Neurexins: Synaptic Cell Surface Proteins Related to the α-Latrotoxin Receptor and Laminin

Yuri A. Ushkaryov, Alexander G. Petrenko, Martin Geppert, Thomas C. Südhof*

A family of highly polymorphic neuronal cell surface proteins, the neurexins, has been identified. At least two genes for neurexins exist. Each gene uses alternative promoters and multiple variably spliced exons to potentially generate more than a 100 different neurexin transcripts. The neurexins were discovered by the identification of one member of the family as the receptor for α-latrotoxin. This toxin is a component of the venom from black widow spiders; it binds to presynaptic nerve terminals and triggers massive neurotransmitter release. Neurexins contain single transmembrane regions and extracellular domains with repeated sequences similar to sequences in laminin A, slit, and agrin, proteins that have been implicated in axon guidance and synaptogenesis. An antibody to neurexin I showed highly concentrated immunoreactivity at the synapse. The polymorphic structure of the neurexins, their neural localization, and their sequence similarity to proteins associated with neurogenesis suggest a function as cell recognition molecules in the nerve terminal.

One fundamental question in understanding the nervous system is how synapses are formed and maintained. The specificity of synaptic connections is a major determinant of brain function. Synapses generally do not contain a basement membrane separating pre- and postsynaptic sites. However, morphologically an undefined "fuzzy" material is found in the synaptic cleft of most synapses (*1*), suggesting that there may be a basement membrane equivalent. Cell-cell contacts at the synapse are presumably mediated by specific cell adhesion molecules or extracellular matrix-like proteins. The simplest way to explain the specificity of synaptic connections would be that the proteins responsible for their establishment and maintenance might be polymorphic and localized to nerve endings. At present, there seem to be no candidate proteins for such a role.

Complementary DNA cloning of α-latrotoxin receptor-related molecules. α-Latrotoxin, a component of the venom from black widow spiders, causes massive neurotransmitter release from the presynaptic nerve terminals of vertebrates (*2*). α-Latrotoxin binds to specific receptors in the presynaptic plasma membrane and initiates secretion of neurotransmitters by an unknown mechanism (*3*). High-affinity receptors for α-latrotoxin have been purified and characterized (*4*). For our study, we obtained extensive peptide sequences from the purified bovine α-latrotoxin receptor (*5*). Degenerate oligonucleotides were synthesized on the basis of these peptide sequences and used in polymerase chain reactions (PCR) to clone cDNA sequences encoding these peptides (*6*). The PCR products and oligonucleotides were then used as probes to screen rat brain cDNA libraries. For the initial screening, we used uniformly ^{32}P-labeled PCR-products as probes. Specific oligonucleotides were used to obtain additional, full-length clones, allowing us to isolate multiple overlapping cDNA's (*7*).

Analysis of the cDNA clones obtained in this manner gave two unexpected results (Fig. 1). First, we isolated two sets of overlapping cDNA's that encoded homologous but distinct proteins, only one of which contained the peptide sequences obtained from the isolated α-latrotoxin receptor. We refer to the proteins encoded by these transcripts as neurexins I and II because they constitute neuron-specific cell surface proteins. Second, each set of cDNA clones was highly polymorphic. Sequence analysis of multiple independent cDNA clones for the neurexins revealed two principal forms that differed by their 5′ coding and untranslated regions. In addition, all messages were alternatively spliced at several internal positions.

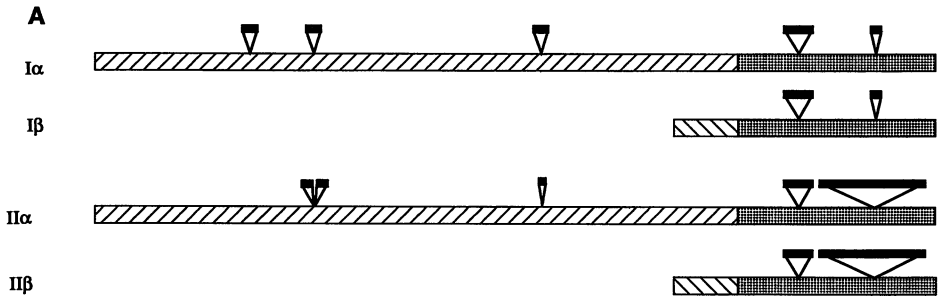
Each of the longer cDNA's, referred to as neurexins Iα and IIα, contained more than 3 kilobases (kb) of coding region that are absent from the shorter forms (referred to as neurexins Iβ and IIβ) (Fig. 1A). Neurexins Iβ and IIβ instead exhibited different and shorter amino-terminal coding regions. In addition to the coding regions, cDNA cloning revealed long 5′ untranslated regions of at least 858 base pairs (bp) for neurexin Iα, 815 bp for neurexin Iβ, 215 bp for neurexin IIα, and 450 bp for neurexin IIβ, with no sequence similarity to each other. This suggests that the α- and β-neurexins contain unrelated 5′ untranslated regions. The α and β forms of the neurexins are most likely transcribed from alternative promoters because they are encoded by separate messenger RNA's and because no sequence overlap could be found in their long 5′ untranslated and coding regions.

Each form of neurexin can undergo multiple differential splicing events. As a criterion for alternative splicing, we used the presence or absence of the respective sequences in independent cDNA clones (mostly in multiple overlapping clones) (*7*). In addition, PCR experiments were performed on single-stranded cDNA that was generated from rat brain poly(A)$^+$-enriched RNA with oligo(dT) as a primer (*6*). In these PCR experiments, specific oligonucleotides were used for the neurexin I splice sites and for most of the neurexin II splice sites, confirming the alternative splice map constructed on the basis of multiple cDNA clones (Fig. 1A). The PCR's also showed a differential distribution of splice sites between cerebrum and cerebellum, suggesting regional differences in expression. Most of the alternatively spliced sequences in the neurexins are small, except for a long sequence (600 bp) in the COOH-terminal region of neurexin II at residue 1421. At this position, neurexin I is also alternatively spliced but only with a 9-bp sequence (Fig. 1). In addition, the COOH-terminus of neurexin II is differentially spliced, resulting in the production of two alternative COOH-termini, only one of which is homologous to that of neurexin I. The alternatively spliced sequences are well separated from each other and likely to be independent. Therefore both neurexins Iα and IIα are probably present in at least 48 different forms, and neurexins Iβ and IIβ in at least 4 and 8 different forms, respectively, resulting in more than 100 different neurexins.

The different forms of neurexins I and II are highly homologous but diverge at the NH_2-termini (Fig. 1B). In spite of this homology, all but one of the peptide sequences obtained from the purified α-latrotoxin receptor were found in the sequence of neurexin Iα but not IIα (underlined in Fig. 1B) (*5*). More than 22 percent of the deduced sequence of neurexin Iα was determined from tryptic fragments of the purified α-latrotoxin receptor. Minor differences were sometimes observed between the bovine peptide sequences and the translated rat sequence, and these were probably due to species differences or peptide sequencing

The authors are with the Howard Hughes Medical Institute and the Department of Molecular Genetics, The University of Texas Southwestern Medical Center, Dallas, TX 75235. A. G. Petrenko is also at the Institute of Bioorganic Chemistry, Moscow 117871, Russia.

*To whom correspondence should be addressed.



B

A. Amino termini of NEUREXINS Iα and IIα

```
I  MGTALVQHGGGCLLGLSLLLLGCWAELGSGLEFPGAEGQWTRFPKW-NACCESEMSFQLKTRSARGLVLYFDDEGFCDFLELILTRG   86
II .ALGSRWRPPPQLPPL-....ALV.-GVR....G.GP...A.YAR.AG.ASTG.L..S.R.NAT.A.L..L..G.D......-.LVD   84

I  GRLQLSFSIFCAEPATLLADTPVNDGAWHSVRIRRQFRNTTLYIDRAEAKWVEVKSKRRDMTVFSGLFVGGLPPELRAAALKLTLAS  173
II ...R.R.TLS.......QL....A.DR..M.LLT.DA.R.A.AV.-G..RAA..R....E.Q.A.D.....I..DV.LS.--...ST  168

I  VRERERPFKGWIRDVRVNSSQALPVDGSEVKLDEEPPNSGG--GSPCEAGD---EGDGGVCLNGGVCSVVDDQAV-CDCSRTGFRGKD  251
II .KYEP..R.LLAN--------------L..G.R..ALL.SQ.LRGA.A.PLCAPARNP.A...L.T.LAPGE.G....H...G..F  237

I  CSQEDNNVEGLAHLMMGDQCKS-------------------KGKEEYIATFKGSEYFCYDLSQNPIQSSSDEITLSFKTLQRNGLM  321
II ..E.EHPM..P...TLNSEVG.LLFSEGGAGRGGAGNVHQPT.....FV.....N.S......H......T.....A.R.......  327

I  LHTGKSADYVNLALKNGAVSLVINLGSGAFEALVEPVNGKFNDNAWHDVKVTRNLRQHSGIGHANV----TISVDGILTTTGYT  401
II .....N....S..S...W.I.............................R.....A.....NKIHYTV.............  414

I  QEDYTMLGSDDFFYVGGSPSTADLPGSPVSNNFMGCLKEVVYKNNDVRLELSRLAKQGDPKMKIHGVVAFKGENVATLDPITFETPE  488
II .............I....N.................D.......FK........E......LQ.DLS.R..D..A...V...S..  501

I  SFISLPKWNAKKTGSISFDFRTTEPNGLILFSHGKPRHQKDAKHPQMIKVDFFAIEMLDGHLYLLLDMGSGTIKIKALQKKVNDGEW  575
II A.VA..R.S..R....L.........L...Q.RRAGAGVGS.SSSQRA.Y..M.L...Y.........G..LR.SSR.......  588

I  YHVDFQRDGRSGTISVNTLRTPYTAPGESEILDLDDELYLGGLPENKAG-LVFPTEVWTALLNYGYVGCIRDLFIDGQSKDIRQMAE  661
II C.........K.S....SRS..FL.T....V...ES.........GGRVD.PL.P.....A.RA.....V.......R.R.L.GL..  675

I  IQSTAGVKPSCSRETAKPCLSNPCKNNGMCRDGWNRYVCDCSGTGYLGRSCEREATVLSYDGSMFMKIQLPVVMHTEAEDVSLRFRS  748
II A.GAV..A.F....L.Q.A.A..R.G.I..E....F....I...F...V.............Y...M..NA............M.  762

I  QRAYGILMATTSRDSADTLRLELDAGRVKLTVNLDGTRINGNSS---KGPETLFAGYNLNDNEWHTVRVVRRGKSLKLTVDDQQANT  832
II .....LM......E..........G.QM.........L.VG.AP.AAA.........HK.........L.........Q.S.-.NVTVE  848

I  GQMAGDHTRLEFHNIETGIITERRYLSSVPSNFIGHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIIADPVTFKTKSSYVALA  919
II .....A.............M....FI.V.........SG.V...QP.M.Q..D...T.........L.A.V.......SR...L...  935

I  TLQAYTSMHLFFQFKTTSLDGLILYNSGDGNDFIVVELVKGYLHYVFDLGNGANLIKGSSNKPLNDNQWHNVMISRDTSNLHTVKID 1006
II .....A...........AP...L.F...N......I......I.........SPS.M..N.D..V.........VV...PG.V..L... 1022

I  TKITTQITAGARNLDLKSDLYIGGVAKETYKSLPKLVHAKEGFQGCLASVDLNGRLPDLISDALFCNGQIERGCEGPSTTCQEDSCS 1093
II SRTV..HSN........GE.....LS.NMFSN.....ASRD.....................A...HRI..V....D......T.E..A 1109

I  NQGVCLQQWDGFSCDCSMTSFSGPLCNDP                                                            1122
II .............T...T...YG..V....                                                            1138
```

B. Amino termini of NEUREXINS Iβ and IIβ

```
I  MYQRMLRCGAELGSPGGGSSGGAGGRLALLWIVPLTLSGLLGVAWGA---SSLGA-HHIHHFHGSSKHHSVPIAIYRSPASLRGGHA   83
II .PPGGSGQ.GCPRR.PALAGPLPPPPPPPPLPLL.G.LL...A.E..RVS...STT..V....-...GT.....N.M.FLT.S...   85
```

C. Carboxyl terminal region common to NEUREXINS Iα, Iβ, IIα and IIβ

```
I  GTTYIFSKGGGQITYKWPPNDRPSTRADRLAIGFSTVQKEAVLVRVDSSSGLGDYLELHIHQGKIGVKFNVGTDDIAIEESNAIIND 1209
II ......G...AL...T.........M....V....H.RS........A.......Q...D..TV..I.........T.D.P...VS. 1225

I  GKYHVVRFTRSGGNATLQVDSWPVIERYPADNNDNRRLATARQRIPVRLGRVVDPVTLDVKGRQLTIFNSQATIIIGGKEQGQPFQGQ 1296
II .........................N.....F.....................................A.K...RD..R..... 1312

I  LSGLYVNGLKVLNMAAENDANIAIVGNVRLVGEVPSSMT-TESTATAMQSEMSTSIMETTTTLATSTARRGKPPT-KEPISQTTDDI 1381
II V..........AL...S.P.VRTE.HL....G..VLLSA.T...TLLAD.A.T.......M..T.T...RS..MRDSTT.N...L 1399

I  LVASAECPSDDEDIDPCEPSSGGS---------------------------------------------------------------- 1405
II .............LEE....T..ELILPITTEDSLDPFPVATRSPFVPPPPTEVPFETGVGATQTDLPPEAARRPSSGQPQQAEKDD 1486

I  ---------------------------------------------------------------------------------------- ----
II SDGEEPVEASGPASGKVFDSSLFPTDDEDFYTTFPLVTDRTTLLSPRKPAPRPNLRTDGATGAPGVLLAPSAPAPNLPAGKHNHRDP 1573

I  ----------------------------------ANPTRVGGREPYPGSAEVIRESSSTTGMVVGIVAAAALGILILLLYA 1451
II EQPLEKNFPLGPQVVTAFEPRRPPPLKPGVTSVPGFPPLPT....GP.E.G.-..AV.....|..................... 1659

I  MYKYRNRDEGSYHVDESRNYISNSAQSNGAVVKEKQPSSAKSANKNKKNKDKEYYV*                                1507
II .|.....Q..Q.................A.AAP.TPS.A.....*                                            1715
```

D. Alternatively spliced carboxyl termini of NEUREXINS IIα and IIβ

```
II MYKYRKRCRKSPREEKLLPGSAQGLGLDLAKACCVCRGRATCIAGKPLEERGGGRGEGERQMQIYIKNK*                  1728
```

Fig. 1. Primary structures of different forms of neurexins. (**A**) Bar diagram of the different forms of neurexins I and II as predicted from the sequences of multiple cDNA's (*7*). The NH_2-terminal domains unique to the α forms or β forms of the neurexins are shown in a right- or left-bound hatch pattern, respectively. Regions shared by α- and β-neurexins are shown in a finely stippled pattern. Triangles with horizontal solid lines denote alternatively spliced segments. (**B**) Amino acid sequences of the different forms of neurexins I and II. The variable amino termini of the α- and β-neurexins are shown in sections A and B and the variant COOH-terminus of neurexin II is shown in section D. The middle region common to all neurexins is shown in section C. The putative transmembrane sequence is boxed, and alternatively spliced sequences are shaded. The differentially spliced sequences at position 254 of neurexin Iα and at position 385 in neurexin IIα exhibited three forms (*7*). Dots signify identities between the neurexin I and II sequences. Numbering of the sequences in regions where α- and β-neurexins are identical refers to the α forms. The underlined sequences starting at residues 31, 265 and 269 (asterisks) were identified in the mixed NH_2-terminal sequence of the receptor (*5*). Thickly underlined sequences of neurexin I denote amino acid sequences determined from purified bovine α-latrotoxin receptor protein, with sequence ambiguities or discrepancies shown as thin lines. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

ambiguities (*8*). In addition to the internal peptide sequences, a mixed NH_2-terminal sequence obtained from the purified protein was also identified in translated amino acid sequence. Finally, we tested whether neurexins and α-latrotoxin could be co-immunoprecipitated. Solubilized rat brain membrane proteins that had been incubated with 0.4 nM ^{125}I-α-latrotoxin were subjected to immunoprecipitation with two independent antibodies to the COOH-terminus of neurexin Iα (Table 1). Specific co-immunoprecipitation of receptor-bound α-latrotoxin was observed with the test antibodies but not with preimmune sera or control antibodies. Together the immunological and structural data suggest that cloned neurexin Iα is either identical with or closely related to the α-latrotoxin receptor.

Domain structure of the neurexins. Analysis of the primary structures of the neurexins reveals a distinctive domain structure (Fig. 2A). The NH_2-termini of neurexins Iα and IIα contain hydrophobic sequences characteristic of signal peptides (*9*). These signal sequences are followed by three copies of a repeat of approximately 450 amino acids (labeled A, B, and C). Each repeat is composed of a central epidermal growth factor (EGF) domain that is flanked by a left arm and a right arm. The right and left arms of each repeat are themselves weakly homologous to each other (Fig. 2B). Accordingly, most of the extracellular domains of neurexins Iα and IIα are composed of three larger repeats encompassing six smaller repeats punctuated by interspersed EGF domains (*10*).

In contrast, the NH_2-termini of neurex-

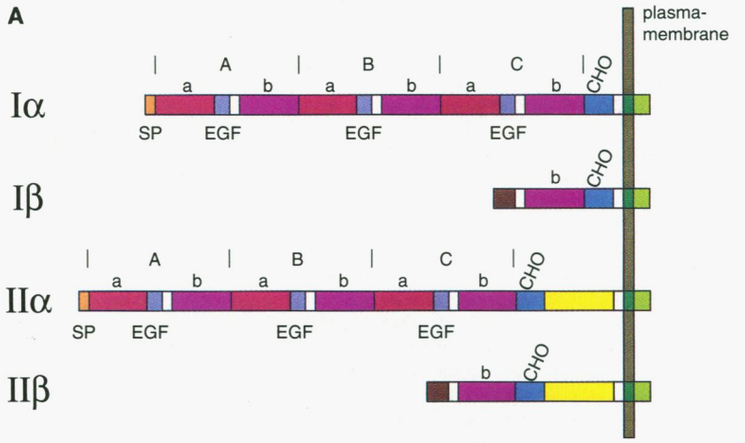
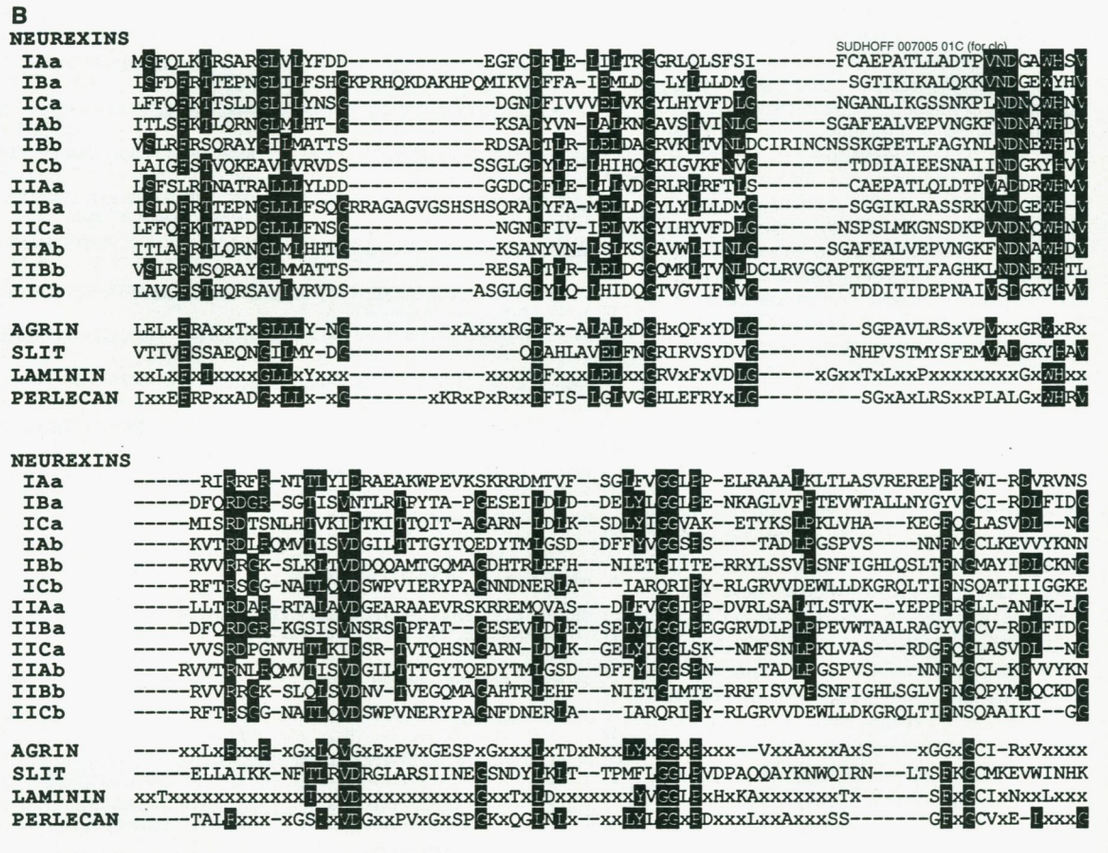
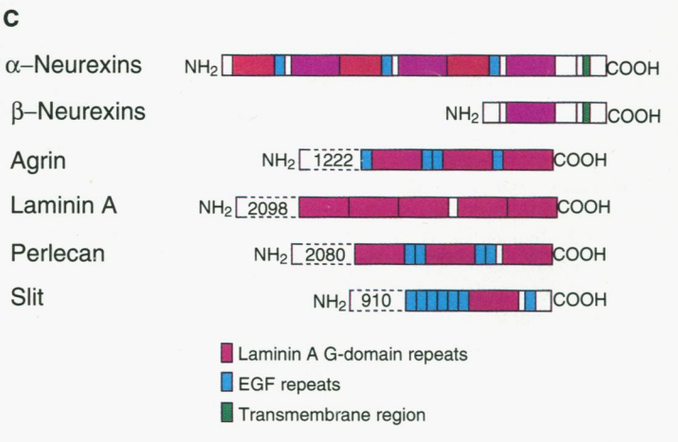


Fig. 2. Domain structure of neurexins. (**A**) Protein domains and transmembrane structure of the neurexins. Proteins are shown as bar diagrams with NH_2-termini on the left and COOH-termini on the right. In neurexins Iα and IIα, the signal peptide (labeled SP) is followed by three repeats designated A, B, and C. Each repeat contains a central EGF-domain (shown in blue) flanked by right and left arms (named a and b and shown in two shades of red) that are weakly homologous to each other. Neurexins Iβ and IIβ contain only the right arm of the third repeat (C) preceded by sequences different from these occurring in α-neurexins. The putative O-linked sugar domain common to all neurexins is labeled CHO. The single transmembrane regions and cytoplasmic tails are shown in dark and light green, respectively. The major large alternatively spliced fragment in neurexins IIα and IIβ is shown in yellow. (**B**) Alignment of the sequences of the neurexin repeats with each other and with COOH-terminal sequences from rat agrin, mouse laminin A, mouse perlecan, and the *Drosophila* mutant *slit* (*27*). Residues that occur at least in half of the neurexin repeats are shown in white on a black background and are so indicated if they also occur in the other proteins at the same position. For agrin, laminin A, and perlecan, consensus sequences derived from their repeats are shown, with x denoting variable residues. For the neurexin sequences, IAa and IAb denote the left and right arms of the A repeat of neurexin Iα, respectively, with the other repeats named accordingly. (**C**) Relative location of the neurexin or laminin A G-domain repeats (shown in red) in different proteins. Agrin, *slit*, perlecan and laminin A represent secreted extracellular matrix proteins with COOH-terminal repeats whereas the neurexins are cell surface proteins with a transmembrane region (depicted in green) and an NH_2-terminal localization of these repeats. The numbers in the dashed boxes for laminin A, agrin, perlecan, and *slit* indicate the numbers of residues preceding the domains shown. The dashed COOH-terminal domain in the neurexins corresponds to the major alternatively spliced segment in neurexin II.





ins Iβ and IIβ are much shorter than those of the α-neurexins and lack most of the extracellular repeats. At the NH_2-terminus, the β-neurexins contain sequences of 83 and 85 amino acids, respectively, that are hydrophobic but unusually rich in proline, glycine, and histidine. After this sequence, neurexins Iβ and IIβ splice into the sequences of the α-forms precisely at the end of the EGF-domain in the third repeat. As a consequence, the β-neurexins contain only the last one of the six repeats that make up most of the extracellular domains of the α-neurexins (Fig. 2A).

After the repeats, all neurexins contain a serine- and threonine-rich sequence. Databank searches indicate a similarity between this sequence and serine- and threonine-rich sequences in the low density lipoprotein receptor and Alzheimer precursor proteins that have been suggested to serve as O-linked sugar domains in these proteins (*11*). Digestion of the purified α-latrotoxin receptor with endoglycosidase F or with neuraminidase followed by *O*-glycanase demonstrated that the receptor is extensively post-translationally modified with both N-linked and O-linked sugars (Fig. 3). Analysis of neurexin Iβ transfected into COS cells showed that it is also extensively O-glycosylated (*12*). Since the putative O-linked sugar region is shared by α- and β-neurexins, these results suggest that, like in other cell surface receptors, the threonine- and serine-rich sequence of the neurexins may serve as an O-linked sugar domain.

After the O-linked sugar domain, the neurexins contain a hydrophobic sequence that appears to constitute their only transmembrane region as judged by hydrophobicity plots (*13*). The transmembrane region is followed by a short, highly charged cytoplasmic tail. In neurexin II, alternative splicing creates two different cytoplasmic tails of similar size, only one of which is homologous to that of neurexin I (Fig. 2B).

The evidence for the proposed transmembrane orientation of the α-neurexins consists of the presence of a signal sequence, EGF domains (which are characteristic of extracellular proteins), and glycosylation domains. The β-neurexins are proposed to have the same transmembrane orientation

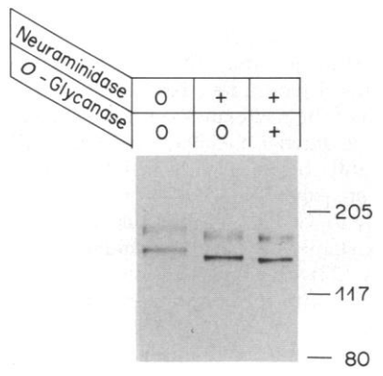


Fig. 3. O-linked glycosylation of the α-latrotoxin receptor. Purified α-latrotoxin receptor was analyzed by SDS–polyacrylamide gel electrophoresis (PAGE) and immunoblotting either in native form or after digestion with neuraminidase or with neuraminidase followed by *O*-glycanase (*12*). Numbers on the right indicate positions of molecular markers (in kilodaltons). Purified receptor consists of two major bands of large molecular mass that are both glycosylated (*12*).

Table 1. Immunoprecipitation of receptor-bound ^{125}I-α-latrotoxin with antibodies to neurexin I. Proteins from rat brain membranes solubilized in CHAPS were incubated with 0.4 nM ^{125}I-α-latrotoxin (6.5 Ci/μmol) in 50 mM tris-HCl, pH 8.0, 2 mM EDTA, 3 mM $CaCl_2$, and 0.2 mM PMSF on ice. Portions of this solution (70 μg of protein) were added to protein A–Sepharose that had been coated with the indicated immune or preimmune sera. Samples were incubated at 4°C overnight with agitation. Protein A–Sepharose was then centrifuged, washed with phosphate-buffered saline, and the radioactivity associated with the protein A–beads was counted. Each experiment was performed in triplicate. The antibody to α-latrotoxin was included in the experiment to assess the total amount of precipitable ^{125}I-α-latrotoxin.

Antibody*	Immunoprecipitated ^{125}I-labeled α-latrotoxin (10^3 cpm ± SD)	
	Immune serum	Preimmune serum
Anti-neurexin I (a)	0.76 ± 0.03	0.23 ± 0.06
Anti-neurexin I (b)	0.72 ± 0.07	0.22 ± 0.03
Control antibody	0.28 ± 0.03	0.23 ± 0.04
Anti-α-latrotoxin	10.65 ± 1.56	0.37 ± 0.01

*Antibodies used: anti-neurexin I (a) and (b) correspond to two independently produced sera to the COOH-terminus of neurexin I (*22*). Purified toxin was used to produce the antibody to α-latrotoxin in rabbits.

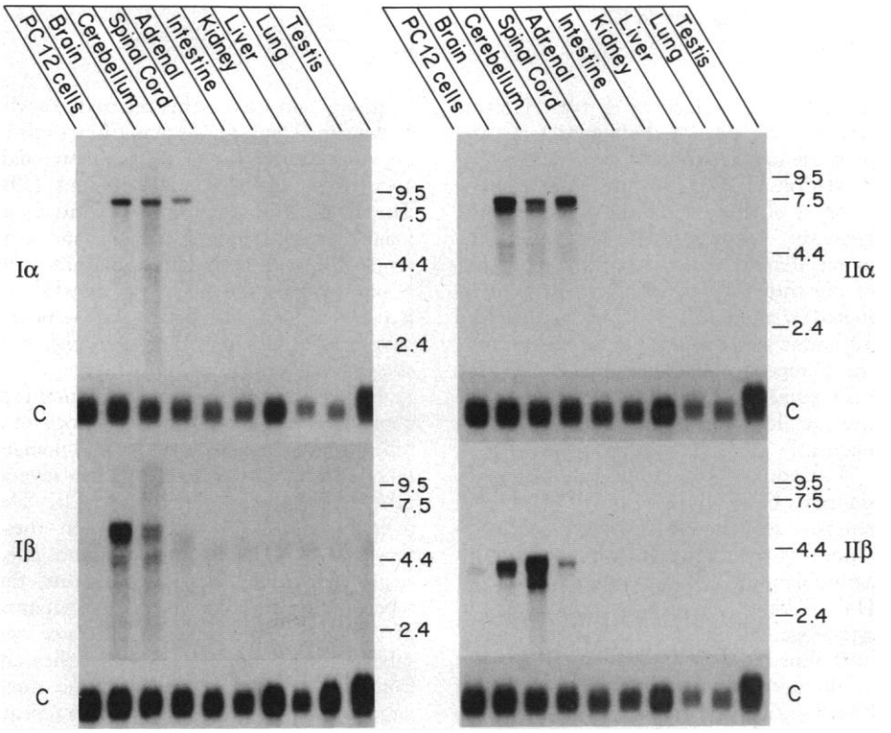


Fig. 4. Analysis of the neurexin mRNA's by RNA blotting. Total RNA (10 μg) from the indicated tissues was subjected to electrophoresis, blotted, and hybridized to uniformly ^{32}P-labeled probes specific for neurexins Iα (top left), Iβ (bottom left), IIα (top right) or IIβ (bottom right) (*28*). After exposure, the blots were rehybridized with a uniformly labeled probe for cyclophilin (lower panels labeled C) to control for RNA loading. Numbers on the right denote the relative positions of molecular markers in kilobases. Exposure times (with screen at −70°C): Iα and IIα, 48 hours; Iβ, 4 days; IIβ, 9 hours; cyclophilin, 16 hours. The faint 5-kb band observed in all lanes in the Iβ blot corresponds to 28*S* RNA.

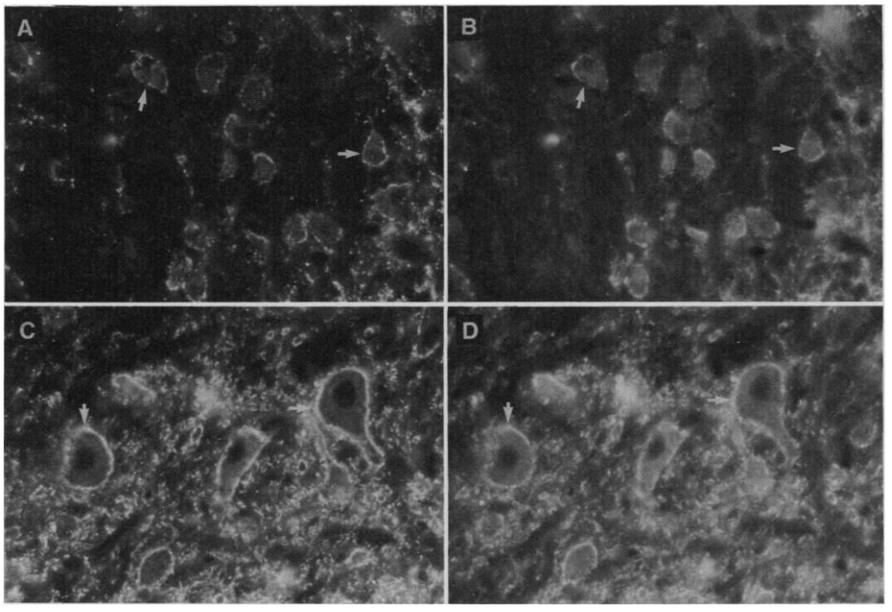


Fig. 5. Immunocytochemical localization of neurexin I in rat brain. Sections from rat brain (**A** and **B**: brain stem; **C** and **D**: anterior horn of the spinal cord) were double-labeled with a monoclonal antibody to synaptophysin (A and C) and an affinity-purified polyclonal antibody to the cytoplasmic tail of neurexin I (B and D) (*22*). Antibody-reactive sites were visualized by immunofluorescence with fluorescein-labeled goat antibody to mouse protein (A and C) and rhodamine-labeled goat antibody to rabbit protein (B and D). Arrows indicate cell bodies decorated by synapses that are identically stained by the antibodies to synaptophysin and neurexin I. Identical staining patterns were observed in single-label immunofluorescence experiments whereas controls with preimmune serum showed no staining (*22*).

in spite of the lack of a typical signal sequence because of their partial identity with the α-neurexins and their similar glycosylation.

Most of the differential splicing in the neurexins involves small sequences in the extracellular repeats. Each of the right arms of the three overall large repeats is interrupted by alternative splicing. A conserved 30–amino acid sequence in the right arm of the C repeat is differentially spliced in all neurexins. Additional alternative splice sites are located between the protein domains.

Databank searches revealed a sequence similarity between the neurexin repeats and repeated sequences found in the COOH-termini of the extracellular matrix proteins agrin, laminin A, and perlecan (Fig. 2B) (*14*). Agrin is a neuronal protein active in synaptogenesis and causes clustering of acetylcholine receptors on muscle cells (*15*); it is alternatively spliced in the region of homology with the neurexins. This splicing regulates the agrin activity (*16*), suggesting that its active site is located in the region of sequence similarity with neurexins. Laminin and perlecan are components of basement membranes (*17, 18*). Laminin is a trimeric complex formed by laminins A and B. Only laminin A contains the COOH-terminal repeats that are homologous to the neurexins and are together referred to as the G domain. Laminin has major functions in organogenesis, particularly in axon guidance during neuronal development (*19*). The G domain of laminin A constitutes its major heparin-binding domain and may function in cell adhesion of laminin (*20*). A single copy of the repeats described above is also found in the sequence of *slit*, a neural developmental mutant from *Drosophila melanogaster* (Fig. 2, B and C) (*21*).

In consequence, α-neurexins contain six copies and β-neurexins a single copy of a repeat that is homologous to the G-domain repeats of laminin A and found in several other proteins (Fig. 2, B and C). The overall sequence identity between these repeats is low, and clusters at conserved sequence motifs. In the α-neurexins, the repeats coincide with the left and right arms of the three larger repeats and form two subclasses, depending on whether they are from the left or right arms of the three larger repeats. Each of the right arm repeats but none of the left arm repeats contains an alternative splice site. Together, the repeats account for more than half of the neurexin Iα and IIα proteins. In the neurexins, the G-domain repeats form the extracellular NH_2-termini of a transmembrane protein, whereas in the other proteins they constitute the COOH-termini of secreted polypeptides. In all proteins except laminin A, EGF domains are associated with these repeats but with a different arrangement in each protein (Fig. 2C).

Neural expression of neurexins. RNA blotting experiments were used to investigate the tissue distribution of neurexin transcripts (Fig. 4). The expression of the neurexins was brain-specific. Even adrenal gland, which expresses many otherwise neuron-specific genes, had no detectable mRNA. This finding correlates well with the absence of α-latrotoxin binding sites from adrenal membranes. Neurexins Iα and Iβ each exhibited single-sized mRNA's that were diffuse in appearance, presumably as a result of multiple alternatively spliced transcripts. Several transcripts of different sizes were observed for neurexins IIα and IIβ as is expected from their more extensive differential splicing.

A heterogeneous distribution of the neurexin mRNA's between different brain regions was observed. For example, the cerebellum was rich in Iα compared to Iβ but contained much less IIα than IIβ (Fig. 4). Conversely, spinal cord contained much more of both α forms than the β forms. Two different neurexins can also be coexpressed in the same cell: PC12 cells (which are related to neurons) exhibited mRNA's for Iα and IIβ at low levels, whereas no Iβ or IIα transcripts were detected (Fig. 4). These results suggest that the different promoters for the α- and β-neurexins may be differentially regulated as is the case for alternative splicing of the neurexins.

Parallel binding measurements of ^{125}I-labeled α-latrotoxin to membranes from PC12 cells and rat brain indicated that PC12 cells contain approximately 10 percent of the specific α-latrotoxin binding sites present in rat brain, a result in agreement with the RNA levels for neurexin Iα.

Although the RNA blotting experiments demonstrated brain-specific expression of the neurexins even to the exclusion of the adrenal medulla, this result did not identify the cell type in which the neurexins are expressed. To achieve this, an antibody to the COOH-terminus of neurexin I was affinity-purified for immunocytochemistry (*22*). Rat brain sections were double-labeled with the affinity-purified polyclonal antibody to neurexin I (Fig. 5, B and D) and a monoclonal antibody to synaptophysin as a synaptic marker (Fig. 5, A and C). Neurexin I immunoreactivity was heterogeneously distributed in brain, suggesting differential expression. In those regions where neurexin I immunoreactivity could be detected, the signal obtained with two independent antibodies precisely coincided with the synaptophysin signal, suggesting an exclusive synaptic localization. Control stains with preimmune serum showed no

reactivity (*22*). Because of the strong homology between neurexins I and II in the COOH-terminus, it cannot be concluded with certainty whether the observed signal corresponds exclusively to neurexin I. However, preliminary results with an antibody raised to the COOH-terminus of neurexin II also revealed a concentration of the immunoreactivity in the nerve terminal in a pattern that was distinct from that observed with the antibody raised to neurexin I.

Neurexins: Cell recognition molecules of the nerve terminal. Starting with peptide sequences derived from purified α-latrotoxin receptor, we define a family of cell surface proteins named neurexins. Structurally the neurexins resemble cell surface receptors. The neurexins are highly polymorphic, with two principal forms generated by the use of alternative promoters and multiple alternatively spliced regions creating potentially more than a hundred forms. The neurexins are brain-specific and show differential expression of distinct forms in different brain regions. At least one neurexin appears to be highly enriched in the synapse as judged by immunocytochemistry presented and by the previous localization of the α-latrotoxin receptor (*3*). Together these properties correspond to those expected of cell surface recognition molecules involved in synapse formation.

A role for the neurexins in shaping cell-cell interactions is supported by their homology to the extracellular matrix proteins agrin, laminin, and *slit*. These proteins are thought to participate in axon guidance and synaptogenesis (*18–21*). The regions of homology with these proteins comprise most of the extracellular domains of the neurexins, suggesting that the same protein motif may be used in different proteins for similar functions with different specificities. Finally, a specific interaction of a synaptic vesicle protein, synaptotagmin, with at least one member of the neurexin family, the α-latrotoxin receptor, was previously observed (*23*). It is possible that the neurexins may align synaptic vesicles in the nerve terminal with structures on the extracellular side of the synaptic membrane.

Several structural features of the neurexins are unusual. Other currently described proteins that contain sequences homologous to the G-domain repeats of laminin A are secreted proteins, whereas the neurexins incorporate these repeats within the structure of a cell surface receptor (Fig. 2C). The restricted expression of the neurexins in the nervous system distinguishes them from other cell surface proteins involved in cell recognition such as N-CAM or cadherins that are often widely expressed (*19*). Although many genes contain alternative promoters, these usually represent developmental genes in which alternative promoters regulate expression of similar coding regions (*24*). The α and β forms of the neurexins are also differentially expressed but their encoded proteins differ by more than 1000 amino acids. This large difference in the gene products is unusual.

As to the relation between the α-latrotoxin receptor and neurexin Iα, the identity of the peptide sequences obtained from the α-latrotoxin receptor with more than a fifth of neurexin Iα suggests that the α-latrotoxin receptor represents a variant of neurexin Iα. This conclusion is supported by their coexpression in PC12 cells, their absence from adrenal gland, and their common synaptic localization. Furthermore, the purified α-latrotoxin receptor reacts with several independent antibodies against different parts of neurexin Iα (*25*). Finally, antibodies to the COOH-terminus of neurexin Iα (*22*) coimmunoprecipitate α-latrotoxin from total rat brain membrane proteins (Table 1). Although these data support the suggested identity between the α-latrotoxin receptor and a splice variant of neurexin Iα, we have been unable to obtain α-latrotoxin binding to neurexin Iα expressed in COS cells, and the distribution of the immunoreactivity and the mRNA of neurexin Iα in brain does not precisely correspond to the distribution observed for α-latrotoxin binding (*26*). Although further evidence will be required to define the exact relation of the receptor to neurexin Iα, it is clear that the α-latrotoxin receptor represents a neurexin and constitutes a neural cell surface protein most likely involved in cell recognition.

REFERENCES AND NOTES

1. A. Peters, S. L. Palay, H. deF. Webster, *The Fine Structure of the Nervous System* (Oxford Univ. Press, New York, 1991).
2. H. E. Longenecker, W. P. Hurlbut, A. Mauro, A. W. Clark, *Nature* **225**, 701 (1970); N. Frontali *et al.*, *J. Cell Biol.* **68**, 462 (1976); M.-C. Tzeng and P. Siekevitz, *J. Neurochem.* **33**, 263 (1979).
3. F. Valtorta, L. Maddedu, J. Meldolesi, B. Ceccarelli, *J. Cell Biol.* **99**, 124 (1984).
4. H. Scheer and J. Meldolesi, *EMBO J.* **4**, 323 (1985); A. G. Petrenko *et al.*, *ibid.* **9**, 2023 (1990).
5. Tryptic fragments from the purified α-latrotoxin receptor were isolated by high-performance liquid chromatography and sequenced (A. G. Petrenko *et al.*, in preparation). The receptor preparation was subjected to NH_2-terminal sequencing, resulting in a mixed sequence that could tentatively be assigned to NH_2-terminal and internal positions of the translated cDNA sequence (Fig. 1B). All but one of the peptide sequences were identified in the translated nucleotide sequence of neurexin Iα. The unidentified sequence corresponded to FNDNAWELTxDVExQ (x denotes unidentified residues). The peptide sequence corresponding to residues 362 to 371 in Fig. 1B contained two insertions and was FNDNAWAHVDVK.
6. R. K. Saiki *et al.*, *Science* **230**, 1350 (1984). All PCR were performed on single-stranded cDNA's generated from poly(A)$^+$-enriched RNA from different brain regions. The PCR conditions with redundant oligonucleotides consisted of two initial cycles at 94°C (5 minutes), 40°C (2 minutes), and 72°C (10 minutes), followed by 33 cycles of 94°C (1 minute) and 65°C (5 minutes). The PCR products were purified by acrylamide gel electrophoresis, cloned into M13 vectors, and sequenced. The PCR conditions with nonredundant oligonucleotides consisted of two cycles at 95°C (5 minutes), 62 to 68°C (2 minutes), and 72°C (10 minutes) followed by 25 cycles of 95°C (1 min), 62 to 68°C (2 minutes), and 72°C (2 minutes). All DNA sequencing was performed with M13 subclones and an automated ABI DNA sequencer model 373A.
7. Complementary DNA screening with rat and bovine brain libraries constructed in λZAP (Stratagene) was done essentially as described [T. C. Südhof *et al.*, *Science* **238**, 1142 (1987); T. C. Südhof, *J. Biol. Chem.* **265**, 7849 (1990)]. Twenty-two clones were partially or fully sequenced for rat neurexin Iα, seven clones for neurexin Iβ, six clones for rat neurexin IIα, and six clones for rat neurexin IIβ. All nucleotide sequences have been submitted to GenBank (accession numbers M96374 to M96377). Of the fully sequenced cDNA clones, the following distribution of splice sites was observed. For neurexin I, position 254: three clones contained the full insert, three clones had only a deletion of the Glu at position 255, and three clones lacked the complete insert; position 379: two clones had the complete insert and six lacked it; position 783: ten clones contained the full insert and two lacked it; position 1239: four clones contained the insert and seven lacked it; position 1403: five clones each contained or lacked the insert. For neurexin II, position 385: one sequenced clone each completely lacked or contained the insert, whereas two clones contained the first eight amino acids but lacked the remaining seven; position 807: two clones contained the complete insert and one lacked it, thereby changing the Ser at 808 to Thr; position 1256: three clones each contained or lacked the insert; position 1421: five cDNA's contained and two lacked the insert. Of the COOH-termini, only one clone had the alternatively spliced cytoplasmic tail whereas five clones contained the COOH-terminus homologous to neurexin I. Two consistent sequence polymorphisms resulting in amino acid changes were observed in neurexin I and one in neurexin II: the Gly at position 850 of neurexin Iα was a Gly in five and a Ser in two clones; the Tyr at position 1137 of neurexin Iα was a Tyr in four and a His in two clones; and the Asn at position 434 of neurexin IIα was an Asn and Lys in two clones each.
8. In addition to the rat cDNA sequences, partial bovine cDNA sequences encompassing residues 462 to 910 of neurexin Iα were determined from PCR products and cDNA clones. The bovine sequences contained the same splicing patterns as the rat sequence and was identical to it except for two positions (I at position 662 was a Val, and Gly at position 833 was a Ser in the bovine sequence).
9. G. von Heijne, *Nucleic Acids Res.* **14**, 4683 (1986).
10. The three EGF repeats in neurexins Iα and IIα are of the A-type according to the classification of Doolittle *et al.*, [R. F. Doolittle, D. F. Feng, M. S. Johnson, *Nature* **307**, 558 (1984)]. They are most closely homologous to the EGF repeats in the drosophila neurogenic loci *Notch, Delta*, and *Slit* [S. Kidd *et al.*, *Cell* **34**, 421 (1983); K. A. Wharton *et al.*, *ibid.* **43**, 567 (1985); H. Vassin *et al.*, *EMBO J.* **6**, 3431 (1987); (*21*)].
11. T. C. Südhof, J. L. Goldstein, M. S. Brown, D. W. Russell, *Science* **228**, 815 (1985); J. Kang *et al.*, *Nature* **325**, 733 (1987).
12. For enzymatic deglycosylation of purified α-latrotoxin receptor and of neurexins Iα and Iβ expressed in COS cells, we used neuraminidase (Sigma), endoglycosidase F (Boehringer-Mannheim), *N*-glycanase, and *O*-glycanase (Genzyme) according to the manufacturer's suggestions. Controls for proteolysis consisted in parallel incu-

bations in the absence of enzyme. Of the two major polypeptides in purified bovine and rat α-latrotoxin receptor preparations (*2*), the larger band has the same size as glycosylated neurexin Iα expressed in COS cells. The smaller band could be a result of currently unidentified further alternative splicing or of in vivo or in vitro proteolysis. The latter possibility is supported by the mixed NH_2-terminal sequence, part of which was assigned to an internal sequence in neurexin Iα (Fig. 1B).
13. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982); P. Petr, M. Kanehisa, C. DeLisi, *Biochim. Biophys. Acta* **815**, 46 (1985).
14. GenBank release 70 and NBRF release 30 were searched for homologies both at the amino acid and protein levels using standard programs.
15. U. J. McMahan and B. G. Wallace, *Dev. Neurosci.* **11**, 227 (1989); F. Rupp, D. G. Payan, C. Magill-Solc, D. M. Cowan, R. H. Scheller, *Neuron* **6**, 811 (1991); J. T. Campanelli, W. Hoch, F. Rupp, T. Krieger, R. H. Scheller, *Cell* **67**, 909 (1991).
16. M. A. Ruegg *et al.*, *Neuron* **8**, 691 (1992); M. Ferns *et al.*, *ibid.*, in press.
17. M. Sasaki, H. K. Kleinman, H. Huber, R. Deutzmann, K. Yamada, *J. Biol. Chem.* **263**, 16536 (1988); G. R. Martin and R. Timpl, *Annu. Rev. Cell Biol.* **3**, 57 (1987).
18. D. M. Noonan *et al.*, *J. Biol. Chem.* **266**, 22939 (1991).
19. U. Rutishauser and T. M. Jessell, *Physiol. Rev.* **68**, 819 (1988); L. F. Reichardt *et al.*, in *Molecular Neurobiology*, S. J. Zalcman and R. Scheller, Eds. (U.S. Government Printing Office, Washington, DC, 1989), pp. 110–127; J. R. Sanes, *Annu. Rev. Neurosci.* **12**, 491 (1989).
20. U. Ott, E. Odermatt, J. Engel, H. Furthmayr, R. Timpl, *Eur. J. Biochem.* **123**, 63 (1982); P. D. Yurchenco, Y.-S. Cheng, J. C. Schittny, *J. Biol. Chem.* **265**, 3981 (1990); A. P. N. Skubitz, P. C. Letourneau, E. Wayner, L. T. Furcht, *J. Cell Biol.* **115**, 1137 (1991).
21. J. M. Rothberg, J. R. Jacobs, C. S. Goodman, S. Artavanis-Tsakonas, *Genes Dev.* **4**, 2169 (1990).
22. A cDNA fragment encoding amino acids 1451 to 1507 of neurexin I was prepared by PCR (*4*) with oligonucleotide primers flanked by restriction sites. The fragment was cloned into a pGEX vector. Purified recombinant fusion protein was used for antibody production in rabbits [D. B. Smith and K. S. Johnson, *Gene* **67**, 31 (1988); P. A. Johnston, R. Jahn, T. C. Südhof, *J. Biol. Chem.* **264**, 1268 (1989)]. Antibody from serum was affinity-purified on purified fusion protein immobilized on CNBr-activated Sepharose 4B. For immunofluorescence staining, paraformaldehyde-fixed frozen sections were double-labeled with affinity-purified antibodies to neurexin and with monoclonal antibodies to synaptophysin, with controls for specificity as described [T. C. Südhof *et al.*, *Science* **245**, 1474 (1989)]. Additional controls included (i) single-label immunofluorescence experiments with only one antibody to estimate light bleeding between the fluorescein and rhodamine channels, (ii) preimmune sera, and (iii) use of antisera to other pGEX-fusion proteins. Two independently raised sera gave similar results.
23. A. G. Petrenko *et al.*, *Nature* **353**, 65 (1991).
24. U. Schibler and F. Sierra, *Annu. Rev. Genet.* **21**, 237 (1987).
25. We prepared antibodies to peptides and recombinant proteins corresponding to the following residues of the neurexin Iα sequence (Fig. 1B): 476 to 679; 712 to 728; 905 to 1100; 1451 to 1507; and 1491 to 1507. All of these antibodies reacted with purified α-latrotoxin receptor on immunoblots but only the antibodies to the COOH-terminus recognized the native protein.
26. A. Malgaroli, P. DeCamilli, J. Meldolesi, *Neuroscience* **32**, 393 (1989).
27. The aligned sequences were from the following residues: (i) Neurexins (Fig. 1B): IAa, 54 to 191; IBa, 504 to 651; ICa, 929 to 1060; IAb, 308 to 446; IBb, 742 to 886; ICb, 1152 to 1288; IIAa, 53 to 185; IIBa, 517 to 665; IICa, 945 to 1076; IIAb, 314 to 458; IIBb, 756 to 902; and IICb, 1168 to 1272. (ii) *Slit* (*21*): 1176 to 1309. (iii) Agrin consensus was derived from an alignment of the three repeats corresponding to residues (*15*): 1287 to 1418; 1555 to 1690; and 1788 to 1919. (iv) Laminin A consensus (*17*): 2126 to 2268; 2314 to 2450; 2495 to 2638; 2724 to 2854; and 2901 to 3033. (v) Perlecan consensus (*18*): 3009 to 3045; 3270 to 3401; and 3547 to 3679.
28. RNA blotting was done as described [M. S. Perin *et al.*, *EMBO J.* **7**, 2697 (1988)] with 10 μg of total RNA from the indicated rat tissues. Blots were hybridized with multiple uniformly ^{32}P-labeled probes specific for the different neurexins. To control for RNA loading, blots were rehybridized with a cyclophilin probe. Very long exposures (>14 days) of the blots revealed a low level of neurexin Iα expression in lung which was attributed to contaminating sympathetic ganglia.
29. We thank I. Leznicki, A. Hopkins, and E. Borowicz for technical assistance; M. S. Brown, J. L. Goldstein, and D. W. Russell for advice; and S. Artavanis-Tsakonas and J. M. Rothberg for alerting us to the unpublished sequence similarity of agrin with laminin A G-domain repeats. This study was supported by the Perot Family Foundation and a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft to M.G.

Chaotic Evolution of the Solar System

Gerald Jay Sussman and Jack Wisdom

The evolution of the entire planetary system has been numerically integrated for a time span of nearly 100 million years. This calculation confirms that the evolution of the solar system as a whole is chaotic, with a time scale of exponential divergence of about 4 million years. Additional numerical experiments indicate that the Jovian planet subsystem is chaotic, although some small variations in the model can yield quasiperiodic motion. The motion of Pluto is independently and robustly chaotic.

Advances in computer technology have made it possible to begin to directly address the age-old question of the nature of the long-term evolution of the solar system, with startling results. Sussman and Wisdom (*1*) presented numerical evidence that the motion of Pluto is chaotic, with a time scale for exponential divergence of nearby trajectories of only about 20 million years. Subsequently, Laskar (*2*) found numerical evidence of the chaotic evolution of the solar system excluding Pluto, with a time scale for exponential divergence of only about 5 million years. Laskar's calculation was feasible because he analytically averaged the equations of motion to remove the rapid variations with time scales of the order of the orbital period. The averaged equations are perturbative and necessarily truncated after a particular order in eccentricity, inclination, and mass ratio. An integration of the whole solar system without these approximations was required.

Direct integrations of the whole planetary system are computationally expensive. Notable long-term integrations of the outer solar system include: the classic 1-million-year integration of Cohen, Hubbard, and Oesterwinter (*3*), the 5-million-year integration of Kinoshita and Nakai (*4*), the 210-million-year integration performed on the Digital Orrery (*5*), the 100-million-year integration of the LONGSTOP project (*6*), and the 845-million-year Digital Orrery integration of Sussman and Wisdom (*1*). Studies of the long-term evolution of the whole solar system have been more limited because the computational resources required are significantly larger, by about two orders of magnitude. Integrations of the whole solar system include: the 3-million-year Digital Orrery integration (which excluded Mercury) (*5*), the 2-million-year integration of Richardson and Walker (*7*), and the recent ±3-million-year integration of Quinn, Tremaine, and Duncan (*8*, *9*) (hereafter QTD).

We have developed computational techniques and computer hardware to make possible a direct integration of the whole solar system spanning a significantly longer interval than previously achieved. Our direct integration of the equations of motion spans 36,000,000,000 days, or about 98.6 million years. Our earlier result concerning the chaotic motion of Pluto, as well as the result of Laskar that the solar system is chaotic, are both confirmed. In order to localize the sources of the chaotic behavior we have carried out numerous additional long-term integrations. We have found that the evolution of the Jovian planets is independently chaotic, as is the motion of Pluto.

Method of integration. We use the symplectic n-body mapping method of Wisdom and Holman (*10*) to integrate the planetary

G. J. Sussman is in the Department of Electrical Engineering and Computer Science, and J. Wisdom is in the Department of Earth Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.