 weeks depending on the probe used for the experiment. For emulsion autoradiography, dried slides were coated with Kodak NTB-2 autoradiography emulsion under Kodak safelight filter #2, dried in the dark for 2 hr, and exposed in light-tight containers at 4°C for 1–4 months. Slides were developed at 14°C in Kodak D19, fixed in cold Kodak Rapid fixer, and photographed using microscopes equipped for bright- and dark-field viewing. The oligonucleotides used had the following sequences (all are antisense probes except where specifically noted): Pan-neurexin probes, GGAGATCTAGTTGCGGTACTTGTACATGGCGTAGAGGAGGAT, GTACTTGTACATGGCATACAGAGGATAAGGATGCAGAG, and GTTTCGGTACTTGTACATGGCGTAGAGAAGGAT; neurexin Iα, TCCACTTGGGGAAGCGCGTCCACTGGCCCTCAGCGCCTGGGAACTC; neurexin Iα sense probe, GAGTTCCCAGGCGCTGAGGGCCAGTGGACGCGCTTCCCCAAGTGGA; neurexin Iβ, ATCCCAGCTCGGCACCGCACCGGAGCATCCTCTGGTACATGGC and CACTCCTAGGAGGCCGCTGAGGGTGAGCGGGACTATCCAGA; neurexin IIα, CCAGCGCGCGTAGCGAGCCGACTGCCCGGGGCCGCCGCCAA and GAAAGGCGGCTCGTACTTGACGGTGCTGAGCGTGAGCGCA; neurexin IIβ, CGCGGGCACCCCCCTGCCCGCTCCCCCCGGGGGGCATCTC and CCGCTGCGGGTGAGGAAGGGCATGCGGTTGATGGCGAT; neurexin IIIα, CAGCGTAGATAGCGGGCCCACTGGTTAGGGAGGCCCATGAAC and CAAATACAGGAGCAGCCCAGTGGAAACATTGGTCTTGAAC; neurexin IIIβ, CAAGTGTTTCCCTCTCCCCAGCCTCCAGACATCGGATCATGTA; synapsin I and II, CCAGTCGGTGTGCGGCTCGTCGATTACCAGCAGCACCCTGGC.

Acknowledgments

All correspondence should be addressed to T. C. S. We would like to thank I. Leznicki and A. Roth for excellent technical support and Drs. M. S. Brown and J. L. Goldstein for advice and comments on this work. This study was supported by grants from the Perot Family Foundation and a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft (to B. U.).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *"advertisement"* in accordance with 18 USC Section 1734 solely to indicate this fact.

Received November 28, 1994; revised January 16, 1995.

References

Buck, L., and Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell *65*, 175–187.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823–834.

Gehlsen, K. R., Sriramarao, P., Furcht, L. T., and Skubitz, P. N. (1992). A synthetic peptide derived from the carboxy terminus of the laminin A chain represents a binding site for the $\alpha_3\beta_1$ integrin. J. Cell Biol. *117*, 449–459.

Geppert, M., Ushkaryov, Y. A., Hata, Y., Davletov, B., Petrenko, A. G., and Südhof, T. C. (1992). Neurexins. Cold Spring Harbor Symp. Quant. Biol. *58*, 483–490.

Gerfen, C. R., Young, W. S., and Emson, P. (1992). *In situ* hybridization histochemistry with oligonucleotides for localization of messenger RNA in neurons. In Experimental Neuroanatomy, A Practical Approach, J. P. Bolam, ed. (Oxford: IRL Press), pp. 173–186.

Petrenko, A. B., Lazaryeva, V. D., Geppert, M., Tarasyik, T. A., Moomaw, C., Khokhlatchev, A. D., Ushkaryov, Y. A., Slaughter, C., Nasimov, I. V., and Südhof, T. C. (1993). Polypeptide composition of the α-latrotoxin receptor. J. Biol. Chem. *268*, 1860–1867.

Shepherd, G. M. (1990). The Synaptic Function of the Brain, 3rd edition (New York: Oxford University Press).

Südhof, T. C., Goldstein, J. L., Brown, M. S., and Russell, D. W. (1985). The LDL receptor gene: a mosaic of exons shared with different proteins. Science *227*, 815–822.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581.

Ushkaryov, Y. A., and Südhof, T. C. (1993). Neurexin IIIα: extensive alternative splicing generates membrane-bound and soluble forms in a novel neurexin. Proc. Natl. Acad. Sci. USA *90*, 6410–6414.

Ushkaryov, Y. A., Petrenko, A. G., Geppert, M., and Südhof, T. C. (1992). Neurexins: synaptic cell surface proteins related to the α-Latrotoxin receptor and laminin. Science *257*, 50–56.

Ushkaryov, Y. A., Hata, Y., Ichtchenko, K., Moomaw, C., Afendis, S., Slaughter, C. A., and Südhof, T. C. (1994). Conserved domain structure of β-neurexins. J. Biol. Chem. *269*, 11987–11992.

Zorn, A. M., and Krieg, P. A. (1991). PCR analysis of alternative splicing pathways: identification of artifacts generated by heteroduplex formation. Biotechniques *11*, 181–183.

GenBank Accession Number

The accession number for the sequence reported in this paper is L14855.