UNIVERSITY OF CALIFORNIA, SAN DIEGO

A Structural Analysis of the Synaptic Adhesion Properties of alpha- and beta-Neurexins

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Biomedical Sciences

by

Meghan T. Miller

Committee in charge:

Professor Palmer Taylor, Chair
Professor Mark Ellisman
Professor Tracy Handle
Professor Vivian Hook
Professor Teru Nakagawa

2011

This Dissertation Meghan T. Miller is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2011

DEDICATION

Dedicated to

Mom, Dad and Tally

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AChE…… Acetylcholinesterase

AMPA……... α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate

ASD…….Autism spectrum disorders

AUC…….Analytical ultracentrifugation

Caspr/CNTNAP……..……Contactin associated protein

CASK……...Calcium/calmodulin-dependent serine protein kinase

CNS……..……Central nervous system

CNV…….Copy number variation

EGF……..……Epidermal growth factor

EM…….……Electron microscopy

ER……..……..……..Endoplasmic reticulum

FBS……..……Fetal bovine serum

FPLC……..……Fast performance liquid chromatography

GABA……..Gamma-aminobutyric acid

GlcNAc…….……...N-acetylglucosamine

DMEM……..……Dulbecco's modified Eagle's medium

HBS……..……HEPES saline buffer

HEK…..…….…….Human embryonic kidney

iGluRs……..…….……..Ionotropic glutamate receptors

kD..……..……..Kilodalton

LRRTM……..Leucine-rich repeat transmembrane protein

LNS…….……..……Laminin neurexin sex hormone-binding globulin

Man……...Mannose

mGluRs……..……..Metabotropic glutamate receptors

MINT…… Munc 18 interacting protein

NaCl……..Sodium chloride

NMDA……..…. *N*-methyl-D-aspartic acid

NL or NLGN……Neuroligin

NX or Nrx or NRXN……Neurexin

PDB……Protein Data Bank

PSD-95……...…….Postsynaptic density protein-95

RMSD……..…….Root mean square deviation

SAM……..……Synaptic adhesion molecule

SAXS……..……...Small angle X-ray scattering

SHBG……..…… Sex hormone binding globulin

SNP……..Single-nucleotide polymorphism

SPR……..……..……..……Surface plasmon resonance

SS……Splice site

SSM……..……Secondary structure matching

AMINO ACID RESIDUES

A or Ala...……Alanine

C or Cys……...……..Cysteine

D or Asp……Asparagine

E or Glu……...……Glutamate

F or Phe……Phenylalanine

G or Gly..……Glycine

H or His…...……...Histidine

I or Ile……..……..Isoleucine

K or Lys……..Lysine

L or Leu……Leucine

M or Met……Methionine

N or Asn…..……Asparagine

P or Pro……...……..Proline

Q or Gln.. ……...……..……..Glutamine

R or Arg...……Arginine

S or Ser……..Serine

T or Thr……Threonine

V or Val……..Valine

W or Trp……..……..Tryptophan

Y or Tyr……..……...…….……Tyrosine

ACKNOWLEDGEMENTS

This work would not have been possible without the guidance and support of many friends, family and colleagues. I want to especially thank Dr. Palmer Taylor for providing me with a supportive, stimulating environment for the duration of my graduate career. You gave me the freedom to explore many areas of interest and always supported me in my endeavors. You inspired me to take on challenging projects and guided me through the, sometimes difficult, but always valuable experiences. I would also like to thank Dr. Davide Comoletti for being a mentor to me and supporting me throughout the years. You spent many hours helping me in the lab and getting me started with my project. It has been great working along side you for the past five years, and I am sure that you are headed for great things. I am also thankful to Dr. Teru Nakagawa for the many hours he dedicated to the EM project. You patiently worked with me and mentored me as I navigated through this experience.

I also want to thank Dr. Mauro Mileni for his support, guidance and hard work on the crystallography project. You offered your assistance when things weren't working, first through discussions, then by using some of your own beam time to screen crystals for me, and finally by setting up a collaboration between our labs. This facilitated access to equipment and expertise that were undoubtedly essential to the success of this project. Towards the end you put in many hours to help me through the final stages of this work, which you did not only as my colleague, but also as my friend.

I am especially grateful to Jennifer Wilson and Dr. Antonella De Jaco for their endless support and compassion. You have been amazing friends to me, helping me through many struggles. Antonella, I admire your strength and perseverance. Jennifer, you are strong and caring. I will be very sad when we no longer work together, but I know that we will always be close friends.

I want to thank Artie Suckow, Akos Nemecz and John Yamauchi for their support, great conversations about science and life, and good times in the lab and at science meetings. It has been wonderful to work along side such accomplished and intelligent graduate students. I know you will all be great in whatever path you end up taking.

I am very grateful to all of the past and present members of the Taylor lab. Many thanks goes to Ryan Hibbs, Scott Hansen, Shelly Camp, Todd Talley, Helen Newlin, Kate Rodgers, Giulia Fallivalli, Anne Valle, Limin Zhang, Zoran Radic, Wenru Yu, and Joannie Ho for all their support and advice throughout the years.

I would also like to thank my past advisor Dr. Lawrence Brody at the National Human Genome Research Institute, *Bethesda, MD*, for providing me with a stimulating scientific environment and facilitating my transition to bench science. Your enthusiasm and never-ending support inspired me to take the leap towards getting my Ph.D. This would not have been possible without you.

Finally, I would like to sincerely thank my mom, dad and sister, whose support throughout the years has been my backbone. Thank you for teaching me to preserver and believing in me.

Chapter 2 is a reprint of the material as it appears in Fabrichny, I.P., Leone, P., Sulzenbacher, G., Comoletti, D. **Miller, M.T.,** Taylor, P., Bourne, Y., Marchot, P. (2007) *Neuron* **56**, 979-991. This dissertation author contributed to the conceptual design of the study and worked on the protein purification and crystallization experiments in Palmer Taylor's lab.

Chapter 3 is a reprint of the material as it appears in Comoletti, D.*, **Miller, M.T.*,** Jeffries, C.M., Willson, J., Demeler, B., Taylor P., Trewhella J., Nakagawa T. (2010) *Structure* **18,** 1044-1053. * These authors contributed equally to this work. The dissertation author contributed to the overall concept of the work, was the leading investigator for the Electron Microscopy experiments and author of this section of the paper, designed one of the protein constructs, and did much of the protein expression and purification.

A manuscript including **Chapter 4** is authored by Meghan T. Miller, Mauro Mileni, Davide Comoletti, Raymond C. Stevens, Michal Harel, and Palmer Taylor and has been submitted for publication. The dissertation author was the primary investigator in the development and execution of this study, and the principal author of the paper. This work encompassed the construction of expression vectors for various truncations of the α-NRXN extracellular region, crystallization of the protein, structure determination of the α-NRXN_2-6 construct and analysis of NLGN-NRXN association by SPR.

CURRICULUM VITAE

Education

Doctor of Philosophy, Biomedical Sciences

University of California, San Diego *San Diego, CA* 2004 – 2011

Palmer Taylor Laboratory

Dissertation Title: A Structural Analysis of the Synaptic Adhesion Properties of α- and β-Neurexins

Bachelors of Art in Mathematics, *Magna Cum Laude*

Colgate University *Hamilton, NY* 1998 – 2002

Professional Experience

National Human Genome Research Institute, Genome Technology Branch, Laboratory of Lawrence C. Brody June 2003- July 2004
Pre-IRTA fellowship program

- SNP/VNTR Genotyping (MALDI_TOF mass spectrometry, capillary electrophoresis (ABI 3100), Agarose Gel Electrophoresis)
- Analysis of polymorphisms for risk associations with neural tube defects
- Cell culture & cell proliferation assay using tritiated thymidine
- Assay optimization

National Institutes of Health, Diagnostic Radiology Department, Laboratory of Ronald M. Summers June 2002- June 2003
Pre-IRTA fellowship program

- Development of Computer-Aided CT Colonography
- Pattern Recognition using statistical machine learning algorithms
- Optimizing classification and feature selection techniques
- Data analysis

Teaching Assistant, Colgate University Mathematics Department, September 2001-May 2002.

- Graded homework sets for upper level math courses

Laboratory for Satellite Altimetry, National Oceanographic Data Center, National Oceanic and Atmospheric Administration, Silver Spring, MD. June-August 1998.

- Co-developed web page, "An Interactive Global Map of Sea Floor Topography Based on Satellite Altimetry &Ship Depth Soundings," (http://ibis.grdl.noaa.gov/cgi-bin/bathy/bathD.pl)
- Helped debug Generic Mapping Tools – GMT version 3.1 (http://www.soest.hawaii.edu/gmt)

FELLOWSHIPS AND AWARDS:

2009	Graduate Student Best Abstract Award, *second place*. Experimental Biology 2009 *New Orleans*
2009	ASBMB Travel Award to Experimental Biology 2009 *New Orleans*
2008	Med-into-Grad Training Fellowship, *HHMI*
2005	Chancellor's Fellowship Award, *UCSD*
2002-2004	NIH pre-IRTA fellowship
2002	Magna cum laude
1999	Phi Beta Kappa

SCIENTIFIC PUBLICATIONS:

- **Miller MT,** Mileni M, Comoletti D., Stevens RC, Harel M, Taylor P, *The crystal structure of the α-NRXN-1 extracellular region reveals a hinge point for mediating synaptic adhesion and function.* Submitted for publication.

- De Jaco A, Lin MZ, Dubi N, Comoletti D, **Miller MT**, Camp S, Ellisman M, Butko MT, Tsien RY, Taylor P. *Neuroligin trafficking deficiencies arising from mutations in the α/β-hydrolase fold protein family*. **(2010)** *J. Biol. Chem.*; 285(37), 28674-82.

- Comoletti D*, **Miller MT***, Jeffries CM, Wilson J, Demeler B, Taylor P, Trewhella J, Nakagawa T. *The macromolecular architecture of the extracellular domain of αNRXN1: domain organization, flexibility, and insights into trans-synaptic disposition***. (2010)** *Structure*; 18, 1044-1053. *These authors contributed equally to the work.

- Fabrichny IP, Leone P, Sulzenbacher G, Comoletti D, **Miller MT**, Taylor P, Bourne Y, and Marchot P. *Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion.* **(2007)** *Neuron*; 56, 979-991.

- Summers RM, Franaszek M, **Miller MT**, Pickhardt PJ, Choi JR, Schindler WR. *Computer-aided detection of polyps on oral contrast-enhanced CT colonography.* **(2005)** *Am J Roentgenol*.; 184(1): 105-8.

- **Miller MT**, Jerebko AK, Malley JD, Summers RM. *Feature Selection for Computer-Aided Polyp Detection Using Genetic Algorithms.* Medical Imaging **2003**: Physiology and Function: Methods, Systems, and Applications Proc. SPIE Vol. 5031, 102-110,

SCIENTIFIC PRESENTATIONS:

- **Miller MT**, Comoletti D, Jeffries CM, Wilson J, Trewhella J, Nakagawa T, and Taylor P. **Structural properties of the extracellular domain of the synaptic adhesion protein α-neurexin.** Experimental Biology, Anaheim, CA 2010. *Received second place in ASPET best abstract in neuropharmacology

- **Miller MT**, Comoletti D, Wilson J, Nakagawa T, and Taylor P. **Structural characterization of the extracellular domain of alpha-neurexin: insights into synapse assembly.** Experimental Biology, New Orleans, LA 2009

- **Miller MT**, Comoletti D, Wilson J, Demeler B, Trewhella J, Nakagawa T, and Taylor P. **Structural features of the extracellular domain of alpha-neurexin by single particle EM and small angle x-ray scattering.** Gordon Research Conference: Neurobiology of Brain Disorders, Magdalen College, Oxford, UK 2008.

- **Miller MT**, Comoletti D, Wilson J, Nakagawa T, and Taylor P. **Insights into the structure and function of the extracellular domain of alpha-neurexin by single particle EM and surface plasmon resonance.** Experimental Biology, SanDiego, CA 2008.

- **Miller MT**, De Jaco A, Comoletti D, Ellisman M, and Taylor P. **Distinct Neuroligin Mutants Implicated in Autistic Spectral Disorders Undergo Similar Defective Endoplasmic Reticulum Trafficking.** Neuroscience, San Diego, CA 2007.

- **Miller MT, Hansen SB, McIntosh JM, Olivera BM, Taylor P. Structural insights into competitive and non-competitive nicotinic antagonists.** Experimental Biology. San Francisco, CA 2006.

- **Miller MT, Jerebko AK, Malley JD, Summers RM. Feature Selection for Computer-Aided Polyp Detection Using Genetic Algorithms. SPIE Medical Imaging 2003:** International Conference on Physiology and Function: Methods, Systems, and Applications.

- **Miller MT**, Jerebko AK, Malley JD, Summers RM. **Comparison of Two Feature Subset Selection Techniques for Improving Classification Accuracy**. NIH Research Festival, 2002.

ABSTRACT OF THE DISSERTATION

A Structural Analysis of the Synaptic Adhesion Properties of alpha- and beta-Neurexins

by

Meghan T. Miller

Doctor of Philosophy in Biomedical Science

University of California, San Diego, 2011

Professor Palmer Taylor, Chair

Synaptic function and integrity in the nervous system requires the expression of synaptic adhesion molecules (SAMs) linking pre- and post-synaptic sites. Studies indicate that SAMs participate in the formation, maturation, function and plasticity of synaptic connections, and thus are essential for trans-cellular signaling. Alterations in SAMs lead to susceptibility to neurological diseases including the autism spectrum disorders, schizophrenia, and addiction. Neurexins compose a family of highly polymorphic type I transmembrane proteins that are expressed on the pre-synaptic membrane at excitatory

glutamatergic and inhibitory GABAergic synapses. β-neurexins have a single folding domain in the extracellular region, while α-neurexins have a larger extracellular region containing nine independently folding domains and multiple protein interaction sites. They both function as adhesion molecules through a trans-synaptic complex with post-synaptic neuroligins.

Synaptogenesis and synapse function requires the precise assembly of pre- and post-synaptic protein complexes. The work described herein uses structural and biophysical techniques to discern the molecular properties of the neurexin and neuroligin proteins that mediate their complex formation. The first aim included solving the X-ray crystal structure of the β-neurexin:neuroligin complex, which showed a stable neuroligin dimer and two monomeric β-neurexin molecules bound on either side of the dimer. This structure revealed the molecular adhesion properties of the complex, including a Ca^{2+}-coordination site.

α- and β-neurexins have the same binding domain for neuroligins, yet they are likely to act as functionally distinct molecules in the synapse. To consider their distinctive adhesion properties, as well as how the multi-domain α-neurexins assemble in the limited space of a synapse, the second and third aims were directed at solving the 3D structure of α-neurexin. First, small angle X-ray scattering and single-particle negative-stain electron microscopy provided information on the overall domain organization and flexibility of the protein. This work led to the high-resolution crystal structure of a major portion of the α-neurexin extracellular region. The crystal structure reveals molecular details that

suggest a multi-functional mechanism of the α-neurexin extracellular region in the synapse.

Overall, this work contributes to the understanding of how synapses assemble through a complex network of protein-protein interactions and provides structural templates for the development of molecular tools to study function and potentially therapeutic applications.

CHAPTER 1:

Introduction

In the central nervous system, information is exchanged between neurons at sites of contact known as synapses. Synapse formation requires the assembly of highly ordered multi-protein complexes that link the pre- and post-synapse. In turn, the complexes also form a unique collection of cytoplasmic proteins at the synaptic membrane. The distinct assembly of receptors, signaling molecules and scaffolding proteins determines synaptic specification. A growing body of evidence indicates that adhesion molecules not only function as a physical link between cells, but also play an integral role in synaptogenesis and synapse maintenance through the bi-directional recruitment and stabilization of defining molecules [1-3]. Research suggests that the diversity of adhesion molecules may be the key to achieving synaptic specification, and at the same time maintain some redundancy in order to impart synaptic resilience [4, 5]. The work presented here is based on the previously identified heterophilic cell adhesion molecules, neurexins and neuroligins, which form trans-synaptic complexes at excitatory and inhibitory synapses. Using a structural approach, I aim to gain insight into the distinguishing characteristics that may contribute to synapse specification and function. The introduction that follows offers a glimpse into the complex assembly of the synapse and serves to orient the reader to the role of neurexins and neuroligins in this assembly.

1.1 The Architecture of Excitatory and Inhibitory Synapses

The balance between excitation and inhibition is essential to brain function. Excitatory and Inhibitory synapses are characterized by the expression of distinct pre-

and post-synaptic proteins. On the pre-synaptic side, synaptic vesicles containing predominantly glutamate or γ-aminobutyric acid, the primary endogenous excitatory or inhibitory neurotransmitter, respectively, are assembled. Upon membrane depolarization, voltage-dependent Ca^{2+} channels on the pre-synaptic membrane trigger vesicle docking and exocytosis of neurotransmitter into the synaptic cleft.

On the post-synaptic side the complementary expression of excitatory glutamatergic or inhibitory GABAergic receptors completes the chemical circuit by functioning as receivers for the exocytosed neurotransmitter. The excitatory glutamate receptors are divided into the ionotropic glutamate receptors (iGluRs), including the NMDA (*N*-Methyl-D-aspartic acid), AMPA (α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate) and kainate receptors, and metabotropic glutamate receptors (mGluRs). The NMDA and AMPA receptors, named after their high- affinity agonists, are ligand-gated non-selective cation channels and are the primary glutamate receptors responsible for fast excitatory neurotransmission. They are clustered through their interaction with the post-synaptic scaffolding protein PSD-95 and appear to be important modulators of synaptic plasticity [6] The recruitment of these receptors to the cell membrane appears to be linked to the expression of neurexin-neuroligin adhesion complexes at the membrane [7-9].

Inhibitory GABA receptors are also classified into the ionotropic ($GABA_A$) and metabotropic ($GABA_B$) sub-types. $GABA_A$ receptors fall into the larger class of Cys-loop ligand-gated ion channels, which includes nicotinic acetylcholine receptors, $GABA_A$ receptors, glycine and $5\text{-}HT_3$ receptors. They selectively control the flow of Cl^- ions into the cell producing a rapid inhibitory response. They are clustered through their interaction with the scaffolding protein gephyrin, enriched at inhibitory post-synaptic

membranes. Neurexins and neuroligins also play a crucial role in organizing $GABA_A$ receptors at the synapse [10-12]. Recent evidence also indicates that the pre-synaptic neurexins physically interact with $GABA_A$ receptors [10-12].

The synapse is not a static structure, but rather a dynamic interface that changes in morphology and motility based its molecular constituents. The flexibility of synaptic contacts may provide the basis for synaptic plasticity and function. Efficient neurotransmission relies on the temporal expression and spatial alignment of the pre- and post-synaptic elements in the active zone, which is mediated by a network of proteins that is ultimately linked through the synaptic cleft by synaptic adhesion molecules (SAMs).

1.2 Synaptic Adhesion Molecules: Neurexins and Neuroligins

There are several distinct classes of SAMs, which function through either homo- or heterophilic interactions and seem to control discrete aspects of synaptic development and function. These predominantly included the neurexins (NRXNs) and neuroligins (NLGNs), integrins, Ephrin receptors, immunoglobulin (Ig)-containing cell adhesion molecules and cadherins [1].

NRXNs and NLGNs are type I transmembrane protein that form a calcium-dependent heterophilic complex through their extracellular domains. Targeting of different gene transcripts and splice variants of NRXNs and NLGNs shows distinction between excitatory and inhibitory synapse localization [11, 13-17]. Studies *in vitro* suggest that NRXNS and NLGNs are capable of triggering synapse initialization, and moreover are able to direct excitatory or inhibitory synaptogenesis in an isoform-dependent manner [2, 11, 15, 18, 19]. However, studies *in vivo* lean towards a more prominent role for NRXN/NLGN interactions later in synapse development [20, 21]. Both

NRXNs and NLGNs contain an intracellular PDZ binding domain that mediates interactions with synaptic scaffolding proteins, and thereby mediates the bi-directional assembly of pre- and post-synaptic elements [22-24]. On the pre-synaptic side NRXNs bind to the scaffolding proteins CASK (calcium/calmodulin-dependent serine protein kinase) and MINT (Munc 18 interacting protein; lin-10/X11) through their intracellular PDZ-binding domain, which couples NRXN signaling to synaptic vesicle exocystosis. On the post-synaptic side, NLGNs interact through their intracellular PDZ-binding domain with PSD-95 (postsynaptic density protein-95) or gephyrin, which couples NLGNs to the post-synaptic neurotransmitter receptors.

The NRXN and NLGN families are extensive, consisting of multiple genes and spliced variants. A detailed description of the structural characteristics of the NRXN and NLGN proteins follows.

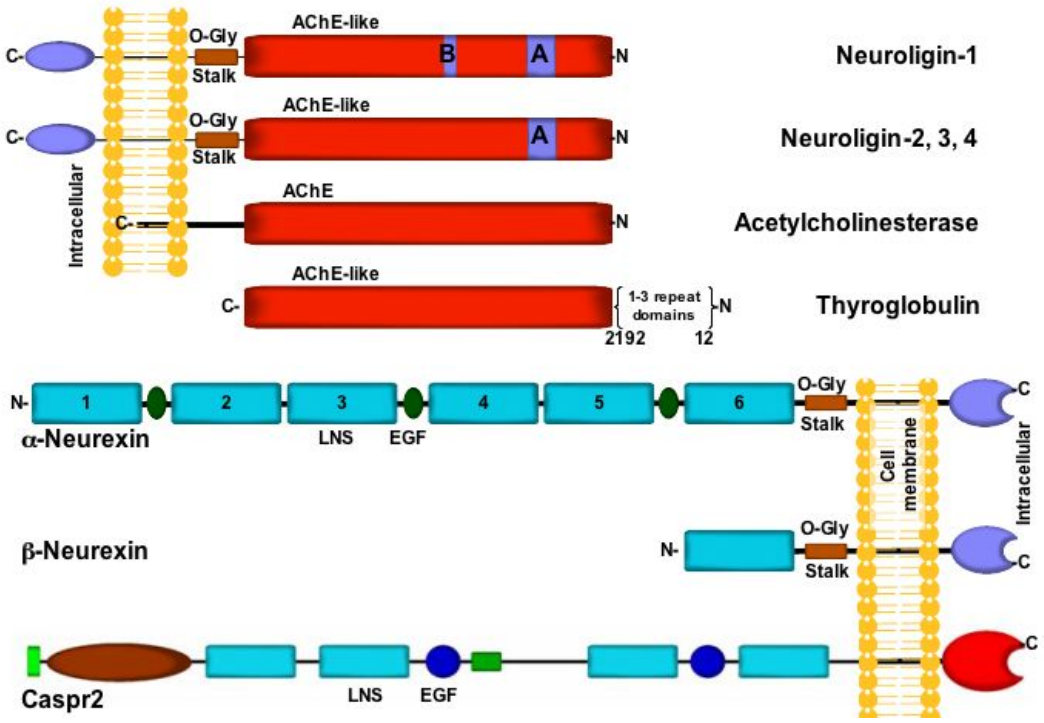

Figure 1.1: Domain organization of neuroligins and neurexins shown with homologous proteins. The neuroligin family consists of 5 genes, NLGN-1, -2, -3, -4X and -4Y. The mature proteins all have a large globular extracellular domain that is a member of the α/β-hydrolase fold family, which includes acetylcholinesterase and thyroglobulin, among others. There are three neurexin genes in mammals, each with an α- and β- promoter, encoding the α- and β-NRXN proteins. Their extracellular domains are made up of individually folding sub-domains, which include LNS and EGF-like domains. The NRXNs are structurally homologous to the Caspr/CNTNAP family of proteins.

1.3 The Neurexin Family

In vertebrates there are three highly conserved *NRXN* genes (*NRXN 1-3*), each encoding either the longer α-NRXN or the shorter β-NRXN via alternative promoter regions. For each gene, the α- and β-NRXNs yield class I transmembrane proteins that share the same C-terminal sequence and differ only in the extent of their N-terminal extracellular region. After post-translational cleavage of an α-specific signal peptide, the extracellular region of the mature α-NRXNs contain nine independently folded domains, including six laminin neurexin sex hormone-binding globulin (LNS) domains separated by three epidermal growth factor (EGF)-like domains, and a C-terminal *O*-glycosylated stalk domain (~100 amino acids) that forms an extended linker to the transmembrane domain. The much smaller extracellular region of β-NRXNs contains a β-specific signal peptide (cleaved in the mature protein) and a short β-specific N-terminal sequence followed by a single LNS domain, which corresponds to the LNS 6 domain of α-NRXN, and the same *O*-glycosylated stalk domain. NRXN mRNAs are subject to extensive alternative splicing. There are five alternative splice sites in total: SS#1-5. α-NRXNs have all five splice sites, whereas β-NRXNs have only SS#4 and #5 on the LNS domain and stalk domain, respectively. The diversity of NRXNs, which is achieved by different genes, alternative promoters and extensive splicing, results in the potential for thousands of isoforms [25]. Alternative splicing is thought to function in the regulation of protein-protein interactions and is potentially governing the maintenance of synapse specificity [26].

Although the NRXNs are primarily thought to function as cell-adhesion molecules through their extracellular interaction with the post-synaptic NLGNs [27, 28], they recently have been identified to also associate with the post-synaptic leucine-rich repeat

transmembrane proteins (LRRTMs) [27-31], the α4β2 nicotinic acetylcholine receptor [32], and the $GABA_A$ receptor [12].

α-NRXN-1 was originally identified as a high-affinity receptor for the spider neurotoxin, α-latrotoxin, which binds to pre-synaptic receptors and causes massive neurotransmitter release [33, 34]. Studies in mice identified the extracellular region of α-NRXN-1 as an essential component for the regulation of Ca^{2+}-dependent exocytosis in neurons [35, 36]. In addition to the post-synaptic partner proteins described above, early works showed that α-NRXNs interact with α-dystroglycan through the LNS 2 and LNS 6/β-NRXN domains [37] and to the soluble neurexophilin proteins through the LNS 2 domain [38]. The precise role of each of these interactions in governing synapse maintenance remains to be investigated.

LNS domains

Laminin Neurexin Sex Hormone-Binding Globulin (LNS) domains, also called Laminin-G like (LG) domains, are common in extracellular proteins. They bind to a diverse range of ligands and are involved in a host of biological functions including cell adhesion and migration, steroid transport, blood coagulation, and mammalian spermatogenesis [39]. Despite only a 20-25% sequence identity, these domains have a highly conserved tertiary structure, as revealed by X-ray crystallography [40-45]. They furthermore share structural homology with pentraxins and lectins, despite only 10-15% sequence identity [39].

LNS domains are ~200 amino acids and are characterized by a β sandwich formed by two seven-stranded anti-parallel β-sheets, with N- and C-termini on the same side and juxtaposed in space. The globular domain has a convex and concave side of

the sandwich. On the concave side, a loop, referred to as the β11-β12 loop, fills the depression to varying degrees in different proteins. At least in pentraxins and lectins this loop is important for ligand specificity [46, 47]. In other proteins containing LNS domains, including the NRXNs, laminins and sex hormone binding globulin (SHBG), the functional ligand-binding site is found at the rim of the β sheet sandwich where the β10-β11 loop runs across. Along this surface there is a conserved Ca^{2+}-binding site found in laminin, perlecan, agrin and neurexin domains, the steroid-binding site and proximal zinc-binding in SHBG, and the location of various alternative splice inserts that mediate protein-protein interactions [39]. Therefore, despite a highly conserved structural fold, the variation in topological and physiochemical properties at specific sites on the LNS domain are enough to confer very distinct biological functions.

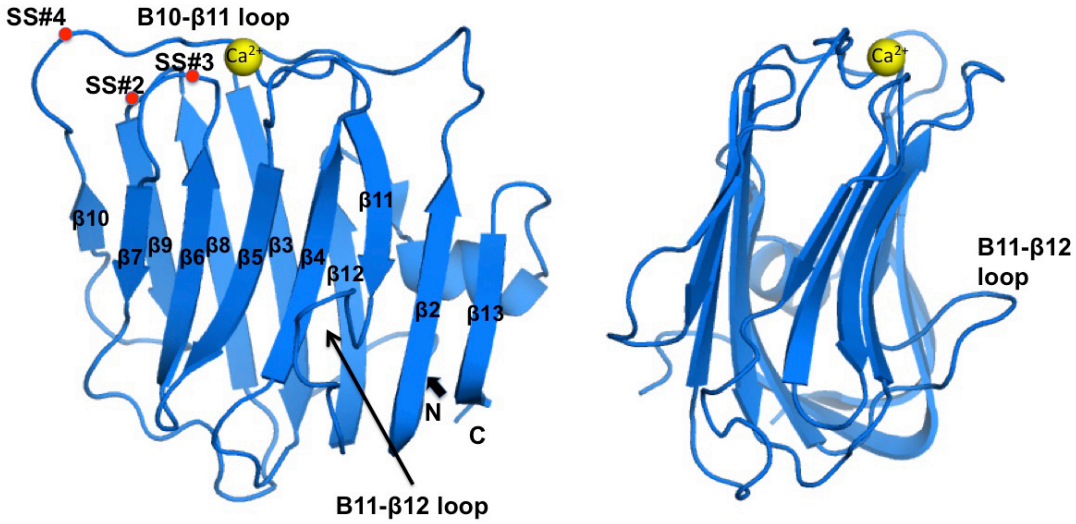

Figure 1.2: Conserved structural fold of an LNS domain depicted with the β-NRXN-1 structure (PDB: 3bod). The domain is comprised of a β-sheet sandwich with N- and C-termini on the same corner and juxtaposed in space. Regions of variability and functional distinction among homologous proteins include the β11-β12 loop, which is variable in size and important for ligand specificity in lectins and pentraxins, and the surface containing the β10-β11 loop, which contains the cation binding site in laminins, neurexins and SHBG, and sites of splice variation, including SS#2, #3 and #4 in neurexins. The site of steroid binding in SHBG is within the β-sheets just below the β10-β11 loop.

EGF-like domains

Epidermal growth factor (EGF)-like domains are structurally conserved domains found in a number of extracellular proteins in mammals and associated with a diverse set of functions, including blood coagulation, fibrinolysis, cell-cell adhesion, cell signaling in development and embryonic cell fate [48-50]. They are only ~30-40 amino acids in length and contain a conserved pattern of six cysteines, which form three disulfide bonds making a compact domain with three anti-parallel strands. A subset of EGF-like domains contain a conserved Ca^{2+}-binding motif at the N-terminus, which may be crucial for protein-protein interactions [48, 51]. In other cases, three types of unusual post-translational modifications have been found within conserved amino acid motifs in EGF-like domains [49, 52]. These include two *O*-linked glycosylation sites and a β-hydroxylation site at either an Asp or Asn residue. The biological role of one of the *O*-linked modifications, consisting of the attachment of an *O*-fucose moiety to a Ser or Thr residue, is best described in the Notch signaling pathway. Here, the elongation of the *O*-fucose to a poly-glycan by the glycotransferase, Fringe, regulates binding to extracellular ligands, which determine the proteolytic processing of Notch and subsequent intracellular signaling [49, 53].

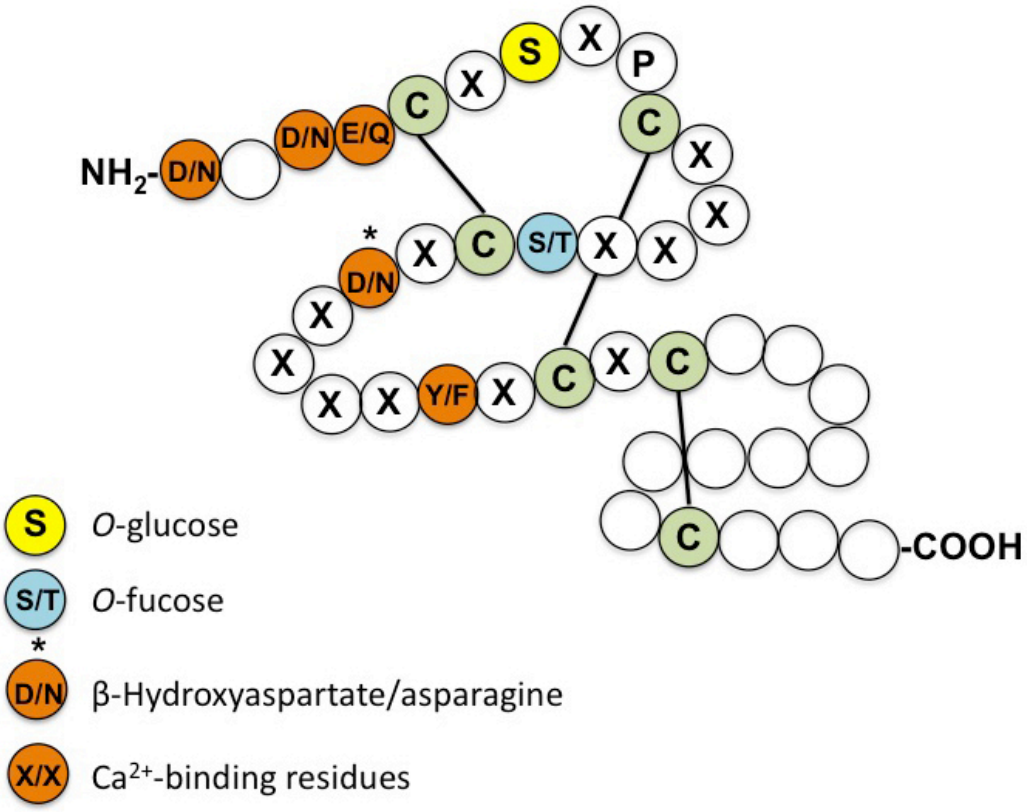

Figure 1.3: Schematic of an epidermal growth factor (EGF)-like domain. All EGF domains contain a conserved pattern of six Cys residues that make three disulfide bonds (shown as connecting lines). A subset of EGF-like domains bind Ca^{2+} in their N-terminal fold (the conserved residues are shown in orange). In some cases, EGF domains contain consensus sequence motifs that result in post-translational modifications. These include two *O*-glycosylation (yellow and blue) sites and a β-hydroxylation site (asterisk) of either an Asp or Asn residue. This image was adapted from Varki, A. *et al.* Essential of Glycobiology. 2nd Edition.

The Neuroligin Family

The NLGN molecules constitute another family of class I transmembrane cell-adhesion proteins. There are five NLGN genes (*NLGN1,2, 3, 4X and 4Y*), which are expressed post-synaptically at either glutamatergic (NLGN-1, -3, -4) or GABAergic (NLGN-2) synapses [16, 17]. The mature NLGN proteins are composed of an extracellular, N-linked glycosylated domain with strong sequence homology to acetylcholinesterase (AChE), a Ser-Thr-rich stalk domain that carries both N- and O-linked oligosaccharides [28, 54], a single transmembrane domain, and a small, intracellular C-terminal domain with a PDZ-binding motif.

NLGNs belong to the super-family of α/β-hydrolase fold proteins, but lack the characteristic enzymatic activity due to mutation of one or more residues in the catalytic triad. Instead, they have evolved to have adhesive properties through a negatively charged surface (unique to NLGNs) on the opposite face from the catalytic site found in AChE [55]. At the negatively charged surface, they form a Ca^{2+}-dependent heterophilic cell-adhesion complex with the NRXNs [56-59]. Similar to the other α/β-hydrolase fold proteins, NLGNs form homodimers through two parallel hydrophobic α-helices at the C-terminal region of each monomer. Therefore, the full complex consists of a NLGN dimer bound to two NRXNs on opposite sides of the long axis of the dimer [56-59].

NLGNs have high amino acid identity (~70%) within the family, as well as across species (>98% for homologous genes) [60]. The NRXN-binding interface on all NLGNs is highly conserved; however, non-conserved residues on the edge of the interaction surface appear to explain some of the variability between the different NLGN affinities for NRXNs [57]. The distinctive binding properties of the various NLGNs for β-NRXNs have, in part, been attributed to alternative splicing in both NLGNs and β-NRXNs [13, 60, 61].

Comparative binding studies on affinities between the NLGNs and α-NRXNs remains to be determined.

1.4 Neurexins and Neuroligins: Relations to Disease

An increasing number of genetic studies suggest that perturbations in the *NRXN and NLGN* genes lead to the susceptibility to several neurological diseases, including the autism spectrum disorder (ASD) [62-72], schizophrenia [73], addiction [74-78], and rarer forms of mental retardation [79, 80].

Autistic spectrum disorders (ASD) represent a neurobiologically diverse group of conditions with heterogeneous clinical representation. Characteristic signs of patients with ASD include (1) social impairment, (2) language or communication deficits, and (3) patterns of restricted and/or repetitive behavior. Beyond these core symptoms, autism can present itself with highly variable pathological symptoms from one individual to the next, representing a continual spectrum rather than discrete diagnostic categories.

There is a clear genetic predisposition to idiopathic autism [81, 82], beginning with a disproportion of cases in males compared to females (4:1), suggesting X-linked inheritance and/or genetic imprinting. Furthermore, studies of monozygotic and dizygotic twins show that autism is highly heritable with a monozygotic concordance rate between 70-90% and a dizygotic rate around 7%. A sibling recurrence risk of about 5-6% is also suggestive of genetic predisposition. The data are consistent with a multifactorial inheritance pattern, where multiple genetic loci, genetic heterogeneity, epitasis and gene-environment interactions contribute to the pathology of the diseases. In recent years, epidemiological studies indicate that the incidence of autism and associated disorders is approaching 1% in the United States and about 0.1% world-wide [81, 83].

The increase is thought to be due to a combination of the changes in diagnostic and reporting practices as well as environmental and epigenetic factors acting on a genetically predisposed population. There are a number of characterized syndromes whose pathophysiology lie within the autism spectrum, including Asperger, Fragile X, Angelman, Rett, Williams, Prader-Willi, Hyperlexia, tuberous sclerosis complex and Landau-Kleffner syndrome. Most of these diseases have genetic mutations that have been identified as the underlying cause. It is also important to note that about 40-60% of patients diagnosed with ASD show some form of mental retardation, and about 30% suffer from seizures.[84].

Genome studies of idiopathic ASD are revealing an increasing number of chromosomal abnormalities, ranging from CNVs to chromosomal rearrangements to SNPs. Although each individual mutation accounts for only ~1% of the cases in the study, all together it is thought that these rare mutations could represent up to 10-20% of all cases [82, 83]. Researchers are starting to identify a trend in the cellular mechanisms that are coming out of studies of genetic defects related to ASD. In particular, evidence is pointing to genes encoding proteins related to synapse development and integrity (see Table 1.1).

Table 1.1: Some of the genes that have been implicated in ASD or ASD-related disorders. These genes all play a role in synapse function, either directly or indirectly through activity-dependent regulation of protein levels. Where possible, reported rates of incidence in genetic abnormalities are indicated.

Gene	Gene Function	Rate in Autism	References
	Related to Synapse Function		
NLGN3 *NLGN4X*	*Synaptic adhesion molecules expressed at excitatory and inhibitory synapses*	<1%	**[68-72, 85, 86]**
NRXN1	*Synaptic adhesion molecule expressed at excitatory and inhibitory synapses. Ca^{2+}-dependent exocytosis*	<1%	**[62-67]**
CDH9/10/15/18	*Synaptic adhesion molecules*		**[87, 88]**
CNTN3 *CNTN4*	*Axon-associated cell adhesion molecule of the Ig superfamily*	<1%	**[63, 89, 90]**
CNTNAP2 (Casp2)	*Cell adhesion molecule in NRXN superfamily. Localized at the juxtaparanodes of myelinated axons and associated with potassium channels.*	<1%	**[91-93]**
PCDH9/10/19	*Synaptic adhesion molecules*		**[87, 90, 94]**
NRCAM	*Neuronal cell adhesion molecule*		**[95-97]**
RELN	*Neuronal migration. Secreted protein*		**[98, 99]**
SHANK3	*PSD scaffolding protein localized at the membrane, binds to NLGNs*	<1%	**[100-103]**
GRIN2A	*NMDA receptor subunit*		**[104]**
GRIK2	*Kainate receptor subunit*		**[105]**
GABAR	*GABA receptor subunit*		**[106]**
	Regulation of synaptic protein levels		
MECP2	*Transcriptional repressor. Mutated in Rett's syndrome*	2%	**[107-115]**
EN2	*Transcriptional regulator*		**[116-118]**
PTEN	*Inhibitor of PIsK/mTOR signaling. Loss of function leads to increased activity of mTORC1 [119]. Mutated in ASD with macrocephaly*	1%	**[120-126]**
FMR1	*Translational Repressor. Mutated in Fragile X syndrome. Estimated to interact with >400 distinct mRNAs [127].*	2-5%	**[128-130]**
TSC1/2	*Inhibitor of mTOR-raptor complex (mTORC1), important for translational regulation. Mutated in tuberous sclerosis complex*	1-4%	**[131]**
UBE3A	*E3 ubiquitin ligase. Regulates ubiquitin-dependent protein turnover. Mutated in Angelman's syndrome*	1%	**[63, 132, 133]**

NLGN mutations were one of the first reported mutations in genetic studies of ASD patients [69, 70]. The initial reports pointed to mutations in the *NLGN3* and *NLGN4* genes, of which one, an Arg to Cys mutation in the *NLGN3* gene, has prompted multiple follow-up studies to characterize its phenotype. Initial studies *in vitro*, showed that the single mutation was sufficient to trap the majority of the partially processed protein in the endoplasmic reticulum (ER), where it is then sent to the proteosome for degradation [134-136]. Further studies *in vitro* validated this observation in neurons and also showed an upregulation of ER resident chaperone proteins responsible for the protein retention [137, 138] (see Figure 1.4). A subsequent study using a mouse knockin model showed that the Arg to Cys substitution in NLGN 3 resulted in an increase in inhibitory synaptic transmissions [139]. Together, these studies revealed a molecular mechanism that suppresses the expression of the NLGN protein on the cell surface, where it is needed in order to interact with its partner NRXN proteins, and acts as a gain of function mutation *in vivo* by disrupting the excitatory/inhibitory ratio through an observed increase in inhibitory signaling. Future studies on other rare mutations in the NLGN, NRXN and other autism candidate genes may reveal similar trafficking defects that yield the protein ineffective.

Along with *NLGN3 and 4, t*he *NRXN1* gene has been strongly implicated in ASD. Identified genetic abnormalities include copy number variations [63, 65, 66, 73], chromosomal alterations [64, 67, 80] and a few rare sequence mutations [62, 64, 67, 79]. To date, no studies have been done on the phenotype of specific *NRXN1* mutations.

1.5 Objectives of the Dissertation

In the brain, there is a delicate balance between excitation and inhibition of electrical currents leading to cell signaling events. Synapses are the major gateway that determine cell-signaling fate, and the spatial and temporal organization of proteins at the synapse is a crucial element of its functionality. Revealing the molecular organization and interaction characteristics of proteins involved in synapse assembly provides insight into the ability of different proteins to specify synaptic function. The major objectives of my dissertation work were to (1) determine the interaction between NLGN and β-NRXN at high resolution, (2) characterize the macromolecular assembly of the multi-domain α-NRXN protein, and (3) solve the high-resolution structure of the α-NRXN extracellular region. Objective 1 was accomplished through collaboration with Drs. Yves Bourne and Pascale Marchot, in which we solved the structure of the complex of NLGN 4 with β-NRXN (Chapter 2). I then successfully characterized the macromolecular structure of the extracellular region of α-NRXN by single particle electron microscopy, which complemented work done by Dr. Comoletti using small angle X-ray scattering (Chapter 3). Finally, by the characterization of α-NRXN using the low-resolution structural techniques, I designed an optimal construct that facilitated the crystallization and medium/high-resolution (3.0 Å) structure determination of a major portion of the α-NRXN extracellular region (Chapter 4). Together, these studies provide novel structural characterizations of the α- and β-NRXNs, and offer insights into the defining roles of each protein in the synapse. However, as is the case with most structural studies, the work described here opens the door to a new room of questions and future studies that will inevitably lead to a better understanding of synapse biology and the role of neurexins and neuroligins.

1.7 REFERENCES:

1. Dalva, M.B., A.C. McClelland, and M.S. Kayser, *Cell adhesion molecules: signalling functions at the synapse.* Nat Rev Neurosci, 2007. **8**(3): p. 206-20.

2. Washbourne, P., et al., *Cell adhesion molecules in synapse formation.* J Neurosci, 2004. **24**(42): p. 9244-9.

3. Dean, C., et al., *Neurexin mediates the assembly of presynaptic terminals.* Nat Neurosci, 2003. **6**(7): p. 708-16.

4. Craig, A.M., E.R. Graf, and M.W. Linhoff, *How to build a central synapse: clues from cell culture.* Trends Neurosci, 2006. **29**(1): p. 8-20.

5. Piechotta, K., I. Dudanova, and M. Missler, *The resilient synapse: insights from genetic interference of synaptic cell adhesion molecules.* Cell Tissue Res, 2006. **326**(2): p. 617-42.

6. Dityatev, A., M. Schachner, and P. Sonderegger, *The dual role of the extracellular matrix in synaptic plasticity and homeostasis.* Nat Rev Neurosci, 2010. **11**(11): p. 735-46.

7. Thyagarajan, A. and A.Y. Ting, *Imaging activity-dependent regulation of neurexin-neuroligin interactions using trans-synaptic enzymatic biotinylation.* Cell, 2010. **143**(3): p. 456-69.

8. Barrow, S.L., et al., *Neuroligin1: a cell adhesion molecule that recruits PSD-95 and NMDA receptors by distinct mechanisms during synaptogenesis.* Neural Dev, 2009. **4**: p. 17.

9. Heine, M., et al., *Activity-independent and subunit-specific recruitment of functional AMPA receptors at neurexin/neuroligin contacts.* Proc Natl Acad Sci U S A, 2008. **105**(52): p. 20947-52.

10. Huang, Z.J. and P. Scheiffele, *GABA and neuroligin signaling: linking synaptic activity and adhesion in inhibitory synapse development.* Curr Opin Neurobiol, 2008. **18**(1): p. 77-83.

11. Kang, Y., et al., *Induction of GABAergic postsynaptic differentiation by alpha-neurexins.* J Biol Chem, 2008. **283**(4): p. 2323-34.

12. Zhang, C., et al., *Neurexins physically and functionally interact with GABA(A) receptors.* Neuron, 2010. **66**(3): p. 403-16.

13. Chih, B., L. Gollan, and P. Scheiffele, *Alternative splicing controls selective trans-synaptic interactions of the neuroligin-neurexin complex.* Neuron, 2006. **51**(2): p. 171-8.

14. Graf, E.R., et al., *Structure function and splice site analysis of the synaptogenic activity of the neurexin-1 beta LNS domain.* J Neurosci, 2006. **26**(16): p. 4256-65.

15. Graf, E.R., et al., *Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins.* Cell, 2004. **119**(7): p. 1013-26.

16. Song, J.Y., et al., *Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses.* Proc Natl Acad Sci U S A, 1999. **96**(3): p. 1100-5.

17. Varoqueaux, F., S. Jamain, and N. Brose, *Neuroligin 2 is exclusively localized to inhibitory synapses.* Eur J Cell Biol, 2004. **83**(9): p. 449-56.

18. Chih, B., H. Engelman, and P. Scheiffele, *Control of excitatory and inhibitory synapse formation by neuroligins.* Science, 2005. **307**(5713): p. 1324-8.

19. Scheiffele, P., et al., *Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons.* Cell, 2000. **101**(6): p. 657-69.

20. Dudanova, I., et al., *Deletion of alpha-neurexins does not cause a major impairment of axonal pathfinding or synapse formation.* J Comp Neurol, 2007. **502**(2): p. 261-74.

21. Varoqueaux, F., et al., *Neuroligins determine synapse maturation and function.* Neuron, 2006. **51**(6): p. 741-54.

22. Irie, M., et al., *Binding of neuroligins to PSD-95.* Science, 1997. **277**(5331): p. 1511-5.

23. Hata, Y., S. Butz, and T.C. Sudhof, *CASK: a novel dlg/PSD95 homolog with an N-terminal calmodulin-dependent protein kinase domain identified by interaction with neurexins.* J Neurosci, 1996. **16**(8): p. 2488-94.

24. Biederer, T. and T.C. Sudhof, *Mints as adaptors. Direct binding to neurexins and recruitment of munc18.* J Biol Chem, 2000. **275**(51): p. 39803-6.

25. Tabuchi, K. and T.C. Sudhof, *Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing.* Genomics, 2002. **79**(6): p. 849-59.

26. Ullrich, B., Y.A. Ushkaryov, and T.C. Sudhof, *Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons.* Neuron, 1995. **14**(3): p. 497-507.

27. Ichtchenko, K., et al., *Neuroligin 1: a splice site-specific ligand for beta-neurexins.* Cell, 1995. **81**(3): p. 435-43.

28. Ichtchenko, K., T. Nguyen, and T.C. Sudhof, *Structures, alternative splicing, and neurexin binding of multiple neuroligins.* J Biol Chem, 1996. **271**(5): p. 2676-82.

29. de Wit, J., et al., *LRRTM2 interacts with Neurexin1 and regulates excitatory synapse formation.* Neuron, 2009. **64**(6): p. 799-806.

30. Ko, J., et al., *LRRTM2 functions as a neurexin ligand in promoting excitatory synapse formation.* Neuron, 2009. **64**(6): p. 791-8.

31. Geschwind, D.H. and P. Levitt, *Autism spectrum disorders: developmental disconnection syndromes.* Curr Opin Neurobiol, 2007. **17**(1): p. 103-11.

32. Cheng, S.B., et al., *Presynaptic targeting of alpha4beta 2 nicotinic acetylcholine receptors is regulated by neurexin-1beta.* J Biol Chem, 2009. **284**(35): p. 23251-9.

33. Geppert, M., et al., *Neurexin I alpha is a major alpha-latrotoxin receptor that cooperates in alpha-latrotoxin action.* J Biol Chem, 1998. **273**(3): p. 1705-10.

34. Davletov, B.A., et al., *High affinity binding of alpha-latrotoxin to recombinant neurexin I alpha.* J Biol Chem, 1995. **270**(41): p. 23903-5.

35. Missler, M., et al., *Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis.* Nature, 2003. **423**(6943): p. 939-48.

36. Zhang, W., et al., *Extracellular domains of alpha-neurexins participate in regulating synaptic transmission by selectively affecting N- and P/Q-type Ca2+ channels.* J Neurosci, 2005. **25**(17): p. 4330-42.

37. Sugita, S., et al., *A stoichiometric complex of neurexins and dystroglycan in brain.* J Cell Biol, 2001. **154**(2): p. 435-45.

38. Missler, M., R.E. Hammer, and T.C. Sudhof, *Neurexophilin binding to alpha-neurexins. A single LNS domain functions as an independently folding ligand-binding unit.* J Biol Chem, 1998. **273**(52): p. 34716-23.

39. Rudenko, G., E. Hohenester, and Y.A. Muller, *LG/LNS domains: multiple functions -- one business end?* Trends Biochem Sci, 2001. **26**(6): p. 363-8.

40. Grishkovskaya, I., et al., *Crystal structure of human sex hormone-binding globulin: steroid transport by a laminin G-like domain.* EMBO J, 2000. **19**(4): p. 504-12.

41. Hohenester, E., et al., *The crystal structure of a laminin G-like module reveals the molecular basis of alpha-dystroglycan binding to laminins, perlecan, and agrin.* Mol Cell, 1999. **4**(5): p. 783-92.

42. Rudenko, G., et al., *The structure of the ligand-binding domain of neurexin Ibeta: regulation of LNS domain function by alternative splicing.* Cell, 1999. **99**(1): p. 93-101.

43. Sheckler, L.R., et al., *Crystal structure of the second LNS/LG domain from neurexin 1alpha: Ca2+ binding and the effects of alternative splicing.* J Biol Chem, 2006. **281**(32): p. 22896-905.

44. Shen, K.C., et al., *Regulation of neurexin 1beta tertiary structure and ligand binding through alternative splicing.* Structure, 2008. **16**(3): p. 422-31.

45. Tisi, D., et al., *Structure of the C-terminal laminin G-like domain pair of the laminin alpha2 chain harbouring binding sites for alpha-dystroglycan and heparin.* EMBO J, 2000. **19**(7): p. 1432-40.

46. Emsley, J., et al., *Structure of pentameric human serum amyloid P component.* Nature, 1994. **367**(6461): p. 338-45.

47. Rini, J.M., *Lectin structure.* Annu Rev Biophys Biomol Struct, 1995. **24**: p. 551-77.

48. Rao, Z., et al., *The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions.* Cell, 1995. **82**(1): p. 131-41.

49. Haines, N. and K.D. Irvine, *Glycosylation regulates Notch signalling.* Nat Rev Mol Cell Biol, 2003. **4**(10): p. 786-97.

50. Bork, P., et al., *Structure and distribution of modules in extracellular proteins.* Q Rev Biophys, 1996. **29**(2): p. 119-67.

51. Downing, A.K., et al., *Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders.* Cell, 1996. **85**(4): p. 597-605.

52. Harris, R.J. and M.W. Spellman, *O-linked fucose and other post-translational modifications unique to EGF modules.* Glycobiology, 1993. **3**(3): p. 219-24.

53. Jafar-Nejad, H., J. Leonardi, and R. Fernandez-Valdivia, *Role of glycans and glycosyltransferases in the regulation of Notch signaling.* Glycobiology, 2010. **20**(8): p. 931-49.

54. Hoffman, R.C., et al., *Structural characterization of recombinant soluble rat neuroligin 1: mapping of secondary structure and glycosylation by mass spectrometry.* Biochemistry, 2004. **43**(6): p. 1496-506.

55. Leone, P., et al., *Structure-function relationships of the alpha/beta-hydrolase fold domain of neuroligin: a comparison with acetylcholinesterase.* Chem Biol Interact, 2010. **187**(1-3): p. 49-55.

56. Leone, P., et al., *Structural insights into the exquisite selectivity of neurexin/neuroligin synaptic interactions.* EMBO J, 2010. **29**(14): p. 2461-71.

57. Arac, D., et al., *Structures of neuroligin-1 and the neuroligin-1/neurexin-1 beta complex reveal specific protein-protein and protein-Ca2+ interactions.* Neuron, 2007. **56**(6): p. 992-1003.

58. Chen, X., et al., *Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions.* Nat Struct Mol Biol, 2008. **15**(1): p. 50-6.

59. Fabrichny, I.P., et al., *Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion.* Neuron, 2007. **56**(6): p. 979-91.

60. Comoletti, D., et al., *Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for beta-neurexins.* Biochemistry, 2006. **45**(42): p. 12816-27.

61. Koehnke, J., et al., *Splice form dependence of beta-neurexin/neuroligin binding interactions.* Neuron, 2010. **67**(1): p. 61-74.

62. Feng, J., et al., *High frequency of neurexin 1beta signal peptide structural variants in patients with autism.* Neurosci Lett, 2006. **409**(1): p. 10-3.

63. Glessner, J.T., et al., *Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.* Nature, 2009. **459**(7246): p. 569-73.

64. Kim, H.G., et al., *Disruption of neurexin 1 associated with autism spectrum disorder.* Am J Hum Genet, 2008. **82**(1): p. 199-207.

65. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.

66. Szatmari, P., et al., *Mapping autism risk loci using genetic linkage and chromosomal rearrangements.* Nat Genet, 2007. **39**(3): p. 319-28.

67. Yan, J., et al., *Neurexin 1alpha structural variants associated with autism.* Neurosci Lett, 2008. **438**(3): p. 368-70.

68. Blasi, F., et al., *Absence of coding mutations in the X-linked genes neuroligin 3 and neuroligin 4 in individuals with autism from the IMGSAC collection.* Am J Med Genet B Neuropsychiatr Genet, 2006. **141B**(3): p. 220-1.

69. Jamain, S., et al., *Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism.* Nat Genet, 2003. **34**(1): p. 27-9.

70. Laumonnier, F., et al., *X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family.* Am J Hum Genet, 2004. **74**(3): p. 552-7.

71. Talebizadeh, Z., et al., *Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism.* J Med Genet, 2006. **43**(5): p. e21.

72. Zhang, C., et al., *A neuroligin-4 missense mutation associated with autism impairs neuroligin-4 folding and endoplasmic reticulum export.* J Neurosci, 2009. **29**(35): p. 10843-54.

73. Rujescu, D., et al., *Disruption of the neurexin 1 gene is associated with schizophrenia.* Hum Mol Genet, 2009. **18**(5): p. 988-96.

74. Bierut, L.J., et al., *Novel genes identified in a high-density genome wide association study for nicotine dependence.* Hum Mol Genet, 2007. **16**(1): p. 24-35.

75. Hishimoto, A., et al., *Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms.* Hum Mol Genet, 2007. **16**(23): p. 2880-91.

76. Novak, G., et al., *Association of a polymorphism in the NRXN3 gene with the degree of smoking in schizophrenia: a preliminary study.* World J Biol Psychiatry, 2009. **10**(4 Pt 3): p. 929-35.

77. Nussbaum, J., et al., *Significant association of the neurexin-1 gene (NRXN1) with nicotine dependence in European- and African-American smokers.* Hum Mol Genet, 2008. **17**(11): p. 1569-77.

78. Sato, N., et al., *Association between neurexin 1 (NRXN1) polymorphisms and the smoking behavior of elderly Japanese.* Psychiatr Genet, 2010. **20**(3): p. 135-6.

79. Zweier, C., et al., *CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila.* Am J Hum Genet, 2009. **85**(5): p. 655-66.

80. Zahir, F.R., et al., *A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha.* J Med Genet, 2008. **45**(4): p. 239-43.

81. El-Fishawy, P. and M.W. State, *The genetics of autism: key issues, recent findings, and clinical implications.* Psychiatr Clin North Am, 2010. **33**(1): p. 83-105.

82. Abrahams, B.S. and D.H. Geschwind, *Advances in autism genetics: on the threshold of a new neurobiology.* Nat Rev Genet, 2008. **9**(5): p. 341-55.

83. Geschwind, D.H., *Advances in autism.* Annu Rev Med, 2009. **60**: p. 367-80.

84. Walsh, C.A., E.M. Morrow, and J.L. Rubenstein, *Autism and brain development.* Cell, 2008. **135**(3): p. 396-400.

85. Thomas, N.S., et al., *Xp deletions associated with autism in three females.* Hum Genet, 1999. **104**(1): p. 43-8.

86. Yan, J., et al., *Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients.* Mol Psychiatry, 2005. **10**(4): p. 329-32.

87. Marshall, C.R., et al., *Structural variation of chromosomes in autism spectrum disorder.* Am J Hum Genet, 2008. **82**(2): p. 477-88.

88. Wang, K., et al., *Common genetic variants on 5p14.1 associate with autism spectrum disorders.* Nature, 2009. **459**(7246): p. 528-33.

89. Roohi, J., et al., *Disruption of contactin 4 in three subjects with autism spectrum disorder.* J Med Genet, 2009. **46**(3): p. 176-82.

90. Morrow, E.M., et al., *Identifying autism loci and genes by tracing recent shared ancestry.* Science, 2008. **321**(5886): p. 218-23.

91. Alarcon, M., et al., *Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene.* Am J Hum Genet, 2008. **82**(1): p. 150-9.

92. Arking, D.E., et al., *A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism.* Am J Hum Genet, 2008. **82**(1): p. 160-4.

93. Bakkaloglu, B., et al., *Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders.* Am J Hum Genet, 2008. **82**(1): p. 165-73.

94. Dibbens, L.M., et al., *X-linked protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment.* Nat Genet, 2008. **40**(6): p. 776-81.

95. Bonora, E., et al., *Mutation screening and association analysis of six candidate genes for autism on chromosome 7q.* Eur J Hum Genet, 2005. **13**(2): p. 198-207.

96. Marui, T., et al., *Association of the neuronal cell adhesion molecule (NRCAM) gene variants with autism.* Int J Neuropsychopharmacol, 2009. **12**(1): p. 1-10.

97. Sakurai, T., et al., *Association analysis of the NrCAM gene in autism and in subsets of families with severe obsessive-compulsive or self-stimulatory behaviors.* Psychiatr Genet, 2006. **16**(6): p. 251-7.

98. Hutcheson, H.B., et al., *Defining the autism minimum candidate gene region on chromosome 7.* Am J Med Genet B Neuropsychiatr Genet, 2003. **117B**(1): p. 90-6.

99. Persico, A.M., et al., *Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder.* Mol Psychiatry, 2001. **6**(2): p. 150-9.

100. Durand, C.M., et al., *Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders.* Nat Genet, 2007. **39**(1): p. 25-7.

101. Moessner, R., et al., *Contribution of SHANK3 mutations to autism spectrum disorder.* Am J Hum Genet, 2007. **81**(6): p. 1289-97.

102. Qin, J., et al., *Association study of SHANK3 gene polymorphisms with autism in Chinese Han population.* BMC Med Genet, 2009. **10**: p. 61.

103. Sykes, N.H., et al., *Copy number variation and association analysis of SHANK3 as a candidate gene for autism in the IMGSAC collection.* Eur J Hum Genet, 2009. **17**(10): p. 1347-53.

104. Barnby, G., et al., *Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT.* Am J Hum Genet, 2005. **76**(6): p. 950-66.

105. Jamain, S., et al., *Linkage and association of the glutamate receptor 6 gene with autism.* Mol Psychiatry, 2002. **7**(3): p. 302-10.

106. Ma, D.Q., et al., *Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism.* Am J Hum Genet, 2005. **77**(3): p. 377-88.

107. Beyer, K.S., et al., *Mutation analysis of the coding sequence of the MECP2 gene in infantile autism.* Hum Genet, 2002. **111**(4-5): p. 305-9.

108. Carney, R.M., et al., *Identification of MeCP2 mutations in a series of females with autistic disorder.* Pediatr Neurol, 2003. **28**(3): p. 205-11.

109. Lam, C.W., et al., *Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome.* J Med Genet, 2000. **37**(12): p. E41.

110. Lasalle, J.M. and D.H. Yasui, *Evolving role of MeCP2 in Rett syndrome and autism.* Epigenomics, 2009. **1**(1): p. 119-130.

111. Longo, I., et al., *Three Rett patients with both MECP2 mutation and 15q11-13 rearrangements.* Eur J Hum Genet, 2004. **12**(8): p. 682-5.

112. Nagarajan, R.P., et al., *MECP2 promoter methylation and X chromosome inactivation in autism.* Autism Res, 2008. **1**(3): p. 169-78.

113. Orrico, A., et al., *MECP2 mutation in male patients with non-specific X-linked mental retardation.* FEBS Lett, 2000. **481**(3): p. 285-8.

114. Scala, E., et al., *MECP2 deletions and genotype-phenotype correlation in Rett syndrome.* Am J Med Genet A, 2007. **143A**(23): p. 2775-84.

115. Zappella, M., et al., *Study of MECP2 gene in Rett syndrome variants and autistic girls.* Am J Med Genet B Neuropsychiatr Genet, 2003. **119B**(1): p. 102-7.

116. Petit, E., et al., *Association study with two markers of a human homeogene in infantile autism.* J Med Genet, 1995. **32**(4): p. 269-74.

117. Gharani, N., et al., *Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder.* Mol Psychiatry, 2004. **9**(5): p. 474-84.

118. Benayed, R., et al., *Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus.* Am J Hum Genet, 2005. **77**(5): p. 851-68.

119. Kwon, C.H., et al., *Pten regulates neuronal arborization and social interaction in mice.* Neuron, 2006. **50**(3): p. 377-88.

120. Butler, M.G., et al., *Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations.* J Med Genet, 2005. **42**(4): p. 318-21.

121. Buxbaum, J.D., et al., *Mutation screening of the PTEN gene in patients with autism spectrum disorders and macrocephaly.* Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(4): p. 484-91.

122. Goffin, A., et al., *PTEN mutation in a family with Cowden syndrome and autism.* Am J Med Genet, 2001. **105**(6): p. 521-4.

123. Herman, G.E., et al., *Increasing knowledge of PTEN germline mutations: Two additional patients with autism and macrocephaly.* Am J Med Genet A, 2007. **143**(6): p. 589-93.

124. McBride, K.L., et al., *Confirmation study of PTEN mutations among individuals with autism or developmental delays/mental retardation and macrocephaly.* Autism Res, 2010. **3**(3): p. 137-41.

125. Redfern, R.E., et al., *A mutant form of PTEN linked to autism.* Protein Sci, 2010 . **19**(10): p. 1948-56.

126. Varga, E.A., et al., *The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly.* Genet Med, 2009. **11**(2): p. 111-7.

127. Brown, V., et al., *Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome.* Cell, 2001. **107**(4): p. 477-87.

128. Qin, M., J. Kang, and C.B. Smith, *A null mutation for Fmr1 in female mice: effects on regional cerebral metabolic rate for glucose and relationship to behavior.* Neuroscience, 2005. **135**(3): p. 999-1009.

129. Reiss, A.L., et al., *Neurodevelopmental effects of the FMR-1 full mutation in humans.* Nat Med, 1995. **1**(2): p. 159-67.

130. Hatton, D.D., et al., *Autistic behavior in children with fragile X syndrome: prevalence, stability, and the impact of FMRP.* Am J Med Genet A, 2006. **140A**(17): p. 1804-13.

131. Baker, P., J. Piven, and Y. Sato, *Autism and tuberous sclerosis complex: prevalence and clinical features.* J Autism Dev Disord, 1998. **28**(4): p. 279-85.

132. Greer, P.L., et al., *The Angelman Syndrome protein Ube3A regulates synapse development by ubiquitinating arc.* Cell, 2010. **140**(5): p. 704-16.

133. Bucan, M., et al., *Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes.* PLoS Genet, 2009. **5**(6): p. e1000536.

134. De Jaco, A., et al., *A single mutation near the C-terminus in alpha/beta hydrolase fold protein family causes a defect in protein processing.* Chem Biol Interact, 2005. **157-158**: p. 371-2.

135. De Jaco, A., et al., *A mutation linked with autism reveals a common mechanism of endoplasmic reticulum retention for the alpha,beta-hydrolase fold protein family.* J Biol Chem, 2006. **281**(14): p. 9667-76.

136. Comoletti, D., et al., *The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing.* J Neurosci, 2004. **24**(20): p. 4889-93.

137. De Jaco, A., et al., *Neuroligin trafficking deficiencies arising from mutations in the alpha/beta-hydrolase fold protein family.* J Biol Chem, 2010. **285**(37): p. 28674-82.

138. De Jaco, A., et al., *Trafficking of cholinesterases and neuroligins mutant proteins. An association with autism.* Chem Biol Interact, 2008. **175**(1-3): p. 349-51.

139. Tabuchi, K., et al., *A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice.* Science, 2007. **318**(5847): p. 71-6.

CHAPTER 2:

Neuron
Article

Structural Analysis of the Synaptic Protein Neuroligin and Its β-Neurexin Complex: Determinants for Folding and Cell Adhesion

Igor P. Fabrichny,[1,4] Philippe Leone,[2] Gerlind Sulzenbacher,[2] Davide Comoletti,[3] Meghan T. Miller,[3] Palmer Taylor,[3] Yves Bourne,[2,*] and Pascale Marchot[1,*]
[1]Biochimie des Interactions Moléculaires et Cellulaires, CNRS FRE-2738, Institut Fédératif de Recherche Jean Roche, Université de la Méditerranée, Faculté de Médecine Secteur Nord, F-13916 Marseille Cedex 20, France
[2]Architecture et Fonction des Macromolécules Biologiques, CNRS UMR-6098, Universités Aix-Marseille I & II, Campus Luminy - Case 932, F-13288 Marseille Cedex 09, France
[3]Department of Pharmacology, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093-0657, USA
[4]Alternate spelling: Fabrichniy.
*Correspondence: pascale.marchot@univmed.fr (P.M.), yves.bourne@afmb.univ-mrs.fr (Y.B.)
DOI 10.1016/j.neuron.2007.11.013

SUMMARY

The neuroligins are postsynaptic cell adhesion proteins whose associations with presynaptic neurexins participate in synaptogenesis. Mutations in the neuroligin and neurexin genes appear to be associated with autism and mental retardation. The crystal structure of a neuroligin reveals features not found in its catalytically active relatives, such as the fully hydrophobic interface forming the functional neuroligin dimer; the conformations of surface loops surrounding the vestigial active center; the location of determinants that are critical for folding and processing; and the absence of a macromolecular dipole and presence of an electronegative, hydrophilic surface for neurexin binding. The structure of a β-neurexin-neuroligin complex reveals the precise orientation of the bound neurexin and, despite a limited resolution, provides substantial information on the Ca^{2+}-dependent interactions network involved in *trans*-synaptic neurexin-neuroligin association. These structures exemplify how an α/β-hydrolase fold varies in surface topography to confer adhesion properties and provide templates for analyzing abnormal processing or recognition events associated with autism.

INTRODUCTION

Synapse formation requires the assembly of highly ordered complexes containing ion channel, receptor, neurotransmitter storage, cell adhesion, and scaffolding proteins. Interactions between cell adhesion molecules contribute to the formation of complementary presynaptic and postsynaptic sites (Lardi-Studler and Fritschy, 2007). Various signaling events including neurotransmitter release, soluble factor secretion, and direct *trans*-synaptic interactions between cell adhesion molecules have roles in synapse formation and development (Waites et al., 2005). The neuroligins are a recently characterized family of postsynaptic cell adhesion proteins shown to be critical for synapse maturation and maintenance in vivo (Nguyen and Sudhof, 1997; Varoqueaux et al., 2006) and sufficient for triggering the initial formation of synaptic contacts in cultured neurons (Scheiffele et al., 2000). In vitro, the neuroligins bind in a Ca^{2+}-dependent fashion to the laminin-neurexin-sex hormone binding globulin (LNS) domains of α-neurexins (which encompass six LNS domains) and β-neurexins (one domain only), thus establishing an intricate, heterophilic *trans*-synaptic recognition code (Boucard et al., 2005; Comoletti et al., 2006; Graf et al., 2006; Nguyen and Sudhof, 1997). Since the neuroligins are expressed in a variety of tissues, they likely have recognition and adhesions functions extending to peripheral secretory tissues.

The five mammalian neuroligin genes (numbered *NLGN1* to *4* and *4Y*) encode membrane-anchored proteins with more than 70% sequence identity in their N-terminal, extracellular domains (Figure 1). This region is linked through a highly O-glycosylated region to a single transmembrane span and a C-terminal, small intracellular domain. The neuroligin extracellular domain exhibits 32%–36% sequence identity and presumably shares overall shape with the globular domain of acetylcholinesterase (AChE) (Ichtchenko et al., 1996) (Figure 1). This makes it a member of the α/β-hydrolase fold superfamily of proteins (Ollis et al., 1992; Cygler et al., 1993), which includes catalytically active cholinesterases, lipases, and carboxylesterases, along with the noncatalytic adhesion proteins neurotactin (de la Escalera et al., 1990), glutactin (Olson et al., 1990), and gliotactin (Auld et al., 1995) (see Figure S1 available online), and the thyroglobulin C-terminal domain (Schumacher et al., 1986).

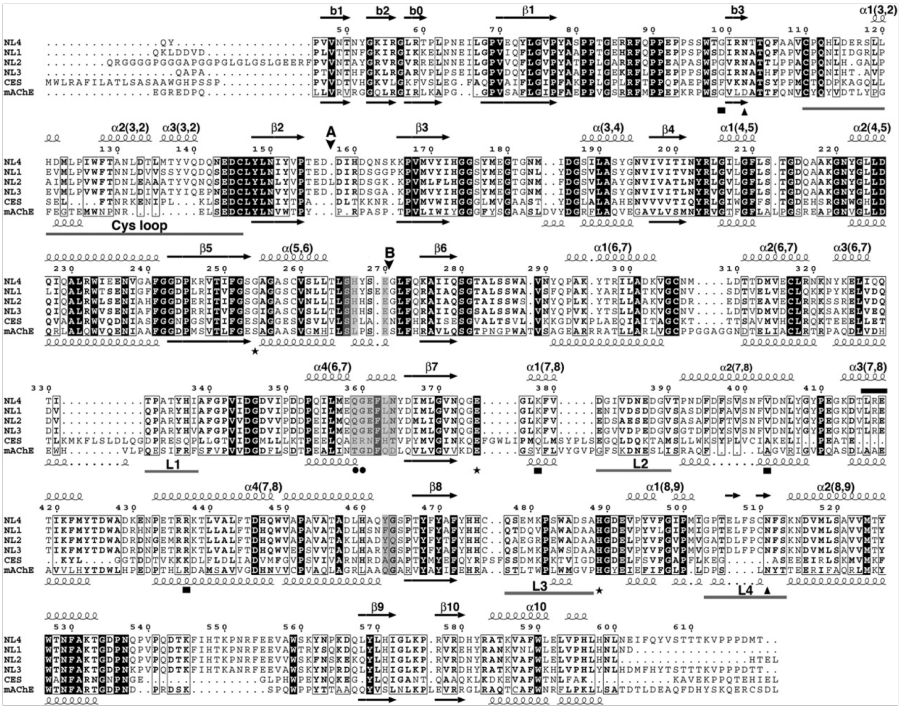

Figure 1. Sequence Conservation within the Neuroligins and Comparison with Human Liver Carboxylesterase and mAChE

The NL4 sequence displayed is that of the recombinant Gln44-Thr619 protein used for crystallization; the FLAG epitope is not shown. The positions for alternatively spliced inserts A and B are indicated by vertical arrowheads. The NL4 and mAChE numbering and their secondary structure elements are displayed above and below the alignment, respectively. The Cys-loop and loops L1 ($\alpha^3_{6,7}$-$\alpha^4_{6,7}$, Thr332-Ile338 in NL4), L2 ($\alpha^1_{7,8}$-$\alpha^2_{7,8}$, Asp381-Val390), L3 (β_8-$\alpha^1_{8,9}$, Gln477-Ser487) and L4 ($\alpha^1_{8,9}$-$\alpha^2_{8,9}$, Gly503-Ser513) are indicated by gray bars under the alignment. The NL4 residues buried at the Nrxβ1 binding interface are shaded as gray boxes. The mAChE catalytic triad residues are indicated by asterisks; the two NL4 glycosylation sites by triangles; the residue pair involved in Ca^{2+} binding by filled circles; and the NL3/NL4 residues mutated in autism patients by filled squares. The LRE motif is indicated with a black bar above the sequences.

However, several characteristics of the neuroligin molecule extend beyond predictions based solely on AChE homology. For example, one finds a Gly to Ser substitution in the catalytic triad at the active center, sufficient for eliminating catalytic function; a distinct linkage for the third disulfide bond, perhaps defining a different modular organization of the molecule; divergent patterns of consensus sites for N-linked glycosylation, between neuroligins and their α/β-hydrolase fold relatives as well as between the five neuroligins (Hoffman et al., 2004); and the possibility for insertion of two alternatively spliced sequences of 20 (insert A) and 9 (insert B) residues to generate several additional isoforms (Bolliger et al., 2001; Ichtchenko et al., 1996).

Mutations in the coding regions of the genes for neuroligin-3 and -4 (NL3 and NL4), including point mutations, truncations, and sequence deletions, were found to be associated with autism spectrum disorders (ASD) and mental retardation, indicating a strong genetic influence on developmental disorders (Blasi et al., 2006; Jamain et al., 2003; Laumonnier et al., 2004; Talebizadeh et al., 2006; Yan et al., 2005). These findings suggest that abnormal presentation of neuroligin at the postsynaptic surface, its altered association with neurexin, or both governs synaptic structural and cognitive disorders (Tabuchi et al., 2007).

Crystal structures of α- and β-neurexin LNS domains show a lectin, jelly-roll fold and reveal the positions for alternatively spliced surface loops on the hypervariable loop edge, located opposite the roll N and C termini (Rudenko et al., 1999; Sheckler et al., 2006). The presence or absence of these loops influences neurexin recognition for

neuroligin. The expression of different neuroligin genes and presence of alternatively spliced insert B at the neuroligin surface also dictate affinity for neurexin (Boucard et al., 2005; Comoletti et al., 2006).

Neuroligin association with neurexin and its influence on cell adhesion only require extracellular sequences (Scheiffele et al., 2000). Hydrodynamic analyses established that the neuroligin extracellular domain, isolated through insertion of an early stop codon before the transmembrane span, forms homodimers in solution (Comoletti et al., 2003, 2006). Small-angle X-ray scattering (SAXS) combined with rigid-body modeling data led to experimentally based neuroligin models (Comoletti et al., 2007), using the crystal structure of mouse AChE (mAChE) (Bourne et al., 2003) and data on Cys connectivity and positions for oligosaccharide linkages (Hoffman et al., 2004) as constraints. These data are consistent with neuroligin dimerization through a four-helix bundle, as found for AChE (Bourne et al., 1995; Sussman et al., 1991), and an extended C-terminal linker emerging from the bundle.

A combination of X-ray and neutron scattering data recorded on neuroligin-1 (NL1) in complex with soluble β-neurexin-1 (Nrxβ1) determined that two Nrxβ1 subunits associate on cross-opposite sides of the long axis of the neuroligin dimer (Comoletti et al., 2007). Final positioning of the bound neurexins was achieved using rigid-body modeling and constraints derived from earlier binding and mutagenesis data. This resulted in an initial image of the complex wherein the Nrxβ1 hypervariable loop edge faces the neuroligin surface area proximal to splice site B. However, the low-resolution SAXS data only delineate the overall shape of the complex and the Nrxβ1 binding position, and do not reveal the precise orientation of the bound Nrxβ1 or the surface determinants involved.

A soluble form of NL4, which is naturally devoid of splice inserts and contains only two N-glycosylation sites (Bolliger et al., 2001) (Figure 1), proved to be amenable to obtaining well-diffracting crystals. Herein we present a 2.2 Å resolution crystal structure of the extracellular domain of NL4 (Table 1), also a structure of a synaptic cell adhesion member of the α/β-hydrolase fold family of proteins. This structure permits a detailed analysis of those structural features that are unique to the neuroligins when compared with their catalytically active relatives, and that could not be predicted from homology-based models. The structure also defines positions for surface determinants known to be critical for folding and processing and for neurexin interaction. Moreover, a 3.9 Å resolution structure of a complex between NL4 and a soluble Nrxβ1 form, also devoid of a splice insert, was also solved. This structure reveals the exact position and orientation of the bound neurexin relative to the neuroligin, along with a restricted, Ca^{2+}-mediated, interfacial area consistent with a labile *trans*-synaptic complex. Hence, these structures, as compared with earlier models, provide an improved template for analysis of abnormal processing and/or recognition events that could be associated with ASD.

RESULTS AND DISCUSSION

Sequence-based searches for NL4 catalytically active relatives revealed highest identity with AChE and human liver carboxylesterase (Figure 1), while structure-based searches also pointed to bovine bile salt-stimulated lipase and bacterial *p*-nitrobenzyl esterase as close homologs. Hence mAChE, as a mammalian synaptic protein with the capacity for forming noncovalent dimers (Bourne et al., 1995), is an appropriate relative for comparative description of the NL4 structure.

The Neuroligin Subunit

The extracellular domain of NL4, which adopts an ellipsoidal shape with overall dimensions of 45 × 60 × 65 Å, consists of a twisted, 12-stranded central β sheet surrounded by 14 α helices, typical of the α/β-hydrolase fold (Cygler et al., 1993; Ollis et al., 1992) (Figure 2). Of the three intramolecular disulfide bridges conserved in the neuroligins, the first two, Cys110-Cys146 and Cys306-Cys317 (Cys69-Cys96 and Cys257-Cys272 in mAChE) (Figure 1), are conserved in most members of the fold family. The third bridge, Cys476-Cys510 (as earlier defined by peptide mapping and mass spectrometry analyses), is unique to the neuroligins (Hoffman et al., 2004) (Figure S1). This bridge constrains the included surface loop Gly503-Ser513 (or L4, or $\alpha^1_{8,9}$-$\alpha^2_{8,9}$; cf. Figure 1) into a conformation not observed in AChE that directly stabilizes the tip of the neighboring loop Gln 477-Ser487 (L3, β_8-$\alpha^1_{8,9}$) through a stacking interaction of Trp484 with Arg437, and with Val377, to a more limited extent. In individuals with ASD, residues Arg437 and Lys378 are found to be replaced by Cys in NL3 and Arg in NL4, respectively (Jamain et al., 2003; Yan et al., 2005).

In turn, the neuroligins lack the third bridge found in AChE (Cys409-Cys529 in mAChE), where it connects helix $\alpha^4_{7,8}$ to the C-terminal helix α_{10} (Figure 1). In fact, in NL4, absence of this bridge is compensated for by a set of strong hydrophobic interactions involving Ala451, Val454, Val587, and Leu591 (Cys409, Ala412, Cys529, and Asn533 in mAChE) in this region. Residue Cys259, which is free in all neuroligins, is deeply buried within helix $\alpha_{5,6}$, far from the protein surface. Ala substitution to this Cys in an NL1-soluble form resulted in improved intracellular processing and secretion yields (Comoletti et al., 2004).

GlcNAc moieties are linked to Asn102 and Asn511, which are conserved in all neuroligins. Residues Asp141 and Asn307, which correspond to potential Asn-linked glycosylation sites in neuroligin-2 (NL2) and NL1, respectively, are located at the NL4 surface, as are the positions for the spliced insert A (between Asp157 and Asp158) found in NL1, NL2, and NL3, and the spliced insert B (between Glu270 and Gly271) found in NL1.

The Neuroligin Dimer

Two NL4 subunits, related by a two-fold symmetry axis and linked through a tightly packed four-helix bundle made of helix $\alpha^3_{7,8}$ and the C-terminal helix α_{10} from each subunit

Table 1. Data Collection and Refinement Statistics

	NL4: Low Resolution	NL4: High Resolution	Nrxβ1-NL4 Complex
Data collection			
Beamline at the ESRF	ID23	ID14-EH1	ID29
Wavelength (Å)	0.976	0.934	0.979
Space group	$C2_1$	$P2_12_12$	$P2_12_12$
Cell dimensions a, b, c (Å) α,β,γ (°)	197.9, 63.6, 138.4 112.9	139.7, 154.0, 81.3	158.5, 198.7, 85.7
Resolution range (Å)[a]	15–3.2 (3.3–3.2)	30–2.2 (2.25–2.20)	30–3.9 (4.0–3.9)
Total observations	79,446	587,014	121,593
Unique reflections	25,664	89,603	25,251
Multiplicity	3.2 (3.2)	6.6 (6.6)	4.8 (4.9)
Completeness (%)	96.3 (98.8)	99.9 (100)	99.4 (99.9)
R_{sym}(%)[b]	12.2 (41.5)	6.5 (47.0)	8.2 (55.9)
$< I /\sigma(I) >$	8.9 (3.1)	21.6 (4.5)	16.1 (3.3)
Refinement			
Resolution (Å)		30–2.2	30–3.9
No. reflections		85,129	25,214
R_{work}/R_{free} (%)[c]		16.7/20.1	23.1/32.3
No. atoms		9309	11,341
Protein		8606	11,339
Ligand/sugar/ion		38/56/17	-/-/2
Water		592	-
Average B-factors (Å^2)			
Protein		38.8/41.5	-
Ligand/sugar/ion		78.5/154.5/63.5	-
Water		43.0	-
Rmsd[d]			
Bond (Å)		0.013	0.013
Angles (°)		1.39	1.56
Chiral volume (A^3)		0.087	0.109
Ramachandran plot (%)			
Most favored regions		90.7	71.6
Additionally allowed regions		9.3	26.8
PDB accession code	(not deposited)	3BE8	2VH8

[a] Values in parentheses are those for the last shell.
[b] $R_{sym} = \Sigma_{hkl} \Sigma_i |I_{i(hkl)} - < I_{hkl} >| / \Sigma_{hkl} \Sigma_i I_{i(hkl)}$, where I is an individual reflection measurement and $< I >$ is the mean intensity for symmetry-related reflections.
[c] $R_{cryst} = \Sigma_{hkl} ||Fo| - |Fc|| / \Sigma_{hkl} |Fo|$, where Fo and Fc are observed and calculated structure factors, respectively. R_{free} is calculated for 5% of randomly selected reflections excluded from refinement.
[d] Root mean square deviation from ideal values.

(cf. Figure 1), assemble as a noncovalent antiparallel dimer (Figure 2). The same assembly was found for covalent and noncovalent AChE dimers (Bourne et al., 1995; Sussman et al., 1991). Hydrophobic interactions account for 100% of the ~875 Å^2 interface area buried on each subunit, versus 77% and 870 Å^2 for mAChE, suggesting that the NL4 dimer is even more stable than the mAChE dimer. These values are consistent with analytical centrifugation data that showed that the neuroligin extracellular domain sediments as a dimer, even at low concentration (Comoletti

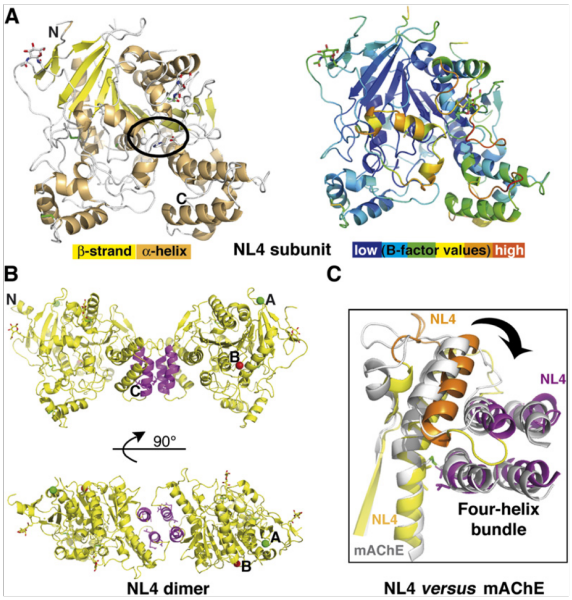

Figure 2. Overall View of the NL4 Structure

(A) (Left) The NL4 subunit is displayed with yellow β strands, orange α helices, green disulfide bridges, and an encircled vestigial catalytic triad. The GlcNAc moieties are visible; the N and C termini are labeled. (Right) The NL4 subunit, oriented as in (A), is colored according to temperature factor (B-factor values). The flexibility of the Cys-loop (orange) is evident.

(B) Overall views of the NL4 dimer with the C termini pointing down (top) and with 90° rotation around the main axis (bottom). The four-helix bundle is displayed in magenta. The positions for splice sites A and B are indicated by green and red spheres, respectively.

(C) Close-up view of the superimposed four-helix bundle regions in the NL4 (yellow, orange, and magenta) and mAChE (gray) dimers, showing the twist of the helical region at the subunit interface.

et al., 2003), whereas diluted mAChE sediments as a monomer (Marchot et al., 1996). In vitro data on mutants within the four-helix bundle of NL1 (Lys578/Val579 and Glu584/Leu585 Ala mutations) (Dean et al., 2003) argue for the requirement of an oligomer for synaptogenic activity.

Compared with mAChE, in NL4 the displacement of the third disulfide bridge and its replacement with noncovalent, hydrophobic interactions, along with the distinct length and conformation of surface loop L2 ($\alpha^1_{7,8}$-$\alpha^2_{7,8}$, Asp381-Val390 in NL4) and shifted positions of loops $\alpha^2_{7,8}$-$\alpha^3_{7,8}$ (Tyr407-Gly412) and $\alpha^3_{7,8}$-$\alpha^4_{7,8}$ (Thr425-Pro433) in the four-helix bundle region (cf. Figure 1), result in a weaker connection between the core of the subunit and the C-terminal $\alpha^3_{7,8}/\alpha_{10}$ motif (Figure 2). In fact, this motif as a rigid body is shifted outward by ~14° from its position in mAChE; this motion is associated with a 15°–20° twist of one subunit relative to the second one around the dimer long axis.

A Leu-Arg-Glu (LRE) laminin binding motif conserved in all neuroligins is found in the two $\alpha^3_{7,8}$ helices in the four-helix bundle. While the Leu416 side chain establishes hydrophobic interactions within the bundle, the charged side chains of Arg417 and Glu418 point to the bulk solvent and are part of a highly polar, conserved network. LRE motifs are found at three distinctive positions in the extracellular domain of neurotactin (de la Escalera et al., 1990) (Figure S1). Whether these motifs in the neuroligins and neurotactin bind laminin or have another function is unknown.

The Neuroligin Central Pocket

The central pocket, which in the cholinesterases contains the active center and oxyanion hole, lies centrosymmetric to the NL4 molecule (Figure 3 and Figure S2). The pocket in NL4, with its 200 Å^3 volume, is ~2-fold smaller than that in mAChE, mainly due to partial occupancy by the tip of the long Cys-loop (cf. below). However, and despite the Gly254 substitution to the active center Ser residue found in the catalytically active hydrolases (Ser203 in mAChE), the topology of the pocket and many of its lining residues are remarkably well conserved. Only small deviations are found upon superimposition of the vestigial triad residues Gly254, Glu375 (shared with AChE and only a few other relatives, instead of an Asp as found in most hydrolases), and His489; the oxyanion hole residues Gly175, Ser176 (Gly in mAChE), and Ala255 (Ile in NL3 but conserved in mAChE); Ser253 (Glu in mAChE); and Trp287 (conserved in mAChE) with their counterparts in mAChE. NL4 residue Ser253 replaces mAChE Glu202, whose mutation changes the catalytic constant (Gibney et al., 1990). A Gly254Ser mutation is not sufficient to confer AChE activity to NL1 (D.C., unpublished data).

In the oxyanion hole, a tightly bound phosphate, positioned similarly to the carbonate/acetate found in mAChE structures (Bourne et al., 2003), forms five hydrogen bonds with NL4 residues (Figure 3). A chloride ion is also observed, 7 Å from the phosphate. Hence the NL4 central pocket, while not serving a catalytic role, may confer

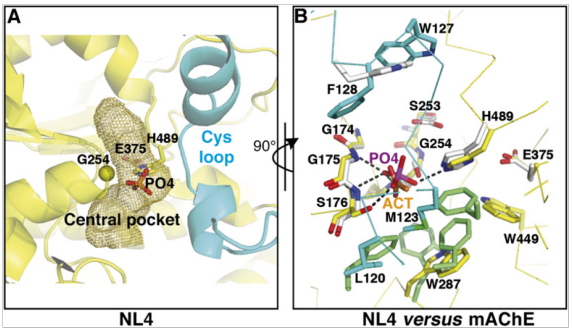

Figure 3. The NL4 Central Pocket Compared with the mAChE Active Center

(A) Close-up view of the tip of the Cys-loop (cyan) occluding access to the central pocket in NL4. The internal pocket, containing the phosphate, is displayed as a mesh surface. Residue Gly254 at the base of the nucleophilic elbow (β_5-$\alpha_{5,6}$ turn) is shown as a yellow sphere.

(B) Close-up superimposition of the vestigial active center in NL4, rotated 90° from (A), and the active center in mAChE (rmsd, 0.3 Å for 34 Cα atoms). The NL4 residues in the pocket are shown in yellow and those from the sealing Cys-loop in cyan. The mAChE residues in the pocket with Trp86 at the top are shown in white and those in the acyl pocket are shown in green. The NL4 Gly254 substitution to mAChE Ser203 is visible at the center. Bound phosphate (PO4) in NL4 mimics bound acetate (ACT) in mAChE. The phosphate oxygen atoms establish five hydrogen bonds with the main chain nitrogen atoms of Gly175, Ser176, and Ala255 (Gly121, Gly122, and Ala204 in mAChE) and the side chains of Ser176 and His489 (His447 in mAChE). The bond with Ala255 lies behind the figure plan and is not visible.

additional flexibility in the NL4 molecule with its capacity for solvent and ion binding. Whether this feature proves advantageous to the neuroligin adhesion function remains to be documented.

Structural Variations within the α/β-Hydrolase Fold

NL4 and mAChE diverge primarily in the conformation and size of five surface loops located on the same face of the molecule: the Cys-loop (Cys110-Cys146); and loops L1 ($\alpha^3_{6,7}$-$\alpha^4_{6,7}$, Thr332-Ile338 in NL4), L2 ($\alpha^1_{7,8}$-$\alpha^2_{7,8}$, Asp381-Val390), L3 (β_8-$\alpha^1_{8,9}$, Gln477-Ser487), and L4 ($\alpha^1_{8,9}$-$\alpha^2_{8,9}$, Gly503-Ser513) (cf. Figure 1). The Cys-loop (Figure 4, cyan) is homologous to the long Ω-loop (Cys69-Cys96) that in AChE forms one side of the rim of the active site gorge and appears to restrict substrate accessibility (Bourne et al., 1995; Sussman et al., 1991). In NL4 this loop, longer by nine residues and of a distinct sequence, folds as two short α helices almost perpendicular to each other and is followed by a short helical turn, in place of the single short α helix in mAChE. The tip of the loop curves oppositely to that in mAChE and bends, with a displacement of up ot 17 Å inward, toward the center of the subunit, thereby occluding access to the central pocket. Several stabilizing polar and nonpolar interactions are found between residues of the loop edge and residues surrounding the central pocket.

The conformation of the Cys-loop in NL4 requires large compensative conformational rearrangements from four other surface loops located on the same face of the NL4 molecule that delineate the gorge rim in mAChE. Loop L1 (Figure 4, yellow) has a shorter edge segment than in mAChE and is ~4 Å removed from the Cys-loop, precluding loop overlap. Loop L2 (Figure 4, orange) is shorter than in mAChE and moves away from the Cys-loop tip. Together with an ~3.5 Å outward shift of helix $\alpha^1_{7,8}$, this positioning yields ~9 Å separation between these loops. Loop L4 (Figure 4, magenta) is longer than in mAChE and adopts a conformation that appears to stabilize the Cys-loop conformation. It stretches toward the side of the Cys-loop, providing the Phe508 side chain for stacking interaction with Trp127 in the Cys-loop edge segment. The particular conformation of loop $\alpha^1_{8,9}$-$\alpha^2_{8,9}$ in NL4 requires conformational adaptation of the surrounding loop L3 (Figure 4, green), while both loops are stabilized by the neuroligin-specific Cys476-Cys510 bridge.

The Cys-loop in NL4 is also a homolog of the lid or flap in the lipases in the α/β-hydrolase fold family (Cygler et al., 1993; Grochulski et al., 1994). Structures of the apo and complexed forms of the *Candida rugosa* lipase show a closed flap occluding the active site cavity in the absence of substrate and an open loop folded backward to accommodate the bound substrate, respectively (Grochulski et al., 1994). In human pancreatic lipase, the open conformation is stabilized by the bound colipase partner (van Tilbeurgh et al., 1992). In *p*-nitrobenzyl esterase, the corresponding loop adopts a closed conformation similar to NL4, but it is 15 residues shorter and turned about 40° away from the active site entry, thus leaving the gorge unoccupied (Spiller et al., 1999). In AChE, the Ω-loop is believed to undergo rapid conformational fluctuations, allowing diffusion-limited substrate entry (Shi et al., 2001). In NL4, substantial mobility in the Cys-loop is illustrated by its greater B-factor values as compared with other regions (Figure 2 and Figure 4C), along with presence of the bound phosphate in the central pocket (Figure 3) and the existence of distinct conformations in the two dimer subunits (cf. Experimental Procedures).

The global change within the Cys-loop of NL4, compared with its enzyme relatives, requires large conformational and positional adaptation of the surrounding loops associated with deletions or insertions to construct a novel surface topography, which is well conserved within the neuroligins. Similarly, sequence alignment among the adhesion proteins, neuroligins, neurotactin, gliotactin, and glutactin reveals length compensation between the

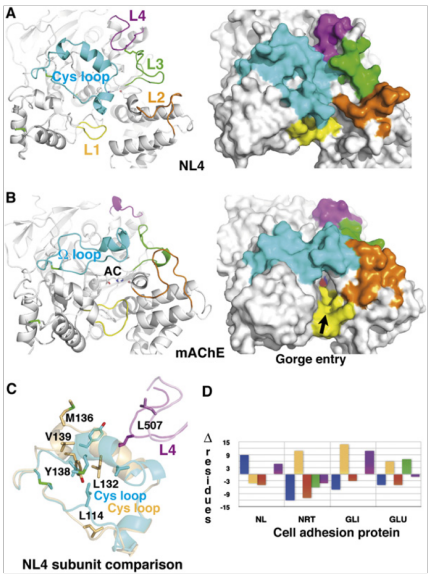

Figure 4. Comparison of the Surface Loops in NL4 and mAChE

(A) The Cys-loop (Cys110-Cys146; cf. Figure 1) in NL4 and (B) the Ω-loop (Cys69-Cys96) in mAChE are displayed in cyan, loops L1 (Thr332-Ile338 in NL4) in yellow, loops L2 (Asp381-Val390) in orange, loops L3 (Gln477-Ser487) in green, and loops L4 (Gly503-Ser513) in magenta. The Cys-loop conformation is stabilized by several polar and nonpolar interactions between Leu120, His121, Asp122, and Trp127 at the loop edge, and Ser176, Glu179, Lys378, Gly490, and Leu507, which surround the central pocket. Label AC on mAChE denotes the active center. Molecular surfaces of NL4 ([A], right) and mAChE ([B], right) show the occluded surface in NL4 and the accessible gorge entry, marked by the arrow, in mAChE. (C) Superimposition of the two Cys-loops (cyan and orange) and interacting loops L4 (magenta and violet) from the two NL4 subunits in the dimer. Residues that undergo large displacements are displayed. (D) Graph of differences in the loop sizes (in numbers of residues) of NL4, neurotactin (NRT), gliotactin (GLI), and glutactin (GLU) compared with mAChE (color codes as in [A]).

Cys-loop and its neighboring surface loops (Figure 4D and Figure S1). Hence, there is considerable variation in the loop sizes and conformations among members of the α/β-hydrolase fold family, where the adhesion proteins add an even greater degree of complexity.

Electrostatic Properties of Neuroligin

NL4 and mAChE show distinctive electrostatic features (Figure 5). NL4 displays an annular motif of electronegative surface potential covering a large area on the Cys-loop face of the molecule. A similar motif is found in AChE, where it surrounds the active site gorge entry and is thought to attract the positively charged ACh substrate (Ripoll et al., 1993), and in other adhesion members of the α/β-hydrolase fold family (Botti et al., 1998). This face in mAChE also encompasses the binding site for the peptidic ligand fasciculin (Bourne et al., 1995).

In contrast to mAChE, NL4 uniquely displays a second region of electronegative surface potential covering half the surface area on the face opposite the Cys-loop (Figure 5). This second motif covers a region more ordered than the Cys-loop region that is conserved in the other neuroligins, but not in neurotactin, gliotactin, or glutactin (Figure S1), as found upon analysis of primary theoretical models. In the symmetric NL4 dimer, the two electronegative motifs on each subunit merge into a continuous belt surrounding the two subunits, whereas in the mAChE dimer, the motif is restricted to one subunit. The second electronegative surface in the neuroligin corresponds to the Nrxβ1 binding face, as defined by the SAXS findings (Comoletti et al., 2007). As a result of the surface charge distribution, the overall net charge and dipole moment of the NL4 subunit (−25*e* and ∼600 Debyes) are significantly larger and slightly lower, respectively, than those of mAChE (−9*e* and ∼900 Debyes). The dipole vectors in the two molecules diverge by ∼30°; the negative pole in each molecule is located close to the central pocket and the positive pole is on the side opposite the NL4 Cys-loop and mAChE Ω-loop regions. Hence, selectivity of the binding surface does not appear to depend on electrostatics. Rather, the balance of surface potential may be related to the adhesive function of neuroligin.

The Neurexin-Neuroligin Complex

To delineate the structural determinants of neurexin specificity for neuroligins, we have determined a crystal structure of the LNS domain of Nrxβ1 bound to NL4. Despite the overall limited resolution, the electron density map unambiguously reveals the position and orientation of the bound Nrxβ1 and those of the interacting side chains at the complex interface (Figure 5 and Figure S2). Nrxβ1 is bound through its hypervariable loop edge to the electronegative surface of the NL4 molecule opposite the Cys-loop region, where it protrudes some 40 Å with its β strands oriented perpendicular to the long axis of the NL4 dimer. This makes the Nrxβ1-NL4 complex a mirror image of the mAChE complex with the β-strand three-fingered peptide fasciculin (Bourne et al., 1995) (Figure 5), exemplifying the variability in ligand selectivity and interaction capability of the α/β-hydrolase fold.

The concave side of the Nrxβ1 molecule faces the dimer four-helix bundle, which results in bound Nrxβ1 being rotated a quarter of a turn from its orientation in the SAXS model (Comoletti et al., 2007). The binding interface is very flat, and it encompasses a buried surface area of ∼560 Å^2 on each partner in the complex (1.4 Å probe radius) (Figure 6). This area, which is consistent with a low-affinity complex, corresponds to only 2.5% and 7% of the molecular surfaces of NL4 and Nrxβ1, respectively. Compared with the unbound NL4 and Nrxβ1 individual structures,

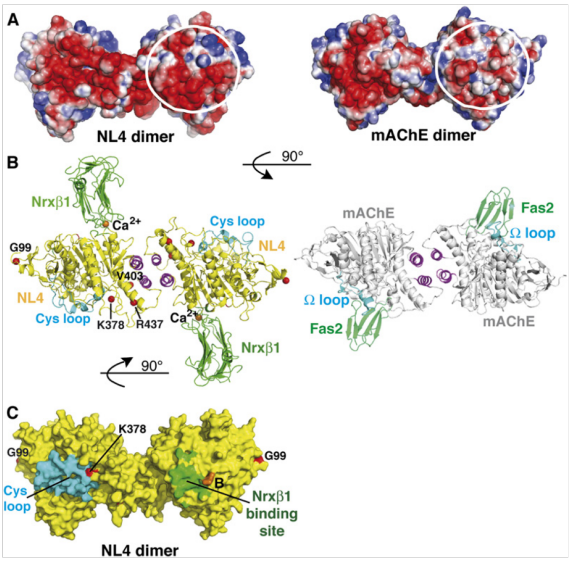

Figure 5. Electrostatic Potentials of the NL4 and mAChE Dimers and Comparison of the Dimeric Nrxβ1-NL4 and Fasciculin-mAChE Complexes

(A) The NL4 (left) and mAChE (right) dimers are oriented with their C termini pointing down (same as in Figure 2B, top). In each dimer, the two-fold symmetry axis places opposite faces of each subunit side by side. The encircled surface area is negatively charged in NL4, but not mAChE. This results in a continuum of electronegative surface potentials from one subunit face to the other face in the NL4 dimer, in contrast to the restricted localization of this potential to a single subunit face in the mAChE dimer. Electrostatic surface potentials are contoured at −3/+3 kT/e, where red describes a negative and blue a positive potential.

(B) (Left) The Nrxβ1-NL4 complex with a yellow NL4, magenta four-helix bundle, cyan Cys-loops, and green Nrxβ1 molecules. The bound Nrxβ1 is oriented such that its curved β sheet is positioned in the prolongation of the curved long helix $\alpha^4_{7,8}$, which crosses the NL4 subunit. Red spheres denote the Gly99, Lys378, Val403, and Arg437 positions for mutations in NL3/NL4 of autism patients. (Right) The fasciculin-mAChE complex with a gray mAChE, magenta bundle, cyan Ω-loops, and green fasciculins. Nrxβ1 (~25 kDa) and fasciculin (~7 kDa) bind on opposite faces of the NL4/mAChE dimer, through loops that connect β strands oriented perpendicular to the binding interfaces. The two complexes are oriented 90° compared with (A).

(C) The NL4 dimer, oriented as in (A), is displayed with a yellow molecular surface, a cyan Cys-loop, and a green area buried by bound Nrxβ1. The position for splice site B in NL1 is displayed in orange, and the Gly99 and Lys378 positions for mutations in autism patients are displayed in red. The other mutations are not visible with this orientation.

no major conformational changes are observed at the interacting surfaces in the two partners. However, at the opposite side of the NL4 molecule, the Cys-loop appears to exhibit an ~3 Å outward displacement. Whether this rigid-body motion is related to Nrxβ1 binding is unknown.

The presence of bound Ca^{2+} is an evident feature at the center of the complex interface (Figure 6). Here, Ca^{2+} is coordinated by the carboxylate side chain of Asp137, the amide of Asn238, and the carbonyl oxygens of Val154 and Ile236 from Nrxβ1 (same numbering as in Ushkaryov et al., 1994), which is consistent with the coordination observed in the Nrxα1 LNS-2 domain structure (Sheckler et al., 2006). The orientation of the Asn238 side chain is constrained by an additional interaction with the backbone carbonyl of Gly360 from NL4. Additional candidates for the coordination sphere include the carbonyl oxygens of NL4 residues Gln359 and Gly360, probably through water-mediated interactions, and a second water molecule in apposition with the Nrxβ1 surface, suggesting that NL4 residues may act as indirect ligands for Ca^{2+} coordination. A contribution from NL4 would compensate for the low Ca^{2+} affinity of neurexin alone (~400 μM for the Nrxα1 LNS-2 domain) (Sheckler et al., 2006). Higher resolution will be required to fully describe the Ca^{2+} coordination geometry and the reasons for Ca^{2+} selectivity over other cations (Comoletti et al., 2003; Nguyen and Sudhof, 1997).

Direct interactions between the complex partners arise principally from the NL4 Gln359-Asn364 stretch within loop $\alpha^4_{6,7}$-β_7 (cf. Figure 1), with a minor contribution from Tyr463-Gly464 in loop $\alpha^1_{7,8}$-β_8, His268/Glu270 in loop $\alpha_{5,6}$-β_6, and residues within the hypervariable loop edge of Nrxβ1 (Table 2). Two salt bridges (Nrxβ1 Arg109 with NL4 Glu361/Glu270; and Nrxβ1 Arg232 with NL4 Asp 351) and a polar interaction (Nrxβ1 Asn103 with NL4 Arg561) are arranged as a triangle at the perimeter of the Ca^{2+} binding site and may serve as boundary clamps for the complex. Overall, these structural features are consistent with a labile or moderate-affinity complex (K_D = 132 nM for Nrxβ1 and NL4) (Comoletti et al., 2006) imparting flexibility in the interaction partners of the membrane-tethered adhesion proteins, and perhaps contributing to plasticity of the synapse. Prior mutagenesis data on the Nrxβ1 Ca^{2+}-binding residues Asp137 and Asn238 and on residues Arg109, Arg232, and Thr235 at the complex interface, but not Ser107, Thr108, Thr156, Asp157, Asn184, and Gln233, show diminished binding or synaptogenic activity (Graf et al., 2006), consistent with the spatial arrangement of partners in the crystalline complex.

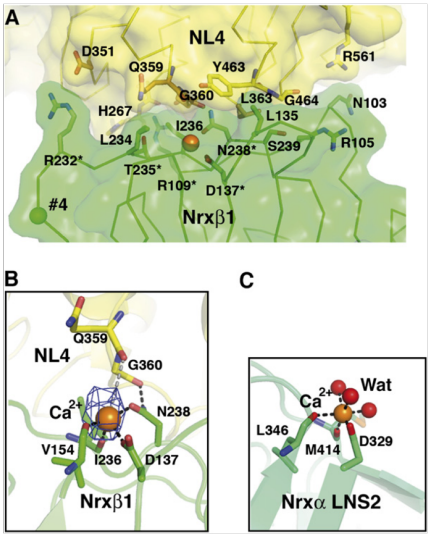

Figure 6. Close-up View of the Ca^{2+}-Mediated Nrxβ1-NL4 Complex Interface

(A) The Nrxβ1 molecular surface and key interacting side chains are displayed in green. The NL4 molecular surface is in yellow. The NL4 interacting side chains are in orange for the Ca^{2+}-binding Gln359-Asn364 residues, and in yellow for the other residues. The Ca^{2+} ion is shown as an orange sphere central to the interface, and the position of splice site 4 in Nrxβ1 is shown as a green sphere. The asterisks denote Nrxβ1 residues whose mutation impairs binding to neuroligin (Graf et al., 2006).

(B) Close-up view of the Ca^{2+} coordination at the complex interface (same orientation as in [A]). The omit difference electron density map (cyan) for Ca^{2+} is contoured at 3.5 σ. Bonds with Nrxβ1 residues and likely bonds with NL4 residues are displayed as black and gray dotted lines, respectively.

(C) Ca^{2+} coordination in the Nrxα LNS-2 domain structure (PDB accession code 2H0B; Sheckler et al., 2006) (same orientation as that of Nrxβ1 in [B]). The three coordinating water molecules are displayed as red spheres.

Table 2. Nrxβ1 and NL4 Residues in Contact or within Contact Distance in the Complex

	Nrxβ1	NL4
Hydrogen bonds	Pro106 O[a]	Asn364 N[a]
	Ser107 O[a]	Asn364 ND2[a]
	Arg109 NH1/NH2	Ser266 O[b]
		Glu270 OE1[b]
		Glu361 OE1/OE2
	Thr235 OG1[a]	*Gln359* NE2[a]
Van der Waals contacts[c]	Ser107	Leu363
	Arg109	His267
		Glu270[b]
	Leu135	Tyr463
	Leu234	*Gln359*
	Ile236	Gly360
		Glu361
	Asn238	Leu363
	Ser239	Leu363
		Gly464
Putative salt bridges	Arg109	Glu270
		Glu361
	Arg232	Asp351
Putative hydrogen bonds	Asn103 OD1	Arg561 NH1/2
	Arg105 NH1/2	Gly464 O
	Arg109 NH1/2	Ser266 O
		His267 O
		Asn364 OD1[a]
	Thr235 OG1[a]	Glu361 N[a]
	Asn238 ND2[b]	*Gln359* O[b]
		Gly360 O[b]
	Ser239 OG[a]	Gly464 N[a]

In italics are those residues that participate to the Ca^{2+} binding site.

[a] Contact observed only in the one or the other subunit.

[b] Contact observed only in the one or the other subunit.

[c] Within a 4 Å distance between atoms from each partner in the complex.

Most interacting residues are conserved among neuroligins (Figure 1 and Figure 6), arguing for the present structure to be a representative model structure for other neuroligins. The only exception refers to Gly464, which is replaced by Gln in NL2. Since the Gly464 Cα is 4.6 Å away from the facing Nrxβ1 Ser239, the bulkier Gln side chain is likely responsible for the lower affinity of NL2 (K_D = 8840 nM) (Comoletti et al., 2006). In fact, an Ala substitution to Gly464 in NL1 abolishes Nrxβ1 binding (D.C., unpublished data). In contrast to the Cys-loop region, the shape of the Nrxβ1 binding region is surprisingly well conserved between NL4 and AChE, with only a small displacement of surface loop $\alpha^4_{6,7}$-β_7. However, the key NL4 interacting residues are neither conserved in AChE nor in neurotactin, glutactin and gliotactin (Figure S1).

NL4 lacks both splice inserts A and B that are typically found in other neuroligins (Figure 1). The positions for these inserts are located on the neuroligin face opposite the Cys-loop where they are separated by 25 Å. The presence of the short insert B, which borders the Nrxβ1

binding site, and its posttranslational processing for N-linked glycosylation, appreciably affect the Nrxβ1 binding affinity (Boucard et al., 2005; Comoletti et al., 2006), consistent with the current Nrxβ1-NL4 complex structure. By contrast, presence of the long insert A, which can display two distinctive sequences (Ichtchenko et al., 1996) with drastically opposite overall net charges, may affect binding indirectly by influencing the binding surface topography and overall electrostatics. In turn, splice site 4 in Nrxβ1 is remote from the binding interface (15 Å from the bound Ca^{2+}) and the presence of the 30-residue insert is likely to have a minor influence on the complex formation, in agreement with biochemical data (Boucard et al., 2005; Comoletti et al., 2006; Graf et al., 2006).

Autism and Neurexin-Neuroligin Mutations

Recently, several point mutations located within the extracellular domain of neuroligins have been reported in human NL3 and NL4 and appear to be associated with ASD. Mapping these mutations onto the molecular surface of the NL4 dimer shows that all are remote from the bound Nrxβ1 (Figure 5). In human NL3, a Cys substitution to Arg451 (NL4 Arg437) gives rise to endoplasmic reticulum retention of the protein and decreases affinity for Nrxβ1 (Comoletti et al., 2004). The fact that a homologous mutation in butyrylcholinesterase (Yen et al., 2003) causes an identical phenotype reveals a common mechanism of intracellular retention for relatives in the α/β-hydrolase fold family (De Jaco et al., 2006). In vivo studies of this mutation using a knockin mouse result in enhanced GABA inhibitory activity (Tabuchi et al., 2007). Other missense mutations identified within the region encoding the extracellular domain of each NL4 gene in unrelated patients with autism disorder are Gly99Ser, Lys378Arg and Val403Met (the latter two residues are invariant in all neuroligins). Gly99 is located at the surface in a turn preceding the short β_3 strand, and its mutation does not seem to affect folding or binding. In contrast, Val403 is located near the end of helix $\alpha^2_{7,8}$ (Asn493-Leu406) and participates in the tight parallel packing of this helix onto the four-helix bundle. Mutation of this residue by a bulkier side chain may affect correct folding of the C-terminal domain and prevent formation of the functional neuroligin dimer. Lys378 is located at the beginning of the short helix $\alpha^1_{7,8}$ (Leu377-Val380) and establishes van der Waals contacts with Asp122 within the Cys-loop. Arg437, located within the long helix $\alpha^4_{7,8}$, is tightly packed against Trp484 and hydrogen bonded to the backbone carbonyls of Asp388 and Trp484 via a solvent molecule. The group of charged residues Asp388, Asp486, Glu434, and Lys338, surrounding Arg437, form a dense network at the surface that could be important for processing events.

In summary, the NL4 structure and its complex with Nrxβ1 reveal several key features for these adhesion molecules. First, the neurexin interaction surface of the neuroligin dimer is on the side opposite that for substrate access in the catalytically active neuroligin relatives. This demonstrates distinctive functional modules for cell adhesion and substrate binding in the α/β-hydrolase fold core. Second, a central pocket containing a vestigial active center is present in NL4, but its occlusion by the Cys-loop would preclude efficient substrate entry. Third, the loops that delineate the entry to the active center in the hydrolases differ greatly in size and positional conformation in NL4, where larger loops compensate for smaller neighboring loops and vice versa. The precise functional role of these loops in neuroligin remains to be investigated. Fourth, the Cys-loop, although tethered to the molecular core, displays conformational flexibility. This feature, along with the presence of entrapped solvent and bound phosphate in the central pocket, argues for potential motion of the Cys-loop in solution. The looser connection between the molecular core and four-helix bundle provides additional flexibility to the neuroligin dimer. Molecular plasticity may be a hallmark for the adhesion function. Fifth, neuroligin displays negatively charged surfaces on both its neurexin binding and Cys-loop faces. In contrast to AChE, a large electrostatic dipole in the neuroligins, which function as surface adhesion molecules, is not required to draw a substrate into the vestigial active center. Rather, absence of a large, permanent dipole for the neuroligin, along with the moderate affinity of the complex, may permit fluctuating interactions, which is a desirable property for adhesion function. All in all, the α/β-hydrolase fold, with its capacity for turning a dense core with a buried active center tailored for trapping and hydrolyzing a substrate into a flexible adhesion molecule with surface recognition properties, gives rise to surprising diversity in both molecular form and function.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

Recombinant NL4, encompassing residues Gln44 to Thr619 of full-length human NL4 (Bolliger et al., 2001) (Figure 1) and N-terminally flanked with a FLAG epitope, numbered (-12)DYKDDDDKLAAA(-1), was expressed at high yields (~2 mg/l) as a soluble exported protein from stable HEK293 cells (Comoletti et al., 2006). It was purified from the culture medium on immobilized anti-FLAG M2 antibody (Sigma-Aldrich) (Comoletti et al., 2003). Further purification was by gel-filtration FPLC on prepacked Superdex-200 (GE Healthcare) equilibrated and eluted with 10 mM Na-HEPES (pH 7.4), 50 mM NaCl, and 0.01% NaN_3 (buffer A). Quantification of NL4 by UV spectrophotometry used an extinction coefficient of 102 680 $M^{-1}cm^{-1}$ at 280 nm, calculated from the protein sequence, and a molecular mass of 72,663 Da, determined by MALDI-TOF spectrometry. Electrophoretic analyses used a PhastSystem apparatus and related gels and buffers (GE Healthcare). Purified NL4 was concentrated by ultrafiltration, filtered on sterile 0.22 μm membranes, and stored on ice.

Soluble recombinant Nrxβ1, encompassing residues Gly81-Val288 of the LNS domain of rat β-neurexin-1 (Ushkaryov et al., 1994) and N-terminally flanked with a 6xHis tag followed by a TEV cleavage site, was obtained by subcloning our initial pGEX construct (Comoletti et al., 2004) into pDEST17 (Invitrogen). This Nrxβ1 does not comprise those N- and C-terminal stretches found to be disordered in the Nrxβ1 structure (Rudenko et al., 1999). Nrxβ1 was expressed at high yields (~70 mg/l) from Rosetta PlyS cells, and purified from the culture medium by nickel affinity chromatography on a HisTrap FF Crude cartridge (GE Healthcare). Removal of the 6xHis tag used a 6xHis-tagged TEV protease (InVitrogen) in a 1:25 TEV-to-Nrxβ1 weight ratio

Neuron
Structures of Neuroligin and Its Neurexin Complex

(overnight incubation, 20°C), and a second nickel affinity chromatography step. A molecular mass of 19,312 Da consistent with efficient tag removal was found by MALDI-TOF spectrometry. Final purification was by gel-filtration FPLC on Superdex 200 in buffer A. The Nrxβ1-NL4 complex was formed in solution using a 1.2 molar excess of Nrxβ1 over NL4 in 10 mM HEPES (pH 7.4), 25 mM NaCl, and 2 mM $CaCl_2$ (overnight incubation, 4°C). The buffer composition and final complex concentration were adjusted by ultrafiltration.

Crystallization, Data Collection, and Processing

Initial crystallization conditions were found from the PEG (Quiagen) and SM1 (Nextalbiotech) screens using vapor diffusion in 96-well Greiner plates and the Cartesian PixSys 4200 (Genomic Solutions) robot. Optimized conditions for crystallization of NL4 used hanging drops of 1.2 μl and a protein-to-well solution ratio of 1:1, with 10% PEG-3000 in 100 mM Na-phosphate-citrate (pH 4.2), 300 mM NaCl, and 0 mM (20°C; 3.2 Å resolution data set) or 10 mM (4°C; 2.2 Å resolution data set) $CaCl_2$ as the well solution. Crystallization of the complex at 20°C used sitting drops and 8% PEG-20000 in 100 mM Mes (pH 6.5) as the well solution. Crystals were briefly transferred to the mother liquor supplemented with 22%–25% glycerol and flash-cooled in liquid nitrogen. Space groups and resolution limits are reported in Table 1. Data were processed, merged, and scaled with XDS (Kabsch, 1993).

Structure Solution and Refinement

An initial 3.2 Å resolution NL4 structure was solved by molecular replacement using, as a search model, one mAChE subunit (PDB accession code 1J06) (Bourne et al., 2003) and MOLREP from the CCP4 Program Suite (CCP4, 1994). The resolution achieved was sufficient to unambiguously position a four-helix bundled dimer of NL4 molecules within the asymmetric unit and build most of the backbone and side chains into the electron density using COOT (Emsley and Cowtan, 2004). This structure without further refinement was used as a template to solve the 2.2 Å structure. The molecular replacement solution again positioned an NL4 dimer within the asymmetric unit. However, after ~90% of the protein model was built automatically with ARP/wARP (Perrakis et al., 1999), the region surrounding the Cys-loop (Cys110-Cys146) (Figure 1) appeared to be highly disordered. The structure was manually corrected using COOT, and structure refinement including TLS refinement was performed with Refmac in CCP4. TLS groups were generated using the TLS Motion Determination server (Painter and Merritt, 2006) and were manually adjusted to five groups per subunit corresponding to residues Ala(-1)-Cys110/Gln142-Tyr291/Phe340-Gln373/Trp449-Phe472/Lys561-Tyr582, Pro111-Asp141, Gln292-Ala339, Gly374-Gln448/Arg583-His598, and Tyr473-Ser560. Data collection and refinement statistics are summarized in Table 1.

The final NL4 structure comprises FLAG residues Asp(-8)-Ala(-1) and NL4 residues Gln44-Glu156, Ser164-Val540, and Glu556-Leu597 in subunit A of the dimer, and residues Ala(-2)-Ala(-1), Gln44-Ile159, Ser164-Pro541, and Val557-Leu597 in subunit B, out of the total 588 residues per subunit of the crystallized protein (Figure 1). GlcNAc moieties linked to Asn102 and Asn511; a phosphate (arising from the crystallization liquor) trapped in the central pocket; a citrate and a glycerol molecule (from the crystallization liquor and flash-cooling solution, respectively) trapped in solvent-accessible cavities at the protein surface; and three chloride and one sodium (from the crystallization liquor) are clearly identified. Despite the presence of $CaCl_2$ in the crystallization conditions, no bound Ca^{2+} is observed, consistent with the ion being mainly coordinated by neurexin. The average root mean square deviation (rmsd) between subunits A and B is 0.5 Å for 518 Cα atoms. Missing electron densities correspond to surface loops Asp157-Asn163, which in NL1 contains splice site A, and Pro541/Gln542-Val557/Ala558 and the C-terminal, ~20-residue peptide that prolongs helix α_{10}. The structure stereochemistry was analyzed with PROCHECK (Laskowski et al., 1993); no residues were found in the disallowed regions of the Ramachandran plot.

The structure of the Nrxβ1-NL4 complex was solved by molecular replacement using MOLREP and, as templates, the NL4 dimer structure and the Nrxβ1 structure (1C4R) (Rudenko et al., 1999). The model was refined with Refmac, including medium NCS restrains for the two NL4 molecules (excluding the Cys-loop) and the two Nrxβ1 molecules, and TLS refinement. The resulting electron density maps were used, when well-defined, to correct the backbone traces along the NL4 and Nrxβ1 molecules and position side chains and Ca^{2+} at the complex interface using COOT. An X-ray fluorescence emission spectrum confirmed the presence of Ca^{2+} in the frozen Nrxβ1-NL4 crystal. TLS groups were the same as for NL4, with an additional group for the entire Nrxβ1 molecule. NL4 residues Asp158-Asn163 in one subunit were built manually using COOT. Due to the limited resolution achieved, the remainder of the model was not refined further. The primary Nrxβ1-NL4 complex structure comprises NL4 residues Asp(-8)/Leu(-4)-Leu597 and Nrxβ1 residues His82-Val288.

Structure Analysis and Comparison with Other Structures

The volume of the NL4 central pocket was calculated using the CASTp server (Liang et al., 1998). Primary theoretical models of the other neuroligins and of neurotactin, gliotactin, and glutactin were designed using the mAChE structure as a template and the SWISS-MODEL server (Kopp and Schwede, 2006), and used without correction. The electrostatic surface potentials were calculated using APBS (Baker et al., 2001) with the PyMOL APBS tools. Dipole moments were calculated from the Protein Dipole Moments Server (Felder et al., 2007). The volume of the internal pocket was calculated with VOIDOO (Kleywegt and Jones, 1994).

Sequence- and structure-based searches for NL4's closest catalytically active relatives used PSI-BLAST and SSM (Krissinel and Henrick, 2004), respectively. Highest structural hits were for mAChE, human liver caboxylesterase (2H7C), bovine bile salt-stimulated lipase (2BCE) and bacterial *p*-nitrobenzyl esterase (1C7J). Superimposition of the NL4 structure with each of the four enzymes, performed with LSQMAN (Kleywegt and Jones, 1997), resulted in an average rmsd in the 1.1–1.4 Å range for at least 380 Cα atoms.

Figure 1 and Figure S1 were generated with ESPript (Gouet et al., 1999) and MUSCLE (Edgar, 2004), and Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, and Figure S2 were generated with PyMOL (DeLano, 2002).

Supplemental Data

The Supplemental Data for this article can be found online at http://www.neuron.org/cgi/content/full/56/6/979/DC1/.

ACKNOWLEDGMENTS

We are grateful to Jennifer Wilson (UCSD), Sandrine Conrod (CNRS FRE-2738), and the beamline staff of the European Synchrotron Research Facility (ESRF, Grenoble, France), for assistance in cell culture and protein expression, protein purification and crystallization, and data collection, respectively. This work was supported by the SPINE2-Complexes Consortium (Y.B., P.M., and a postdoctoral grant to P.L.); the CNRS DREI-SDV (P.M. and Y.B.); the CNRS and Fondation pour la Recherche Médicale (postdoctoral grants to I.F.); the Cure Autism Now Pilot Study (D.C.); and NIH grants R37-GM 18360 and UO-1 ES 10337 (P.T.) and GM07752 (M.M.).

Received: October 17, 2007
Revised: November 19, 2007
Accepted: November 26, 2007
Published: December 19, 2007

REFERENCES

Auld, V.J., Fetter, R.D., Broadie, K., and Goodman, C.S. (1995). Gliotactin, a novel transmembrane protein on peripheral glia, is required to form the blood-nerve barrier in Drosophila. Cell *81*, 757–767.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Blasi, F., Bacchelli, E., Pesaresi, G., Carone, S., Bailey, A.J., and Maestrini, E. (2006). Absence of coding mutations in the X-linked genes neuroligin 3 and neuroligin 4 in individuals with autism from the IMGSAC collection. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *141*, 220–221.

Bolliger, M.F., Frei, K., Winterhalter, K.H., and Gloor, S.M. (2001). Identification of a novel neuroligin in humans which binds to PSD-95 and has a widespread expression. Biochem. J. *356*, 581–588.

Botti, S.A., Felder, C.E., Sussman, J.L., and Silman, I. (1998). Electrotactins: a class of adhesion proteins with conserved electrostatic and structural motifs. Protein Eng. *11*, 415–420.

Boucard, A.A., Chubykin, A.A., Comoletti, D., Taylor, P., and Sudhof, T.C. (2005). A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. Neuron *48*, 229–236.

Bourne, Y., Taylor, P., and Marchot, P. (1995). Acetylcholinesterase inhibition by fasciculin: crystal structure of the complex. Cell *83*, 503–512.

Bourne, Y., Taylor, P., Radic, Z., and Marchot, P. (2003). Structural insights into ligand interactions at the acetylcholinesterase peripheral anionic site. EMBO J. *22*, 1–12.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Comoletti, D., Flynn, R., Jennings, L.L., Chubykin, A., Matsumura, T., Hasegawa, H., Sudhof, T.C., and Taylor, P. (2003). Characterization of the interaction of a recombinant soluble neuroligin-1 with neurexin-1beta. J. Biol. Chem. *278*, 50497–50505.

Comoletti, D., De Jaco, A., Jennings, L.L., Flynn, R.E., Gaietta, G., Tsigelny, I., Ellisman, M.H., and Taylor, P. (2004). The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing. J. Neurosci. *24*, 4889–4893.

Comoletti, D., Flynn, R.E., Boucard, A.A., Demeler, B., Schirf, V., Shi, J., Jennings, L.L., Newlin, H.R., Sudhof, T.C., and Taylor, P. (2006). Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for beta-neurexins. Biochemistry *45*, 12816–12827.

Comoletti, D., Grishaev, A., Whitten, A.E., Tsigelny, I., Taylor, P., and Trewhella, J. (2007). Synaptic arrangement of the neuroligin/beta-neurexin complex revealed by X-ray and neutron scattering. Structure *15*, 693–705.

Cygler, M., Schrag, J.D., Sussman, J.L., Harel, M., Silman, I., Gentry, M.K., and Doctor, B.P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. Protein Sci. *2*, 366–382.

De Jaco, A., Comoletti, D., Kovarik, Z., Gaietta, G., Radic, Z., Lockridge, O., Ellisman, M.H., and Taylor, P. (2006). A mutation linked with autism reveals a common mechanism of endoplasmic reticulum retention for the alpha,beta-hydrolase fold protein family. J. Biol. Chem. *281*, 9667–9676.

de la Escalera, S., Bockamp, E.O., Moya, F., Piovant, M., and Jimenez, F. (1990). Characterization and gene cloning of neurotactin, a Drosophila transmembrane protein related to cholinesterases. EMBO J. *9*, 3593–3601.

Dean, C., Scholl, F.G., Choih, J., DeMaria, S., Berger, J., Isacoff, E., and Scheiffele, P. (2003). Neurexin mediates the assembly of presynaptic terminals. Nat. Neurosci. *6*, 708–716.

DeLano, W. (2002). The PyMOL molecular graphics system (http://www.pymol.org).

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Felder, C.E., Prilusky, J., Silman, I., and Sussman, J.L. (2007). A server and database for dipole moments of proteins. Nucleic Acids Res. *35*, W512–521.

Gibney, G., Camp, S., Dionne, M., MacPhee-Quigley, K., and Taylor, P. (1990). Mutagenesis of essential functional residues in acetylcholinesterase. Proc. Natl. Acad. Sci. USA *87*, 7546–7550.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308.

Graf, E.R., Kang, Y., Hauner, A.M., and Craig, A.M. (2006). Structure function and splice site analysis of the synaptogenic activity of the neurexin-1 beta LNS domain. J. Neurosci. *26*, 4256–4265.

Grochulski, P., Li, Y., Schrag, J.D., and Cygler, M. (1994). Two conformational states of Candida rugosa lipase. Protein Sci. *3*, 82–91.

Hoffman, R.C., Jennings, L.L., Tsigelny, I., Comoletti, D., Flynn, R.E., Sudhof, T.C., and Taylor, P. (2004). Structural characterization of recombinant soluble rat neuroligin 1: mapping of secondary structure and glycosylation by mass spectrometry. Biochemistry *43*, 1496–1506.

Ichtchenko, K., Nguyen, T., and Sudhof, T.C. (1996). Structures, alternative splicing, and neurexin binding of multiple neuroligins. J. Biol. Chem. *271*, 2676–2682.

Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., and Bourgeron, T. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat. Genet. *34*, 27–29.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. *26*, 795–800.

Kleywegt, G.J., and Jones, T.A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr. D Biol. Crystallogr. *50*, 178–185.

Kleywegt, G.J., and Jones, T.A. (1997). Detecting folding motifs and similarities in protein structures. Methods Enzymol. *277*, 525–545.

Kopp, J., and Schwede, T. (2006). The SWISS-MODEL Repository: new features and functionalities. Nucleic Acids Res. *34*, D315–D318.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.

Lardi-Studler, B., and Fritschy, J.M. (2007). Matching of pre- and postsynaptic specializations during synaptogenesis. Neuroscientist *13*, 115–126.

Laskowski, R.A., MacArthur, M., Moss, D., and Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. *26*, 283–291.

Laumonnier, F., Bonnet-Brilhault, F., Gomot, M., Blanc, R., David, A., Moizard, M.P., Raynaud, M., Ronce, N., Lemonnier, E., Calvas, P., et al. (2004). X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. Am. J. Hum. Genet. *74*, 552–557.

Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P.V., and Subramaniam, S. (1998). Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. Proteins *33*, 18–29.

Marchot, P., Ravelli, R.B., Raves, M.L., Bourne, Y., Vellom, D.C., Kanter, J., Camp, S., Sussman, J.L., and Taylor, P. (1996). Soluble monomeric acetylcholinesterase from mouse: expression, purification, and crystallization in complex with fasciculin. Protein Sci. *5*, 672–679.

Nguyen, T., and Sudhof, T.C. (1997). Binding properties of neuroligin 1 and neurexin 1beta reveal function as heterophilic cell adhesion molecules. J. Biol. Chem. *272*, 26032–26039.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., et al. (1992). The alpha/beta hydrolase fold. Protein Eng. *5*, 197–211.

Olson, P.F., Fessler, L.I., Nelson, R.E., Sterne, R.E., Campbell, A.G., and Fessler, J.H. (1990). Glutactin, a novel Drosophila basement membrane-related glycoprotein with sequence similarity to serine esterases. EMBO J. *9*, 1219–1227.

Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D Biol. Crystallogr. *62*, 439–450.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Ripoll, D.R., Faerman, C.H., Axelsen, P.H., Silman, I., and Sussman, J.L. (1993). An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. Proc. Natl. Acad. Sci. USA *90*, 5128–5132.

Rudenko, G., Nguyen, T., Chelliah, Y., Sudhof, T.C., and Deisenhofer, J. (1999). The structure of the ligand-binding domain of neurexin Ibeta: regulation of LNS domain function by alternative splicing. Cell *99*, 93–101.

Scheiffele, P., Fan, J., Choih, J., Fetter, R., and Serafini, T. (2000). Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. Cell *101*, 657–669.

Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S.S., Friedmann, T., and Taylor, P. (1986). Primary structure of Torpedo californica acetylcholinesterase deduced from its cDNA sequence. Nature *319*, 407–409.

Sheckler, L.R., Henry, L., Sugita, S., Sudhof, T.C., and Rudenko, G. (2006). Crystal structure of the second LNS/LG domain from neurexin 1alpha: Ca2+ binding and the effects of alternative splicing. J. Biol. Chem. *281*, 22896–22905.

Shi, J., Boyd, A.E., Radic, Z., and Taylor, P. (2001). Reversibly bound and covalently attached ligands induce conformational changes in the omega loop, Cys69-Cys96, of mouse acetylcholinesterase. J. Biol. Chem. *276*, 42196–42204.

Spiller, B., Gershenson, A., Arnold, F.H., and Stevens, R.C. (1999). A structural view of evolutionary divergence. Proc. Natl. Acad. Sci. USA *96*, 12305–12310.

Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., and Silman, I. (1991). Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. Science *253*, 872–879.

Tabuchi, K., Blundell, J., Etherton, M.R., Hammer, R.E., Liu, X., Powell, C.M., and Sudhof, T.C. (2007). A Neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. Science *318*, 71–76.

Talebizadeh, Z., Lam, D.Y., Theodoro, M.F., Bittel, D.C., Lushington, G.H., and Butler, M.G. (2006). Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. J. Med. Genet. *43*, e21–e28.

Ushkaryov, Y.A., Hata, Y., Ichtchenko, K., Moomaw, C., Afendis, S., Slaughter, C.A., and Sudhof, T.C. (1994). Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences in receptor-like neuronal cell-surface proteins. J. Biol. Chem. *269*, 11987–11992.

van Tilbeurgh, H., Sarda, L., Verger, R., and Cambillau, C. (1992). Structure of the pancreatic lipase-procolipase complex. Nature *359*, 159–162.

Varoqueaux, F., Aramuni, G., Rawson, R.L., Mohrmann, R., Missler, M., Gottmann, K., Zhang, W., Sudhof, T.C., and Brose, N. (2006). Neuroligins determine synapse maturation and function. Neuron *51*, 741–754.

Waites, C.L., Craig, A.M., and Garner, C.C. (2005). Mechanisms of vertebrate synaptogenesis. Annu. Rev. Neurosci. *28*, 251–274.

Yan, J., Oliveira, G., Coutinho, A., Yang, C., Feng, J., Katz, C., Sram, J., Bockholt, A., Jones, I.R., Craddock, N., et al. (2005). Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. Mol. Psychiatry *10*, 329–332.

Yen, T., Nightingale, B.N., Burns, J.C., Sullivan, D.R., and Stewart, P.M. (2003). Butyrylcholinesterase (BCHE) genotyping for post-succinylcholine apnea in an Australian population. Clin. Chem. *49*, 1297–1308.

Accession Numbers

Atomic coordinates and structure factors have been deposited with the Protein Data Bank; accession codes are 3BE8 (NL4) and 2VH8 (Nrxβ1-NL4 complex).

ACKNOWLEDGMENTS

Chapter 2 is a reprint of the material as it appears in Fabrichny, I.P., Leone, P., Sulzenbacher, G., Comoletti, D. **Miller, M.T.,** Taylor, P., Bourne, Y., Marchot, P. (2007) *Neuron* **56**, 979-991. This dissertation author contributed to the conceptual design of the study and worked on the protein purification and crystallization experiments in Palmer Taylor's lab.

CHAPTER 3:

Structure
Article

The Macromolecular Architecture of Extracellular Domain of αNRXN1: Domain Organization, Flexibility, and Insights into *Trans*-Synaptic Disposition

Davide Comoletti,[1,6,*] Meghan T. Miller,[1,6] Cy M. Jeffries,[3] Jennifer Wilson,[1] Borries Demeler,[4] Palmer Taylor,[1] Jill Trewhella,[3,5] and Terunaga Nakagawa[2,*]
[1]Department of Pharmacology, Skaggs School of Pharmacy and Pharmaceutical Sciences
[2]Department of Chemistry and Biochemistry
University of California, San Diego, La Jolla, CA 92093, USA
[3]School of Molecular Bioscience, University of Sydney, Sydney, NSW 2006, Australia
[4]Department of Biochemistry, The University of Texas Health Science Center, San Antonio TX 78229, USA
[5]Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA
[6]These authors contributed equally to this work
*Correspondence: dcomolet@ucsd.edu (D.C.), nakagawa@ucsd.edu (T.N.)
DOI 10.1016/j.str.2010.06.005

SUMMARY

Neurexins are multidomain synaptic cell-adhesion proteins that associate with multiple partnering proteins. Genetic evidence indicates that neurexins may contribute to autism, schizophrenia, and nicotine dependence. Using analytical ultracentrifugation, single-particle electron microscopy, and solution X-ray scattering, we obtained a three-dimensional structural model of the entire extracellular domain of neurexin-1α. This protein adopts a dimensionally asymmetric conformation that is monomeric in solution, with a maximum dimension of ~170 Å. The extracellular domain of α-neurexin maintains a characteristic "Y" shape, whereby LNS domains 1–4 form an extended base of the "Y" and LNS5-6 the shorter arms. Moreover, two major regions of flexibility are present: one between EGF1 and LNS2, corresponding to splice site 1, another between LNS5 and 6. We thus provide the first structural insights into the architecture of the extracellular region of neurexin-1α, show how the protein may fit in the synaptic cleft, and how partnering proteins could bind simultaneously.

INTRODUCTION

Genes encoding neuronal cell-adhesion proteins are essential for development and maintenance of connectivity in the nervous system. Synaptic cell-adhesion proteins constitute a principal pathway contributing to genetic susceptibility of autism spectrum disorders (ASD) (Geschwind and Levitt, 2007). Emerging evidence indicates that variations in copy number and other rare variants within the genes encoding neurexin-1 and -3 (*NRXN1* and *NRXN3*) contribute to ASD susceptibility and mental retardation (Feng et al., 2006; The Autism Genome Project Consortium, 2007; Kim et al., 2008; Yan et al., 2008; Glessner et al., 2009; Zahir et al., 2008; Zweier et al., 2009).

Neurexin-1α (αNRXN1) is a neuronal cell surface receptor that was originally identified as a high-affinity receptor for the spider toxin α-latrotoxin, whereas NRXN2 and 3 were subsequently identified from DNA sequence similarity with αNRXN1. Within each *NRXN* gene, the presence of two promoters, α and β, enables the expression of a longer α and a shorter βNRXN, yielding a total of six NRXN proteins (Missler and Südhof, 1998). Extensive independent alternative splicing of the encoded proteins (Ullrich et al., 1995) could specify a code of interactions between NRXNs and their ligands in different classes of synapses. Moreover, alternative splicing in NRXN3 creates a large diversity of secreted gene products, including those encoding multiple variants with in-frame stop codons (Ushkaryov and Südhof, 1993; Ullrich et al., 1995). Similar to the construct used in this study, all secreted splice variants end between the sixth Laminin, Neurexin, Sex-hormone-binding globulin (LNS) domain and various positions before the beginning of the transmembrane domain.

Currently, four groups of endogenous ligands for α and βNRXNs have been identified: neuroligins (NLGN) (Ichtchenko et al., 1995), neurexophilins (Missler et al., 1998), dystroglycan (Sugita et al., 2001), and leucine-rich repeat transmembrane proteins (LRRTM2) (de Wit et al., 2009; Ko et al., 2009). NRXNs and NLGNs are thought to form a *trans*-synaptic complex meeting near the center of the synaptic cleft, with the C-terminal sequences of either protein extending in opposite directions, tethering them to the pre- and postsynaptic membranes, respectively (Comoletti et al., 2007; Fabrichny et al., 2007; Araç et al., 2007; Chen et al., 2008). Early studies of cell association with neurons in cell culture suggested that NLGNs and NRXNs are sufficient to induce formation of new synapses (Scheiffele et al., 2000; Graf et al., 2004). In particular, all three αNRXNs induce clustering of the GABAergic postsynaptic scaffolding protein gephyrin and NLGN2, but not of the glutamatergic postsynaptic scaffolding protein PSD-95 or NLGN 1/3/4 in an artificial synapse-formation assay (Kang et al., 2008). This suggests that αNRXNs may be mediators of GABAergic synaptic protein

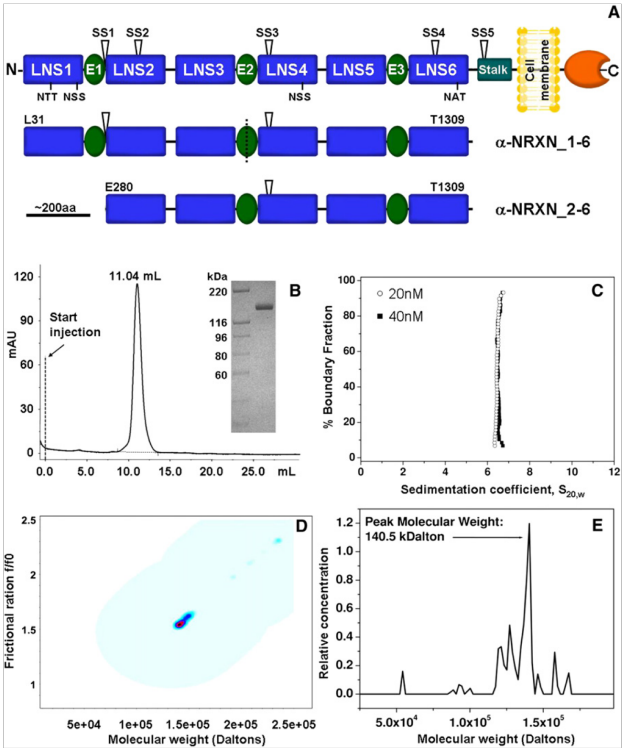

Figure 1. Schematic Diagram of the αNRXN1 Constructs Used and Hydrodynamic Characterization of the Purified Extracellular Domain of αNRXN_1-6

(A) Top: the domain organization of full-length αNRXN1 with respect to the pre-synaptic membrane. The C-terminal intracellular domain is shown to the right of the membrane, whereas the large N-terminal region extending into the synaptic cleft is shown on the left. Stalk, O-lined glycosylated domain; E1 to E3, EGF domains. Potential N-linked glycosylation sites are indicated using single letter code; open arrowheads , sites alternative splicing. Bottom: the extracellular regions of the αNRXN1 used for this study. L31, E280, and T1309 refer to the amino acid number in the protein sequence. The dotted line indicates a potential center of symmetry.

(B) Size exclusion chromatography trace of purified αNRXN_1-6. Inset, Coomassie blue staining of a sample composed of the main peak.

(C) G(s) distribution plots from the enhanced van Holde–Weischet analysis of sedimentation velocity experiments of two concentrations of αNRXN_1-6.

(D) Two-dimensional spectrum/Monte Carlo analysis of the 20 nM velocity experiment. The distribution shows a single, monomeric species with a molecular weight of ~140 kDa. The blue gradient indicates partial concentration.

(E) Concentration histogram of the globally fitted fixed molecular weight distribution analysis of the sedimentation equilibrium data. See also Figure S1.

recruitment and stabilization. Earlier studies of αNRXN-knockout mice revealed only mild variations in synaptic density and ultrastructure (Dudanova et al., 2007). More recently, a study of the αNRXN-1 knockout showed severe impairment of excitatory neurotransmission pathways (Etherton et al., 2009). αNRXNs appear to play a functional role at the synapse, including mediating Ca^{2+}-triggered neurotransmitter release (Missler et al., 2003), but do not seem to participate in synapse formation. These findings suggest that the αNRXNs are a group of *trans*-synaptic cell-adhesion molecules that participate in a modular organization of presynaptic terminals by mediating the localized activation of Ca^{2+} channels.

Structurally, αNRXN1 is a large (~160 kDa) multidomain protein composed of several discernable regions. A cleavable N-terminal signal peptide is responsible for trafficking the protein to the cell membrane. The mature protein contains three homologous repeats, each motif composed of a central epidermal growth factor (EGF) domain flanked upstream and downstream by two LNS domains sharing limited protein identity (Figure 1A). These three repeats, formed by nine independently folded domains that span ~90% of the protein sequence, are followed by a single stalk domain that is likely to be partially rigidified through extensive O-linked glycosylation. The stalk domain connects to a single transmembrane domain and a short cytoplasmic tail containing a classical PDZ recognition motif that appears to target NRXN to the presynaptic region (Fairless et al., 2008). Although the crystal structures of the isolated second, fourth, and sixth LNS domains of αNRXN1 have been solved (Rudenko et al., 1999; Sheckler et al., 2006; Shen et al., 2008), the overall domain complexity of the intact protein has been an impediment to examining structural organization beyond individual LNS domains.

Using a set of complementary biophysical techniques, we developed a three-dimensional structural model of the entire extracellular domain of αNRXN1 in solution. Although some flexibility at the extremities of the extracellular domain is detected, the overall architecture of αNRXN1 is consistent with a semielongated protein with a stable shape resembling the letter "Y." The results reported here represent, to our knowledge, the first three-dimensional structural models of the extracellular domain of αNRXN super-family that include neurexin (1 to 3) and Caspr (Contactin associated-like protein) 1 to 5. Together, these results should facilitate the understanding of how αNRXN might be arranged in the limited space of the synaptic cleft and how this protein may associate with multiple transmembrane and soluble synaptic proteins.

RESULTS

Characterization of the Purified Extracellular Domain of αNRXN

Protein expression and N-terminal sequencing

Two constructs from the extracellular domain of αNRXN1 were expressed as soluble entities in the cell culture medium of HEK293 GnTI- cells: one starting at position Leu31 and encompassing sites of alternative splicing #1 and #3 (αNRXN_1-6) and an N-terminal deletion devoid of the first LNS and EGF domains to yield a protein starting at position Glu280 (αNRXN_2-6) (Figure 1A). By size exclusion chromatography, both constructs elute as single peaks (see Figure 1B; see Figure S1A available online), indicating the presence of homogeneous monomeric species. In SDS-PAGE followed by Coomassie blue staining, αNRXN_1-6 and αNRXN_2-6 appear as single bands of ~140 kDa and ~110 kDa (insets of Figure 1B; Figure S1A), consistent with the calculated molecular weight of the peptides. As we express αNRXN protein with its native leader peptide (mgtallqrggcfllclslllllgcwaelgsgLEFPG), Edman degradation of the first five residues of the mature protein was performed to assign the initial amino acids after the cleavage of the leader sequence. The unambiguously determined sequence Leu-Glu-Phe-Pro-Gly indicated that the mature protein starts at Leu31, consistent with previous findings (Missler et al., 1998) and sequence predictions.

Mass spectrometric analysis of αNRXN

Both αNRXN1 constructs were expressed in the culture medium of HEK293 GnTI- cells. These cells lack N-acetylglucosaminyltransferase I (GnTI) activity, and consequently glycosylation remains restricted to a homogeneous seven-residue oligosaccharide (Reeves et al., 2002), thus simplifying structural analyses. αNRXN1 contains four potential N-linked glycosylation sites at positions N125, N190, N790, and N1223 (Figure 1). Although the peptidic mass of the expressed protein is calculated to be 140,619 Da, MALDI-TOF indicated a MW value of 145,896 Da (data not shown) with a difference of 5278 Da between the two values. Because GnTI-cells only add to each N-linked glycosylation site a Man5GlcNAc2 (mass, 1361 Da), the estimated occupancy of the potential N-linked sugars is 3.87 units per molecule, a value consistent with conjugation by oligosaccharde at all four N-linked glycosylation sites. Functionally, αNRXN_1-6 is fully active in binding neuroligin-1 (Figure S1B) (Boucard et al., 2005) and LRRTM2 (de Wit et al., 2009).

Analytical Ultracentrifugation Analyses of the Extracellular Domain of αNRXN1

Sedimentation velocity and equilibrium measurements provide complementary information useful in determining the globularity and oligomerization state of the extracellular domain of αNRXN1 in solution. To determine whether αNRXN_1-6 forms reversibly self-associating oligomers, we compared sedimentation coefficient distributions extracted using the enhanced van Holde–Weischet analysis (Demeler and van Holde, 2004) of two different loading concentrations. This analysis shows identical monodisperse species for both concentrations (Figure 1C), suggesting an absence of oligomerization at these concentrations. The same data were then analyzed by Monte Carlo analysis (Demeler and Brookes, 2008) together with a two-dimensional spectrum analysis (Brookes et al., 2010) and genetic algorithm analysis (Brookes and Demeler, 2007) to establish a molecular weight (Figure 1D). These analyses indicated that αNRXN_1-6 has a sedimentation coefficient of 6.36 s (6.28 s, 6.39 s) and a molecular weight of 141 kDa (136.5 kDa, 151.4 kDa), with a frictional ratio of 1.54 (1.51, 1.64), consistent with the monomeric mass of αNRXN and indicative of an elongated particle (values in parenthesis are 95% confidence intervals from the Monte Carlo analysis) (Demeler, 2009). The monomeric molecular weight was confirmed by sedimentation equilibrium data and a fixed molecular weight distribution analysis that gave a peak molecular weight of ~140 kDa (Figure 1E). Both sedimentation velocity and equilibrium experiments were in excellent agreement with the expected molecular weight based on amino acid sequence and mass spectrometry analyses.

Single Particle Electron Microscopy

αNRXN_1-6

To obtain structural information on the extracellular domain of αNRXN1, the purified protein was negative-stained and imaged using a transmission EM. αNRXN_1-6 particles were monodisperse and homogeneous in size, although individual particles adopted a variety of conformations (Figure 2A). Approximately 6,000 particles were analyzed using multivariate statistics, image classification, and averaging. In the majority of the class averages, only five globular domains were clearly visible. We presumed that only five of the six LNS domains were uniformly aligned because of extensive conformational heterogeneity intrinsic to the particles (Figure 2B). Analysis of the SDS-PAGE profiles (data not shown) indicated that the protein was intact, and careful inspection of the raw particle images indicate that all six domains were always detectable but appear faintly only in a few class averages (Figure 2C). We concluded that the missing domain was poorly averaged because of its extensive flexibility. These observations provide direct evidence that the LNS1-EGF1 tandem is extremely mobile, because of the flexible linkage of the LNS1 domain respect to the more rigid bulk of the molecule. Molecular labeling was conducted to deduce the N- to C-terminal orientation of the protein. We introduced a FLAG tag at the N terminus and an HA tag at the C terminus and labeled the purified protein with the antigen-binding fragment (Fab) against each of the epitope tags separately. The images of the HA Fab-tagged αNRXN_1-6 particles identified the C terminus (LNS6) in the more structured triangular region (Figure 2D), whereas the FLAG Fab-tagged αNRXN_1-6 identify first LNS domain in the more elongated and flexible region (Figure 2D). Overall, single-particle EM reveals that the extracellular domain of αNRXN adopts a semielongated and asymmetric structure with a shape reminiscent of the letter Y. Although this labeling does not allow a positive identification of each LNS domain, by their sequence in the protein, we infer that LNS domains 1–4 form the longer base of the ''Y'' and LNS5-6 the two arms.

αNRXN_2-6

To reduce the large conformational flexibility in the molecule conferred by the first LNS domain, we removed the N-terminal portion of the protein, which included the first LNS and EGF domains, and the flexible linker region containing the alternative

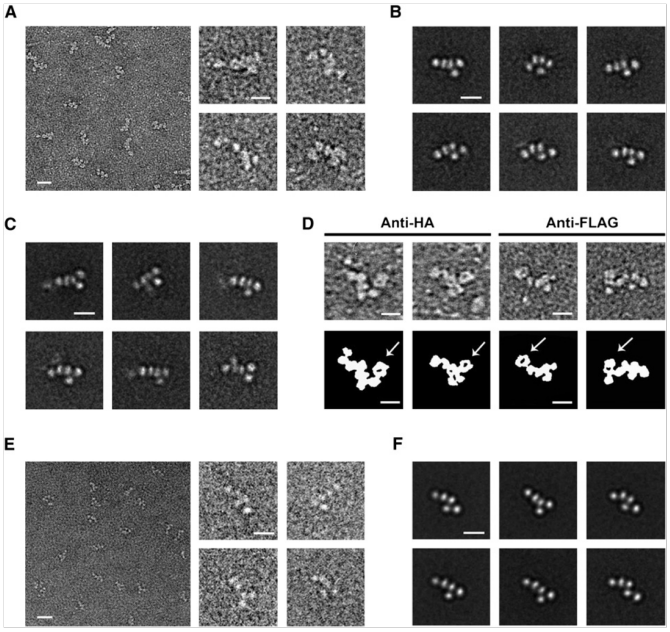

Figure 2. Single-Particle Electron Microscopy Characterization of αNRXN_1-6 and αNRXN_2-6

(A) Raw data images of the particles before alignment (left panel). Scale bar, 20 nm. Boxed particle, scale bar equals 10 nm.

(B) Selection of the highest represented class averages which show only five visibly distinct domains. The panels show the breadth of flexibility of the particles. Scale bar, 10 nm.

(C) Several averages with low representation indicate the presence of a sixth domain. Scale bar, 10 nm.

(D) Labeling of αNRXN 1-6 with Fab fragments against a C-terminal HA tag, and an N-terminal FLAG tag. Raw data images are above and a schematic representation with arrows pointing to the identified Fab fragment is below. Scale bar, 10 nm.

(E) Left panel, raw data of αNRXN_2-6 particles; 17,321 particles were aligned and grouped into 150 class averages. Scale bar, 20 nm.

(F) Six of the highest represented averages show various conformations. Scale bar, 10 nm. See also Figure S2.

splice site 1 (Figure 1A), generating a protein beginning at Glu280 (αNRXN_2-6). Approximately 17,000 particles of αNRXN_2-6 were analyzed, and 150 class averages were generated. The most common class averages clearly define all five LNS domains in αNRXN_2-6. By removing LNS1, our identification of the N and C termini was also confirmed: the particle alignment was greatly improved by the N-terminal truncation, and the structured triangular region was maintained (Figures 2E and 2F). The difference in particle shapes detectable in the class averages show that, although some flexibility remains at the junctions between LNS domains, the general Y shape is conserved. The class averages shown in Figure 2F depict the range of conformational variability. Although little variation is present at the N terminus of the protein, more flexibility is adopted at the C terminus where the Y shape can be disrupted by LNS6 being removed from LNS4. The complete set of class averages shown in Figure S2 yields a more comprehensive view of the conformational heterogeneity of the extracellular domain of αNRXN1. Probably because of their small size, densities corresponding to the three EGF domains remain unresolved. Overall, it appears that αNRXN1 has a tightly packed core composed of LNS2-4 and contains flexible regions that allow the extremities of the protein to extend and retract freely. From a biological perspective, this mobility could be relevant for allowing the domain to fit within the dimensions of the synaptic cleft and interact simultaneously with multiple binding partners. Having established the basic two-dimensional conformation of the extracellular domain of αNRXN1, we proceeded to reconstruct the three-dimensional structure using solution scattering methods.

Small Angle X-Ray Scattering Analysis of the Extracellular Domain of αNRXN1

Reproducible, high-quality scattering data for the entire extracellular domain of αNRXN_1-6 were collected from three independent sample preparations as well as from monodisperse solutions of lysozyme used as a secondary standard for

calibration of scattering intensity (Krigbaum and Kugler, 1970). Single-particle EM results show that the first LNS-EGF pair is significantly flexible. Therefore, to simplify our structural analyses and to strengthen our interpretation of the small angle X-ray scattering (SAXS) data, experiments on the truncated αNRXN_2-6 construct were also performed. The forward scattering intensities ($I(0)$) and radius of gyration (R_g) of each of the αNRXN_1-6 and αNRXN_2-6 samples were derived from the scattering data using Guinier analysis. As expected for monodisperse particles in solution, excellent linear correlations were observed in the Guinier plots for both constructs (Figure S3, top panels), and no significant concentration-dependent change in the R_g or $I(0)$ values were observed (Figure S3, lower panels). Using the known relationship $I(0)/c \propto$ MW (with c in units of mg/mL) and comparing data with that of lysozyme scattering, estimates of the molecular weight of the scattering particles were 164–174 kDa for αNRXN_1-6, and 109–118 kDa for αNRXN_2-6, consistent with hydrodynamic and mass spectrometric measurements (for detailed tabulated results of $I(0)$, R_g and molecular weight estimates, see Table S1).

Indirect Fourier transformation yields the probable interatomic distance distribution $P(r)$ within the scattering molecule, providing an estimate of the maximum dimension of the particle and its shape. The $P(r)$ profiles for αNRXN_1-6 and αNRXN_2-6 (Figure 3), calculated using the program GNOM (Svergun, 1992), indicate that both proteins (Figures 3B and D, respectively) are extended particles in solution (structurally anisotropic), as noted by the skewed distribution of vector lengths. The maximum dimension, D_{max}, of the full-length construct is ~170 Å with an average R_g of 53.0 ± 0.3 Å. As expected, the truncated variant is significantly smaller, with a D_{max} of ~145Å and an average R_g of 44.2 ± 0.6 Å. Removing the LNS-EGF pair shortens the maximum dimension of the protein by ~25 Å and decreases the radius of gyration of ~9 Å without altering the general shape of the $P(r)$ profile. Despite inherent segmental flexibility, the significantly shorter maximum dimension of the deletion mutant αNRXN_2-6 indicates that the domains within the LNS_2-6 region of the protein retain their extended configuration upon removal of the LNS1-EGF1 domain pair. Taken together, these data show that the extracellular domain of αNRXN is monomeric in solution and free of aggregation or interparticle interference. Thus, the scattering data fulfill the requirements necessary for extracting accurate shape information from which reliable three-dimensional structural models of αNRXN_1-6 and αNRXN_2-6 can be constructed.

Three-dimensional Reconstruction of the Extracellular Domain of αNRXN1

Using a combination of high-resolution structures and homology models of the individual LNS and EGF subunits, rigid-body modeling of the SAXS data enabled us to obtain independent three-dimensional structural models of αNRXN that closely resemble the shapes obtained with single-particle EM. The crystal structures of LNS 2, 4, and 6 are available (Rudenko et al., 1999; Sheckler et al., 2006; Shen et al., 2008), and homology models of the remaining individual subunits (LNS1, 3, and 5; EGF1, 2, and 3) were built using various high-resolution templates, as specified in Experimental Procedures. Eight sequences, ranging between 4 and 26 residues, linking various

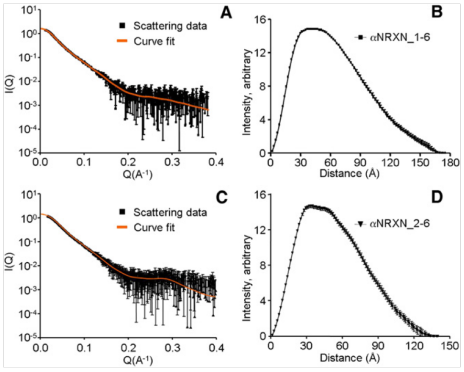

Figure 3. Scattering Intensity and *P(r)* Functions of the αNRXN_1-6 and αNRXN_2-6

(A and C) Scattering profiles of the highest concentrations of αNRXN_1-6 and 2-6 and $P(r)$ fits.

(B and D) $P(r)$ functions of αNRXN_1-6 and 2-6 proteins, indicating the maximum dimension of the particle. Statistical quality of the data in panels B and D can be assessed by the standard error bars; some estimated errors are smaller than the symbols. See also Table S1 and Figure S3.

LNS and EGF domains did not have a suitable three-dimensional template and thus were initially omitted from the calculations performed with the program SASREF (Petoukhov and Svergun, 2005). Distance constraints were imposed between the nine individual rigid bodies to ensure that the N and C termini of the individual domains remain within reasonable distances during refinement (details in Tables S2 and S3). The SASREF refinements were run multiple times against scattering intensity data from αNRXN_1-6 and αNRXN_2-6. The majority (~80%) of solutions converged toward a single class of Y-shaped molecule with LNS1/LNS2 forming the base of the Y and LNS5/LNS6 forming the arms (Figure 4), in excellent agreement with the EM images (Figure 2).

Approximately 11% of the mass of αNRXN_1-6 (primarily in the linkers connecting LNS1/EGF1 and EGF1/LNS2) was not included in the SASREF modeling, whereas only ~4% of the mass was missing for αNRXN_2-6. We therefore used the program BUNCH that, although similar to SASREF, can additionally account for contributions from regions of the model with unknown structure. Initial BUNCH refinements were performed using similar distance constrains as those used in SASREF against both αNRXN_1-6 and αNRXN_2-6 data sets and the fit to the data improved for both constructs (Figure 4A) (further details on BUNCH refinement strategy are available in the Supplemental Information). As expected, all of the refined BUNCH models maintain the characteristic Y-shape obtained by the other methods. In addition, they show the likely average positioning of the linkers of unknown structure.

Because of the inherent flexibility of the multidomain architecture, although we present the "best-fit" SASREF and BUNCH models for both constructs (Figure 4A), it is most pertinent to

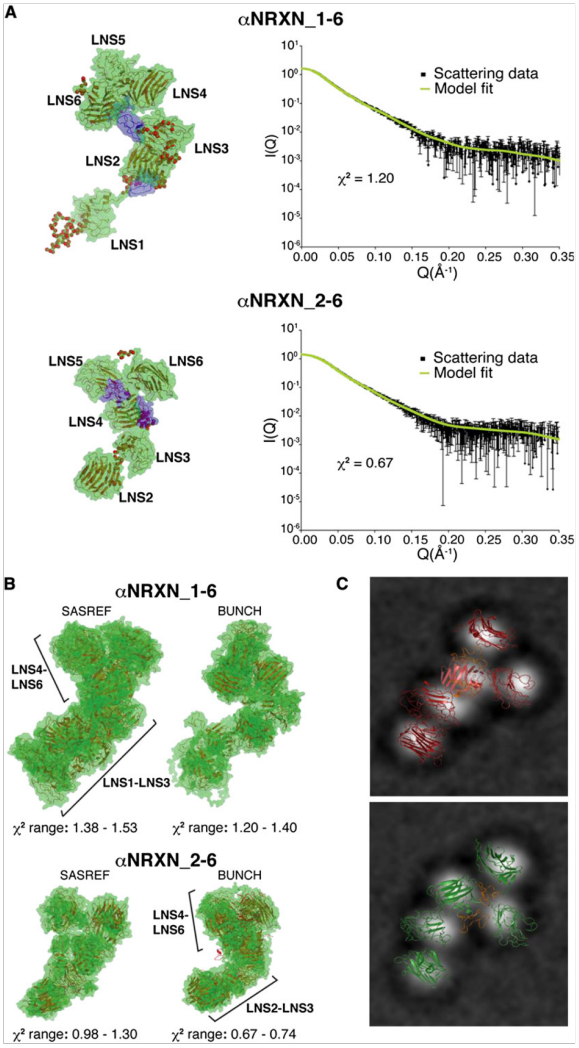

Figure 4. Rigid Body Modeling of the αNRXN_1-6 and 2-6 with the SAXS Data and αNRXN_2-6 SAXS Models Overlay on Selected EM particles

(A) Best-fit models of αNRXN_1-6 and αNRXN_2-6 derived from BUNCH rigid-body refinement and their respective fits to the data. The EGF domains are colored blue (EGF3 is occluded in αNRXN_1-6).

(B) Structural ensembles represented as semitransparent green surfaces and ribbons of both αNRXN constructs generated using SASREF and BUNCH rigid body modeling. Brackets indicate different domains of the ensemble. For clarity, brackets and domain labels have been omitted for some ensembles.

(C) Green and red are two different SASREF reconstructions manually superimposed to two similar class averages to show their degree of identity. Atomic models were made using PyMol (http://www.pymol.org). See also Tables S2 and S3.

view the results in terms of ensembles of structures that share a common Y-shaped topology (Figures 2 and 4B). For example, the arms of the Y (LNS5 or LNS6) can be spatially swapped without greatly affecting the general Y-shape of the model. Furthermore, upon comparing each member across the SASREF and BUNCH ensembles, the individual domains can undergo a limited localized tilting or rotation relative to their domain neighbors without affecting the fits to the data. Consequently, although the general architecture of the protein is maintained, the RMSD C_α across the ensembles is broad, ranging from ~7 to 25 Å. Although this variability could be due to inherent limitations in modeling a protein with multiple domains, the solution scattering results are consistent with the electron microscopy data that indicate individual domains can reorient via flexible interdomain linkers.

DISCUSSION

The small volumes and spanning dimensions of synaptic clefts are critical for rapidity and fidelity of synaptic transmission. With a typical span of ~24 nm between the pre- and postsynaptic membrane, it is unclear how large multi-domain proteins such αNRXN (~1400 amino acids in its entire extracellular domain), L1CAM (~1300 amino acids), and protocadherin (~1100 amino acids), along with other large synaptic receptors and channels, are structured to coexist within the cleft and maintain synaptic structure and function. Furthermore, the extracellular domains of these large proteins are generally composed of a sequential arrangement of several individually folded domains connected by flexible linkers, and they associate with multiple partnering proteins. Whether these proteins are completely flexible or have a restricted interdomain segmental motion is unknown. The αNRXNs are not only bulky multidomain molecules, but between the sixth LNS and transmembrane domains they contain a sequence of ~100 residues that is relatively rich in Ser and Thr and shown to be O-linked glycosylated (Ushkaryov et al., 1992). The presence of oligosaccharides, combined with a relative abundance of Pro residues, presumably rigidifies and elongates the peptide chain, as demonstrated in

neuroligin-1 and other cell-surface receptors (Li et al., 1996; Merry et al., 2003; Comoletti et al., 2007). The length of this single chain tether may be advantageous to extend the sixth LNS domain so that it can approach the center of the synaptic space, permitting association with postsynaptic proteins such as neuroligin or LRRTM2 (Comoletti et al., 2007; de Wit et al., 2009; Ko et al., 2009). The presence of the stalk domain, however, extends the overall length of αNRXN beyond what a semielongated amino acid sequence would predict.

Sedimentation velocity and equilibrium analyses unambiguously show that the extracellular domain of αNRXN is a semielongated monomer. Although analytical ultracentrifugation experiments employ low protein concentration (between 10 nM and 7 μM, equivalent to 0.0014 and 1 mg/mL, respectively) that could favor dissociation to the monomeric species, SAXS experiments were conducted at concentrations up to ~6 mg/mL (~40 μM). Under these conditions, higher order oligomers were not detected, indicating that the extracellular domain of αNRXN, similar to β-neurexin (Comoletti et al., 2006), does not self-associate. Thus, unless oligomerization occurs through the short intracellular domain, αNRXN is likely present as a monomer on the cell surface.

Single-particle EM shows that the LNS1-EGF1 pair of αNRXN_1-6 has extensive interdomain flexibility with respect to the rest of the protein. This flexibility is likely due to the length of the linker (27 amino acids, S256–Y282) that contains splice insert 1. Conversely, the excellent averaging of the particles for αNRXN_2-6 indicates that the rest of the molecule maintains a more stable conformation (Figure 2). The central region of the protein shows LNS2 to 4 in a linear arrangement, whereas LNS5 lies at various angles in relation to the previous domains. LNS6 normally folds back on the protein, making a triangular or Y shaped arrangement with LNS4, but displays a large degree of flexibility.

Building on the crystal structures of LNS domains 2, 4, and 6 (Rudenko et al., 1999; Sheckler et al., 2006; Shen et al., 2008) and from homology models for the remaining domains, we constructed three dimensional structures of the entire extracellular domain of αNRXN1, including the linker regions, and optimized them against X-ray solution scattering data. Consistent with hydrodynamic and electron microscopy findings, the three dimensional best-fit models show that αNRXN maintains a semielongated structure, with LNS domains 1–4 arranged linearly and LNS 4, 5, and 6 adopting a triangular "clover leaf" conformation (Figure 4). Scattering data record the time and rotationally averaged structural information from molecules in solution, rather than a "snap shot" of a single molecule on a grid as it is observed by EM. The rotational averaging inherent to the solution scattering experiment reduces the information content to one-dimension and thus, when interpreting three-dimensional models, some basic starting assumptions must be satisfied. First, on comparing the $P(r)$ profiles of αNRXN_1-6 and αNRXN_2-6 and from a direct overlay of the scattering data, a level of structural preservation must be maintained within the LNS2-LNS6 region of the protein that is not affected by the removal of the LNS1-EGF1 domain pair. Second, in the jelly roll fold of the LNS domain the metal-binding pocket, where neuroligins bind, is located at the rim of the β sheet sandwich opposite the N and C termini (Rudenko et al., 1999) that reside

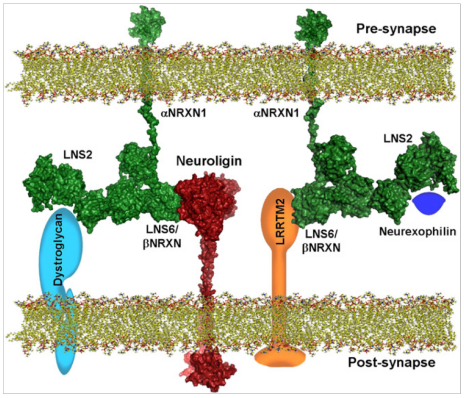

Figure 5. Schematic Model of the Complex Between αNRXN and its Ligands in the Context of the Synapse

The αNRXN stalk domain connects the presynaptic membrane to LNS6 of one of the structural models obtained with SASREF. Information on the contact surface between NLGN and LNS6 of αNRXN was taken from the available crystal structure. Stalk domains are drawn extended because of their likely semirigid structure. Intracellular domains of both NLGN and NRXN have no conformational assignment. Schematic models of the other known αNRXN ligands (dystroglycan, neurexophilin, and LRRTM2) are added to show how multiple ligands can associate simultaneously to their respective LNS domains. Structural models have depth cue visual information. Approximate distances and dimensions are to scale (pre- and postsynaptic gap is maintained at ~22 nm).

close together. Consequently when one LNS domain is connected to the next (or to an EGF domain) by relatively short ~5–7 amino acid linkers, as in the case of the LNS4-6, these domains will be constrained toward a "clover leaf" spatial arrangement as opposed to a linear "beads on a string" conformation (Tisi et al., 2000; Carafoli et al., 2009). Third, although the extracellular domain of αNRXN_1-6 could be considered symmetrical with EGF2 at the center of a two-fold symmetry (Figure 1A), the N and C termini identification with Fab tagging and the deletion construct αNRXN_2-6 yields a precise orientation of the protein. Together, the three-dimensional reconstructions and electron microscopy micrographs show extensive interaction between domains 2, 3, and 4, thus explaining the relative rigidity of this part of the molecule. Consistent with the higher degree of flexibility evident in the raw particle images in the EM micrographs, fewer contacts appear between LNS1 and 2 and between LNS5 and 6. Together, these results indicate that the protein maintains a stable core architecture that likely exists in an extracellular milieu and anchors its biological functions (Figure 5).

Proteins comprising a large number of independently folded domains, such as the αNRXN, laminin-G, and others are normally flexible because interdomain motions are likely linked to their biological activities. The high-resolution structure of LNS 1-3 of laminin α2 (Carafoli et al., 2009) shows that inter-LNS domain linkers maybe extended and flexible. In the case of αNRXN, multiple interacting proteins have been isolated and

characterized. In particular, neurexophilin appears to bind to the second LNS domain of αNRXN (Missler et al., 1998), whereas the neuroligins and LRRTM2 associate with the sixth LNS domain (Ichtchenko et al., 1995; Boucard et al., 2005; de Wit et al., 2009; Ko et al., 2009) and dystroglycan associates with both the second and the sixth LNS domains (Sugita et al., 2001). The mobility of LNS1 provides greater surface accessibility to the second LNS domain. LNS6, belonging to the arms of the "Y" shape, tends to fold back toward LNS4, creating a more compact structure, potentially limiting the accessibility to binding partners (Reissner et al., 2008). However, single-particle EM data show that LNS6 retains some flexibility, and the clover-leaf arrangement suggests that the metal binding rim of each LNS domain is likely solvent exposed rather than being confined to interdomain stabilization. Regarding the decreased affinity that αNRXN shows with the neuroligins (Boucard et al., 2005), it is possible that the extensive flexibility and segmental motion of LNS6 may require that optimal binding is achieved only after a conformational change, thus acting as a factor limiting NLGN binding.

The three-dimensional structural models of αNRXN we present here are, to our knowledge, the first experimentally derived models of the entire extracellular domain of the αNRXN super-family or other LNS modular multidomain proteins. Taking advantage of the multiple crystal structures available of the complex between neuroligin 1 and 4 (Fabrichny et al., 2007; Araç et al., 2007; Chen et al., 2008) with βNRXN, and using our previous findings on the structure of the stalk domain of the neuroligins, we assembled a model of the complex between one of the αNRXN_1-6 models and neuroligin 1 in the context of the synaptic cleft. In the case of αNRXN, flexibility, partial elongation, and the clover-leaf arrangement of LNS4-6 domains are ideal for binding multiple ligands because each LNS subunit can associate with partnering proteins independent of the neighboring domains. As shown in the model of the complex (Figure 5), the presence of the stalk region and the elongated nature of this model explain how multiple ligands can bind simultaneously to the extracellular domain of αNRXN, and how it may orient in the synaptic cleft.

As hypothesized in an earlier work (Tabuchi and Südhof, 2002), mutations in *NRXN* genes likely confer subtle phenotype changes that may result in widespread dysfunction during the development of a complex nervous system, particularly in a polygenic disorder involving other proteins affecting neuronal system development. In fact, new evidence indicates that variations in copy number and rare variants within the genes encoding neurexin-1 and -3 (*NRXN1* and *NRXN3*) contribute to ASD susceptibility, mental retardation (Feng et al., 2006; The Autism Genome Project Consortium, 2007; Kim et al., 2008; Yan et al., 2008; Glessner et al., 2009), schizophrenia (Rujescu et al., 2009), and in altering addiction and reward behaviors to nicotine (Nussbaum et al., 2008).

EXPERIMENTAL PROCEDURES

Expression of αNRXN1

The construct encoding the secreted, soluble extracellular domain of αNRXN1 IgG fusion protein (Ig-N1α-1) (Boucard et al., 2005) was a kind gift of Dr. Thomas Südhof (Stanford, CA). We adapted this construct by introducing a 3C protease cleavage site (LEVLFQ/GP) between residue T1309 of αNRXN and the beginning of the hIgG sequence. HEK293 GnTI cells were transfected with the appropriate plasmids and were selected by growth in G418 (Geneticin, Sigma) (Comoletti et al., 2003). For protein expression, cells were maintained at 37°C and 10% CO_2 in Dulbecco's modified Eagle's medium containing up to 2% fetal bovine serum.

Analytical Ultracentrifugation

All analytical ultracentrifugation experiments were performed in a Beckman/Coulter XL-I ultracentrifuge using An60Ti and An50Ti rotors at the Center for Analytical Ultracentrifugation of Macromolecular Assemblies, San Antonio, TX. Sedimentation equilibrium and velocity experiments were analyzed with the UltraScan software, version 9.9, release 847 (Demeler, 2009). All samples were run in 10 mM sodium phosphate buffer (pH 7.4) with 137 mM NaCl and 2.7 mM KCl, at 4°C. Hydrodynamic corrections were made according to Laue et al. (1992) as implemented in UltraScan.

Negative-Stain Single-Particle Electron Microscopy

Purified αNRXN1 (0.05–0.1 mg/ml) in 10 mM HEPES (pH 7.4) and 150 mM NaCl solution was applied to glow-discharged carbon-coated grids and was negatively stained with 0.75% uranyl formate, as described elsewhere (Ohi et al., 2004). EM images for class averages were collected using an FEI 200KV Sphera microscope equipped with an LaB_6 electron filament using low-dose procedures on SO-163 Kodak film at a magnification of 50,000× and nominal defocus of −1.5 μm. Micrographs were digitized with a Nikon scanner, and particles were selected interactively using the WEB display program. A second pass selection of properly centered particles was done interactively, and particles were aligned and classified by reference-based alignment and the K-means classification (100–150 classes) using the SPIDER suite (Frank et al., 1996).

Small-Angle X-Ray Scattering Data Acquisition of αNRXN1

Data were collected from the proteins and their solvent blanks (ultrafiltrate buffers for the αNRXN and last step dialysate for lysozyme) at 20°C, using an Anton Paar SAXSess line collimation instrument at the University of Utah on 2D position-sensitive image plates (10 mm slit and integration width) as described by Jeffries et al. (2008).

Structure Modeling from the Scattering Data

Rigid body modeling was performed using the programs SASREF and BUNCH (Petoukhov and Svergun, 2005). Both techniques refine the domain positions within the protein against the scattering data using calculated partial scattering amplitudes derived from the atomic structures of the individual component domains.

More information on some of the methods can be found in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, three tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.str.2010.06.005.

ACKNOWLEDGMENTS

This work was supported by USPHS (grant R37 GM-18360 to P.T.), NIEHS (grant P42ES10337 to P.T.), U.S. Department of Energy (grant DE-FG02-05ER64026 to J.T.), Autism Speaks (grant 2617 to D.C.), and the John Merck Fund and Hellman Foundation (support to T.N.). We acknowledge the use of the UCSD Cryo-Electron Microscopy Facility which was supported by NIH grants 1S10RR20016 and GM033050 to Timothy S. Baker and a gift from the Agouron Institute to UCSD. AUC supercomputer analyses were supported by NSF Teragrid allocation TG-MCB070038 (B.D.). UltraScan development is supported by NIH-RR022000 (B.D.). Calculations on Lonestar were supported by NSF TeraGrid allocation TG-MCB070038 (BD). We thank Dennis Winge (University of Utah, UT) for quantitative amino acid analysis, and Majid Ghassemian, Department of Chemistry and Biochemistry, for MALDI-TOF analysis. We thank A. G. Porter of the National University of Singapore for the kind gift of the 3C protease plasmid. We thank Michael Baker (Protein Data Bank) for

helpful discussion on homology modeling, and Greg Fuchs for excellent technical help during the preparation of the cleavable αNRXN construct.

Received: March 20, 2010
Revised: June 14, 2010
Accepted: June 17, 2010
Published: August 10, 2010

REFERENCES

Araç, D., Boucard, A.A., Ozkan, E., Strop, P., Newell, E., Südhof, T.C., and Brunger, A.T. (2007). Structures of neuroligin-1 and the neuroligin-1/neurexin-1 beta complex reveal specific protein-protein and protein-Ca2+ interactions. Neuron *56*, 992–1003.

The Autism Genome Project Consortium. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat. Genet. *39*, 319–328.

Boucard, A.A., Chubykin, A.A., Comoletti, D., Taylor, P., and Südhof, T.C. (2005). A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. Neuron *48*, 229–236.

Brookes, E., Cao, W., and Demeler, B. (2010). A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur. Biophys. J. *39*, 405–414.

Brookes, E., and Demeler, B. (2007). Parsimonious regularization using genetic algorithms applied to the analysis of analytical ultracentrifugation experiments. GECCO Proceedings ACM 978-1-59593-697-4/07/0007.

Carafoli, F., Clout, N.J., and Hohenester, E. (2009). Crystal structure of the LG1-3 region of the laminin alpha2 chain. J. Biol. Chem. *284*, 22786–22792.

Chen, X., Liu, H., Shim, A.H., Focia, P.J., and He, X. (2008). Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions. Nat. Struct. Mol. Biol. *15*, 50–56.

Comoletti, D., Flynn, R., Jennings, L.L., Chubykin, A., Matsumura, T., Hasegawa, H., Südhof, T.C., and Taylor, P. (2003). Characterization of the interaction of a recombinant soluble neuroligin-1 with neurexin-1beta. J. Biol. Chem. *278*, 50497–50505.

Comoletti, D., Flynn, R.E., Boucard, A.A., Demeler, B., Schirf, V., Shi, J., Jennings, L.L., Newlin, H.R., Südhof, T.C., and Taylor, P. (2006). Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for beta-neurexins. Biochemistry *45*, 12816–12827.

Comoletti, D., Grishaev, A., Whitten, A.E., Tsigelny, I., Taylor, P., and Trewhella, J. (2007). Synaptic arrangement of the neuroligin/beta-neurexin complex revealed by X-ray and neutron scattering. Structure *15*, 693–705.

Demeler, B. (2009) UltraScan version 9.9, release 847. Analytical ultracentrifugation data analysis software. The University of Texas Health Science Center at San Antonio, Dept. of Biochemistry. http://www.ultrascan.uthscsa.edu

Demeler, B., and van Holde, K.E. (2004). Sedimentation velocity analysis of highly heterogeneous systems. Anal. Biochem. *335*, 279–288.

Demeler, B., and Brookes, E. (2008). Monte Carlo analysis of sedimentation experiments. Prog. Colloid Polym. Sci. *286*, 129–137.

de Wit, J., Sylwestrak, E., O'Sullivan, M.L., Otto, S., Tiglio, K., Savas, J.N., Yates, J.R., 3rd, Comoletti, D., Taylor, P., and Ghosh, A. (2009). LRRTM2 interacts with Neurexin1 and regulates excitatory synapse formation. Neuron *64*, 799–806.

Dudanova, I., Tabuchi, K., Rohlmann, A., Südhof, T.C., and Missler, M. (2007). Deletion of alpha-neurexins does not cause a major impairment of axonal pathfinding or synapse formation. J. Comp. Neurol. *502*, 261–274.

Etherton, M.R., Blaiss, C.A., Powell, C.M., and Südhof, T.C. (2009). Mouse neurexin-1alpha deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. Proc. Natl. Acad. Sci. USA *106*, 17998–18003.

Fabrichny, I.P., Leone, P., Sulzenbacher, G., Comoletti, D., Miller, M.T., Taylor, P., Bourne, Y., and Marchot, P. (2007). Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion. Neuron *56*, 979–991.

Fairless, R., Masius, H., Rohlmann, A., Heupel, K., Ahmad, M., Reissner, C., Dresbach, T., and Missler, M. (2008). Polarized targeting of neurexins to synapses is regulated by their C-terminal sequences. J. Neurosci. *28*, 12969–12981.

Feng, J., Schroer, R., Yan, J., Song, W., Yang, C., Bockholt, A., Cook, E.H., Jr., Skinner, C., Schwartz, C.E., and Sommer, S.S. (2006). High frequency of neurexin 1beta signal peptide structural variants in patients with autism. Neurosci. Lett. *409*, 10–13.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. *116*, 190–199.

Geschwind, D.H., and Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. Curr. Opin. Neurobiol. *17*, 103–111.

Glessner, J.T., Wang, K., Cai, G., Korvatska, O., Kim, C.E., Wood, S., Zhang, H., Estes, A., Brune, C.W., Bradfield, J.P., et al. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature *459*, 569–573.

Graf, E.R., Zhang, X., Jin, S.X., Linhoff, M.W., and Craig, A.M. (2004). Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. Cell *119*, 1013–1026.

Ichtchenko, K., Hata, Y., Nguyen, T., Ullrich, B., Missler, M., Moomaw, C., and Südhof, T.C. (1995). Neuroligin 1: a splice site-specific ligand for beta-neurexins. Cell *81*, 435–443.

Jeffries, C.M., Whitten, A.E., Harris, S.P., and Trewhella, J. (2008). Small-angle X-ray scattering reveals the N-terminal domain organization of cardiac myosin binding protein C. J. Mol. Biol. *377*, 1186–1199.

Kang, Y., Zhang, X., Dobie, F., Wu, H., and Craig, A.M. (2008). Induction of GABAergic postsynaptic differentiation by alpha-neurexins. J. Biol. Chem. *283*, 2323–2334.

Kim, H.G., Kishikawa, S., Higgins, A.W., Seong, I.S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K., et al. (2008). Disruption of neurexin 1 associated with autism spectrum disorder. Am. J. Hum. Genet. *82*, 199–207.

Ko, J., Fuccillo, M.V., Malenka, R.C., and Südhof, T.C. (2009). LRRTM2 functions as a neurexin ligand in promoting excitatory synapse formation. Neuron *64*, 791–798.

Krigbaum, W.R., and Kugler, F.R. (1970). Molecular conformation of egg-white lysozyme and bovine alpha-lactalbumin in solution. Biochemistry *9*, 1216–1223.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Analytical Ultracentrifugation in Biochemistry and Polymer Science, S.E. Harding, A.J. Rowe, and J.C. Horton, eds. (Cambridge: Royal Society of Chemistry).

Li, F., Erickson, H.P., James, J.A., Moore, K.L., Cummings, R.D., and McEver, R.P. (1996). Visualization of P-selectin glycoprotein ligand-1 as a highly extended molecule and mapping of protein epitopes for monoclonal antibodies. J. Biol. Chem. *271*, 6342–6348.

Merry, A.H., Gilbert, R.J., Shore, D.A., Royle, L., Miroshnychenko, O., Vuong, M., Wormald, M.R., Harvey, D.J., Dwek, R.A., Classon, B.J., et al. (2003). O-glycan sialylation and the structure of the stalk-like region of the T cell co-receptor CD8. J. Biol. Chem. *278*, 27119–27128.

Missler, M., and Südhof, T.C. (1998). Neurexins: three genes and 1001 products. Trends Genet. *14*, 20–26.

Missler, M., Hammer, R.E., and Südhof, T.C. (1998). Neurexophilin binding to alpha-neurexins: a single LNS domain functions as an independently folding ligand-binding unit. J. Biol. Chem. *273*, 34716–34723.

Missler, M., Zhang, W., Rohlmann, A., Kattenstroth, G., Hammer, R.E., Gottmann, K., and Südhof, T.C. (2003). Alpha-neurexins couple Ca^{2+} channels to synaptic vesicle exocytosis. Nature *423*, 939–948.

Nussbaum, J., Xu, Q., Payne, T.J., Ma, J.Z., Huang, W., Gelernter, J., and Li, M.D. (2008). Significant association of the neurexin-1 gene (NRXN1) with nicotine dependence in European- and African-American smokers. Hum. Mol. Genet. *17*, 1569–1577.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative staining and image classification—powerful tools in modern electron microscopy. Biol. Proced. Online *6*, 23–34.

Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys. J. *89*, 1237–1250.

Reeves, P.J., Callewaert, N., Contreras, R., and Khorana, H.G. (2002). Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. Proc. Natl. Acad. Sci. USA *99*, 13419–13424.

Reissner, C., Klose, M., Fairless, R., and Missler, M. (2008). Mutational analysis of the neurexin-neuroligin complex reveals essential and regulatory components. Proc. Natl. Acad. Sci. USA *105*, 15124–15129.

Rudenko, G., Nguyen, T., Chelliah, Y., Südhof, T.C., and Deisenhofer, J. (1999). The structure of the ligand-binding domain of neurexin Ibeta: regulation of LNS domain function by alternative splicing. Cell *99*, 93–101.

Rujescu, D., Ingason, A., Cichon, S., Pietiläinen, O.P., Barnes, M.R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E., et al. (2009). Disruption of the neurexin 1 gene is associated with schizophrenia. Hum. Mol. Genet. *18*, 988–996.

Scheiffele, P., Fan, J., Choih, J., Fetter, R., and Serafini, T. (2000). Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. Cell *101*, 657–669.

Sheckler, L.R., Henry, L., Sugita, S., Südhof, T.C., and Rudenko, G. (2006). Crystal structure of the second LNS/LG domain from neurexin 1alpha: Ca2+ binding and the effects of alternative splicing. J. Biol. Chem. *281*, 22896–22905.

Shen, K.C., Kuczynska, D.A., Wu, I.J., Murray, B.H., Sheckler, L.R., and Rudenko, G. (2008). Regulation of neurexin 1beta tertiary structure and ligand binding through alternative splicing. Structure *16*, 422–431.

Sugita, S., Saito, F., Tang, J., Satz, J., Campbell, K., and Südhof, T.C. (2001). A stoichiometric complex of neurexins and dystroglycan in brain. J. Cell Biol. *154*, 435–445.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Cryst. *25*, 495–503.

Tabuchi, K., and Südhof, T.C. (2002). Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing. Genomics *79*, 849–859.

Tisi, D., Talts, J.F., Timpl, R., and Hohenester, E. (2000). Structure of the C-terminal laminin G-like domain pair of the laminin alpha2 chain harbouring binding sites for alpha-dystroglycan and heparin. EMBO J. *19*, 1432–1440.

Ullrich, B., Ushkaryov, Y.A., and Südhof, T.C. (1995). Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. Neuron *14*, 497–507.

Ushkaryov, Y.A., and Südhof, T.C. (1993). Neurexin III alpha: extensive alternative splicing generates membrane-bound and soluble forms. 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 alpha-latrotoxin receptor and laminin. Science *257*, 50–56.

Yan, J., Noltner, K., Feng, J., Li, W., Schroer, R., Skinner, C., Zeng, W., Schwartz, C.E., and Sommer, S.S. (2008). Neurexin 1alpha structural variants associated with autism. Neurosci. Lett. *438*, 368–370.

Zahir, F.R., Baross, A., Delaney, A.D., Eydoux, P., Fernandes, N.D., Pugh, T., Marra, M.A., and Friedman, J.M. (2008). A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha. J. Med. Genet. *45*, 239–243.

Zweier, C., de Jong, E.K., Zweier, M., Orrico, A., Ousager, L.B., Collins, A.L., Bijlsma, E.K., Oortveld, M.A., Ekici, A.B., Reis, A., et al. (2009). CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. Am. J. Hum. Genet. *85*, 655–666.

Structure *18*

Supplemental Information

The Macromolecular Architecture of Extracellular Domain of αNRXN1: Domain Organization, Flexibility, and Insights into *Trans*-Synaptic Disposition

Davide Comoletti, Meghan Miller, Cy M. Jeffries, Jennifer Wilson, Borries Demeler, Palmer Taylor, Jill Trewhella, and Terunaga Nakagawa

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

N-terminal sequencing and mass spectrometry determinations

Edman degradation was performed by the Molecular Structure Facility core at the University of California, Davis. Mass Spectrometry determination of αNRXN mass was determined on a PE Biosystems Voyager DE-STR instrument (Framingham, MA) at the Biomolecular Mass Spectrometry Facility at the University of California, San Diego.

Purification and Analysis of αNRXN

Proteins were affinity purified using Protein-A Sepharose 4 fast flow resin (GE Healthcare) and subsequently cleaved with 3C protease to remove the IgG fragment. Affinity purified proteins were re-purified and buffer exchanged by gel filtration using a Superdex 200-10/300 DE column (GE Healthcare) in 150mM NaCl, 10mM Hepes, pH 7.4. Fractions corresponding to αNRXN were collected and concentrated up to ~6 mg/mL. Aliquots of the purified proteins were separated by SDS-PAGE to check for purity and integrity **(Figure S1B, S1A).** Lysozyme, for use as a scattering standard, was solubilized in 150 mM NaCl, 40 mM sodium acetate, pH 3.8, and dialyzed against the same buffer with the final dialysate used for solvent blank measurements.

Analytical ultracentrifugation

Sedimentation velocity measurements of αNRXN_1-6 at two loading concentrations, 0.25 and 0.5 OD at 230 nm, corresponding to 20 and 40 nM, respectively, were performed at 40,000 rpm in intensity mode. Time invariant noise was corrected with the two-dimensional spectrum analysis (Brookes et al. 2010) for further processing with the enhanced van Holde - Weischet analysis (Demeler and van Holde, 2004) and genetic algorithm optimization (Brookes and Demeler, 2007). All fitting statistics were obtained by performing Monte Carlo analyses (Demeler and Brookes, 2008). Data were analyzed on the Lonestar supercomputer at the Texas Advanced Computing Center and on the Jacinto cluster at the Bioinformatics Core Facility at the University of Texas Health Science Center, San Antonio, TX.

For equilibrium experiments, 125 μL of αNRXN at 6 different protein concentrations corresponding to OD values of 0.3, 0.5, and 0.7 measured at 230 nm and 280 nm (ranging between 10 nM to 7μM) and ran at five rotor speeds (10.7, 14.1, 16.9, 19.3 and 21.4 krpm), resulting in 30 conditions scanned at equilibrium. For equilibrium analysis, data exceeding 1.3 OD were excluded from the fits. Data at equilibrium were recorded in radial step mode with 0.001 cm step size setting with 20 replicate absorption measurements at each step. Equilibrium experiments were globally fitted using a fixed molecular weight distribution-Monte Carlo analysis with 100 species ranging between 10 - 200 kDa (Demeler, 2009). The extinction coefficient at 230 nm (1.027×10^6 OD_{230} M^{-1} cm^{-1}) was determined with the global extinction spectrum fitter from UltraScan (Demeler, 2009) using wavelength scans from the 6 loading concentrations. The global spectrum was calibrated at 280 nm with a value determined from the protein sequence

(136,790 OD_{280} M^{-1} cm^{-1}) as implemented in UltraScan. The partial specific volume, $\bar{\upsilon}$ of αNRXN_1-6 was calculated accounting for total carbohydrates (see result section) and found to be 0.728 ml/mg using Sednterp ver 1.09 (http://www.jphilo.mailway.com/default.htm)

Negative-stain Single particle EM epitope tagging

We engineered a construct with an N-terminal FLAG tag (DYKDDDDKL) and C-terminal HA tag (YPYDVPDYA). Expressed proteins were tested by western blotting, probing for FLAG and HA to verify functionality of both tags. Fab fragments that bind to HA and FLAG epitopes were purified and prepared using the methods described elsewhere (Shanks et al. 2010). Purified protein was incubated at 4°C overnight in a 2:1, 4:1 and 8:1 ratio (Fab:αNRXN) with the Fab fragments of either the anti-FLAG or anti-HA. Negative-staining was performed as described in the Methods section in main text and raw data images were collected on a FEI Sphera microscope using a Gatan CCD camera.

Quantitative Amino Acid Analysis

For small angle X-ray scattering experiments the protein concentrations were determined by quantitative amino acid analysis on a Beckman 6300 analyzer after hydrolysis in 5.7 N HCl containing 0.1% phenol in vacuum, at 110ºC. Only reliably quantified amino acids (Ala, Arg, Glx (Gln+Glu), Ile, Leu, Lys, Phe, and Val) were used to estimate protein concentration of samples analyzed by SAXS. For all other applications protein concentrations were determined by UV 280 using the extinction coefficient reported above.

Small angle X-ray scattering data analysis of αNRXN

The scattering intensity functions of each protein in solution ($I(q)$ versus q, where $q = \frac{4\pi}{\lambda}\sin\theta$, λ = 1.5418 Å (Cu Kα), θ = half the scattering angle) were determined by subtracting the scattering from the solvent blank from the respective sample scattering. Data reduction was performed using SAXSQuant1D (Anton Paar proprietary software). Guinier plots ($\ln(I(q))$ vs q) were initially evaluated for linearity using the program SAXSQuant1D that can output 'desmeared' data, i.e. data that have been corrected for the slit geometry of the instrument which can then be evaluated using PRIMUS (Konarev et al., 2003). The R_g and $I(0)$ values were estimated using extrapolations of $\ln(I(q))$ vs q in the linear region $q_{max}R_c \leq 1.3$.

The probable inter-atomic distance distribution of vector lengths of each of the proteins, or pair distance distribution function, $P(r)$, were calculated using the indirect Fourier transform method of Svergun as implemented in the program GNOM (Svergun, 1992). This method also incorporates a geometric correction factor that takes into account the slit geometry of the instrument to produce a real-space vector length distribution. Structural parameters derived from the $P(r)$, including the forward scattering intensity ($I(0)$, radius of gyration (R_g) and maximum dimension (D_{max}) were determined for αNRXN proteins and lysozyme standards at various protein concentrations (**Table S1**).

The determination of $I(0)$, from the protein samples was used to evaluate sample monodispersity against a lysozyme standard. Dilute monodisperse proteins of the same X-ray scattering (electron) density in solution adhere to the relationship:

$$K = \frac{I(0)}{M_r c}$$

where M_r is the calculated (expected) molecular mass of the scattering protein (kDa), c is the protein concentration (mg/mL^{-1}) and K is a constant. Lysozyme is known to be monodisperse under the conditions measured (Krigbaum and Kugler 1970). Thus the K value derived from the lysozyme $I(0)$ data can be used as a standard to evaluate the monodispersity of the αNRXN protein samples (provided that protein concentrations are known within ~5 %). The extrapolated $I(0)$ values derived from Guinier analyses and the $P(r)$ profiles for the lysozyme standards collected at the same time as αNRXN_1-6 result in an average K value set to 1. Under the instrumental setup described, all monodisperse, monomeric proteins should have K-values equivalent to 1; monodisperse multimers will have multiples of 1. For αNRXN_1-6 the average K-value across all of the protein concentrations tested is 1.155 (based on an expected

monomeric molecular weight of 145.9kDa) which is ~15% higher than that determined for lysozyme. This higher *K* value is within experimental uncertainty (see results.) The uncertainties arise from the glycosylation state of the protein that could add unaccounted for mass to the expected molecular weight and/or alter the scattering density of the protein relative to the non-glycosylated lysozyme standard. For αNRXN_2-6 (collected at a different time with a slightly altered X-ray flux) the average K value of lysozyme was again set to 1 and for αNRXN_2-6 the average K-value resulted in 0.925 (based on a MW of 117 kDa). This is a decrease in K of only ~7% relative to the lysozyme standard, demonstrating that the truncated variant (that has half the oligosaccharide content of the full length protein) is a monodisperse system of monomers in solution. Using the experimentally determined *I*(0) values, the average molecular weights for αNRXN_1-6 and αNRXN_2-6 across the concentration ranges tested are 168 and 113 kDa, respectively (see **table S1**)

Homology modeling of αNRXN

Crystal structures of the LNS 2, 4 and 6 are available in the Protein Data Bank (PDB ID: 2H0B, 2R16, 1C4R respectively) (Rudenko et al., 1999; Sheckler et al., 2006; Shen et al., 2008). Structural templates for LNS-1, -3 and -5 and EGF1, 2, and 3 were retrieved using the PDB web site using as a query the amino acid sequence of each domain. For LNS1, 3 and 5, portions of crystal structures of rat neurexin 1beta (PDB id: 1C4R), agrin (1PZ7), and laminin (2JD4) respectively, were used as templates. For EGF 1, 2 domains, portions of crystal structures of COX-2 (1CVU, and 1PXX) were used. A portion of prostaglandin H2 synthase-1 (1CQE) was used as model template for EGF3 homology modeling. Sequence identity between template and the target sequence varied from 24% (mouse laminin α1 chain) to 53% (COX-2). Sequence alignments were conducted with the module HOMOLOGY of the program INSIGHT II (Accelrys Inc.) and manually adjusted where necessary. HOMOLOGY was also used to assign coordinates during the homology modeling procedures. The structure of each domain was then energy minimized for 10,000 iterations using the distance-dependent dielectric constant with the program DISCOVER (Accelrys Inc.). A total of nine sequences (ranging between 2 and 35 residues) linking various LNS and EGF domains did not have a suitable three-dimensional template.

Structure Modeling from the Scattering Data

Crystal structures and homology models (described above) were used as inputs for SASREF and BUNCH. BUNCH also incorporates regions of unknown structure within the models as dummy atoms and refines the position and shape of these regions during refinement. The total number of missing residues accounts for 11% of the mass of αNRXN_1-6 and 4.5% of the mass of the truncated construct. The similarities seen in the results generated by multiple SASREF or BUNCH refinements (presented as ensembles in **Figure 4**) suggests that the omission (SASREF) or inclusion of the missing mass (BUNCH) does not greatly affect the overall shapes of the refined models, rather the BUNCH models, which more completely account for the molecular mass, are able to refine to significantly lower χ^2 values.

To model such large multi-domain proteins against the SAXS data we defined a number of assumptions as outlined in the Results. Distance constrains were maintained between the domains during multiple SASREF refinements against the 5.86 and 5.7 $mgml^{-1}$ αNRXN_1-6 and αNRXN_2-6 datasets to ensure that nonsensical domain swapping or drifting did not occur (for a list of residue ranges and distances between domains, see **Table S2**). The first rounds of BUNCH refinement also maintained distance constraints and the models were refined in parallel with αNRXN_1-6 and αNRXN_2-6 data. At the completion of this initial modeling, the region of the protein spanning EGF2-LNS6 was defined as one rigid body while the remainder of the domains were left as individual rigid bodies and refined, without distance constraints, against the 5.86, 4.65 and 3.49 mg/mL data (for αNRXN_1-6) or 5.7, 4.56 and 3.42 mg/mL data (for αNRXN_2-6 .) At the end of this second round of refinement, either LNS1, EGF1, LNS2, LNS3 were defined as one rigid body (for αNRXN_1-6 refinement) or LNS2, LNS3 and EGF2 were defined as one rigid body (for αNRXN_2-6 refinement) while (EGF2)-LNS4-LNS5-EGF3-LNS6 were left to refine their individual positions against the multiple concentration datasets. The procedure of fixing the C-terminal EGF2-LNS6 region of the protein as one rigid body, then

refining the positions of the individual N-terminal domains LNS1, EGF1, LNS2, LNS3 then vice versa — letting EGF2 to LNS6 refine as individual bodies and keeping N-terminal domains as one rigid body — was repeated. Finally, the refined positions of the domains comprising the 'clover-leaf' head of both constructs (EGF2-LNS4-LNS5-EGF3-LNS6) were defined as a single rigid body and the short linkers between the domains in the head were manually built in based on their BUNCH dummy atom substitutes. Ten BUNCH refinements were then performed in parallel for both constructs using the clover-leaf head as one rigid body and the LNS1 EGF1 LNS2, LNS3 domains as individual rigid bodies to produce the BUNCH ensembles presented in **Figure 7** that show the top 5 models fits to the data for both αNRXN_1-6 and αNRXN_2-6. The goodness-of-fit for each SASREF and BUNCH model were assessed against the desmeared experimental scattering data using the χ^2 discrepancy calculated in CRYSOL (Svergun et al., 1995).

$$\chi^2 = (N-1)^{-1}\sum\left[\frac{I(q)-kI_m(q)}{\sigma(q)}\right]^2$$

where N is the number of experimental points, k is a scaling factor, $I(q)$ and $I_m(q)$ are the experimental and atomic model scattering intensities respectively and $\sigma(q)$ represents the experimental error.

Surface Plasmon Resonance Analysis of the αNRXN_1-6/neuroligin-1 binding

Binding experiments was done at 25ºC in 10mM Hepes buffer, pH 7.4, containing 150mM NaCl, 2mM $CaCl_2$ and 0.005% (v/v) surfactant P20, on a BIAcore 3000 apparatus. αNRXN was covalently bound to the carboxymethylated dextran matrix of a CM5 chip (BIAcore, GE Healthcare). To obtain specific αNRXN binding, the first flow cell was mock coupled with buffer only for real time background subtraction. Neuroligin-1 was injected in random order over the αNRXN surface as a set of 6 concentrations extending from 3000nM to 12 nM in three-fold dilutions at a flow rate of 50µL/min. After each injection the binding surfaces were regenerated by injecting for 30 sec a buffer containing 1M NaCl and 5mM EGTA to chelate free Ca^{2+} to disrupt the association.

SUPPLEMENTAL REFERENCES

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Cryst. 36, 1277–1282.

Shanks N.F., Maruo T., Farina, A.N., Ellisman, M.H., and Nakagawa, T. (2010). Contribution of the global subunit structure and stargazin on the maturation of AMPA receptors. J. Neurosci. 30, 2728–2740.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995) CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J Appl. Cryst. 28, 768–773.

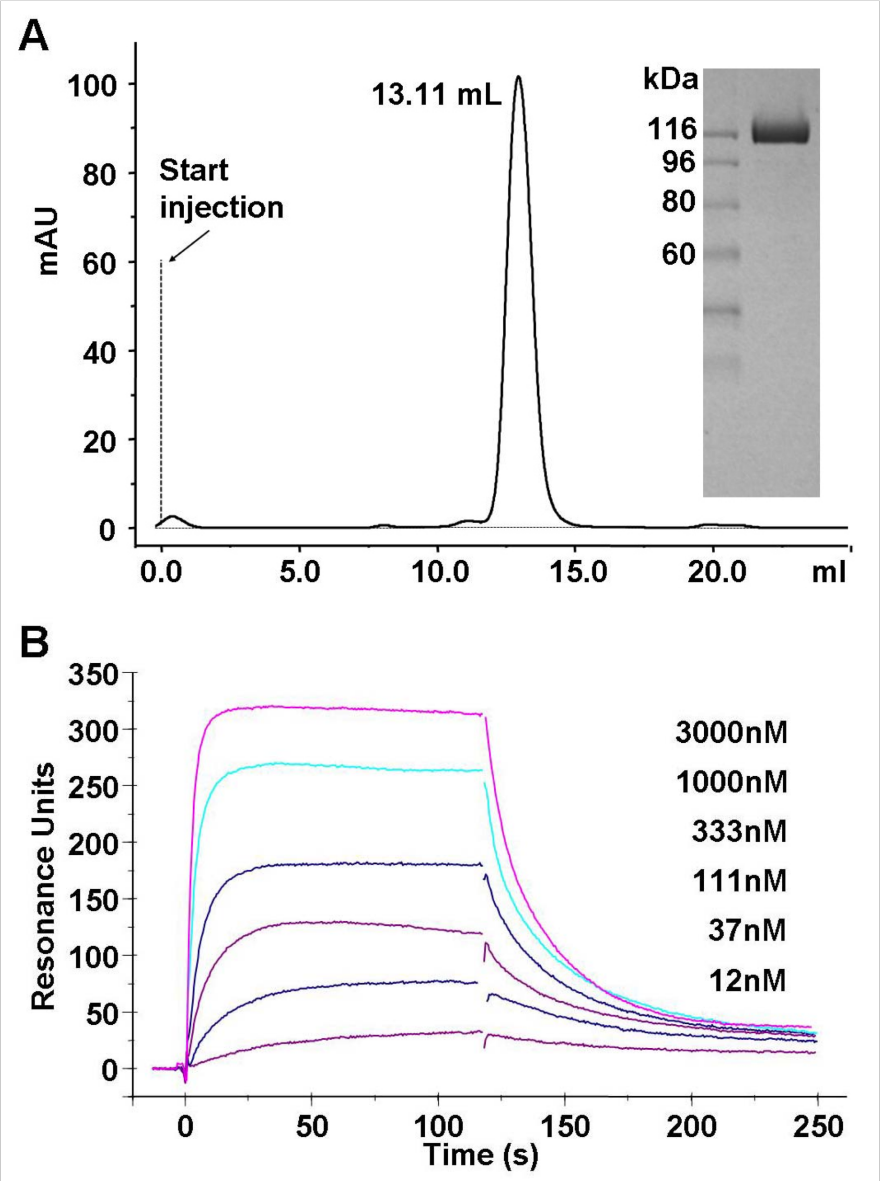

Figure S1, related to Figure 1.
A: Size exclusion chromatography characterization of the purified extracellular domain of αNRXN_2-6 - Size exclusion chromatography trace of purified αNRXN_2-6. Inset, Coomassie blue staining of a sample composed of the main peak.

B: Binding data between αNRXN1_1-6 and neuroligin-1: Surface plasmon resonance curves of the association between αNRXN1_1-6 and NLGN-1 without both splice inserts A&B. Neuroligin-1 was injected over immobilized αNRXN1_1-6 for approximately 120 seconds followed by a wash interval of 150 seconds with running buffer alone. The sensorgrams demonstrate a concentration dependent association, as anticipated for a bimolecular association.

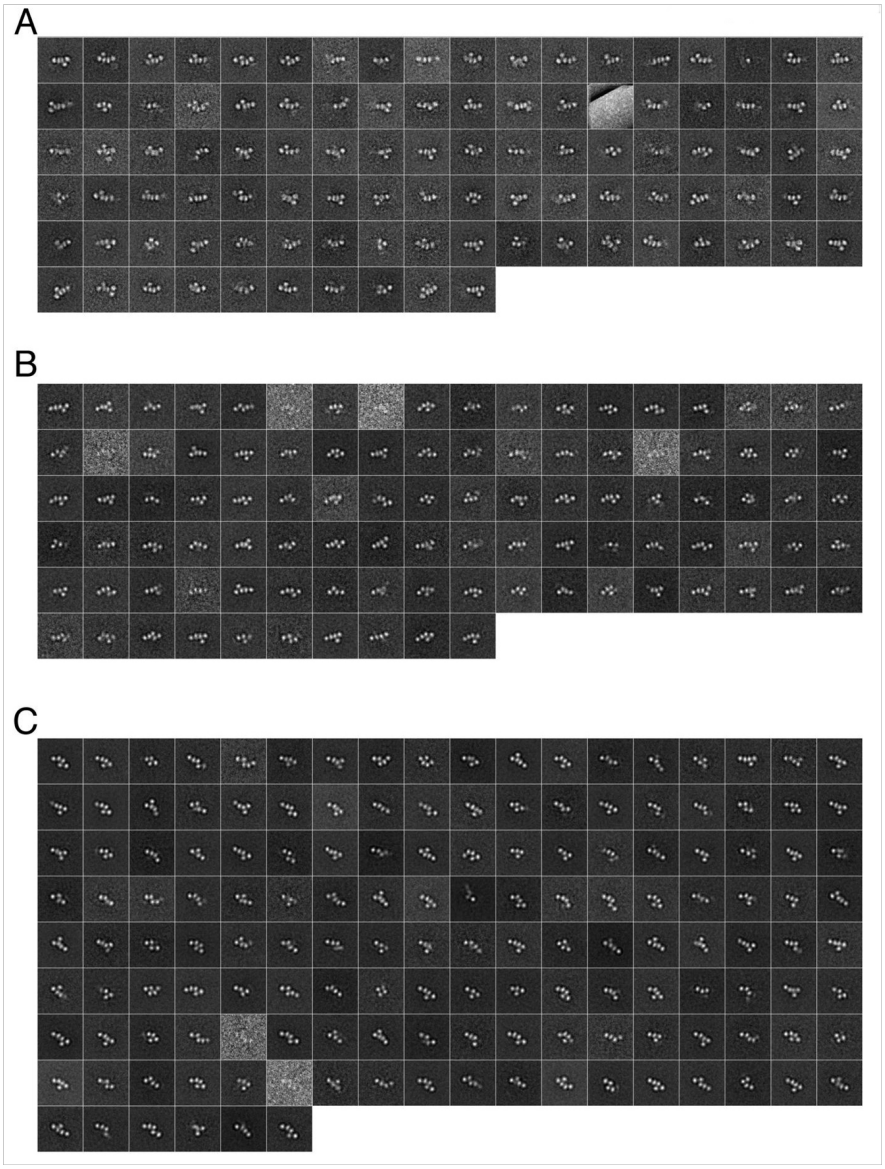

Figure S2, related to Figure 2. Class averages of the αNRXN_1-6 and αNRXN_2-6
A) Class averages of the negative stained αNRXN_1-6 particles. Most of the particles show either six or five distinct domains. In some of the class averages, less than four domains were clearly resolved due to conformational heterogeneity and viewing orientation of the particles.
B) Class averages of the negative stained αNRXN_2-6 particles obtained by analyzing approximately 5,200 particles. This set of class averages was obtained from a data set smaller than the previous one.
C) The class averages of the negative stained αNRXN_2-6 particles obtained from analyzing approximately 17,000 particles. Increasing the number of the particles has significantly improved the signal-to-noise of the class averages compared to the results shown in **A** and **B.** The overall trends of the distribution of the different particle shapes were similar to the results form the analysis that was obtained from analyzing 5,200 particles, as shown in **B**.
Scale: The size of each box of the class averages in **A-C** is 38 x 38 nm

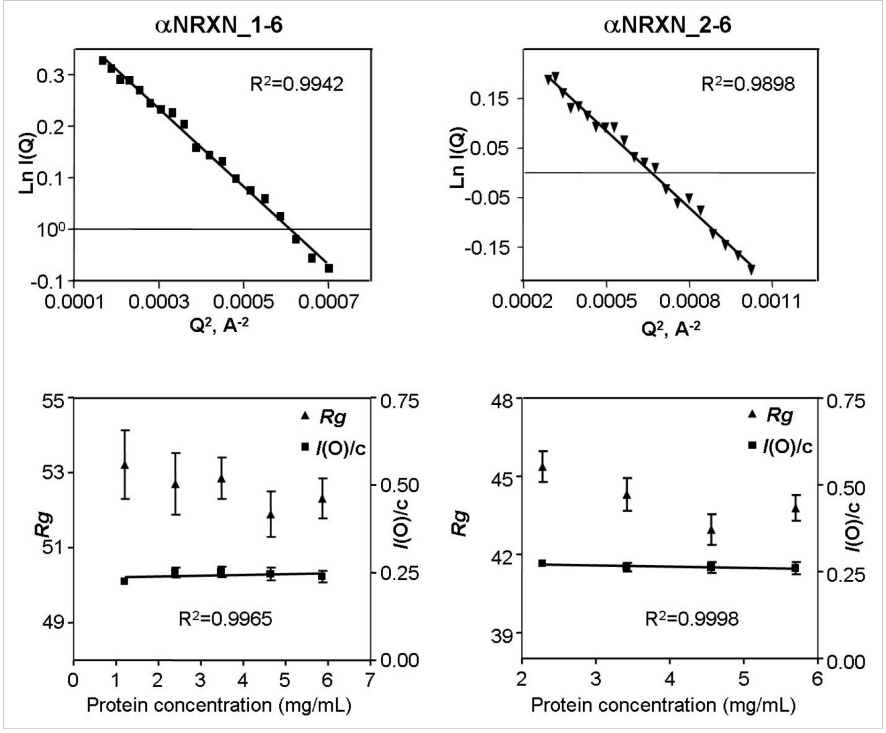

Figure S3, related to Figure 3. Basic scattering parameters for αNRXN_1-6 and αNRXN_2-6
Top panels, Guinier analyses. Solid lines represent the best fit through the Guinier regions of both neurexin constructs.
Bottom panels, dependence of *I*(O)*/c* and *Rg* at different protein concentrations. For αNRXN_1-6, protein concentration range was 1.21 to 5.86 mg/mL; for αNRXN_2-6 protein concentration range was 2.28 to 5.7 mg/mL. Error bars represent standard deviation values.

Table S1, related to Figure 3. Basic SAXS parameters

Sample name	*Dmax*	*Rg*	*I(O)*	Concentration mg/ml	*I(O)/C*	*I(O)/C* /MW
αNRXN-_1-6	170	53.39±0.24	34.75±0.31	5.861	5.93±0.05	0.042
	170	53.04±0.28	27.92±0.27	4.647	6.01±0.06	0.042
	170	53.22±0.30	21.92±0.24	3.488	6.28±0.07	0.044
	170	52.93±0.34	14.11±0.17	2.401	5.88±0.07	0.041
	170	52.63±0.40	6.53±0.12	1.206	5.42±0.10	0.038
average	**170**	**53.04±0.31**				**0.042**
	Dmax	*Rg*	*I(O)*	Concentration mg/ml	*I(O)/C*	*I(O)/C* /MW
Lysozyme	40	14.6±0.03	10.37±0.04	20.62	0.50±0.002	0.035
	40	14.61±0.03	7.578±0.03	15.46	0.49±0.002	0.034
	40	14.76±0.04	5.081±0.02	10.31	0.49±0.002	0.034
	40	14.59±0.09	2.888±0.02	5.15	0.56±0.004	0.039
average	**40**	**14.64±0.05**				**0.036**

Sample name	*Dmax*	*Rg*	*I(O)*	Concentration mg/ml	*I(O)/C*	*I(O)/C* /MW
αNRXN-_2-6	140	43.78±0.49	1.45±0.10	5.700	0.26±0.02	0.022
	140	42.95±0.59	1.16±0.06	4.560	0.26±0.01	0.022
	150	44.30±0.63	0.90±0.01	3.420	0.26±0.01	0.022
	150	45.37±0.59	0.62±0.08	2.280	0.27±0.01	0.023
average	**145**	**44.19±0.57**				**0.022**
	Dmax	*Rg*	*I(O)*	Concentration mg/ml	*I(O)/C*	*I(O)/C* /MW
Lysozyme	42	14.42±0.03	0.239±0.06	7.39	0.32±0.008	0.002

Table S2, related to Figure 4. SASREF calculation parameters

LNS1			EGF1	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
L66	P209	18; 24	C228	Q257
EGF1			**LNS2**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
C228	Q257	22; 24	E280	V460
LNS2			**LNS3**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
E280	V460	5; 24	N466	Q643
LNS3			**EGF2**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
N466	Q643	26; 24	P670	C703
EGF2			**LNS4**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
P670	C703	5; 20	V709	G889
LNS4			**LNS5**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
V709	G889	5; 20	G895	C1071
LNS5			**EGF3**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
A895	C1071	5; 20	T1077	C1110
EGF3			**LNS6**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
T1077	C1110	4; 24	T1115	V1291

Table S3, related to Figure 4. BUNCH calculation parameters

LNS1			EGF1	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
L31	P209	18; 47	C228	Q257
EGF1			**LNS2**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
C228	Q257	22; 40	E280	V460
LNS2			**LNS3**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
E280	V460	5; 25	N466	Q643
LNS3			**EGF2**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
N466	Q643	26; 24	P670	C703
EGF2			**LNS4**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
P670	C703	5; 27	V709	G889
LNS4			**LNS5**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
V709	G889	5; 10	G895	C1071
LNS5			**EGF3**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
G895	C1071	5; 18	T1077	C1110
EGF3			**LNS6**	
STARTS	**ENDS**	**Missing residues; Å**	**STARTS**	**ENDS**
T1077	C1110	4; 27	T1115	V1291

ACKNOWLEDGMENTS

Chapter 3 is a reprint of the material as it appears in Comoletti, D.*, **Miller, M.T.*,** Jeffries, C.M., Willson, J., Demeler, B., Taylor P., Trewhella J., Nakagawa T. (2010) *Structure* **18,** 1044-1053. * These authors contributed equally to this work. The dissertation author contributed to the overall concept of the work, was the leading investigator for the Electron Microscopy experiments and author of this section of the paper, designed one of the protein constructs, and did much of the protein expression and purification

CHAPTER 4:

The crystal structure of the α-neurexin-1 extracellular region reveals a hinge point for mediating synaptic adhesion and function

4.1 SUMMARY

α- and β-neurexins (NRXNs) are transmembrane cell adhesion proteins that localize to pre-synaptic membranes in neurons and interact with the post-synaptic neuroligins (NLGNs). Their gene mutations are associated with the autistic spectrum disorders. The extracellular region of α-NRXNs, containing nine independently folded domains, has structural complexity and unique functional characteristics, distinguishing it from the smaller β-NRXNs. We have solved the X-ray crystal structure of seven contiguous domains of the α-NRXN-1 extracellular region at 3.0 Å resolution. The structure reveals an arrangement where the N-terminal five domains adopt a more rigid linear conformation and the two C-terminal domains form a separate arm connected by a flexible hinge. In an "open" state the molecule is suitably configured to accommodate a bound NLGN molecule, as supported by structural comparison and surface plasmon resonance. These studies provide the structural basis for a multi-functional synaptic adhesion complex mediated by α-NRXN-1.

4.2 INTRODUCTION

Specialized synaptic adhesion proteins play an important role in synapse specification, development and maintenance. Genetic evidence implicates several synaptic adhesion proteins in the onset of neuro-developmental disorders, such as the autism spectrum disorders (ASD) [1]. The *NRXN-1* gene has been identified as one

candidate for susceptibility to ASD in multiple genetic linkage studies, ranging from whole-genome scans to single-nucleotide polymorphism (SNP) screens [2-6]. The neurexins (NRXNs) constitute a family of pre-synaptic transmembrane proteins, which are developmentally and spatially expressed at neuronal GABAergic and glutamatergic synapses [7].

In vertebrates there are three *NRXN* genes (*NRXN 1-3*), each encoding either the longer α-NRXN or the shorter β-NRXN via alternative promoter regions. For each gene, the α- and β-NRXNs yield class I transmembrane proteins that share the same C-terminal sequence and differ only in the extent of their N-terminal extracellular region. After post-translational cleavage of an α-specific signal peptide, the extracellular region of the mature α-NRXN contains nine independently folded domains, including six laminin-neurexin-sex hormone-binding globulin (LNS) domains separated by three epidermal growth factor (EGF)-like domains, and a C-terminal *O*-glycosylated stalk domain (~100 amino acids) that forms an extended linker to the transmembrane domain. The much smaller extracellular region of β-NRXNs contains a β-specific signal peptide (cleaved in the mature protein) and a short β-specific N-terminal sequence followed by a single LNS domain, which is identical to the LNS 6 domain of α-NRXN, and the same *O*-glycosylated stalk domain (**Figure. 4.1a**). NRXN mRNAs are subject to extensive alternative splicing. There are five alternative splice sites in total named SS#1 through 5. α-NRXNs are subject to splicing at all five splice sites, and β-NRXNs only at SS#4 and #5. The diversity of NRXNs, which is achieved by different genes, alternative promoters and extensive splicing, results in the potential for thousands of isoforms [8]. Alternative splicing is thought to function in the regulation of protein-protein interactions and is potentially governing synapse specificity [7].

α- and β-neurexins are thought to function primarily as cell-adhesion molecules through their extracellular interaction with the post-synaptic neuroligins (NLGNs) [9]. More recently they have been found to also associate with the post-synaptic leucine-rich repeat transmembrane proteins (LRRTMs) [10, 11], the α4β2 nicotinic acetylcholine receptor [12], and the $GABA_A$ receptor [13]. α-NRXN-1 was originally identified as a high-affinity receptor for a spider neurotoxin, α-latrotoxin, which binds to pre-synaptic receptors causing massive neurotransmitter release [14]. Studies in mice identified the extracellular region of α-NRXN-1 to be essential for the regulation of Ca^{2+}-dependent exocytosis in neurons [15, 16]. In addition to the post-synaptic partner proteins described above, earlier studies showed that α-NRXNs interact with α-dystroglycan through the LNS 2 and LNS 6/β-NRXN domains [17] and to the soluble neurexophillin proteins through the LNS 2 domain [18]. The precise roles of each of these interactions in governing synaptic assemblies remain to be investigated.

The NLGN molecules constitute another family of class I transmembrane cell-adhesion proteins that are expressed post-synaptically at either glutamatergic (NLGN-1, -3, -4) or GABAergic (NLGN-2) synapses [19, 20]. Structurally, the interaction between the NLGN and β-NRXNs has been extensively characterized. Two β-NRXN molecules bind in a Ca^{2+}-dependent manner on opposing sides of the long axis of the dimeric NLGN molecule. Crystal structures of the complexes between NLGN-1 or NLGN-4 and β-NRXN-1 establish Ca^{2+}-coordination at the binding interfaces and support a regulatory mechanism at the alternative splice site B of NLGN-1 [21-24].

Compared to β-NRXN, the extracellular region of α-NRXN is structurally far more complex. It presents multiple independent folding domains including two known functional protein-binding domains, LNS 2 and LNS 6, which bind to multiple endogenous partner proteins (described above). However, questions persist: what is the

synaptic disposition of extracellular region of α-NRXN in the limited space of the synaptic cleft, and how does the association of the NRXNs with partners mediate synaptic function through protein-protein interactions? We recently demonstrated by single particle electron microscopy (EM) and small angle X-ray scattering (SAXS) that the extracellular region of α-NRXN-1, comprised of LNS 1 through LNS 6, adopts a semi-elongated shape with varying degrees of flexibility between the independent LNS domains, of which the LNS 1 domain displayed the most pronounced flexibility [25]. In fact, removal of the LNS 1 and EGF 1 domains permitted the crystallization of the remainder of the structured extracellular domain.

Herein we present the crystal structure, determined at 3.0 Å resolution, of seven contiguous domains of the α-NRXN-1 extracellular region, including the LNS 2-6 domains and intervening EGF 2 and 3 domains. The domains are organized in an L-shaped conformation with LNS 2-5 forming the longer arm, and EGF 3 and LNS 6 making up the short arm. The achievement of a high-resolution picture of the domain arrangement and distinguishing binding interfaces provides novel insights into the structural basis for biological function. By a comparison with the structure of the β-NRXN:NLGN complex, the α-NRXN structure is suitably shaped to accommodate a NLGN molecule bound to LNS 6. The structural analysis is supported by surface plasmon resonance (SPR) studies in which we quantitatively compare the binding affinities of β-NRXN-1 and various α-NRXN-1 truncations for NLGN-1. As previously seen from affinity pull-down assays [26], we demonstrate exclusive binding of NLGN-1 to the LNS 6/β-NRXN domain. Furthermore, using a series of α-NRXN constructs with incremental truncations of each LNS domain, we show that the addition of LNS 2-5 domains results relatively small changes to the binding affinity for NLGN-1. These data confirm the accessibility of the LNS 6 binding site in solution and support the argument

for a high affinity binding interaction between NLGN-1 and the full-length α-NRXN molecule.

4.3 RESULTS

Three Dimensional Structure of α-NRXN-1

The crystal structure of the glycosylated α-NRXN-1 extracellular region, LNS 2 through LNS 6 (α-NRXN_2-6), was solved at 3.0 Å resolution by molecular replacement and refined (**Table 4.1**). It reveals a unique arrangement of domains, which form an L-shape with the longer arm composed of the LNS 2-5 and EGF 2 domains and the short arm formed by the EGF 3 and LNS 6 domains (**Figure 4.1b**). The LNS domains of the long arm are stacked in a similar orientation such that the concave β-sheet face of the N-terminal LNS domain faces the convex β-sheet face of the proceeding C-terminal LNS domain, making extensive interfacial contacts through abundant hydrophobic and electrostatic interactions, and organizing all of the predicted Ca^{2+}-binding sites on the same side of the molecule with similar inter-Ca^{2+} site distances (26-42 Å) (**Figure 4.1b and Figure 4.2a**). The relatively small EGF-like domains (~35 residues) maintain a compact structural fold through a conserved cysteine sequence pattern that results in the formation of three disulfide bonds. The EGF 2 domain is located in proximal apposition with the LNS 3-LNS 4 interface, opposite the predicted Ca^{2+}-binding sites of both LNS domains. It further stabilizes the LNS 3-LNS 4 interface by forming hydrogen bonds at either of its termini with the LNS 3 and 4 domains. At the junction between LNS 5 and EGF 3 there are no stabilizing inter-domain contacts, making an apparent hinge point where the C-terminal EGF 3 and LNS 6 domains form a distinct arm of the protein. The EGF 3 domain separates LNS 5 and LNS 6 by ~35 Å, displacing LNS 6 from the

linear arrangement of LNS 2-5 (**Figure 4.1b**). The extension of the EGF 3 and LNS 6 domains is stabilized by crystal packing of a symmetry related LNS domain, demonstrating an energetically favorable arrangement for the domains in solution (**Figure 4.2b**)

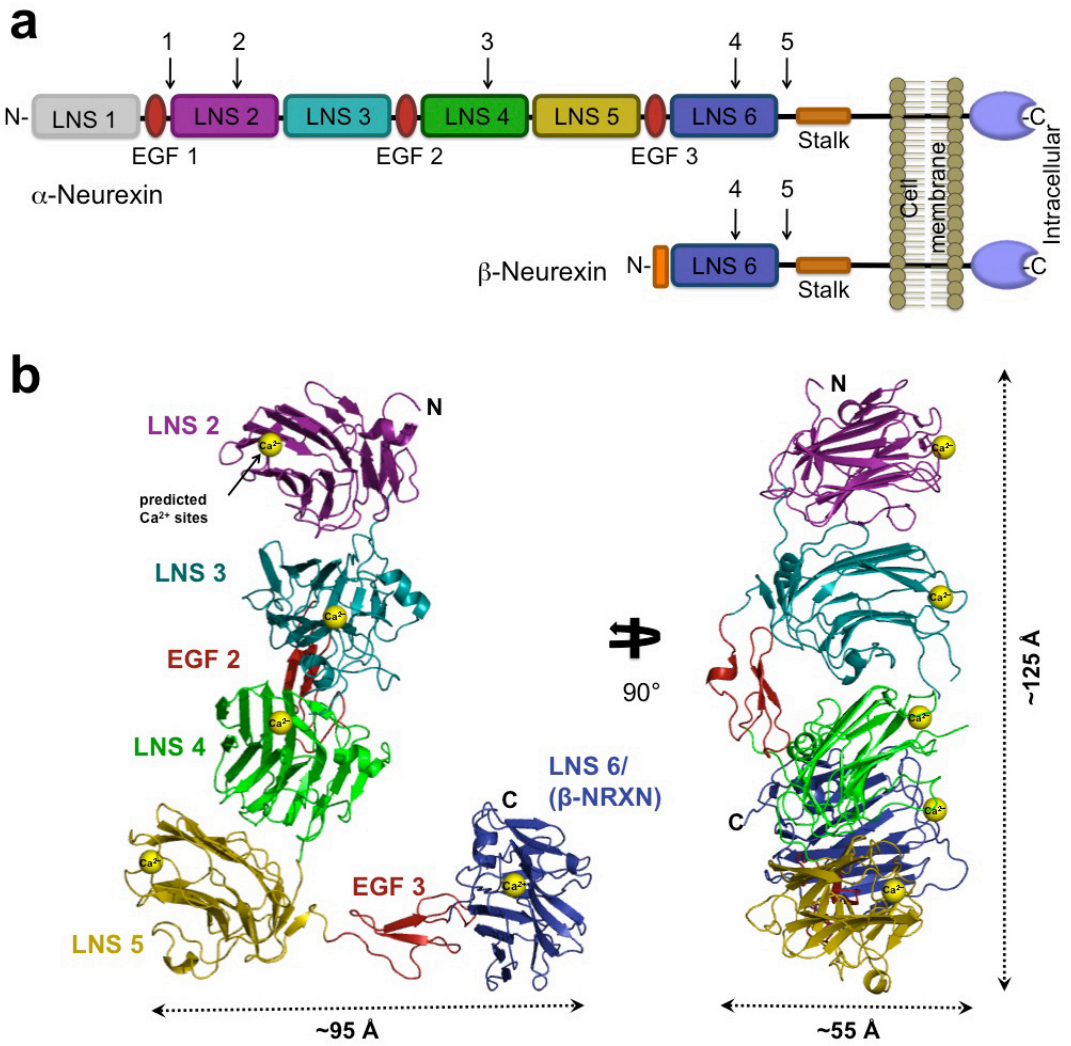

Figure 4.1: The structure of the α-NRXN-1 LNS 2-6 extracellular region. (a) Schematic representation of the mature α- and β-NRXN class I transmembrane proteins. The larger α-NRXNs differ from the shorter β-NRXNs in the extent of their N-terminal extracellular region, with α-NRXNs having six LNS domains interspersed by three EGF domains and β-NRXNs having only one LNS domain, identical to LNS 6 in α-NRXN. The positions for alternative splicing inserts are numbered and indicated with arrows.
(b) Ribbon representation of the crystal structure of α-NRXN-1 LNS 2-6 in two orthogonal orientations. The predicted Ca2+-binding sites, displayed as yellow spheres, were modeled from the Ca2+-bound structures of LNS 2, 4 and 6. Approximate dimensions of the asymmetric molecule are indicated.bound structures of LNS 2, 4 and 6. Approximate dimensions of the asymmetric molecule are indicated.

Table 4.1: Data Collection and Refinement Statistics

X-ray source	SSRL BL11-1
	Crystal data
Space group	C2
Cell dimensions	
a, *b*, *c* (Å)	199.74, 61.24, 155.58
α, β, γ (°)	90.00, 121.21, 90.00
	Data collection
Processing software	XDS package
Wavelength (Å)	0.97945
Resolution (Å)	30.0-3.02(3.13-3.02)
R_{sym} (%)	14.0(62.8)
I / sI	8.7(2.0)
Completeness (%)	97.9(99.5)
No. of unique reflections	31504
Redundancy	3.4(3.4)
	Refinement
Resolution (Å)	3.02(3.10-3.02)
R_{work} / R_{free} (%)	21.4(31.2)/26.9(39.4)
No. atoms	7878
Protein	7795
Ligand/ion	50
Water	33
Average *B* overall (Å^2)	47.6
R.m.s.d. bond length (Å)	0.013
R.m.s.d. bond angle (°)	1.369
Ramachandran Plot	
Preferred (%)	93.6
Allowed (%)	6.1
Outliers (%)	0.3

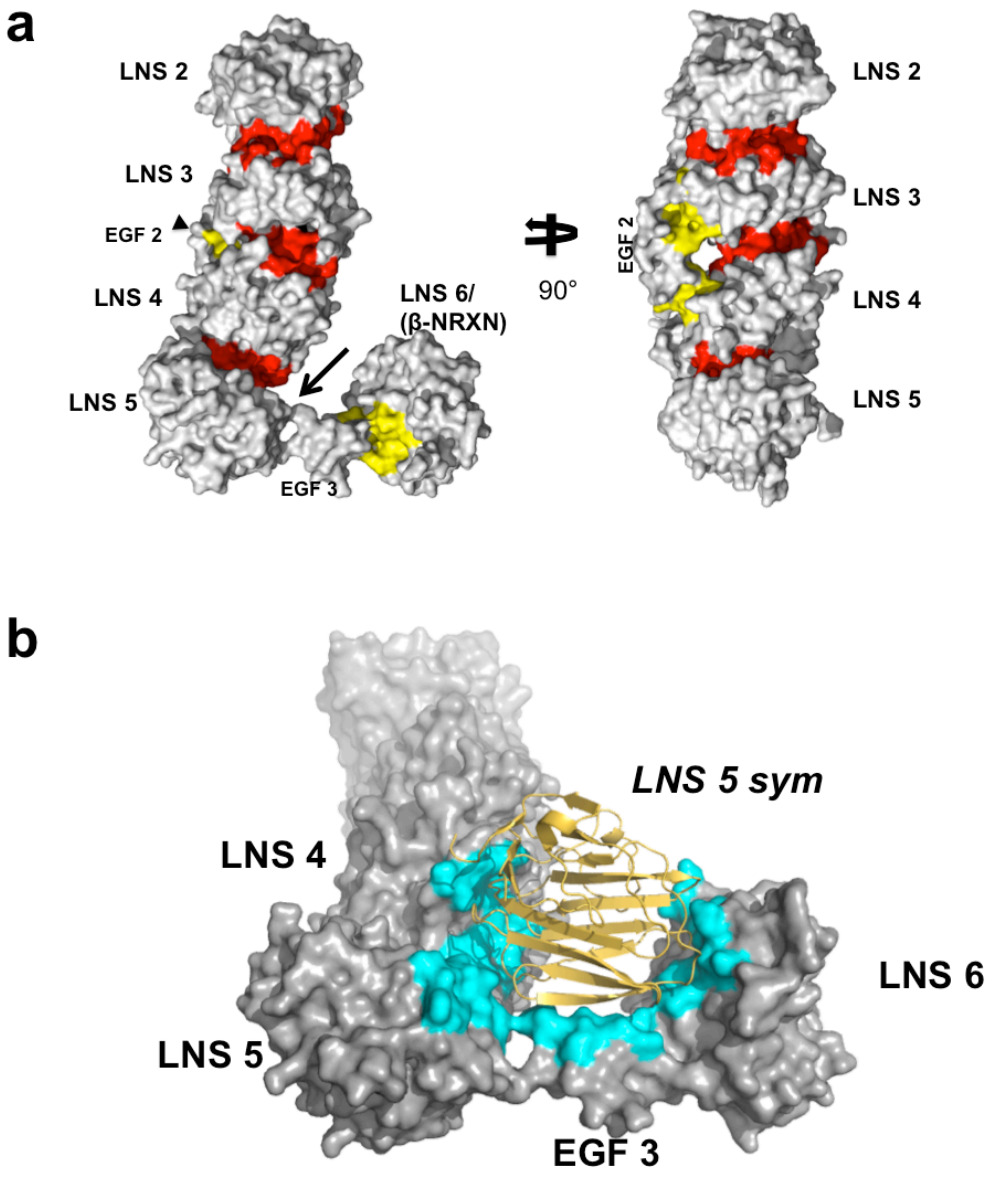

Figure 4.2: α-NRXN_2-6 interfaces
(a) The α-NRXN_2-6 molecule is displayed with interfacing residues between two LNS domains colored in red, and those between an EGF and a LNS domain colored in yellow. Extensive inter-domain contacts are made between the LNS 2-5 domains forming the linear portion of the molecule. The lack of any inter-domain contacts between LNS 5 and EGF 3 indicates a region of flexibility (black arrow).
(b) Crystal packing contacts determine the extension of the short arm (EGF 3-LNS 6) of the α-NRXN_2-6 molecule to form the "L" shape. The reference α-NRXN_2-6 molecule is shown with a gray surface. A symmetry-related LNS 5 domain (*LNS 5 sym*), shown in yellow, makes crystal contacts with the LNS 4, LNS 5, EGF 3 and LNS 6 domains in the first molecule. Interfacing α-NRXN_2-6 residues with *LNS 5 sym* are colored cyan. Residues forming hydrogen bonds or salt bridges with *LNS 5 sym* (within 3.9 Å distance) are colored orange.

There are two predicted *N*-linked glycosylation sites in the sequence of α-NRXN_2-6, one on SS#3 in LNS 4 at N813 and one at N1246 on LNS 6. The Fourier difference maps revealed part of an N-linked glycan bound to N1246 on LNS 6, fitting the electron density of two GlcNAc and one Man moieties. Density is missing for most of SS#3 including the N813 glycosylation site. In addition to the N-linked glycosylation, the Fourier difference maps revealed extra density extending away from the hydroxyl group on S705 in EGF 2. This serine residue belongs to a unique *O*-glycosylation consensus motif (C1-X- S/T-X-P-C2) found between the first and second conserved cysteines in EGF-like domains in several other extracellular proteins, including the coagulation factor IX and the Notch cell surface receptor signaling protein [27, 28]. The complete form of the carbohydrate structure is reported as Xylα1-3Xylα1-3Glcβ1-*O*-Ser [29]. In the α-NRXN_2-6 structure the density accommodates a β-3Glc monosaccharide linked to S705 (**Figure 4.3**).

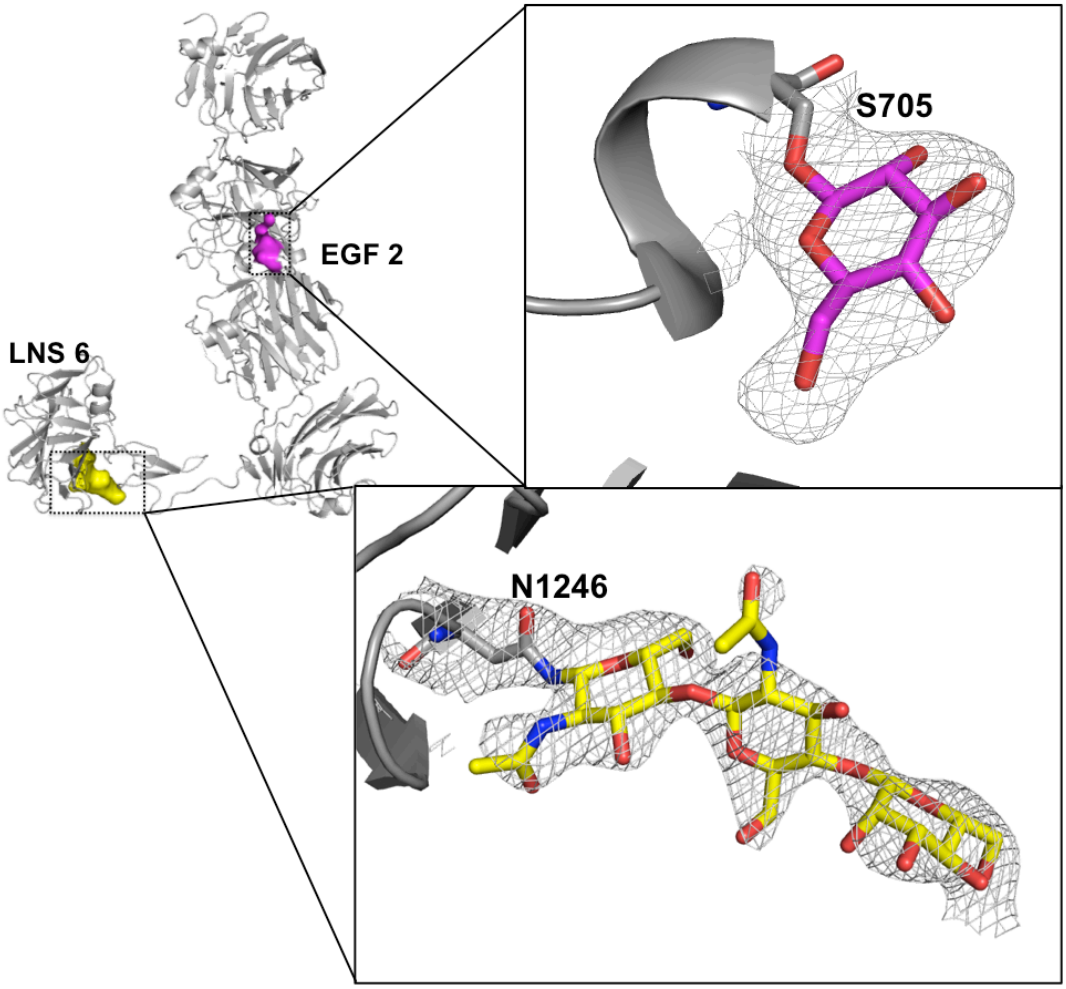

Figure 4.3: α-NRXN N- and O-Glycosylation
The α-NRXN_2-6 molecule bears a β-Glc monosaccharide O-linked to S705 in EGF 2 and a β-GlcNAcβ1,4GlcNAcβ1,4Man trisaccharide N-linked to N1246 in LN 6, respectively. The sequence that includes S705 aligns with a glycosylation consensus sequence found in other extracellular proteins with EGF domains, including Notch and Factor IX. N-linked glycosylation on N1246 matches the consensus glycosylation site found in β-NRXN (PDB:3mw2). The framed figures show the density maps for both sugars. The 2Fo-Fc maps (gray) are contoured at 1.2 σ around S705 and at 1.0 σ around N1246.

Comparison of LNS Domains

The independently solved NRXN-1 LNS domains 2, 4 and 6/β-NRXN show high structural homology despite low sequence identity (~20%) [30-34]. Similarly, a comparison of each of the individual domains with the corresponding domain in the α-NRXN_2-6 molecule shows high structural similarity (average rmsd of 0.5-0.9 Å for all atoms). The five LNS domains in α-NRXN_2-6, which include the previously unsolved LNS 3 and 5, all maintain the characteristic β-sheet sandwich containing 13 β strands making up the core of the domain [31] (**Figure 4.4a**). The face containing the predicted Ca^{2+}-binding site, previously coined the "hyper-variable" surface [31], also contains the sites for SS#2, 3 and 4 in LNS 2, 4, and 6, respectively, and a non-spliced insert that creates a distinct flexible loop at the surface in LNS 3 (**Figure 4.4a**). In the structure, the LNS 3 loop shows high flexibility (average B-factor 88 $Å^2$) and is oriented towards the LNS 4 domain with proximity to the LNS 4 Ca^{2+}-binding site. Long-range electrostatic attractions between the positively charged loop and negatively charged LNS 4 surface are likely to influence the orientation of the flexible LNS 3 loop (**Figure 4.5**). The β11-β12 loop, which extends across the concave side of the β-sheet, diverges in structure for all of the LNS domains (**Figure 4.4a**). Compared to the LNS 6 domain, the β11-β12 loop is significantly larger in the LNS 2-5 domains, covering a greater surface area of the concave side of the β-sheet (**Figure 4.4b**). In LNS 2 and 3 the β11-β12 loop makes stabilizing contacts between the adjacent LNS domains, LNS 3 and 4, respectively. In LNS 4 the loop also makes extensive contacts, predominantly with the symmetry-related LNS 5 domain that participates in the organization of the EGF 3-LNS 6 arm. In LNS 5 the homologous loop is solvent accessible, as is the small β11-β12 loop in LNS 6. Similar to the plant lectins, which have a related fold and homologous β11-β12 loop

involved in carbohydrate binding specificity, the variation in the β11-β12 loop in the different LNS domains of α-NRXN is likely to confer selectivity for the interacting surfaces [31, 35]. The structure demonstrates that this variable surface loop in each domain functions in distinctive ways in the assembly of the full-length α-NRXN protein.

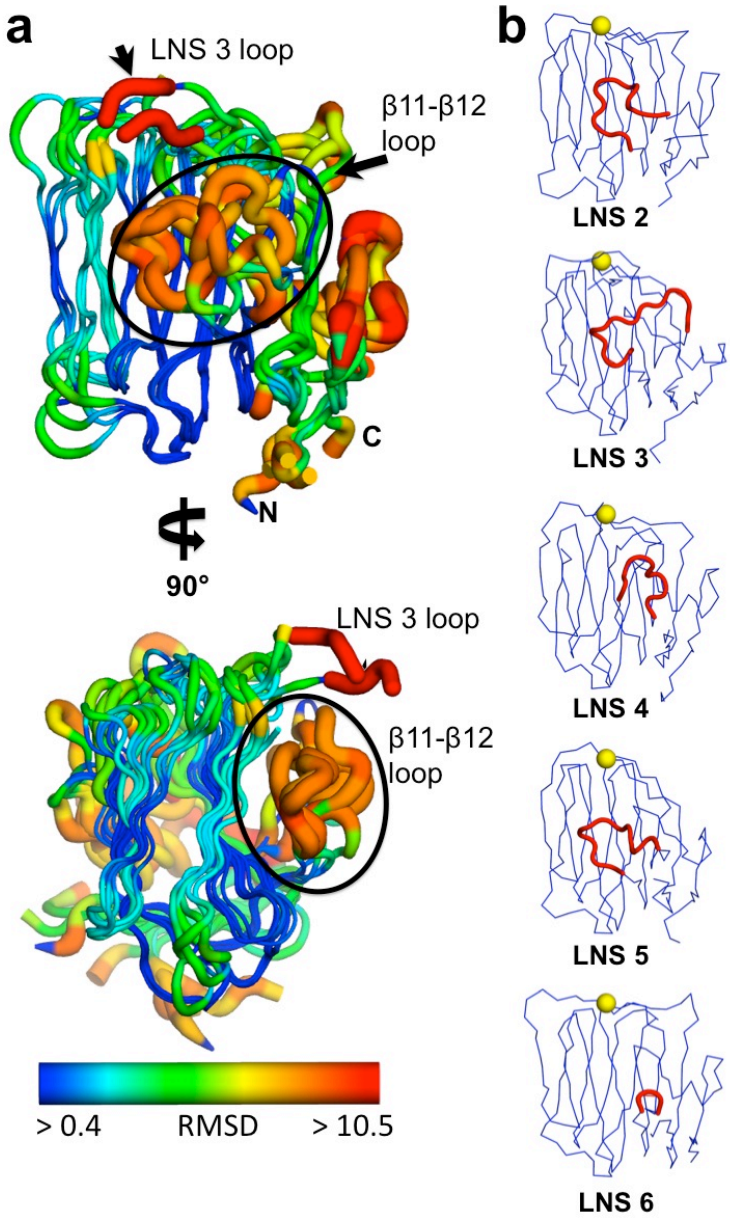

Figure 4.4: Comparison of the α-NRXN_2-6 LNS domains
(a) The five LNS domains share a conserved conformation with a characteristic 13 β-sheet fold. Regions of high variation include the β11-β12 loop (circled) and the LNS 3 loop (black arrow), where non-conserved residues are colored in red. In the structure, the positively charged LNS 3 loop extends towards the Ca2+-binding face of the LNS 4 domain (cf. Supplemental Fig. 3).
(b) The individual LNS domains are displayed in the same orientation with their β11-β12 loops colored in red. The loops play a variable role in stabilizing inter-domain contacts. Predicted Ca2+-binding sites are shown as yellow spheres.

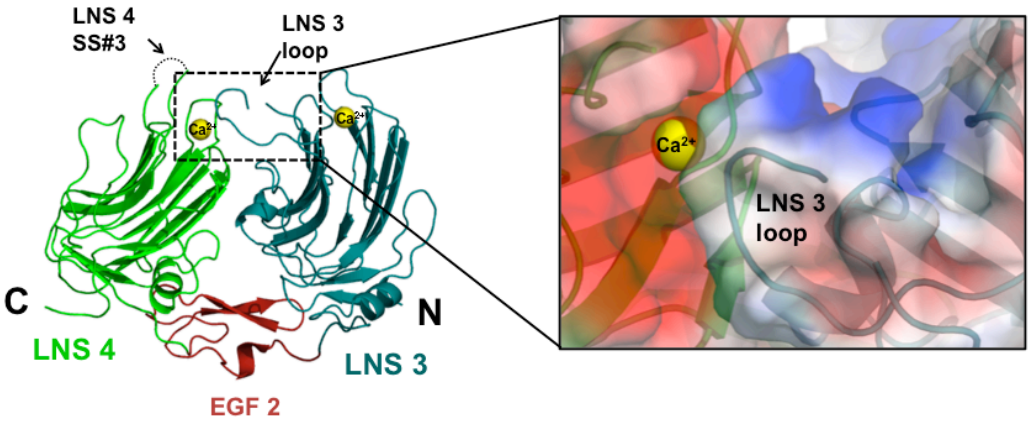

Figure 4.5: LNS 3 loop
The EGF 2 domain, located between LNS 3 and 4, is in proximal apposition to the two interfacing LNS domains. On the opposite side of the interface, LNS 3 has a positively charged flexible loop (LNS 3 loop) that points toward a negatively charged pocket, also the predicted Ca^{2+} binding site, in the LNS 4 domain (framed close-up). The predicted Ca^{2+}-binding sites are shown as yellow spheres.

Structural Comparison with the β-NRXN/NLGN-1 Complex

α-NRXN-1 has two previously identified protein-binding domains, LNS 2 and LNS 6, which bind to several endogenous partner proteins (see Introduction). The previously described NLGN binding site on LNS 6 is solvent accessible in the stabilized conformation found in the α-NRXN_2-6 structure. To model the complex between α-NRXN and NLGN-1, we superimposed the structure of α-NRXN_2-6 onto the structures of the β-NRXN-1 molecules bound to NLGN-1 [21]. The overlaid NLGN molecule interfaces with the α-NRXN_2-6 molecule without significant steric obstruction, indicating that NLGN could bind to α-NRXN_2-6 in the crystallized conformation (**Figure 4.6a**). The LNS 6 domain of α-NRXN_2-6 shows high overlap at the binding interface to the NLGN molecule, as seen in the β-NRXN:NLGN complex structure. Despite the absence of bound Ca^{2+}, the predicted Ca^{2+}-coordinating residues in LNS 6 maintain the same conformational arrangement as observed in Ca^{2+}-bound β-NRXNs either in complex with NLGN or in their free form (**Figure 4.6b&c**). The Fourier difference maps for α-NRXN_2-6 clearly show solvent density in place of the Ca^{2+} in LNS 6 and not the other LNS domains (**Figure 4.6c**). While LNS 6 presents the primary binding site for NLGN [26], the overlay shows the bound NLGN to be proximal to the LNS 4 domain of α-NRXN with no significant steric obstruction between the α-NRXN and NLGN molecules. In the overlay, the interfacing surface of NLGN contains a flexible loop that includes the site for splice insert A. The insert introduces an additional 20 residues and would likely influence any interactions potentially mediated by this surface (**Figure 4.7)**. Similar results were produced in a comparison with the structure of NLGN-4 bound to β-NRXN [22].

In the structure, the Ca^{2+} binding site on LNS 2, previously shown to regulate binding to α-dystroglycan [17], is fully solvent accessible. However, absence of domains LNS 1 and EGF 1 in the α-NRXN_2-6 structure precludes a complete analysis of the exposed interfaces of LNS 2.

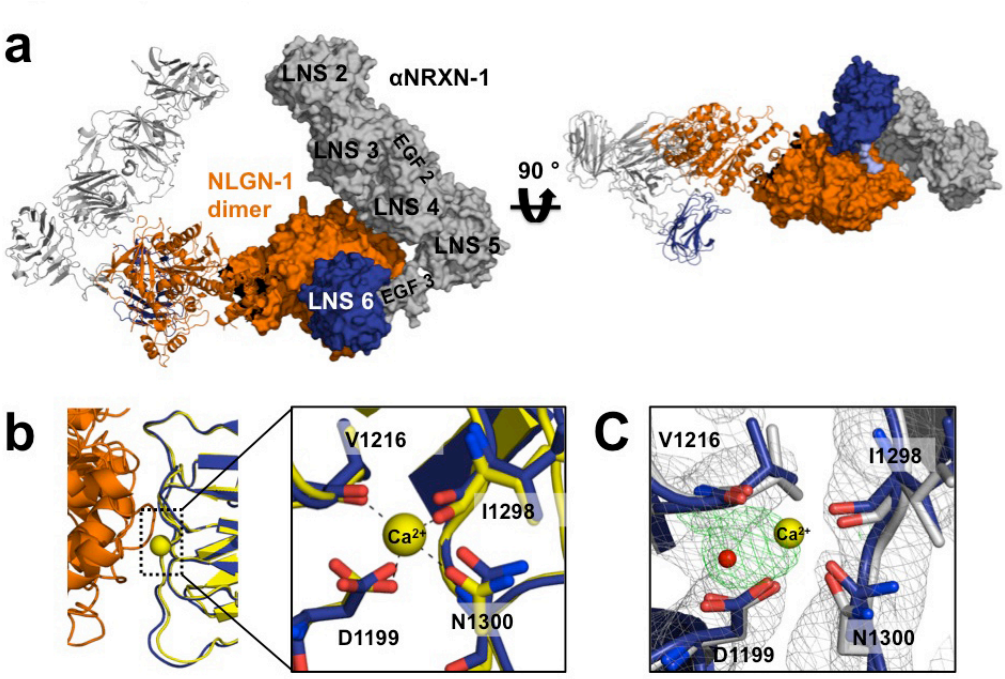

Figure 4.6: Overlay of α-NRXN_2-6 with β-NRXN bound to NLGN-1
(a) The NLGN-1:α-NRXN_2-6 complex, as modeled by overlaying the LNS 6 domain in α-NRXN_2-6 onto the β-NRXN-1 molecule as bound to NLGN-1 α-NRXN_2-6 can accommodate the NLGN molecule without steric constraint, although with close proximity to LNS 4. The NLGN-1 dimer is displayed in orange and α-NRXN_2-6 is shown in gray (LN 2-EGF 3) and blue (LNS 6). On the right panel, the N-linked glycan on LNS 6 is shown in light blue.
(b) The NLGN binding interface of α-NRXN_2-6 is conserved, as shown by the structural overlay with the NGLN-1:β-NRXN structure. In particular, the Ca2+-coordinating residues in β-NRXN (yellow) retain their conformation in the α-NRXN_2-6 LNS 6 domain (blue) despite the absence of Ca2+. Residue numbers are those of α-NRXN_2-6.
(c) Overlay of the Ca2+-free α-NRXN_2-6 molecule with Ca2+-bound β-NRXN-1. The Ca2+and water molecules in the β-NRXN structure are shown as yellow and red spheres, respectively. The 2Fo-Fc maps contoured at 1.2σ are displayed in gray and the Fo-Fc maps contoured at 3.0σ in green. In the absence of Ca2+, the LNS 6 domain in α-NRXN clearly coordinates an ion, which is likely to be sodium. Residue numbers are those of α-NRXN_2-6.

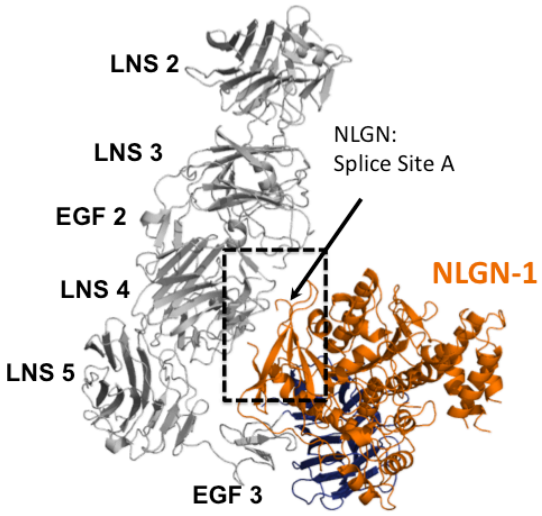

Figure 4.7: Surface proximity in the α-NRXN_2-6 overlay with the NLGN:β-NRXN complex

Overlay of α-NRXN_2-6 on the NLGN1:β-NRXN complex (PDB: 3biw (shown), 3b3q, 2xb6) shows that the α-NRXN molecule (gray) can accommodate a NLGN molecule (orange) without significant steric constraint. However, NLGN proximity to the LNS 4 domain in α-NRXN indicates a region of the α-NRXN molecule that may affect binding kinetics. Furthermore, the position of the NLGN splice insert A (arrow) at the proximal interface is another variant to consider.

Binding Affinity Measurements of the α-NRXN:NLGN-1 Complex

α- and β-NRXNs share the same binding domain for NLGNs [26] and both demonstrate synaptogenic properties in cell culture [36-38]. Yet their evolutionary conservation suggests that they function as independent molecules at the synapse. To address the question of whether α- and β-NRXNs exhibit distinguishing interactions with NLGN, we studied their individual binding properties by SPR. In these studies, the NLGN-1 dimer was immobilized through a C-terminal Fc domain onto a Protein A-coupled sensor chip and the monomeric NRXNs were injected as free ligands.

Soluble constructs of β-NRXN-1 and subsets of the α-NRXN-1 molecule with sequential deletions of each LNS domain (α-NRXN_2-6, α-NRXN_3-6, α-NRXN_4-6, α-NRXN_5-6) were used to analyze binding affinities with NLGN-1 lacking splice inserts A and B (NLGN-1-ΔAB) (**Figure 4.8**). β-NRXN-1 (ΔSS#4) binds to NLGN-1ΔAB with ~30 nM K_D, comparable with previously published results [21, 39]. The α-NRXN_5-6, α-NRXN_4-6 and α-NRXN_3-6 constructs all bind with a ~2-3 fold lower affinity (higher K_D) and similar kinetics involving rapid bimolecular association and unimolecular dissociation (**Figure 4.9 and 4.10**). The α-NRXN_2-6 construct consistently yielded a K_D of ~10-30 nM, reversing the trend of a slight decrease in K_D with the additional domains. When normalized against the maximum binding response (R_{max}), compared with the other truncated α-NRXNs, α-NRXN_2-6 shows the same fast bimolecular association, but displays a different dissociation profile consisting of an initial fast dissociation phase of greater amplitude, followed by a slow secondary dissociation (**Figure 4.10**). Tight clustering in the calculated K_D values across multiple surfaces with various densities of immobilized NLGN-1 attests to reproducibility of the measurements. Additionally, as was previously demonstrated by pull-down studies [26], no binding to NLGN-1 was observed

using the α-NRXN_2-5 construct at concentrations approaching 3 μM (**Figure 4.10**). Although the specific recognition properties of the individual LNS domains remain to be resolved, the comparative equilibrium dissociation constants between the β-NRXN and truncated α-NRXNs demonstrate that both molecules can bind with low nM K_D's (high affinity). The 2-3-fold increase in K_D (decrease in affinity) for α-NRXN corresponds to only minimal differences in free energy requirements for binding and therefore supports the model of a flexible hinge between LNS 5 and EGF 3 in α-NRXN-1.

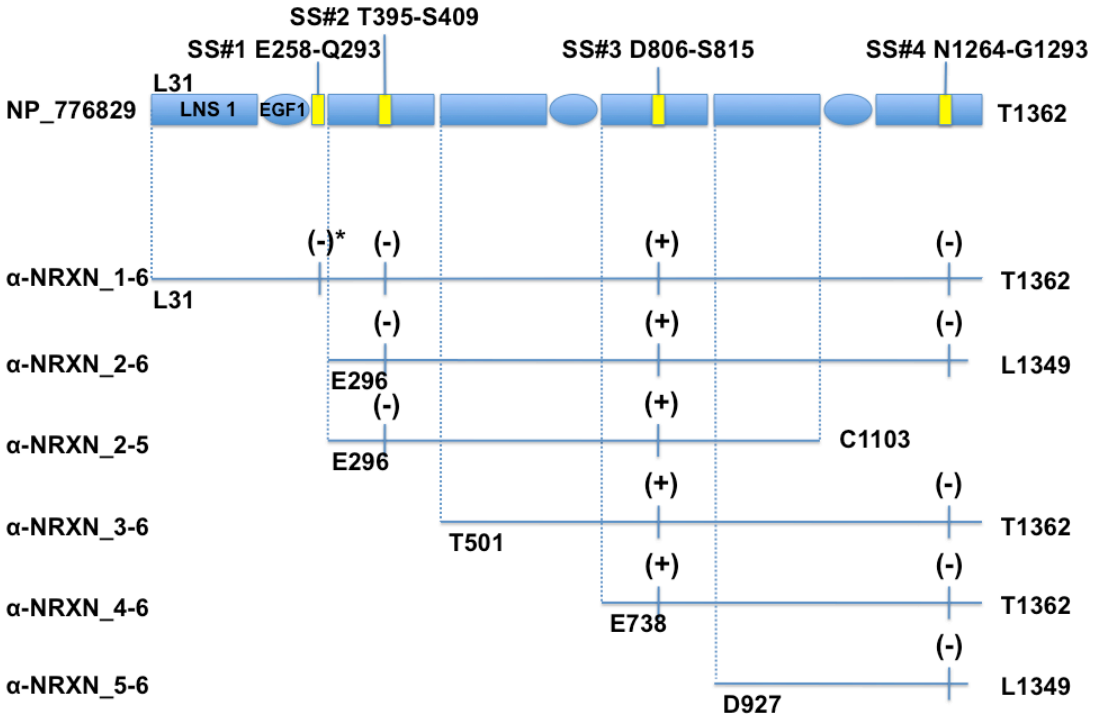

Figure 4.8: The six α-NRXN constructs used in this study
Numbering of the α-NRXN constructs is that of the reference sequence from NP_776829

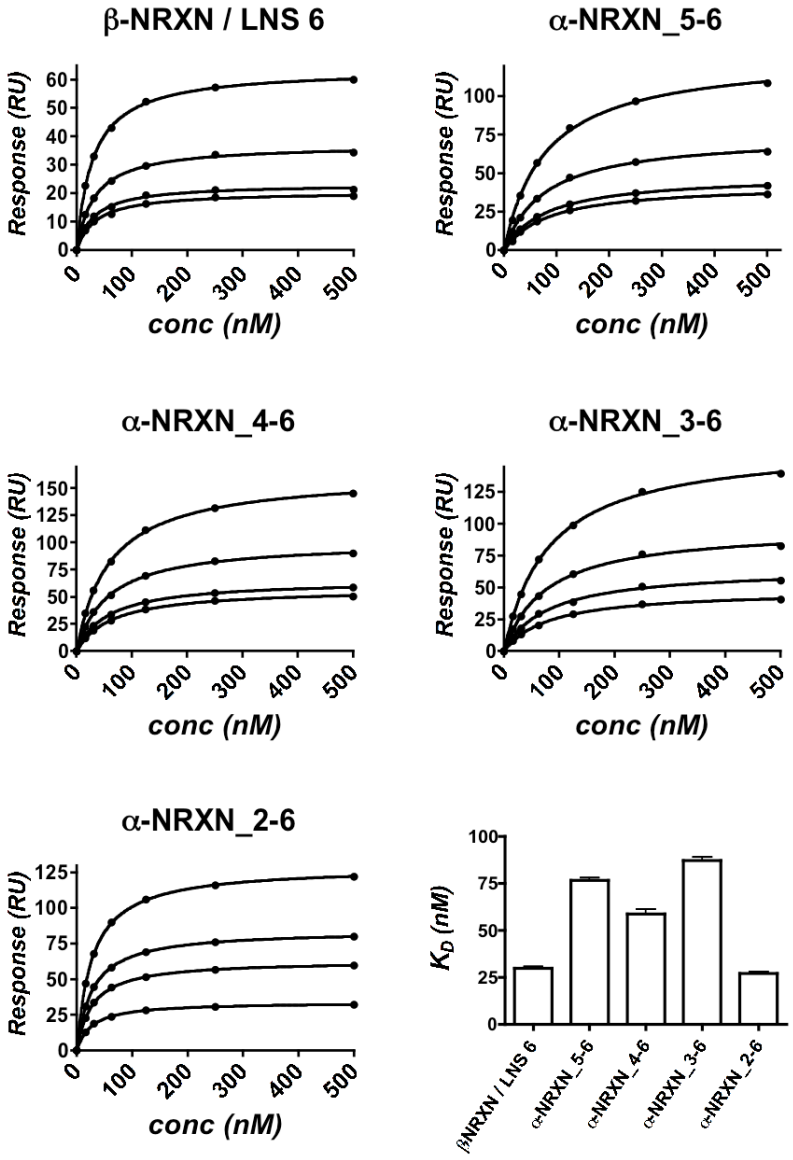

Figure 4.9: NRXN-NGLN association at equilibrium measured by SPR
Each of the five NRXN constructs was injected over four surfaces with a variable density of immobilized NLGN-1. Equilibrium dissociation constants calculated from each of the saturation binding curves are similar across the different NLGN surfaces, as shown by the tight error bars in the lower right graph. All five constructs bind to NLGN-1 with a similar affinity (low nM), indicating accessibility of the NLGN binding site on LNS 6 of the truncated α-NRXN molecules.

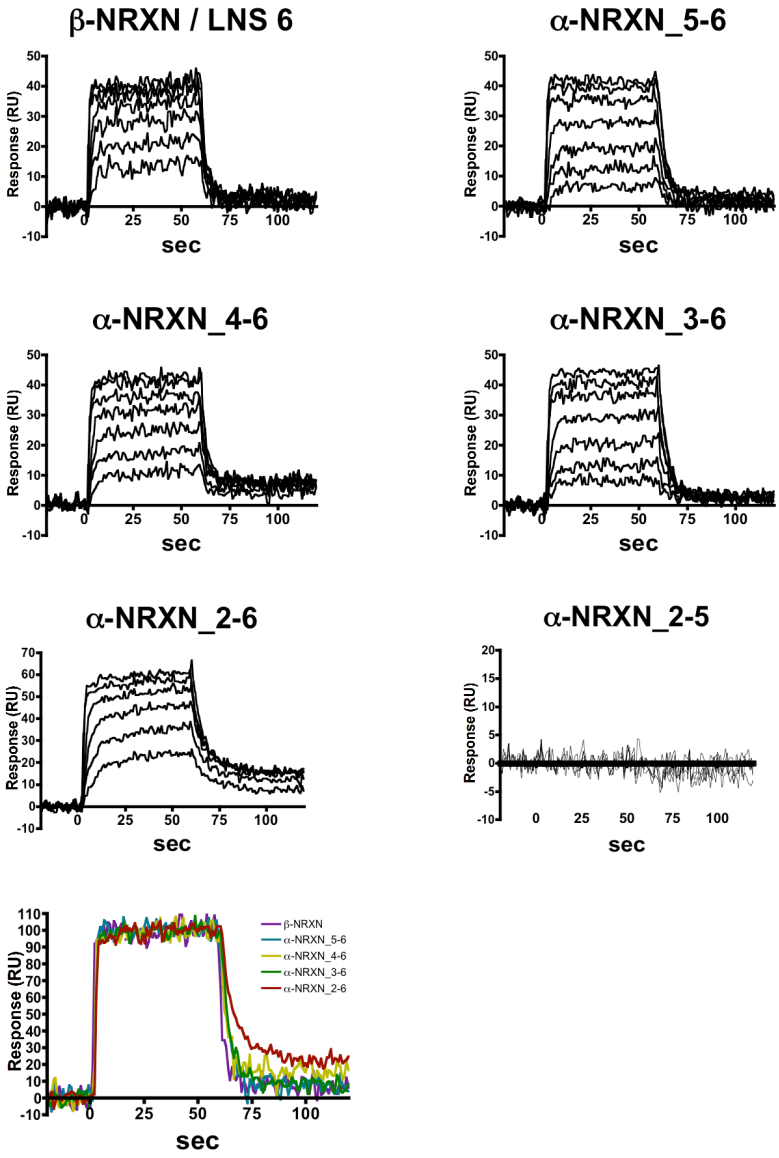

Figure 4.10: SPR analysis of NLGN binding by the a-NRXN variants
Representative binding curves generated for each of the NRXN constructs over a NLGN-1 immobilized surface. Different amounts of NLGN-1 were immobilized to achieve an optimal signal range for the different truncated NRXNs. However, the calculated K_D's did not change significantly based on the surface density (see Figure 4.8). Each graph shows sequential injections of a two-fold dilution gradient from 15.6 nM – 1μM. The α-NRXN_2-5, which shows no binding, was analyzed at a maximum concentration of 3μM. Binding curves for each NRXN truncation normalized against the maximum response (lower panel) highlight similar association and dissociation rates, except for α-NRXN_2-6 which has a secondary slow dissociation phase not seen in the other constructs. All injections intervals were 60s, followed by a 120s dissociation.

4.4 DISCUSSION

The structural characterization of the α-NRXN-1 extracellular region reveals how the concatenated five LNS and two EGF domains assemble into a L-shaped molecule and provides insights into its function as a receptor for several endogenous partnering proteins in the limited dimensions of a synaptic cleft. A complementary SPR analysis shows that α-NRXN binding to NLGN-1 is mostly triggered by the LNS 6/β-NRXN domain, which is likely to be fully accessible in solution due to a flexible linker between LNS 5 and EGF 3. The high resolution structure reveals the accessibility to known protein interaction sites, provides details on the atomic interactions leading to the assembled domains and offers more insight on the flexible properties of the protein compared to the previously determined EM and SAXS structures [25]. Overlaying the α-NRXN_2-6 structure on the 2-D EM images of α-NRXN_1-6 demonstrates a similar overall arrangement of the domains despite the distinctive methodologies, and shows possible orientations of the flexible LNS 1 domain in relation to the other domains (**Figure 4.11**).

Evident in the crystal structure, the extent of inter-domain contacts between LNS 2-5 suggests limited flexibility in their linear arrangement. On the other hand, a lack of stabilizing interactions at the connection between LNS 5 and EGF 3 (P1106) is indicative of innate flexibility at this position in the molecule and suggests that the EGF 3-LNS 6 domains creates a separate arm that can open and close in relation to the rest of the molecule. The remote position of the LNS 6 domain is stabilized by contacts with a symmetry-related molecule in the crystal, and suggests a favorable conformation with increased solvent accessibility to the LNS 6 domain. In agreement with the SPR studies, the dynamic arrangement of the two-arms is likely to influence the binding kinetics with

partner proteins such as NLGN, with the stabilization of an “open” state granting a favorable conformation for association.

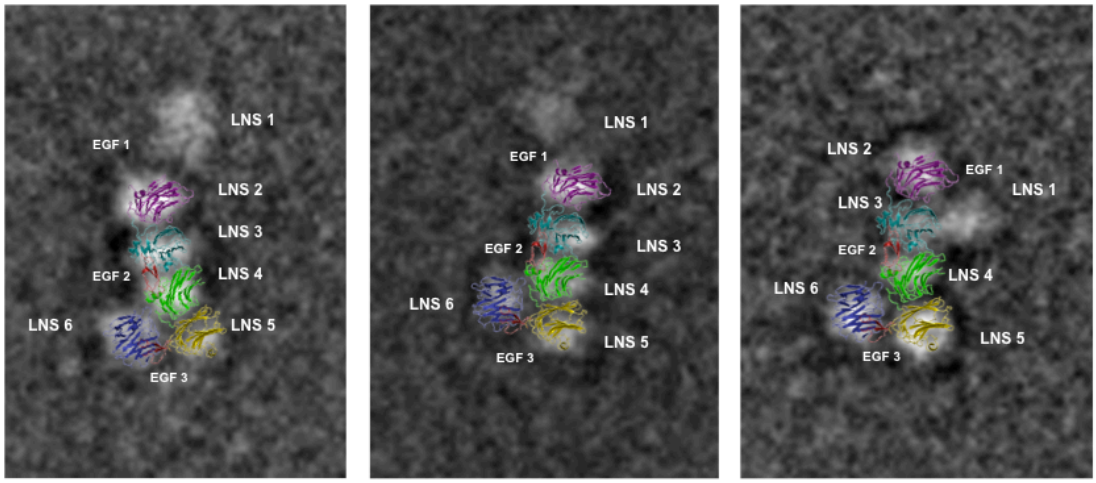

Figure 4.11: 2-D overlay of the α-NRXN_2-6 structure on negative stain EM images
A 2-D image of the α-NRXN_2-6 structure was overlaid onto three previously published 2-D negative stain single particle EM images of the α-NRXN_1-6 purified protein [25]. The overall arrangement of the LNS 2-6 domains is conserved and the multiple placements of the flexible LNS 1 domain relative to the crystal structure are observed.

Predicted Ca^{2+}-Binding Sites in the α-NRXN-1 Extracellular Region

α-NRXN-1 interacts with the post-synaptic NLGNs, LLRTM and α-dystroglycan proteins in a Ca^{2+}-dependent manner through the LNS 2 and/or LNS 6 domains [11, 17, 21]. In the structure, all of the predicted Ca^{2+}-binding sites are aligned on the outer surface of the molecule with similar inter-site distances. The coordinating atoms at the Ca^{2+}-binding sites in LNS domains 2, 3 and 6 have a conserved spatial arrangement based on a structural overlay with the Ca^{2+}-bound structures of the LNS 2, 4 and 6/β-NRXN domains [30, 31, 33]. For the LNS 6 domain, the Fourier difference maps show solvent coordination at the previously reported Ca^{2+}-binding site [32, 33], and may indicate a higher affinity site for divalent cation coordination compared to the other domains. On the other hand, based on a structural overlay, the inclusion of an insert at SS#3 in LNS 4 near the Ca^{2+}-binding site, may reduce the Ca^{2+} affinity by altering the conformation of the coordinating L805 carbonyl oxygen [33]. The loss of measurable Ca^{2+}-binding to the LNS 2 domain due to an insert at SS#2 was previously demonstrated by isothermal titration calorimetry [31]. The predicted Ca^{2+}-binding site in LNS 5 appears intact, however, unlike the LNS 2, 3 and 6 domains that show the conserved spatial distances between the coordinating atoms, the LNS 5 domain has a more open site with increased inter-atomic distances. A definitive analysis of Ca^{2+}-binding of α-NRXN awaits a structure with bound Ca^{2+}.

O-Glycosylation of EGF 2 in α-NRXN-1

The α-NRXN_2-6 structure reveals a previously uncharacterized *O*-linked glycosylation site on the first Ser residue in the EGF 2 domain. The identification of a single *O*-β-glucose moiety was based on the sequence alignment with a conserved *O*-glycosylation consensus motif previously identified in EGF-like domains in other cell-

surface and secreted proteins, including factor VII, factor IX, protein Z and Notch [27, 28]. Two unusual *O*-linked glycans have been mapped to consensus sequences in EGF-like domains, yielding either an *O*-glucose or *O*-fucose post-translational modification [27]. The Notch receptor is made up of a series of EGF-like repeats in which the *O*-fucosylation pattern functions to modulate protein interactions that determine signal transduction capabilities [28, 29]. Many of the EGF domains in Notch contain the *O*-glucose modification, but a specific biological role remains to be determined [40]. In α-NRXN_2-6 the EGF 2 domain contains the consensus sequence for the *O*-glucose modification site. Accordingly, the structure shows a single glucose moiety attached to the EGF 2 domain. Although the preservation of such a moiety in α-NRXN-1 expressed in neurons remains to be determined, we illustrate here the possibility for a new glycosylation pattern of α-NRXN-1 that is likely to influence its function.

Disease-Linked Mutations in the α-NRXN-1 Extracellular Region

An increasing number of genetic studies implicate the *NRXN-1* gene and its partnering *NLGN* genes as important determinants for the pathogenesis of several diseases of the central nervous system, including ASD, schizophrenia, Pitt-Hopkins-like syndrome-2, and milder forms of mental retardation [2, 3, 5, 41-43]. Identified genetic abnormalities include copy number variations [3, 4, 6, 41], chromosomal alterations [2, 5, 42] and a few rare sequence mutations [2, 5, 43, 44]. Reported missense mutations in the NRXN-1 gene include four sites in the α-NRXN-1 leader sequence [2, 5], two in the β-NRXN-1 leader sequence [44], one in the EGF 2 domain [2, 5], two flanking the EGF 2 domain, in LNS 3 and 4 [5] and one in the LNS 5 domain that leads to a premature stop codon [43]. The ASD-linked mutations found on the mature protein are clustered around the EGF 2 domain. They include T688I, L731I and E738K (**Figure 4.12**). Mapping these

disease-linked mutations does not reveal a defining role that should be disrupted by the mutations; however, their proximity may point to an important region of the protein for processing and/or functional regulation. A number of rare structural variants have also been identified in the partnering NLGN proteins [45-47]. One reported NLGN mutation (NLGN 4 G99S), is exposed to a solvent surface of the NLGN dimer and shows proximity to the overlaid α-NRXN molecule to which it may interact in the bound state (**Figure 4.12**).

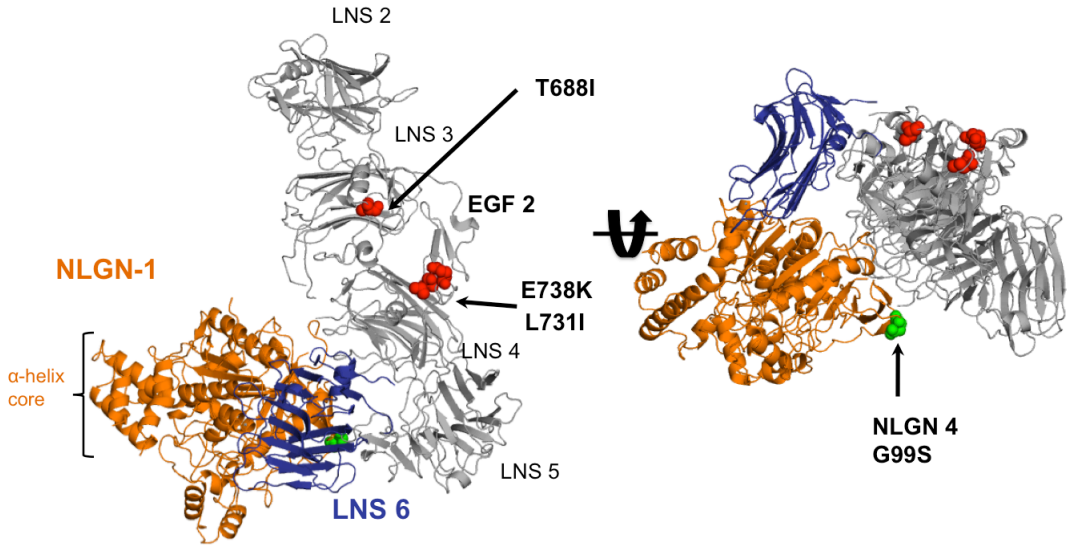

Figure 4.12: Autism-related point mutations at the α-NRXN_2-6 surface.
Overlay of the α-NRXN_2-6 structure onto the NLGN1:β-NRXN complex (PDB: 3biw). NLGN1 is colored in orange, α-NRXN LNS 6 in blue and α-NRXN LNS 2- EGF 3 in gray. The point mutations T665I, L708I, and E715K, which have been found in a small percentage (<1%) of patients with autism spectrum disorders, are colored in red and shown in space-filling representation on the α-NRXN_2-6 molecule. Several point mutations have also been found on the NLGN genes. A glycine to serine substitution on NLGN 4, displayed in green, is proximal to the bound α-NRXN molecule in the overlay.

The α-NRXN:NLGN Complex

α- and β-NRXNs share the same binding domain for the NLGNs; however, it is not fully understood if these comparable interactions are functionally distinct. Binding properties between the NLGNs and β-NRXNs in their various splice isoforms have been extensively studied using SPR [21, 26, 34, 39, 48]. Information on the α-NRXN binding properties with NLGN has been limited to data from affinity pull-down assays and a single SPR experiment comparing the relative affinities of α-NRXN-1 with NLGN-1 [49]. Pull-down experiments concluded that α-NRXN-1, with or without splice insert #4, binds to NLGN-1-3 and is regulated by splice insert B in NLGN-1 and that binding is exclusively mediated by the LNS 6 domain [26]. Another pull-down experiment suggested that α-NRXN-1 binding to NLGN-1 containing splice insert B is regulated by the structural constraints imposed by the disulfide bonds in the EGF 3 domain [49]. By SPR, the relative binding affinity of α-NRXN-1 compared to β-NRXN-1 with NLGN-1 was reported to be ~2-fold less when α-NRXN lacks SS#4 and ~4-fold less with SS#4 [26].

We suggest that the flexibility of the EGF 3-LNS 6 arm has functional implications on the accessibility of the LNS 6 binding site to receptor proteins, including the NLGNs. By superimposing the α-NRXN_2-6 structure on the β-NRXN:NLGN complexes, it becomes apparent that the EGF 3-LNS 6 arm must be in an extended conformation in order to accommodate the NLGN molecule without steric hindrance. Our SPR data show that the sequential addition of LNS domains 2-5 to LNS 6 results in, at most, a 2-3-fold decrease in affinity for NLGN-1 compared to β-NRXN. These variations in the K_D values indicate minimal differences in free energy requirements for binding of the truncated NRXNs. Hence modest conformational adjustments are needed for α-NRXN to

accommodate the NLGN molecule, thereby supporting the propensity for α-NRXN to form an “open” conformation as revealed in the crystal structure.

In the superimposed crystal structure, the LNS 4, 5 and 6 domains fit snugly around the NLGN molecule without evident hindrance, yet with close proximity to the LNS 4 domain. The position of splice insert A in NLGN-1 and -2 at the proximal interface with LNS 4 prompts the question of whether there is a functional role for splice site A in NLGN binding affinities for the α-NRXNs. A previous study showed that the presence of splice insert A in NLGN-1, in the absence of splice insert B, or in NLGN-2 promotes targeting of either neuroligin to GABAergic synapses [50]. The insert at splice site A adds 20 amino acid residues to the structure and is heavily populated with both cationic and anionic residues. The proximity of this loop region in the structural overlay suggests that the combination of bulk and charge of the added sequence would impact NLGN binding to the α-NRXN_2-6 in the conformation shown here. Overall, it remains to be determined if the α-NRXN:NLGN complex stabilizes the L-shape configuration of the domains, or if the two arms would preferentially open up to form an obtuse configuration and if there is a role for splicing in establishing selectivity of interaction. In either case the EGF 3 domain separates the LNS 6 domain from the other five LNS domains in order to accommodate the NLGN molecule.

A Working Model of the α-NRXN:NLGN Complex at the Synapse

Both the α-NRXN and NLGN molecules have long, semi-flexible stalk domains connecting their extracellular binding domains to the cell membrane. The stalk regions contain a relatively large number of serine and threonine residues that are *O*-glycosylated [51]. The glycosylation pattern confers some hydration to the chain and the combination of oligosaccharides and proline amino acid residues is thought to impart

some rigidity to the peptide chain, as demonstrated in NLGN-1 and other cell surface receptors [52, 53].

The crystal structure of α-NRXN_2-6 molecule has a maximum length of ~125 Å. The full extracellular region of α-NRXN has an additional N-terminal LNS domain (LNS 1) and EGF domain (EGF 1), which elongates the soluble protein by ~35 Å, but also shows extensive flexibility [25]. Therefore, depending on the orientations of the LNS 1 domain and the stalk domain, the maximum dimension of the full extracellular region of α-NRXN approaches or possibly exceeds the membrane-to-membrane distance of a typical synaptic cleft (~200 Å). However, in a similar case, structural studies on the large cell-adhesion proteins, cadherins, indicate that they may create repulsive forces that increase the distance between pre- and post-synaptic membranes to approximately 250 Å [54].

Crystal structures of the complexes between the extracellular domains of NLGN and β-NRXN show a NLGN dimer is formed through a hydrophobic four-helix bundle involving the C-terminal α-helices in each subunit with the two C-termini pointing in one direction and two β-NRXN molecules bound on opposite sides of the long axis of the NLGN dimer with their C-termini pointing opposite to the NLGN C-termini [21-23]. The opposing directionality of the C-termini of β-NRXN and NLGN is indicative of their tethering to the pre- and post-synaptic membranes, respectively. Overlaying α-NRXN onto the β-NRXN:NLGN complex positions the long LNS 2-5 arm of α-NRXN perpendicular to the membrane (**Figure 4.13**). If stabilized in this conformation, the N-terminal domains of α-NRXN would lie in close proximity to the pre-synaptic membrane. α-NRXNs are distinct from β-NRXNs in their capacity to regulate Ca^{2+}-dependent exocytosis at the pre-synapse, a role that is believed to be a function of the α-specific

domains [15, 16]. Furthermore, α-NRXNs are receptors for the α-latrotoxin protein and function by recruiting the toxin to the pre-synaptic membrane where it can position itself to stimulate exocytosis [55]. Therefore, a spatial arrangement of α-NRXN with its N-terminus close to the pre-synaptic membrane could favor the observed regulation of Ca^{2+}-dependent neurotransmitter release.

On the other hand, the N-terminal domains of α-NRXN may also function by interacting with proteins on the post-synaptic membrane, as evident from LNS 2 binding to α-dystroglycan [17]. The extent of torsional flexibility in the region between LNS 5 and EGF 3, as dictated by the α-NRXN:NLGN complex, may allow for the long arm of α-NRXN to open into a more linear arrangement, lateral to the membrane, thereby allowing the α-NRXN to form additional contacts with post-synaptic receptor proteins (**Figure 4.13**). The two identified protein-binding domains, LNS 2 and LNS 6, are separated by ~100 Å, which indicates that with the appropriate alignment, both sites could be occupied at the same time. The potential of α-NRXNs to host multiple partnering proteins suggests a possible scaffold-like functionality that would determine the architectural spacing of the molecular network in the trans-synaptic space. Therefore, future studies to address the possible conformations of the α-NRXN domains when bound to NLGN will be of interest.

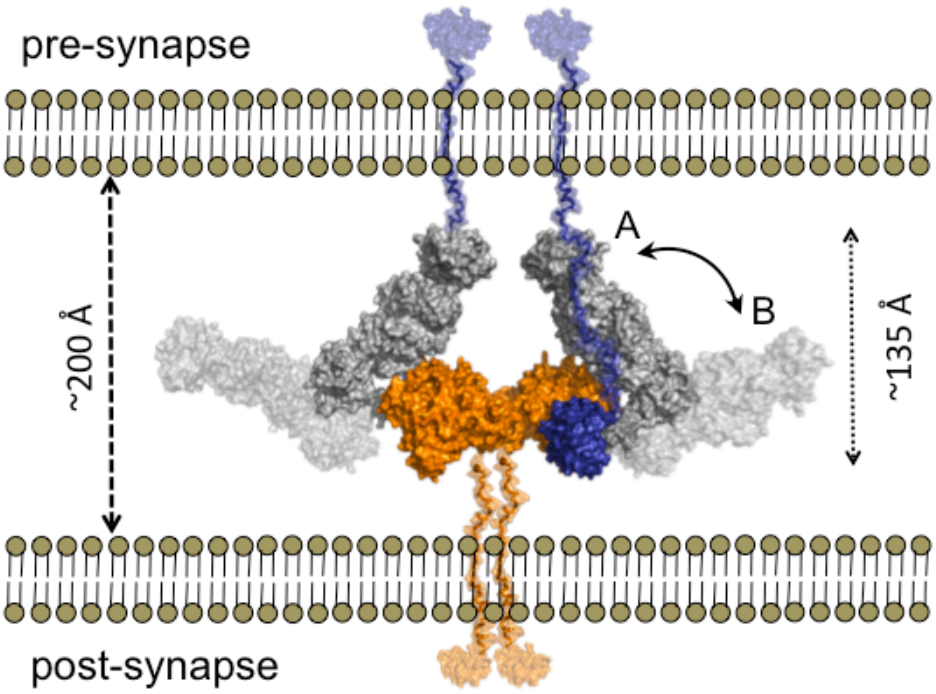

Figure 4.13: Model of the α-NRXN:NLGN complex at the synapse.
Overlays of α-NRXN_2-6 onto β-NRXN-1 as bound to NLGN-1 or -4 suggests a possible position for the α-NRXN:NLGN complex in the synaptic cleft. The overlaid complex is shown with the C-termini of α-NRXN and NLGN oriented towards the pre- and post-synaptic membranes, respectively. In this conformation the N-terminal domains lay to the pre-synaptic membrane (conformation A, non-transparent surface,). Flexibility at the hinge point between LNS 5 and EGF 3 could allow segmental motion of the α-NRXN_2-6 long arm towards a fully linear conformation of the extracellular region, moving the N-terminal domains in closer proximity to the post-synaptic membrane (conformation B, transparent surface). The relative position of the N-terminal domains is likely to confer selectivity for additional binding partners. The NLGN dimer is colored in orange, the α-NRXN LNS 2-EGF 3 in gray and LNS 6 in blue. Stalk domains for NLGN and α-NRXN_2-6, shown in semi-transparent orange and blue, respectively, were modeled into the figure.

4.5 EXPERIMENTAL PROCEDURES

Cloning, Expression and Protein Purification

The original bovine cDNA encoding the secreted, soluble extracellular region of α-NRXN-1 fused with a C-terminal IgG Fc domain to permit affinity chromatography, and inserted in expression vector p-cDNA 3.1, was a gift from Dr. Thomas Sudhof [51]. The construct was further modified to introduce a 3Cpro cleavage site upstream from the Fc domain as described previously [25]. All modified α-NRXN constructs contained the original expression vector, signal peptide sequence and Fc sequence. α-NRXN-1 and its various truncated forms were expressed in the mammalian cell line, HEK293 GnT1-, lacking the N-acetyglucosaminyltransferase I gene needed for high-order N-linked glycosylation processing. The soluble recombinant β-NRXN-1 protein used for SPR studies was expressed in bacterial Rosetta Plys cells as an N-terminally 6-His tagged protein as previously described [22].

NLGN1ΔAB (residues Q46-S693 without splice inserts A or B) was cloned into p-cDNA 3.1 between the sequences encoding the α-NRXN signal peptide and the C-terminal 3Cpro cleavage site, itself followed by the sequence encoding the Fc domain. This construct was transfected into HEK293 cells for expression in selected stable cell lines.

HEK293 GNT1- cells were maintained at 37 °C and 10% CO_2 in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS). For protein expression and purification, cells were expanded into triple layer flasks and the medium was collected and replaced every 1-3 days with low serum medium (2% FBS). The medium containing the secreted, soluble α-NRXN protein (either variant) was centrifuged (6500g, 10 min) to remove cell debris and supplemented with a protease

inhibitor cocktail containing 1mM EDTA, 0.02% NaN_3, 0.01% bacitracin and 0.04% benzamidine. Large-scale purification from the medium was carried out at 4 °C by immunoaffinity chromatography on immobilized Protein A Sepharose (GE Healthcare, USA). We used ~1-1.5 ml of resin per liter of media. Resin was first washed with a high-salt buffer (10 mM HEPES pH 7.4, 450 mM NaCl) and then equilibrated in 10mM HEPES pH 7.4, 150 mM NaCl. α-NRXN was bound to the resin overnight by slow gravity flow (~0.5ml/min). The resin was then washed again with high-salt buffer (above) and equilibrated with TNED elution buffer (50mM Tris-HCl pH 8.0, 150 mM NaCl, 10mM EDTA, 1mM DTT) before addition of GST-tagged PreScission protease (GE healthcare, USA). Proteolytic cleavage proceeded for ~12-18 hrs at 4°C, after which α-NRXN was eluted from the column with the addition of 3-5x column volumes of TNED elution buffer and concentrated using a 30 kD MWCO filter. PreScission protease was removed by incubation with Glutathione Sepharose beads (GE healthcare) prior to size exclusion chromotography. The protein was buffer exchanged by size exclusion chromatography on Superdex 200 in HBS buffer using an AKTA FPLC system (GE Healthcare, USA). The purified protein was concentrated to ~2-6 mg/ml by ultrafiltration.

Crystallization and X-ray Data Collection

α-NRXN-1 containing domains LNS 2 through LNS 6 (α-NRXN_2-6), residues 296-1349 (NP_776829), lacking inserts at SS#2 and #4 but containing the 10-residue splice insert at SS #3 (**Supplemental Fig. 5**) was concentrated to ~3 mg/ml in HBS buffer. Crystals grew as thin rods by sitting drop vapor diffusion at 14°C in 96-well plates (Innovaplate SD-2, Innovadyne Technologies). Drops of 0.6 μl were set at a 1:1 (v/v) ratio of protein to reservoir buffer (0.1M Bicine pH 9.0, 12% PEG 20,000, 1mM EDTA). Crystals were soaked into the mother liquor complemented with 10-25% glycerol for

several minutes before being flash cooled in liquid nitrogen. X-ray diffraction data were collected at a temperature of 100 K from the 11-1 beamline at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA). Diffraction data were processed with XDS [56].

The crystals belong to space group C2 (cell dimensions a=199.74Å b=61.24Å c=155.58Å α=90.0º β=121.21º γ=90.0º) and contain one α-NRXN_2-6 molecule per asymmetric unit with a solvent content of 65%. The large solvent channels seen in the crystal packing arrangement probably contributed to the difficulties encountered in obtaining well ordered, highly diffracting crystals (Figure 4.3 A).

Extensive attempts made to crystallize the protein in the presence of Ca^{2+} were unsuccessful. Crystals were grown under the Ca^{2+}-free conditions reported here, both with and without the additive EDTA. Soaking the crystals grown without EDTA in a cryo buffer containing 1-10mM Ca^{2+} did not yield high quality diffraction patterns. The majority of the crystals had a very thin third dimension and were either too fragile to harvest or produced weak, anisotropic diffraction in the 4-8 Å range.

Structure Determination, Refinement and Validation

The structure of α-NRXN_2-6 was determined at 3.0 Å resolution by molecular replacement using the coordinates of previously solved structures for LNS 2 (PDB 2h0b), LNS 4 (PDB 2r16) and β-NRXN (PDB 3bod) as search models for input into PHASER [57]. An initial search using the three PDB structures successfully located the LNS 2, 4 and 6 domains. Using the phase information from the previous search, the LNS 3 and 5 domains were located using the coordinates of the LNS 2 and LNS 4 structures, respectively, as the search models in PHASER. Building of LNS 3 and 5 was achieved through subsequent cycles of modeling and refinement. After rigid body

refinement of all LNS domains, the EGF-like domains were manually built into the difference Fourier maps using Coot. In the final refinement stages each LNS and EGF domain was used to define seven TLS groups for refinement. All model building and refinement was performed in Coot [58] and Refmac5 from the CCP4 program suite [59]. A combination of Whatcheck [60] and Molprobity [61] were used for structure validation.

The structure was refined to a final R_{work} = 0.21 and R_{free} = 0.27 and shows 1,005 amino acid residues. It shows no density for residues K551-D552 in LNS 3 and R809-S815 in LNS 4. The latter sequence corresponds to most of SS#3. The structure also includes 33 water molecules, one β-Glc moiety *O*-linked to S705 in the EGF 2 domain, and part of an N-linked glycan bound to N1246 in LNS 6, fitting the electron density of 2 molecules of GlcNAc and one Man.

Structural Analysis

All analyses of the domains interfaces used the PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) [62]. Average rmsds of individual LNS domains were calculated using the secondary structure matching module in COOT [58]. Overlays of the previously solved LNS 2, 4 and 6/β-NRXN domains (PDB: 2h0b, 2r16 and 3bod) onto their respective domains in the α-NRXN_2-6 structure gave average rmsd values between 0.5 and 0.9 Å between Cα atoms. Comparing the five individual LNS domains in the α-NRXN_2-6 structure yielded average rmsds of 1.42-2.3 Å for Cα atoms. Individual Cα rmsds between the LNS domain in α-NRXN_2-6 were calculated using ClustalW [63] and the THESEUS software package [64]. Structural overlays of α-NRXN_2-6 with the β-NRXN:NLGN-1 or β-NRXN:NLGN-4 complexes (PDB: 3biw and 2xb6, respectively) and of the NLGN-2 dimer (PDB:3bl8) with the β-NRXN:NLGN-1 complex were performed using the secondary-structure matching (SSM) module in Coot.

All structure figures and electrostatics calculations were generated using PyMol (DeLano Scientific, LLC).

Surface Plasmon Resonance

For details of the NRXN proteins used for the SPR studies see figure 4.8. All NRXN proteins were analyzed for binding against the NLGN-1ΔAB isoform. Affinity binding analysis was performed on a ProteON X36 biosensor with the GLC chip platform (Bio-Rad Laboratories, Inc.) in a running buffer made of HBS, 3mM Ca^{2+}, 0.05% Tween 20, 1mg/mL bovine serum albumin, at 25 °C. For more details see supplemental material.

The GLC chip was equilibrated in HBS buffer at 25 °C for immobilization of NLGN to the mono-layered surface of the sensor chip. Protein A was immobilized on the biosensor chip surface in the vertical direction using amine coupling chemistry [65]. The surfaces were deactivated with 1M ethanolamine and washed with glycine pH 2.0. NLGN-1ΔAB-Fc was captured over the Protein A surface from filtered medium of cultured HEK293 cells (injected in vertical orientation at 30 μl/min). One channel was dedicated as a control where medium from non-transfected HEK293 cells was injected. Different dilutions of the medium were used to achieve variable amounts of NLGN-1 on the surface, which was necessary to achieve an optimal signal level (RUs) for the various NRXN analytes, which ranged from ~22 kD for β-NRXN to ~140 kD for α-NRXN_1-6.

Affinity binding analysis was performed in a running buffer made of HBS, 3mM Ca^{2+}, 0.05% Tween 20, 1mg/mL BSA at 25 °C. In the experiment shown here each NRXN protein was prepared in a two-fold dilution sequence in running buffer starting at a concentration of 15.6 nM and ending at 1 μM. Additional experiments were performed

using a three-fold concentration gradient going from 1.2 nM to 900 nM. For each injection series, one channel was injected with running buffer to use as a double reference [66]. Samples were injected orthogonal to the NLGN surfaces for 60 s, at a flow rate of 50 µl/min, followed by a 600 s dissociation phase. For each injection the NRXN analyte was exposed to 5 independent surfaces with NLGN-1ΔAB and one surface with only Protein A. The surface was regenerated with an 18 s injection of 350mM EDTA, at 100 µl/min after each injection, followed by a 60 s injection of running buffer at 100 µl/min. The lack of binding on control Protein A surfaces upon injection of culture medium from non-transfected HEK293 cells or containing α-NRXN_1-6 with a C-terminal Fc fragment verified the absence of appreciable non-specific binding. The graphs shown here were generated using GraphPad Prism v.4.0b (Graphpad Software, La Jolla, CA).

4.6 REFERENCES

1. Abrahams, B.S. and D.H. Geschwind, *Advances in autism genetics: on the threshold of a new neurobiology.* Nat Rev Genet, 2008. **9**(5): p. 341-55.

2. Kim, H.G., et al., *Disruption of neurexin 1 associated with autism spectrum disorder.* Am J Hum Genet, 2008. **82**(1): p. 199-207.

3. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.

4. Szatmari, P., et al., *Mapping autism risk loci using genetic linkage and chromosomal rearrangements.* Nat Genet, 2007. **39**(3): p. 319-28.

5. Yan, J., et al., *Neurexin 1alpha structural variants associated with autism.* Neurosci Lett, 2008. **438**(3): p. 368-70.

6. Glessner, J.T., et al., *Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.* Nature, 2009. **459**(7246): p. 569-73.

7. Ullrich, B., Y.A. Ushkaryov, and T.C. Sudhof, *Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons.* Neuron, 1995. **14**(3): p. 497-507.

8. Tabuchi, K. and T.C. Sudhof, *Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing.* Genomics, 2002. **79**(6): p. 849-59.

9. Ichtchenko, K., T. Nguyen, and T.C. Sudhof, *Structures, alternative splicing, and neurexin binding of multiple neuroligins.* J Biol Chem, 1996. **271**(5): p. 2676-82.

10. de Wit, J., et al., *LRRTM2 interacts with Neurexin1 and regulates excitatory synapse formation.* Neuron, 2009. **64**(6): p. 799-806.

11. Ko, J., et al., *LRRTM2 functions as a neurexin ligand in promoting excitatory synapse formation.* Neuron, 2009. **64**(6): p. 791-8.

12. Cheng, S.B., et al., *Presynaptic targeting of alpha4beta 2 nicotinic acetylcholine receptors is regulated by neurexin-1beta.* J Biol Chem, 2009. **284**(35): p. 23251-9.

13. Zhang, C., et al., *Neurexins physically and functionally interact with GABA(A) receptors.* Neuron, 2010. **66**(3): p. 403-16.

14. Davletov, B.A., et al., *High affinity binding of alpha-latrotoxin to recombinant neurexin I alpha.* J Biol Chem, 1995. **270**(41): p. 23903-5.

15. Missler, M., et al., *Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis.* Nature, 2003. **423**(6943): p. 939-48.

16. Zhang, W., et al., *Extracellular domains of alpha-neurexins participate in regulating synaptic transmission by selectively affecting N- and P/Q-type Ca2+ channels.* J Neurosci, 2005. **25**(17): p. 4330-42.

17. Sugita, S., et al., *A stoichiometric complex of neurexins and dystroglycan in brain.* J Cell Biol, 2001. **154**(2): p. 435-45.

18. Missler, M., R.E. Hammer, and T.C. Sudhof, *Neurexophilin binding to alpha-neurexins. A single LNS domain functions as an independently folding ligand-binding unit.* J Biol Chem, 1998. **273**(52): p. 34716-23.

19. Song, J.Y., et al., *Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses.* Proc Natl Acad Sci U S A, 1999. **96**(3): p. 1100-5.

20. Varoqueaux, F., S. Jamain, and N. Brose, *Neuroligin 2 is exclusively localized to inhibitory synapses.* Eur J Cell Biol, 2004. **83**(9): p. 449-56.

21. Arac, D., et al., *Structures of neuroligin-1 and the neuroligin-1/neurexin-1 beta complex reveal specific protein-protein and protein-Ca2+ interactions.* Neuron, 2007. **56**(6): p. 992-1003.

22. Fabrichny, I.P., et al., *Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion.* Neuron, 2007. **56**(6): p. 979-91.

23. Chen, X., et al., *Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions.* Nat Struct Mol Biol, 2008. **15**(1): p. 50-6.

24. Leone, P., et al., *Structural insights into the exquisite selectivity of neurexin/neuroligin synaptic interactions.* EMBO J, 2010. **29**(14): p. 2461-71.

25. Comoletti, D., et al., *The macromolecular architecture of extracellular domain of alphaNRXN1: domain organization, flexibility, and insights into trans-synaptic disposition.* Structure, 2010. **18**(8): p. 1044-53.

26. Boucard, A.A., et al., *A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins.* Neuron, 2005. **48**(2): p. 229-36.

27. Harris, R.J. and M.W. Spellman, *O-linked fucose and other post-translational modifications unique to EGF modules.* Glycobiology, 1993. **3**(3): p. 219-24.

28. Haines, N. and K.D. Irvine, *Glycosylation regulates Notch signalling.* Nat Rev Mol Cell Biol, 2003. **4**(10): p. 786-97.

29. Okajima, T., A. Matsuura, and T. Matsuda, *Biological functions of glycosyltransferase genes involved in O-fucose glycan synthesis.* J Biochem, 2008. **144**(1): p. 1-6.

30. Rudenko, G., et al., *The structure of the ligand-binding domain of neurexin Ibeta: regulation of LNS domain function by alternative splicing.* Cell, 1999. **99**(1): p. 93-101.

31. Sheckler, L.R., et al., *Crystal structure of the second LNS/LG domain from neurexin 1alpha: Ca2+ binding and the effects of alternative splicing.* J Biol Chem, 2006. **281**(32): p. 22896-905.

32. Koehnke, J., et al., *Crystal structures of beta-neurexin 1 and beta-neurexin 2 ectodomains and dynamics of splice insertion sequence 4.* Structure, 2008. **16**(3): p. 410-21.

33. Shen, K.C., et al., *Regulation of neurexin 1beta tertiary structure and ligand binding through alternative splicing.* Structure, 2008. **16**(3): p. 422-31.

34. Koehnke, J., et al., *Splice form dependence of beta-neurexin/neuroligin binding interactions.* Neuron, 2010. **67**(1): p. 61-74.

35. Rini, J.M., *Lectin structure.* Annu Rev Biophys Biomol Struct, 1995. **24**: p. 551-77.

36. Graf, E.R., et al., *Structure function and splice site analysis of the synaptogenic activity of the neurexin-1 beta LNS domain.* J Neurosci, 2006. **26**(16): p. 4256-65.

37. Graf, E.R., et al., *Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins.* Cell, 2004. **119**(7): p. 1013-26.

38. Kang, Y., et al., *Induction of GABAergic postsynaptic differentiation by alpha-neurexins.* J Biol Chem, 2008. **283**(4): p. 2323-34.

39. Comoletti, D., et al., *Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for beta-neurexins.* Biochemistry, 2006. **45**(42): p. 12816-27.

40. Moloney, D.J., et al., *Fringe is a glycosyltransferase that modifies Notch.* Nature, 2000. **406**(6794): p. 369-75.

41. Rujescu, D., et al., *Disruption of the neurexin 1 gene is associated with schizophrenia.* Hum Mol Genet, 2009. **18**(5): p. 988-96.

42. Zahir, F.R., et al., *A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha.* J Med Genet, 2008. **45**(4): p. 239-43.

43. Zweier, C., et al., *CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila.* Am J Hum Genet, 2009. **85**(5): p. 655-66.

44. Feng, J., et al., *High frequency of neurexin 1beta signal peptide structural variants in patients with autism.* Neurosci Lett, 2006. **409**(1): p. 10-3.

45. Yan, J., et al., *Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients.* Mol Psychiatry, 2005. **10**(4): p. 329-32.

46. Comoletti, D., et al., *The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing.* J Neurosci, 2004. **24**(20): p. 4889-93.

47. Talebizadeh, Z., et al., *Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism.* J Med Genet, 2006. **43**(5): p. e21.

48. Comoletti, D., et al., *Characterization of the interaction of a recombinant soluble neuroligin-1 with neurexin-1beta.* J Biol Chem, 2003. **278**(50): p. 50497-505.

49. Reissner, C., et al., *Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components.* Proc Natl Acad Sci U S A, 2008. **105**(39): p. 15124-9.

50. Chih, B., L. Gollan, and P. Scheiffele, *Alternative splicing controls selective trans-synaptic interactions of the neuroligin-neurexin complex.* Neuron, 2006. **51**(2): p. 171-8.

51. Ushkaryov, Y.A., et al., *Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin.* Science, 1992. **257**(5066): p. 50-6.

52. Comoletti, D., et al., *Synaptic arrangement of the neuroligin/beta-neurexin complex revealed by X-ray and neutron scattering.* Structure, 2007. **15**(6): p. 693-705.

53. Merry, A.H., et al., *O-glycan sialylation and the structure of the stalk-like region of the T cell co-receptor CD8.* J Biol Chem, 2003. **278**(29): p. 27119-28.

54. Craig, A.M., E.R. Graf, and M.W. Linhoff, *How to build a central synapse: clues from cell culture.* Trends Neurosci, 2006. **29**(1): p. 8-20.

55. Sudhof, T.C., *alpha-Latrotoxin and its receptors: neurexins and CIRL/latrophilins.* Annu Rev Neurosci, 2001. **24**: p. 933-62.

56. Kabsch, W., *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants* J. Appl. Cryst., 1993. **26**: p. 795-800.

57. McCoy, A.J., et al., *Phaser crystallographic software.* J. Appl. Cryst. , 2007. **40**: p. 658-674.

58. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

59. CCP4, *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.

60. Hooft, R.W., et al., *Errors in protein structures.* Nature, 1996. **381**(6580): p. 272.

61. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

62. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.

63. Chenna, R., et al., *Multiple sequence alignment with the Clustal series of programs.* Nucleic Acids Res, 2003. **31**(13): p. 3497-500.

64. Theobald, D.L. and D.S. Wuttke, *THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures.* Bioinformatics, 2006. **22**(17): p. 2171-2.

65. Rich, R.L. and D.G. Myszka, *BIACORE J: a new platform for routine biomolecular interaction analysis.* J Mol Recognit, 2001. **14**(4): p. 223-8.

66. Myszka, D.G., *Improving biosensor analysis.* J Mol Recognit, 1999. **12**(5): p. 279-84.

ACKNOWLEDGEMENTS

A manuscript including **Chapter 4** authored by Meghan T. Miller, Mauro Mileni, Davide Comoletti, Raymond C. Stevens, Michal Harel, and Palmer Taylor and has been submitted for publication. The dissertation author was the primary investigator in the development and execution of this study, and the principle author of the paper. This work encompassed the construction of expression vectors for various truncations of the α-NRXN extracellular region, crystallization of the protein, structure determination of the α-NRXN_2-6 construct and analysis of NLGN-NRXN association by SPR.

Accession codes - Atomic coordinates and structure factors have been deposited with the Protein Data Bank: accession code 3POY

CHAPTER 5:

Conclusions and Future Directions

A growing body of evidence indicates that adhesion molecules not only function as a physical link between cells, but also play an integral role in synaptogenesis and synapse maintenance through the bi-directional recruitment and stabilization of defining molecules [1-3]. The overall goal of the work presented here was to gain insights into the structural characteristics of two important synaptic adhesion molecules, NRXN and NLGN, in order to better understand their heterophilic interaction properties. A major question that bears relevance to this work is how the highly conserved NRXN family, which consists of the larger α- and smaller β-NRXN isoforms, is structurally adapted to confer functional distinctions at the synapse. Using a combination of structural techniques, I have examined the relationship between the complexes of the extracellular region of α- and β-NRXN with NLGN.

Pre-synaptic NRXNs and post-synaptic NLGNs form a heterophilic trans-synaptic adhesion complex through their globular extracellular domains. They are localized to excitatory and inhibitory synapses and are thought to distinguish between the two by a complex genetic code that yields a diverse number of isoforms [4-8]. To that end, their role in determining the ratio of excitatory to inhibitory synaptic connections is a major question in the field of neurobiology. Disruption of the excitatory/inhibitory balance is thought to be one of the main determinants for several neurological disorders, in particular, the ASDs [9]. Understanding the structural characteristics that lead to a functional distinction between NRXN:NLGN complexes is prudent for the realization of their individual roles and how perturbations may alter homeostasis.

In the first body of work I have described the molecular interaction between NLGN 4 and β-NRXN-1 which forms a Ca^{2+}-dependant complex consisting of a homodimer of two NLGN molecule bound to two β-NRXN molecules, one on either side of the long axis of the dimer. This work supplements parallel works by other groups, which showed an identical stoichiometric relationship and similar binding interface between NLGN 1 and β-NRXN-1 [10, 11]. The structures revealed key features that distinguish the NLGNs from the homologous AChE enzyme, making them catalytically inactive and imparting their adhesive properties. The organization of the heterotetramer positions the start of the C-terminal stalk region of the NRXN and NLGN molecules at opposite faces, in agreement with the bi-directional tethering to the pre- and post-synaptic membranes, respectively. The structures further elucidate the relative positions of splice site B in NLGN-1 and splice site #4 in β-NRXN, which are shown to regulate binding [5, 6, 12]. This work provides a detailed analysis of the interaction between β-NRXN and NLGN, however, some key questions remain to be answered.

Studies seem to indicate that NLGN 2 and 3 bind to β-NRXN with much lower affinity compared to NLGN 1 and 4 [6]. This is striking due to the highly conserved binding interface between all of the NLGNs [10, 13]. It has been proposed that that the loss of affinity is partially due to modifications at residues on the edge of the binding interface, however structural characterization or mutagenesis studies are needed to confirm this analysis.

In relation to a proposed environmental influence on the etiology of ASD, another general question relates to the ability of exogenous chemical compounds disrupting the Ca^{2+}-dependent interaction and thereby influence synapse connectivity. An in depth study of chemical perturbants may provide new insights into the increasing prevalence of ASD disorders.

The second study in this work describes the macromolecular assembly of the α-NRXN-1 extracellular region using two different techniques, SAXS and single particle EM. Compared to β-NRXN, α-NRXN has a complex extracellular region made up of nine individually folding sub-domains, including six LNS domains interspersed by three EGF-like domains. Prior to this work three of the LNS domains were crystallized as individual entities, LNS 2, 4 and 6, where LNS 6 is represented by the β-NRXN structure [13-16]. The structures revealed a highly conserved β-sheet sandwich, characteristic of LNS domains, and a Ca^{2+}-binding site at analogous positions on the edge of each domain, which in the case of LNS 6/β-NRXN is essential for binding to NLGN. However, the structures of the individual domains do not provide information on how the assembled molecule governs protein-protein interactions. Unlike small molecules, protein-protein interactions are dependent on the macromolecular structure in order to accommodate binding without steric hindrance. Another question surrounding the assembly of the full α-NRXN molecule is how the large, multi-domain extracellular region fits into the limited space of a synaptic cleft and interacts with other large proteins. Using the truncated, soluble extracellular region of α-NRXN-1, SAXS and single particle EM revealed that α-NRXN adopts a semi-elongated shape with varying degrees of flexibility between the independent LNS domains. By directed antibodies against the N- and C-terminal domains, I was able to orient the molecule was further determine that the N-terminal LNS 1 domain is highly flexible. Interestingly, in conjunction with these experiments, analysis on SDS-page gels showed that the α-NRXN_1-6 protein was subject to degradation, over time showing two additional distinct bands at lower molecular weight. Analysis by Edman degradation sequencing identified a specific site of degradation between the EGF 1 and LNS 2 domain, further supporting an open, flexible region in the protein. Removal of the LNS 1 and EGF 1 domain yielded a stable

protein that maintained the structural assembly identified in the full construct. The combination of SAXS and single particle EM offered a complementary analysis in both a solid state and solution based environment, which demonstrated the integrity of the structural results by either method.

In chapter 4, the story is pulled together with the solved crystal structure of α-NRXN_2-6 at a resolution of 3.0 Å. The asymmetric, flexible nature of the α-NRXN protein made crystallization a challenging endeavor. The success of this project was guided by the knowledge from the previous low-resolution studies, which revealed points of extensive flexibility and proteolysis, and led to the design of an optimal construct for crystallization.

The α-NRXN structure, the first of a multi-domain member of the extended NRXN family, reveals a unique asymmetric arrangement of the domains such that the two known protein-interaction domains are located at distal ends of the two arms of the molecule. It is further apparent that the core of the molecule, consisting of LNS 2-5, is arranged in a compact linear formation, held together by extensive inter-domain contacts, and the C-terminal LNS 6 domain is remote from the other LNS domains, separated by the intervening EGF 3 domain. A flexible hinge at the junction between LNS 5 and EGF 3 marks the bifurcation point. Interestingly, the conformation of the LNS 6 domain is aptly suited to accommodate a NLGN molecule bound at the previously described β-NRXN site, although notably with close proximity to the LNS 4 domain. The conformation of the long arm of α-NRXN (LSN 2-5) when bound to NLGN has functional implications on the function of the molecule as a receptor for other endogenous proteins. For a discussion of the arrangement of the domains and functional implications please refer to Chapter 4.

Initial studies of the binding energetics of β-NRXN compared to α-NRXN for NLGN-1 indicate comparable binding affinities, supporting the open conformation of the LNS 6 domain described by the crystal structure. A more in depth analysis of the binding properties of the different α-NRXNs for the different NLGNs is warranted. Such studies may reveal important contributions by the non-binding LNS domains of α-NRXN and/or splice variations. Of particular interest is splice site A in NLGN, which shows proximity to the α-NRXN LNS 4 domain when the α-NRXN structure is overlaid onto the β-NRXN:NLGN complex structures.

Another observation that comes out of the crystal structure of α-NRXN and may indicate a region of functional significance, is the identification of an *O*-glycosylation site on the EGF 2 domain. Unlike the EGF 3 domain, which lies directly between the LNS 5 and LNS 6 domains, the EGF 2 domain is set in proximal apposition to the flanking LNS 3 and LNS 4 domains making it more superficial. The Fourier difference maps of the structure revealed extra density at a Ser residue in the N-terminal sequence of the EGF 2 domain, which aligns with an *O*-glycosylation sequence motif found in other EGF domain of extracellular proteins and is thought to function in protein-protein recognition [17, 18]. It is also of interest to note that a few rare point mutations in the α-NRXN-1 molecule in patients with ASD, all localize to the region surrounding the EGF 2 domain [19, 20]. Mapping these disease-linked mutations does not reveal a defining role in the structural organization of the protein that should be disrupted by the mutations; however, their proximity may point to an important region of the protein for processing and/or functional regulation.

Through the combined structural studies of the α-NRXN extracellular region we now have valuable insights into the spatial arrangement of the assembled multi-domain protein and how this may direct protein-protein interactions. The solution of the crystal

structure of a large portion of the α-NRXN extracellular region is a significant contribution to the study of synapse biology. It provides a new template from which we can design molecular tools to study protein function and possibly one-day therapeutic applications. The structure revealed many new observations and set the stage for future continuation of the exploration of the NRXNs. The remainder of this thesis will focus on a discussion of some of the ongoing questions and directions for future studies.

5.1 Building on the α-NRXN structure

By removing the flexible and unstructured portions of the extracellular region of α-NRXN-1, I was able to obtain crystals that were ordered enough to yield near-atomic resolution diffraction. However, this does not provide the entire picture. The full extracellular region of α-NRXN contains an additional LNS and EGF domain on the N-terminus, and a long (~100 amino acid) stalk domain that extends from the C-terminus of LNS 6 to the transmembrane domain. These domains are likely to have significant features that contribute to the overall function of the protein.

In the SAXS and single particle EM studies of the α-NRXN_1-6 molecule, it was revealed that the LNS 1 domain, and probably also the EGF 1 domain, exhibits a high degree of mobility, moving freely around the other domains and therefore indicating a flexible linker. We propose that this flexibility, which is more distinct than the flexibility between LNS 5 and EGF 3, is imparted by the addition of a splice variant at SS #1 in between the EGF 1 and LNS 2 domains. Through the initial work on the α-NRXN_1-6 construct, I discovered that this splice insert contained a proteolysis site, further suggesting its open, flexible state. Based on this observation, I have designed a construct, which removes the 20 amino acid splice variant. Future, structural studies may reveal a more stable conformation in the absence of a splice insert. Removal of

this highly flexible insert may also facilitate crystallization of all or part of α-NRXN containing the LNS 1 and EGF 1 domains. Because LNS 2 is a site of protein-interaction, there is likely to be a biological significance of the conformation of the LNS 1 and EGF 1 domains. Splice site #1 in α-NRXN-1 is one of the most variable among all splice sites in all NRXN genes, having 24 possible splice variants [8]. This variability may impart a number of functionally distinct properties on the α-NRXN molecule. Studies of the splice isoforms have been limited to the splice variants at SS#2 and #4 on LNS 2 and 6, respectively [4-6, 12, 14]. This is primarily because LNS 2 and 6 are the known binding domains of α-NRXN, however, an inclusive study of the other splice variants is needed in order to determine their role in the function of the molecule.

The stalk domain, which is identical in both the α- and β-NRXNs and is extensively *O*-glycosylated, is thought to impart flexibility in the disposition of the rest of the extracellular region relative to the membrane in order to allow binding to post-synaptic elements, such as the NLGNs [21, 22]. NLGNs have a similar stalk domain, which is described as a physically distinct linker [21]. It is expected that, in α- and β-NRXNs, this domain will display similar structural characteristics.

For α-NRXN, a structural question that remains is whether stabilizing interactions are used to guide the placement of the N-terminal domains towards either the pre- or post-synaptic membrane. The “L” shaped conformation that is observed in the crystal structure is directed by the packing of a symmetry related LNS domain at the bifurcation point, and suggests that there is flexibility in the spatial arrangement between the two arms of the protein. The overlay of α-NRXN_2-6 onto the β-NRX:NLGN complex shows that the α-NRXN molecule can accommodate a NLGN molecule in the observed conformation, and further suggests the potential for a stabilizing interaction between the long arm of α-NRXN and the NLGN molecule. In the discussion in Chapter 4, a possible

functional significance of having the N-terminal LNS domains proximal to the pre-synaptic membrane is proposed. A more in depth study of interactions that may affect the disposition of the N-terminal domains of α-NRXN in the synapse is needed.

The discovery of a novel *O*-glycosylation site on the EGF 2 domain leads to a new pathway of exploration that may reveal an important regulatory region in the α-NRXN protein. EGF domains in a number of extracellular proteins have been shown to carry unusual post-translational modifications, which are thought to impart significant implications on biological function [18]. In the case of the Notch receptor proteins, which carry two conserved glycosylation motifs on some of their extracellular EGF domains, at least one of the glycosylation patterns is implicated in the regulation of cell signaling by mediating ligand interactions [17, 23]. The Notch signaling pathway is widely used in cell-cell interactions during animal development and adult life, and regulates a variety of processes including cell fate specification, differentiation, left–right asymmetry, apoptosis, compartment boundary formation, somitogenesis and angiogenesis [24]. Since α-NRXN is also mediating multiple protein-protein interactions, it is reasonable to suggest that this novel glycosylation site may impart specific functions in protein recognition. Future studies should include verification of the post-translational modification in neurons, followed by functional assays that look at ligand specificity modulated by the *O*-glycosylation.

5.2. Multiple functions for one protein: Does α-NRXN serve as an extracellular scaffold?

α-NRXN has been shown to interact with multiple endogenous proteins. NLGNs, LRRTMs, dystroglycan, $GABA_A$ receptors, and the α4β2 nicotinic acetylcholine receptor have all been shown to bind through the LNS 6/β-NRXN domain [4, 25-29]. A smaller

set of proteins has been identified as receptors for the LNS 2 domain, including dystroglycan and the soluble neuropeptide neurexophilins [29, 30]. Furthermore, α-NRXN-1 is a high affinity receptor for the spider toxin, α-latrotoxin, which localizes to the pre-synapse where it integrates into the membrane and triggers neurotransmitter exocytosis. It has been shown to interact with both the LNS 2 and LNS 6 domains of α-NRXN, possibly in concert [31, 32].

The diversity of protein interactions governed by the NRXN molecules highlights the complex nature of this system. The crystal structure revealed a unique bifurcation in the assembly of the domains that leads to two arms of the protein, each containing one of the known binding domains. It seems likely that the two arms confer distinct functional roles and may act independently or together to bind to their respective partner proteins. In the case of a bi-molecular binding event, it is suggested that α-NRXN serves as an extracellular scaffolding protein to recruit multiple proteins within proximity in the extracellular space. Future studies directed at resolving the possibility for multiple binding events will enhance our understanding of the spatial arrangement of proteins in the synaptic cleft, and potentially reveal new ways in which the α-NRXNs are functionally distinct from the β-NRXNs.

5.3 Studying the interaction between α-NRXN and NLGN.

Crystallography is an invaluable tool for looking at the high-resolution properties of protein complexes, however it is limited by its ability to yield highly ordered crystals in cases where the complexes are dimensionally asymmetric or heterogeneous in nature. The success that was reported here in obtaining the structure of a large portion of the α-NRXN extracellular region did not come without years of experimental refinements, and even then, was partially a result of experimental luck. Ultimately, however, the most

information comes out of an understanding of the molecular interactions between α-NRXN and its partner proteins. Crystallization may be a futile endeavor to study the complex of the full α-NRXN with NLGN. One of the major questions that comes out of the crystal structure of α-NRXN is whether, upon binding to NLGN, the long arm of the molecule is stabilized in a near perpendicular orientation to the membrane or is free to rotate forming different orientations. Based on the overlay of α-NRXN_2-6 on the β-NRXN:NLGN complex, there is the potential for a secondary low affinity interaction site between NLGN and the LNS 4 domain of α-NRXN. Small angle X-ray scattering and/or single particle EM would be good technologies to employ to study the macromolecular assembly of the complex. If low-resolution studies support a secondary interaction site, it may be reasonable to attempt to crystallize the complex of NLGN with a truncated construct of α-NRXN consisting of LNS 4-6.

Other ongoing studies to study the interaction between α-NRXNs and NLGNs include the SPR binding experiments. Binding properties between the NLGNs and β-NRXNs in their various splice isoforms have been extensively studied using SPR [4, 6, 10, 12, 33]. Similar comparative studies for the various α-NRXN splice isoforms are needed. One interesting observation that came out of the crystal structure was the potential for splice site A in NLGN to mediate interactions with α-NRXN (see Chapter 4 for further details). A related finding showed that the presence of splice insert A in NLGN-2 or in NLGN-1 lacking splice insert B promotes targeting of either neuroligin to GABAergic synapses [5]. Other work has suggested a preferential role for α-NRXNs in GABAergic synapses [34]. In initial experiments reported in Chapter 4, I demonstrate by SPR that α-NRXN binds to NLGN-1 lacking splice site A or B, and that the binding affinity is comparable with β-NRXN. As an immediate follow up on these experiments, it

will be interesting to compare the different NLGNs and see if the inclusion of splice site A in any of the NLGNs has an influence on binding.

5.4 Implications for disease: a balancing act between excitatory and inhibitory neurotransmission?

There are a number of theories about the common pathophysiology leading to a range of neuro-developmental diseases, including ASD. One intriguing idea is that the mutual variant is an imbalance between excitatory and inhibitory synaptic pathways, possibly caused by alterations in local and/or long-range synaptic connectivity, which leads to irregular cell signaling [35-38]. Further evidence of this may come from the study of neurological diseases with known genetic causes that also exhibit autistic symptoms. For example, Fragile X syndrome is a single-gene disorder that leads to altered gene expression, which may ultimately affect the same cellular and molecular pathways as those disrupted in patients with ASD. In fact, many of the genes that are implicated in ASD have a direct or indirect role in the cellular pathways leading to protein expression and function at the synapse (see Chapter 1 and Table 1.1). Synapses are made up of a complex network of proteins that promote efficient neurotransmission. NRXNs and NLGNs play a crucial role in the recruitment of the necessary components that specify synapse function. Therefore, the modulation of gene expression and turnover of these proteins would likely yield defective synapse organization and signaling and therefore have a significant impact on overall synapse function.

5.5 The potential for therapeutic applications

We are still a ways off from understanding the precise role of synaptic adhesion proteins in the etiology of neurological disease, such as ASD. However, a short

discussion about the potential for therapeutic applications seems warranted, as this thesis will result in a degree in biomedical sciences.

The U.S. Food and Drug Administration has approved only one pharmacological treatment for ASD, risperidone, an atypical antipsychotic, and it does not target the core symptoms but rather certain maladaptive behaviors. While there is significant evidence to support the efficacy of early cognitive and behavioral therapy intervention, these patients do not see full remission of their symptoms. Therefore, there is a huge unmet need for therapeutic options for patients afflicted with ASD.

In recent years, some remarkable studies on animal models that exhibit autistic symptoms have shown reversibility of the disease-related phenotype by replacing or modulating gene function after birth and into adulthood [9]. These studies included models of Fragile X syndrome and Rett's syndrome, two genetic disorders that have a significant overlap in phenotype with the ASD. In one case, the administration of metabotropic glutamate antagonists helped regulate excessive protein translation caused by mutations the FMR protein [39]. In another example, reinstating a null *MECP2* gene after birth led to a reversal of the disease phenotype and suggested that *MECP2* is not essential for the earliest wiring of the nervous system but instead is required later for activity-dependent processes [40]. Together, these findings highlight a relatively new paradigm shift in our knowledge of developmental disorders. They suggest that if we can design therapeutic that modulate synaptic changes and synaptic related activity we may be able to reverse the symptoms caused by genetic abnormalities. NRXNs and NLGNs are intriguing candidates as they clearly play an important role in synapse function through the life of the synapse. The structures may serve as templates in the design of therapeutic molecules that help modulate protein-protein interactions, and thereby reinforce or prohibit specific interactions.

5.6 REFERENCES

1. Dalva, M.B., A.C. McClelland, and M.S. Kayser, *Cell adhesion molecules: signalling functions at the synapse.* Nat Rev Neurosci, 2007. **8**(3): p. 206-20.

2. Washbourne, P., et al., *Cell adhesion molecules in synapse formation.* J Neurosci, 2004. **24**(42): p. 9244-9.

3. Dean, C., et al., *Neurexin mediates the assembly of presynaptic terminals.* Nat Neurosci, 2003. **6**(7): p. 708-16.

4. Boucard, A.A., et al., *A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins.* Neuron, 2005. **48**(2): p. 229-36.

5. Chih, B., L. Gollan, and P. Scheiffele, *Alternative splicing controls selective trans-synaptic interactions of the neuroligin-neurexin complex.* Neuron, 2006. **51**(2): p. 171-8.

6. Comoletti, D., et al., *Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for beta-neurexins.* Biochemistry, 2006. **45**(42): p. 12816-27.

7. Ullrich, B., Y.A. Ushkaryov, and T.C. Sudhof, *Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons.* Neuron, 1995. **14**(3): p. 497-507.

8. Tabuchi, K. and T.C. Sudhof, *Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing.* Genomics, 2002. **79**(6): p. 849-59.

9. Walsh, C.A., E.M. Morrow, and J.L. Rubenstein, *Autism and brain development.* Cell, 2008. **135**(3): p. 396-400.

10. Arac, D., et al., *Structures of neuroligin-1 and the neuroligin-1/neurexin-1 beta complex reveal specific protein-protein and protein-Ca2+ interactions.* Neuron, 2007. **56**(6): p. 992-1003.

11. Chen, X., et al., *Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions.* Nat Struct Mol Biol, 2008. **15**(1): p. 50-6.

12. Koehnke, J., et al., *Splice form dependence of beta-neurexin/neuroligin binding interactions.* Neuron, 2010. **67**(1): p. 61-74.

13. Koehnke, J., et al., *Crystal structures of beta-neurexin 1 and beta-neurexin 2 ectodomains and dynamics of splice insertion sequence 4.* Structure, 2008. **16**(3): p. 410-21.

14. Sheckler, L.R., et al., *Crystal structure of the second LNS/LG domain from neurexin 1alpha: Ca2+ binding and the effects of alternative splicing.* J Biol Chem, 2006. **281**(32): p. 22896-905.

15. Rudenko, G., et al., *The structure of the ligand-binding domain of neurexin Ibeta: regulation of LNS domain function by alternative splicing.* Cell, 1999. **99**(1): p. 93-101.

16. Shen, K.C., et al., *Regulation of neurexin 1beta tertiary structure and ligand binding through alternative splicing.* Structure, 2008. **16**(3): p. 422-31.

17. Haines, N. and K.D. Irvine, *Glycosylation regulates Notch signalling.* Nat Rev Mol Cell Biol, 2003. **4**(10): p. 786-97.

18. Harris, R.J. and M.W. Spellman, *O-linked fucose and other post-translational modifications unique to EGF modules.* Glycobiology, 1993. **3**(3): p. 219-24.

19. Kim, H.G., et al., *Disruption of neurexin 1 associated with autism spectrum disorder.* Am J Hum Genet, 2008. **82**(1): p. 199-207.

20. Yan, J., et al., *Neurexin 1alpha structural variants associated with autism.* Neurosci Lett, 2008. **438**(3): p. 368-70.

21. Comoletti, D., et al., *Synaptic arrangement of the neuroligin/beta-neurexin complex revealed by X-ray and neutron scattering.* Structure, 2007. **15**(6): p. 693-705.

22. Ushkaryov, Y.A., et al., *Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin.* Science, 1992. **257**(5066): p. 50-6.

23. Moloney, D.J., et al., *Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules.* J Biol Chem, 2000. **275**(13): p. 9604-11.

24. Jafar-Nejad, H., J. Leonardi, and R. Fernandez-Valdivia, *Role of glycans and glycosyltransferases in the regulation of Notch signaling.* Glycobiology, 2010. **20**(8): p. 931-49.

25. de Wit, J., et al., *LRRTM2 interacts with Neurexin1 and regulates excitatory synapse formation.* Neuron, 2009. **64**(6): p. 799-806.

26. Ko, J., et al., *LRRTM2 functions as a neurexin ligand in promoting excitatory synapse formation.* Neuron, 2009. **64**(6): p. 791-8.

27. Zhang, C., et al., *Neurexins physically and functionally interact with GABA(A) receptors.* Neuron, 2010. **66**(3): p. 403-16.

28. Cheng, S.B., et al., *Presynaptic targeting of alpha4beta 2 nicotinic acetylcholine receptors is regulated by neurexin-1beta.* J Biol Chem, 2009. **284**(35): p. 23251-9.

29. Sugita, S., et al., *A stoichiometric complex of neurexins and dystroglycan in brain.* J Cell Biol, 2001. **154**(2): p. 435-45.

30. Missler, M., R.E. Hammer, and T.C. Sudhof, *Neurexophilin binding to alpha-neurexins. A single LNS domain functions as an independently folding ligand-binding unit.* J Biol Chem, 1998. **273**(52): p. 34716-23.

31. Li, G., et al., *N-terminal insertion and C-terminal ankyrin-like repeats of alpha-latrotoxin are critical for Ca2+-dependent exocytosis.* J Neurosci, 2005. **25**(44): p. 10188-97.

32. Ushkaryov, Y.A., A. Rohou, and S. Sugita, *alpha-Latrotoxin and its receptors.* Handb Exp Pharmacol, 2008(184): p. 171-206.

33. Comoletti, D., et al., *Characterization of the interaction of a recombinant soluble neuroligin-1 with neurexin-1beta.* J Biol Chem, 2003. **278**(50): p. 50497-505.

34. Kang, Y., et al., *Induction of GABAergic postsynaptic differentiation by alpha-neurexins.* J Biol Chem, 2008. **283**(4): p. 2323-34.

35. Rubenstein, J.L. and M.M. Merzenich, *Model of autism: increased ratio of excitation/inhibition in key neural systems.* Genes Brain Behav, 2003. **2**(5): p. 255-67.

36. Belmonte, M.K., et al., *Autism and abnormal development of brain connectivity.* J Neurosci, 2004. **24**(42): p. 9228-31.

37. Belmonte, M.K., et al., *Autism as a disorder of neural information processing: directions for research and targets for therapy.* Mol Psychiatry, 2004. **9**(7): p. 646-63.

38. Rippon, G., et al., *Disordered connectivity in the autistic brain: challenges for the "new psychophysiology".* Int J Psychophysiol, 2007. **63**(2): p. 164-72.

39. McBride, S.M., et al., *Pharmacological rescue of synaptic plasticity, courtship behavior, and mushroom body defects in a Drosophila model of fragile X syndrome.* Neuron, 2005. **45**(5): p. 753-64.

40. Guy, J., et al., *Reversal of neurological defects in a mouse model of Rett syndrome.* Science, 2007. **315**(5815): p. 1143-7.