Title
A structural analysis of the synaptic adhesion properties of alpha- and beta-neurexins

Permalink
https://escholarship.org/uc/item/9n32x7g0

Author
Miller, Meghan T.

Publication Date
2011

Peer reviewed|Thesis/dissertation
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

Signature Page .......................................................................................................................... iii
Dedication ................................................................................................................................. iv
Table of Contents .................................................................................................................... v
List of Figures ............................................................................................................................ ix
List of Tables .............................................................................................................................. xi
List of Abbreviations ................................................................................................................ xii
Acknowledgements ................................................................................................................ xvi
Curriculum Vitae ..................................................................................................................... xviii
Abstract of the Dissertation ................................................................................................... xxiii

Chapter 1: Introduction ........................................................................................................... 1

1.1 The Architecture of Excitatory and Inhibitory Synapses ................................................. 1
1.2 Synaptic Cell Adhesion Molecules: Neurexins and Neuroligins ................................. 3
1.3 The Neurexin Family ......................................................................................................... 6
    LNS Domains ....................................................................................................................... 7
    EGF Domains ....................................................................................................................... 10
1.4 The Neuroligin Family .................................................................................................... 12
1.5 Neurexins and Neuroligins: Relations to Disease ............................................................ 13
1.6 Objective of the Dissertation .......................................................................................... 17
1.7 References ....................................................................................................................... 18

Chapter 2: Structural Analysis of the Synaptic Protein Neuroligin and Its β-Neurexin
Complex: Determinants for Folding and Cell Adhesion ....................................................... 28

2.1 Summary .......................................................................................................................... 28
Three-Dimensional Reconstruction of the Extracellular Domain of αNRXN1......46

3.4 Discussion .................................................................47

3.5 Experimental Procedures .............................................49

Expression of αNRXN1 .......................................................49

Analytical Ultracentrifugation .............................................49

Negative-Stain Single-Particle Electron Microscopy ..................49

Small-Angle X-ray Scattering Data Acquisition of αNRXN1 ..........49

Structure Modeling from the Scattering Data ............................49

3.6 References ..................................................................50

3.7 Supplemental Information .............................................52

Supplemental Experimental Procedures .................................52

Supplemental References ....................................................55

Supplemental Figures ..........................................................56

3.8 Acknowledgements .......................................................62

Chapter 4: The crystal structure of the α-NRXN-1 extracellular region reveals a hinge point for mediating synaptic adhesion and function...............................63

4.1 Summary ...................................................................63

4.2 Introduction ..................................................................63

4.3 Results .......................................................................67

Three Dimensional Structure of α-NRXN-1 ..............................67

Comparison of LNS Domains ...............................................74

Structural Comparison with the β-NRXN/NLGN-1 Complex ..........78

Binding Affinity Measurements of the α-NRXN:NLGN-1 Complex ....82

4.4 Discussion ..................................................................87

Predicted Ca^{2+}-Binding Sites in the α-NRXN-1 Extracellular Region ....90
O-Glycosylation of EGF 2 in α-NRXN-1 .................................................................90
Disease-Linked Mutations in the α-NRXN-1 Extracellular Region ......................91
The α-NRXN: NLGN Complex ..........................................................................94
A Working Model of the α-NRXN:NLGN Complex at the Synapse ....................95
4.5 Experimental Procedures ........................................................................99
  Cloning, Expression and Protein Purification ..............................................99
  Crystallization and X-ray Data Collection ...............................................100
  Structure Determination, Refinement and Validation ..................................101
  Structural Analysis ......................................................................................102
  Surface Plasmon Resonance ......................................................................103
4.6 References ................................................................................................105
4.7 Acknowledgements ...................................................................................110
4.7 Accession Codes .......................................................................................110
Chapter 5: Conclusions, Open Question, and Future Directions .....................111
  5.1 Building on the α-NRXN Structure .........................................................116
  5.2 Multiple functions for one protein: Does α-NRXN serve as an extracellular
      scaffold? ......................................................................................................118
  5.3 Studying the interaction between α-NRXN and NLGN ..............................119
  5.4 Implications for disease: a balancing act between excitatory and inhibitory
      neurotransmission .......................................................................................121
  5.5 The potential for therapeutic applications .............................................121
  5.6 References ...............................................................................................123
LIST OF FIGURES

Figure 1.1: Domain organization of neuroligins and neurexins ........................................5
Figure 1.2: Structural characterization of LNS domains ..................................................9
Figure 1.3: Schematic of an epidermal growth factor (EGF)-like domains .....................11
Figure 2.1: Sequence conservation within the neuroligins and comparison with human liver carboxylesterase and mAChE ..............................................................29
Figure 2.2: Overall view of the NL4 structure .................................................................32
Figure 2.3: The NL4 central pocket compared with the mAChE active center ...............33
Figure 2.4: Comparison of the surface loops in NL4 and mAChE .................................34
Figure 2.5: Electrostatic potentials of the NL4 and mAChE dimmers and comparison of the dimeric Nrxβ1-NL4 and Fasciculin-mAChE complexes ........................................35
Figure 2.6: Close-up view of the Ca^{2+}-mediated Nrxβ1-NL4 complex interface ..........36
Figure 3.1: Schematic diagram of the αNRXN1 constructs used and hydrodynamic characterization of the purified extracellular domain of αNRXN_1-6 ..................43
Figure 3.2: Single-particle electron microscopy characterization of αNRXN_1-6 and αNRXN_2-6 .................................................................................................................45
Figure 3.3: Scattering intensity and $P(r)$ functions of the αNRXN_1-6 and αNRXN_2-6 .........................................................................................................................46
Figure 3.4: 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 ..............................47
Figure 3.5: Schematic model of the complex between αNRXN and its ligands in the context of the synapse .......................................................................................48
Figure 3.6: Size exclusion chromatography characterization of the purified extracellular domain of αNRXN_2-6 ..............................................................................56
Figure 3.7: Class averages of the αNRXN_1-6 and αNRXN_2-6 ........................................57
Figure 3.8: Basic scattering parameters for αNRXN_1-6 and αNRXN_2-6 ......................58
Figure 4.1: The structure of the α-NRXN-1 LNS 2-6 extracellular region ......................69
Figure 4.2: α-NRXN_2-6 interfaces ...............................................................................71
Figure 4.3: α-NRXN N- and O-Glycosylation ................................................................73
Figure 4.4: Comparison of the α-NRXN_2-6 LNS domains ........................................76
Figure 4.5: LNS 3 loop ..................................................................................................77
Figure 4.6: Overlay of α-NRXN_2-6 with β-NRXN bound to NLGN-1 ......................80
Figure 4.7: Surface proximity in the α-NRXN_2-6 overlay with the NLGN:β-NRXN complex ........................................................................................................................................81
Figure 4.8: The six α-NRXN constructs used in this study ........................................84
Figure 4.9: NRXN-NGLN association at equilibrium measured by SPR ...................85
Figure 4.10: SPR analysis of NLGN binding by the α-NRXN variants ......................86
Figure 4.11: 2-D overlay of the α-NRXN_2-6 structure on negative stain EM images ...89
Figure 4.12: Autism-related point mutations at the α-NRXN_2-6 surface ..................93
Figure 4.13: Model of the α-NRXN:NLGN complex at the synapse .......................98
LIST OF TABLES

**Table 1.1:** List of common genes implicated in ASD and related disorders ..................15

**Table 2.1:** Data collection and refinement statistics .....................................................31

**Table 2.2:** Nrxβ1 and NL4 residues in contact or within contact distance in the complex .................................................................................................................................36

**Table 3.1:** Basic SAXS parameters ..................................................................................59

**Table 3.2:** SASREF calculation parameters .....................................................................60

**Table 3.3:** BUNCH calculation parameters .......................................................................61

**Table 4.1:** Data Collection and Refinement Statistics......................................................70
LIST OF ABBREVIATIONS

AChE........................................................................................................ Acetylcholinesterase
AMPA........................................................................................................... α-amino-3-hydroxy-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., Trehwella 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 \( \alpha \)- and \( \beta \)-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

**FELLOWSHIPS AND AWARDS:**

<table>
<thead>
<tr>
<th>Year</th>
<th>Award Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>Graduate Student Best Abstract Award, <em>second place</em>. Experimental Biology 2009 <em>New Orleans</em></td>
</tr>
<tr>
<td>2009</td>
<td>ASBMB Travel Award to Experimental Biology 2009 <em>New Orleans</em></td>
</tr>
<tr>
<td>2008</td>
<td>Med-into-Grad Training Fellowship, <em>HHMI</em></td>
</tr>
<tr>
<td>2005</td>
<td>Chancellor’s Fellowship Award, <em>UCSD</em></td>
</tr>
<tr>
<td>2002-2004</td>
<td>NIH pre-IRTA fellowship</td>
</tr>
<tr>
<td>2002</td>
<td>Magna cum laude</td>
</tr>
<tr>
<td>1999</td>
<td>Phi Beta Kappa</td>
</tr>
</tbody>
</table>
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.


SCIENTIFIC PRESENTATIONS:

- **Miller MT**, Comoletti D, Jeffries CM, Wilson J, Trewella 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**, 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
eexpression 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 (\(\alpha\)-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-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 GABAA 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 Ca²⁺-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 \( \beta_{11}-\beta_{12} \) 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 \( \beta \) sheet sandwich where the \( \beta_{10}-\beta_{11} \) loop runs across. Along this surface there is a conserved \( \text{Ca}^{2+} \)-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. 2$^{nd}$ 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 Ca²⁺-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% worldwide [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.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Function</th>
<th>Rate in Autism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Related to Synapse Function</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLGN3</td>
<td>Synaptic adhesion molecules expressed at excitatory and inhibitory synapses.</td>
<td>&lt;1%</td>
<td>[68-72, 85, 86]</td>
</tr>
<tr>
<td>NLGN4X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRXN1</td>
<td>Synaptic adhesion molecule expressed at excitatory and inhibitory synapses.</td>
<td>&lt;1%</td>
<td>[62-67]</td>
</tr>
<tr>
<td></td>
<td>Ca^2+-dependent exocytosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDH9/10/15/18</td>
<td>Synaptic adhesion molecules</td>
<td></td>
<td>[87, 88]</td>
</tr>
<tr>
<td>CNTN3</td>
<td>Axon-associated cell adhesion molecule of the Ig superfamily</td>
<td>&lt;1%</td>
<td>[63, 89, 90]</td>
</tr>
<tr>
<td>CNTN4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNTNAP2</td>
<td>Cell adhesion molecule in NRXN superfamily. Localized at the juxta-</td>
<td>&lt;1%</td>
<td>[91-93]</td>
</tr>
<tr>
<td></td>
<td>paranodes of myelinated axons and associated with potassium channels.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH9/10/19</td>
<td>Synaptic adhesion molecules</td>
<td></td>
<td>[87, 90, 94]</td>
</tr>
<tr>
<td>NRCAM</td>
<td>Neuronal cell adhesion molecule</td>
<td></td>
<td>[95-97]</td>
</tr>
<tr>
<td>RELN</td>
<td>Neuronal migration. Secreted protein</td>
<td></td>
<td>[98, 99]</td>
</tr>
<tr>
<td>SHANK3</td>
<td>PSD scaffolding protein localized at the membrane, binds to NLGns</td>
<td>&lt;1%</td>
<td>[100-103]</td>
</tr>
<tr>
<td>GRIN2A</td>
<td>NMDA receptor subunit</td>
<td></td>
<td>[104]</td>
</tr>
<tr>
<td>GRIK2</td>
<td>Kainate receptor subunit</td>
<td></td>
<td>[105]</td>
</tr>
<tr>
<td>GABAR</td>
<td>GABA receptor subunit</td>
<td></td>
<td>[106]</td>
</tr>
<tr>
<td><strong>Regulation of synaptic protein levels</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MECP2</td>
<td>Transcriptional repressor. Mutated in Rett's syndrome</td>
<td>2%</td>
<td>[107-115]</td>
</tr>
<tr>
<td>EN2</td>
<td>Transcriptional regulator</td>
<td></td>
<td>[116-118]</td>
</tr>
<tr>
<td>PTEN</td>
<td>Inhibitor of PI3K/mTOR signaling. Loss of function leads to increased activity of mTORC1 [119]. Mutated in ASD with macrocephaly</td>
<td>1%</td>
<td>[120-126]</td>
</tr>
<tr>
<td>FMR1</td>
<td>Translational Repressor. Mutated in Fragile X syndrome. Estimated to interact with &gt;400 distinct mRNAs [127].</td>
<td>2-5%</td>
<td>[128-130]</td>
</tr>
<tr>
<td>TSC1/2</td>
<td>Inhibitor of mTOR-raptor complex (mTORC1), important for translational regulation. Mutated in tuberous sclerosis complex</td>
<td>1-4%</td>
<td>[131]</td>
</tr>
<tr>
<td>UBE3A</td>
<td>E3 ubiquitin ligase. Regulates ubiquitin-dependent protein turnover. Mutated in Angelman’s syndrome</td>
<td>1%</td>
<td>[63, 132, 133]</td>
</tr>
</tbody>
</table>
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, the *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:


CHAPTER 2:

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

Igor P. Fabrichny,1* Philippe Leone,2 Gerlind Sulzenbacher,2 DavideComoletti,2 Meghan T. Miller,3 Palmer Taylor,3 Yves Bourne,4* and Pascale Marchot1,2

1Biochimie des Interactions Moleculaires et Cellulaires, CNRS FRE-2738, Institut Federatif de Recherche Jean Roche,Universite de la Mediterranee, Faculte de Medicine Secteur Nord, F-13915 Marseille Cedex 20, France
2Architecture et Fonction des Macromolecules Biologiques, CNRS UMR-6098, Universites Aix-Marseille I & II, Campus Luminy - Case 932, F-13288 Marseille Cedex 09, France
3Department of Pharmacology, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093-0657, USA
4Alternate spelling: Fabricrny.

*Correspondence: pascale.marchot@univmed.fr (P.M.), yves.bourne@alfmb.univ-aix.fr (Y.B.)
DOI 10.1016/neuron.2007.11.013

SUMMARY

The neurogin types are postsynaptic cell adhesion proteins whose associations with presynaptic neurexins participate in synaptogenesis. Mutations in the neurogin and neurexin genes appear to be associated with autism and mental retardation. The crystal structure of a neurogin reveals features not found in its catalytically active relatives, such as the fully hydrophobic interface forming the functional neurogin 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-neurogin complex reveals the precise orientation of the bound neurexin and, despite a limited resolution, provides substantial information on the Ca2+-dependent interactions network involved in trans-synaptic neurexin-neurogin 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 (Landi-Studier 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 (Waltes et al., 2005). The neurogin types are 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., 2000) and sufficient for triggering the initial formation of synaptic contacts in cultured neurons (Scheiffele et al., 2000). In vitro, the neurogin bind to a Ca2+-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 neurogin types are expressed in a variety of tissues, they likely have recognition and adhesion functions extending to peripheral secretory tissues.

The five mammalian neurogin genes (numbered NGLN1 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 neurogin extracellular domain exhibits 32%–36% sequence identity and presumably shares overall shape with the globular domain of acetylcholinesterase (AChE) (Liu et al., 1990) [Figure 1]. This makes it a member of the α/β-hydrolase fold superfamly of proteins (Oliff et al., 1992; Cygler et al., 1993), which include catalytically active cholinesterases, lipases, and carboxylesterases, along with the noncatalytic adhesion proteins neuroactin (de la Escaler et al., 1990), glutaclin (Oliff et al., 1990), and gliotacin (Auld et al., 1995) (see Figure S1 available online), and the thyroglobulin C-terminal domain (Schumacher et al., 1988).
However, several characteristics of the neurogin 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 neurogins and their α/β-hydrolase fold relatives as well as between the five neurogins (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; Lichtenko et al., 1996).

Mutations in the coding regions of the genes for neurogin-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 (Bisoi 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 neurogin 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 N and C termini (Rudenko et al., 1999; Shockey et al., 2006). The presence or absence of these loops influences neurexin recognition for
Neuron

Structures of Neuriligin and Its Neurexin Complex

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

Neuriligin association with neurexin and its influence on cell adhesion only require extracellular sequences (Scheiffele et al., 2000). Hydrodynamic analyses established that the neuriligin extracellular domain, isolated through insertion of an early stop codon before the transmembrane domain or by point mutations (Schoeberl et al., 2003, 2006). Small-angle X-ray scattering (SAXS) combined with rigid-body modeling led to experimentally based neuriligin models (Comolotti et al., 2007). In a crystal of mouse ACHE (mACHE) (Bourne et al., 2003) and data on Cys connectivity and interpretations of oligosaccharide linkages (Hoffman et al., 2004) as constraints. These data are consistent with neuriligin dimerization through a four-helix bundle, as found for ACHE (Bourne et al., 1995; Susman et al., 1991), and an extended C-terminal linker emerging from the bundle.

A combination of X-ray and neutron scattering data recorded on neuriligin-1 (NL1) in complex with soluble β-neurexin-1 (Nrx1β) determined that two Nrx1β subunits associate on cross-bridge sides of the long axis of the neuriligin dimer (Comolotti 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 Nrx1β hypervariable loop edge faces the neuriligin surface area proximal to split site B. However, the low-resolution SAXS data only delineate the overall shape of the complex and the Nrx1β binding position, and do not reveal the precise orientation of the bound Nrx1β or the surface determinants involved.

A soluble form of NL4, which is naturally devoid of splice inserts and contains only two N-glycosylation sites (Boliger 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 molecule of the s1β-hydrolase fold family of proteins. This structure permits a detailed analysis of those structural features that are unique to the neurilignins 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 3.0 Å resolution structure of a complex between NL4 and a soluble Nrx1β 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 neuriligin, along with a restricted, Ca2+-mediated, interfacial area consistent with a stable 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., 1999), is an appropriate relative for comparative description of the NL4 structure.

The Neuriligin 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 s1β-hydrolase fold (Cygler et al., 1999; Ottes et al., 1999) (Figure 2). Of the three intramolecular disulfide bridges conserved in the neurilignins, the first two, Cys110-Cys146 and Cys306-Cys317 (Cys98-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 neurilignins (Hoffman et al., 2004) (Figure S1). This bridge constrains the included surface loop Gly503-Ser513 (or L4, or αL4β-βL4γ2, cf. Figure 1) into a conformation not observed in ACHE that directly stabilizes the tip of the neighboring loop Glu477-Ser487 (L3, or αL3b-βL3γ4) through a stacking interaction of Trp484 with Arg357, 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 (Jamaan et al., 2003; Yan et al., 2005).

In turn, the neurilignins lack the third bridge found in ACHE (Cys409-Cys529 in mACHE), where it connects helix αL4β to the C-terminal helix αL3b (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 Leu691 (Cys409, Ala412, Cys529, and Asn533 in mACHE) in this region. Residue Cys259, which is free in all neurilignins, is deeply buried within helix αL4b, far from the protein surface. Ala substitution to this Cys in an NL1-soluble form resulted in improved intracellular processing and secretion yields (Comolotti et al., 2004). GlcNAc moieties are linked to Asn102 and Asn511, which are conserved in all neurilignins. Residues Asp141 and Asn307, which correspond to potential Asn-linked glycosylation sites in neuriligin-2 (NL2) and NL1, respectively, are located at the NL4 surface, as are the positions for the spilled insert A (between Asp157 and Asp158) found in NL1, NL2, and NL3, and the spilled insert B (between Glu270 and Gly271) found in NL1.

The Neuriligin Dimer

Two NL4 subunits, related by a two-fold symmetry axis and linked through a tightly packed four-helix bundle made of helix αL4β and the C-terminal helix αL3b from each subunit
Table 1. Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th></th>
<th>NL4: Low Resolution</th>
<th>NL4: High Resolution</th>
<th>Nrxn21-NL4 Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline at the ESRF</td>
<td>ID23</td>
<td>ID14-EH1</td>
<td>ID29</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.976</td>
<td>0.934</td>
<td>0.979</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>P2₁,2,2</td>
<td>P2₁,2,2</td>
</tr>
<tr>
<td>Cell dimensions a, b, c (Å)</td>
<td>197.9, 63.6, 138.4</td>
<td>139.7, 154.0, 81.3</td>
<td>158.5, 198.7, 85.7</td>
</tr>
<tr>
<td>h,k,l symmetry</td>
<td>112.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>15–32 (3.3–3.2)</td>
<td>30–22 (2.25–2.20)</td>
<td>30–3.9 (4.0–3.9)</td>
</tr>
<tr>
<td>Total observations</td>
<td>79,446</td>
<td>587,014</td>
<td>121,593</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>25,664</td>
<td>89,603</td>
<td>25,251</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.2 (3.2)</td>
<td>6.6 (6.6)</td>
<td>4.8 (4.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.3 (98.8)</td>
<td>99.9 (100)</td>
<td>99.4 (99.9)</td>
</tr>
<tr>
<td>Rw (Rmerge)%</td>
<td>12.2 (41.5)</td>
<td>6.5 (47.0)</td>
<td>8.2 (55.9)</td>
</tr>
<tr>
<td>&lt;I/σ(I)&gt;</td>
<td>8.9 (3.1)</td>
<td>21.6 (4.5)</td>
<td>16.1 (3.3)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30–2.2</td>
<td>30–3.9</td>
<td></td>
</tr>
<tr>
<td>No. reflections</td>
<td>85,129</td>
<td>25,214</td>
<td></td>
</tr>
<tr>
<td>Rmerge/Rwork (%)</td>
<td>16.7/20.1</td>
<td>23.1/32.3</td>
<td></td>
</tr>
<tr>
<td>No. atoms</td>
<td>9309</td>
<td>11,341</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>8606</td>
<td>11,339</td>
<td></td>
</tr>
<tr>
<td>Ligand/sugar/ion</td>
<td>38/56/17</td>
<td>-/-/2</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>592</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Average B-factors (Å²)</td>
<td>38.8/41.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>78.5/154.5/63.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ligand/sugar/ion</td>
<td>43.0</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td>90.7</td>
<td>71.6</td>
<td></td>
</tr>
<tr>
<td>Most favored regions</td>
<td>90.7</td>
<td>71.6</td>
<td></td>
</tr>
<tr>
<td>Additionally allowed regions</td>
<td>9.3</td>
<td>26.8</td>
<td></td>
</tr>
<tr>
<td>PDB accession code</td>
<td>3BE8</td>
<td>2VH8</td>
<td></td>
</tr>
</tbody>
</table>

\* Values in parentheses are those for the last shell.
\*\* Rmerge = Σ|Iobse−Icalc|/ΣIobs, where I is an individual reflection measurement and <I> is the mean intensity for symmetry-related reflections.
\*\* Rwork = Σ[Fo-|Fc|]/Σ[Fo], where Fo and Fc are observed and calculated structure factors, respectively. Rwork is calculated for 5% of randomly selected reflections excluded from refinement.

(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 Å² interface area buried on each subunit, versus 77% and 870 Å² 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 neurogin extracellular domain sediments as a dimer, even at low concentration (Conioletti)
Neuron
Structures of Neuroligin and Its Neurexin Complex

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 Glu344 residues 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.

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 Å³ 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, 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, Glu275 (shared with mACHE 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); Ser293 (Glu in mACHE); and Tyr287 (conserved in mACHE) with their counterparts in mACHE. NL4 residue Ser253 replaces mACHE Glu202, whose mutation changes the catalytic constant (Gibney et al., 1993). 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...
additional flexibility in the NL4 molecule with its capacity for solvent and ion binding. Whether this feature proves advantageous to the neurologin 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 (Cys10-Cys14); and loops L1 (βαβγδ); L2 (αβαβγδ); L3 (αβαγδ); L4 (αβαγδ); and L5 (αβαγδ). The Cys-loop (Figure 4, cyan) is homologous to the long α-loop (Cys69-Cys96) that in mACHE 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 of 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 compensatory 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 α1β1, 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 α1β1εα2δ in NL4 requires conformational adaptation of the surrounding loop L3 (Figure 4, green), while both loops are stabilized by the neurologin-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 collapse partner (van Tilbeurgh et al., 1993). In α-nitrobenzoyl 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 α-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 4G), 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 neurologin family. Similarly, sequence alignment among the adhesion proteins, neurologin, neurotactin, glicentin, and glutacin reveals length compensation between the
 Structures of Neuroligin and Its Neurexin Complex

Figure 4. Comparison of the Surface Loops in NL4 and mACHE
(A) The Cys-loop (Cys116-Cys146; cf. Figure 1) in NL4 and (B) the Ω-loop (Cys59-Cys68) in mACHE are displayed in cyan, loops L1 (Thr332-His338 in NL4) in yellow, loops L2 (Asp361-Val368) in orange, loops L3 (Gly377-Ser387) in green, and loops L4 (Gly503-Ser513) in magenta. The Cys-loop conformation is stabilized by several polar and nonpolar interactions between Leu125, His121, Asp122, and Thr127 at the loop edge, and Ser176, Glu179, Lys378, Gly400, and Leu607, which surround the central pocket. Label AG 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, neurexin (NRT), glutactin (GLU), 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 ω/γ-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 ω/γ-hydrolase fold family (Botti et al., 1998). This face in mACHE also encompasses the binding site for the peptidic ligand fasciculin (Boume et al., 1993).

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 neurexin, glutactin, 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 neurotin corresponds to the Nrx1/1 binding face, as defined by the SAXS findings (Comelli et al., 2007). As a result of the surface charge distribution, the overall net charge and dipole moment of the NL4 subunit (−2984.6 and −600 Debyes) are significantly larger and slightly lower, respectively, than those of mACHE (−904.9 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 Nrx1/1 bound to NL4. Despite the overall limited resolution, the electron density map unambiguously reveals the position and orientation of the bound NL4/1 and those of the interacting side chains at the complex interface (Figure 5 and Figure S2). Nrx1/1 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 Ω-strands oriented perpendicular to the long axis of the NL4 dimer. This makes the Nrx1/1-NL4 complex a mirror image of the mACHE complex with the Ω-strand three-fingered peptide fasciculin (Boume et al., 1995) (Figure 5), exemplifying the variability in ligand selectivity and interaction capability of the ω/γ-hydrolase fold.

The concave side of the Nrx1/1 molecule faces the dimer four-helix bundle, which results in bound Nrx1/1 being rotated a quarter of a turn from its orientation in the SAXS model (Comelli et al., 2007). The binding interface is very flat, and it encompasses a buried surface area of ~560 Å² 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 Nrx1/1, respectively. Compared with the unbound NL4 and Nrx1/1 individual structures,
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 Nrx1/1 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 Val254 and Ile236 from Nrx1/1 (same numbering as in Ushkaryov et al., 1994), which is consistent with the coordination observed in the Nrx1 LNS-2 domain structure (Scheckler 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 Gin359 and Gly360, probably through water-mediated interactions, and a second water molecule in apposition with the Nrx1/1 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 Nrx1/1 LNS-2 domain) (Scheckler 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 (Comolli et al., 2003; Nguyen and Sudhof, 1997).

Direct interactions between the complex partners arise principally from the NL4 Gin359-Asn364 stretch within loop s\(_{2}^{252}-\phi_{2}\) (cf. Figure 1), with a minor contribution from Tyr453-Gly464 in loop s\(_{2}^{252}-\phi_{2}\). His260/Glu270 in loop s\(_{2}^{252}-\phi_{2}\), and residues within the hypervariable loop edge of Nrx1/1 (Table 2). Two salt bridges (Nrx1/1 Arg109 with NL4 Gin361/Glu270, and Nrx1/1 Arg323 with NL4 Asp351) and a polar interaction (Nrx1/1 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 table or moderate-affinity complex (K\(_{d}\) = 132 nM for Nrx1/1 and NL4) (Comolli et al., 2006) imparting flexibility in the interaction partners of the membrane-inserted adhesion proteins, and perhaps contributing to plasticity of the synapse. Prior mutagenesis data on the Nrx1/1 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 Gin233, show diminished binding or synaptogenic activity (Graf et al., 2006), consistent with the spatial arrangement of partners in the crystalline complex.
Most interesting residues are conserved among neurolins (Figure 1 and Figure 6), arguing for the present structure to be a representative model structure for other neurolins. The only exception refers to Gly464, which is replaced by Gin in NL2. Since the Gly464 Cα is 4.6 Å away from the facing Nnx1/1 Ser239, the bulkier Gin side chain is likely responsible for the lower affinity of NL2 (Kc = 8840 nM) (Cornoletti et al., 2006). In fact, an Ala substitution to Gly464 in NL1 abolishes Nnx1/1 binding (D.C., unpublished data). In contrast to the Cys-loop region, the shape of the Nnx1/1 binding region is surprisingly well conserved between NL4 and AChE, with only a small displacement of surface loop α2-3−β2. However, the key NL4 interacting residues are neither conserved in AChE nor in neuroactin, glutactin and gliactin (Figure S1). NL4 lacks both splice inserts A and B that are typically found in other neurolins (Figure 1). The positions for these inserts are located on the neurogin face opposite the Cys-loop where they are separated by 25 Å. The presence of the short insert B, which borders the Nnx1/1
binding site, and its posttranslational processing for N-linked glycosylation, appreciably affect the Nrxn1 binding affinity (Boucard et al., 2005; Comoliti et al., 2006), consistent with the current Nrxn1-NL4 complex structure. By contrast, presence of the long insert A, which can display two distinctive sequences (Ichtcherenko 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 Nrxn1 is remote from the binding interface (15 Å from the bound Ca<sup>2+</sup>) 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; Comoliti et al., 2006; Graf et al., 2006).

**Autism and Neurexin-Neulin Mutations**

Recently, several point mutations located within the extracellular domain of neurelins 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 Nrxn1 (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 Nrxn1 (Comoliti 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 Jago 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 neurelins). Gly99 is located at the surface in a turn preceding the short β strand, and its mutation does not seem to affect folding or binding. In contrast, Val403 is located near the end of the helix (x7.8, (Leu340-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 neurexin dimer. Lys378 is located at the beginning of the short helix (x7.8, (Leu377-Val386) and establishes van der Waals contacts with Asp122 within the Cys-loop. Arg437, located within the long helix (x14, 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 Nrxn1 reveal several key features for these adhesion molecules. First, the neurexin interaction surface of the neurexin dimer is on the side opposite that for substrate access in the catalytically active neurexin 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 neurexin 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 neurexin dimer. Molecular plasticity may be a hallmark for the adhesion function. Fifth, neurexin displays negatively charged surfaces on both its neu- rexin binding and Cys-loop faces. In contrast to ACH, a large electrostatic dipole in the neurexins, 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 neurexin, 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 Glu44 to Thr919 of full-length human NL4 (Boliger et al., 2001) (Figure 1) and N-terminally flanked with a FLAG epitope, numbered (12)DYKDDDKLAAAAE-1, was expressed at high yields (~2 mg/l) as a soluble exported protein from Stable lysates cells (Comoliti et al., 2006). It was purified from the culture medium on immobilized anti-FLAG M2 antibody (Sigma-Aldrich) (Comoliti et al., 2005). 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<sub>3</sub> (buffer A). Quantification of NL4 by UV spectrophotometry used an extinction coefficient of 102 680 M<sup>−1</sup> cm<sup>−1</sup> 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 Pharmacia 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 Nrxn1, encompassing residues Gly81-Val288 of the LNS domain of rat (μ-neurexin-1) (Shteynman et al., 1994) and N-terminally flanked with a 6His tag followed by a TEV cleavage site, was obtained by subcloning our initial pGEX construct (Comoliti et al., 2004) into pDE3T17 (Invitrogen). This Nrxn1 does not comprise those N- and C-terminal stretches found to be discarded in the Nrxn1 structure (Rudenko et al., 1999). Nrxn1 was expressed at high yields (~70 mg/l) from Rosetta PluS cells, and purified from the culture medium by nickel affinity chromatography on a HisTrap FF Crude cartridge (GE Healthcare). Removal of the 6His tag used a 6His-tagged TEV protease (Invitrogen) in a 1:25 TEV-to-Nrxn1 weight ratio.
(overnight incubation, 20 °C) and a second nick translation chromatography step. A molecular mass of 19,312 Da consistent with efficient tag removal was found by MALDI-TOF mass spectrometry. Final purification was by gel filtration SEC on Superdex 200 in buffer A. The Nn1/k-NL4 complex was formed in solution using a 1:2 molar excess of Nn1/ k over NL4 in 10 mM HEPES (pH 7.4), 25 mM Na2CO3, and 2 mM CaCl2 (overnight incubation, 4 °C). The buffer composition and final complex concentration were adjusted by ultrafiltration.

Crystalization, Data Collection, and Processing
Initial crystalization conditions were found from the PEG (Quiaqen) and SMF (Wizardbicoc) screens using vapor diffusion in 96-well Greiner plates and the Cartesian PFXs 4200 (Serionic 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 4.5), 200 mM NaCl, and 0 mM (20°C; 3.2 Å resolution data set) or 10 mM (4°C; 2.2 Å resolution data set) CaCl2 as the well solution. Crystallography of the complex at 20°C used setting drops and 8% PEG-20000 in 130 mM Mes (pH 6.0) as the well solution. Crystals were briefly transferred to the mother liquor supplemented with 2%-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). Although no trace of electron density was visible for Trx, no 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 ~30% of the protein model was built automatically with ARP/wARP (Peternell et al., 1999), the region surrounding the Cyto-c loop (Cyto-Cys148) (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, 2000) and were manually adjusted to five groups per subunit corresponding to residues Ala-1 to 140, Gly-141 to 297, Phe-340 to Gly-375, Phe-449 to Phe-457, Phe-561, Tyr-582, Phe-611 to Asp-141, Gly-260 to Asa399, Gly-254 to Gly-464, Arg-636 to His-658, and Tyr-737 to Ser-860. Data collection and refinement statistics are summarized in Table 1.

The final NL4 structure comprises PLG residues Asp-8 to Ala-1 and Nn1 residues Gln-44 to Ile-162, Ser-164 to Val-545, and Glu-559 to Leu-697 in subunit A of the dimer, and residues Ala-2 to Ala-1 and Ser-164 to Phe-549, and Val-557 to Leu-697 in subunit B. In total, the model includes 588 residues per subunit of the crystalized protein (Figure 1). Glu-44 residues link to Asn-102 and Asn-511; a phosphatase (arising from the crystallographically independent) trapped in the central pocket; a citrate and a glyceraldehyde (from the crystallographic and flash-cooling solution, respectively) trapped in solvent-accessible cavities at the protein surface; and three chloride and one sodium (from the crystallization buffer) are clearly identified. Despite the presence of CaCl2 in the crystallization conditions, no bound CaCl2 is observed, consistent with the ion being mainly coordinated by neuramin.

The average root mean square deviation (rmsd) between subunits A and B is 0.5 Å for 518 Ca atoms. Missing electron densities correspond to surface loops Asp-157 to Asn-163, which in Nl4 contains splice sites A, and Pro-611 to Gln-642 to Val-657 to Ala-658 and the C-terminal, ~20-residue peptide that protons helix αC. 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 Nn1/k-NL4 complex was solved by molecular replacement using MOLREP and, as templates, the N4 dimer structure and the Nn1/k structure (1C4R) (Fukudeno et al., 1999). The model was refined with Refmac, including solvation NCS (Johnson, for the two Nl4 molecules (excluding the Cys-loop) and the two Nn1/k molecules, and TLS refinement. The resulting electron density maps were used, when well-defined, to correct the backbone traces along the NL4 and Nn1/k molecules and position side chains and Ca2+ at the complex interface using COOT. An X-ray fluorescence emission spectrum confirmed the presence of Ca2+ in the frozen Nn1/k-NL4 crystal. TLS groups were the same as for NL4, with an additional group for the entire Nn1/k molecule. NL4 residues Asp-156 to Asn-163 in one subunit were built manually using O (Jones). Due to the limited resolution achieved, the remainder of the model was not refined further. The primary Nn1/k-NL4 complex structure comprises NL4 residues Asp-8 to Leu-697 and Nn1/k residues His22 to Val-354.

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 neurotrophins and of neurotactin, glutactin, and gulturactin 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 (Elber 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 VDOOD (Kleywegt and Jones, 1994).

Sequence- and structure-based searches for NL4’s closest catalytically active relatives used PSI-BLAST and SSVM (Kossinovsky and Honig, 2004), respectively. Highest structural hits were for mACHe, human liver carboxylesterase (HCE), bovine bile salt-stimulated lipase (BCL2E) and bacterial β-nitrophenol esterase (βNPE). Superposition 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 280 Ca atoms.

Figure 1 and Figure S1 were generated with ESPript (Gouet et al., 1999) and MUSCLE (Edgar, 2004), and Figures 2, Figure 3, Figure 4, Figure 5, Figure 6, and Figure S2 were generated with PyMOL (DeLano, 2000).

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) LS(LS), Soundra Conrad (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.) and the CNRS DR2-5D0 (P.M. and Y.B.) the CNRS and Fondation pour la Recherche Médicale (postdoctoral grants to I.F.) the Cure Autism Now Pilot Study (B.C.) and NIH grants R37-GM 18300 and U01-ES 10037 (P.T.) and GM07572 (M.M.).

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

REFERENCES
Structures of Neurexin and Its Neurexin Complex


Accession Numbers

Atomic coordinates and structure factors have been deposited with the Protein Data Bank; accession codes are 3B83 (N4L4) and 2V48 (N4S1-N4L4 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:

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

Davide Comolli1,2,6, K. Meghan T. Miller,1,6 Cy M. Jeffries,3 Jennifer Wilson,1 Borries Demeier,1 Palmer Taylor,1 Jil Trewella4,6, and Terunaga Nakagawa1,5
1Department of Pharmacology, Skaggs School of Pharmacy and Pharmaceutical Sciences University of California, San Diego, La Jolla, CA 92093, USA
2School of Molecular Bioscience, University of Sydney, Sydney, NSW 2006, Australia
3Department of Biochemistry, The University of Texas Health Science Center, San Antonio TX 78229, USA
4Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA
5These authors contributed equally to this work
6Correspondence: dcomolli@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-1. 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 suggests 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; Giessner 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 (Ulrich 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 (Uskokovic and Südhof, 1999; Ulrich 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) (Ichihara et al., 1998), neurexophilins (Missler et al., 1998), dystroglycan (Sugita et al., 2001), and leucine-rich repeat transmembrane proteins (LRRTM) (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 (Comolli 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 GABAAergic 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 GABAAergic synaptic protein...
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 (Elhott et al., 2009). αNRXNs appear to play a functional role at the synapse, including mediating Ca²⁺-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 Ca²⁺ 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 LGN domains sharing limited protein identity (Figure 1A). These three repeats, formed by nine independently folded domains that span ≈90% of the protein sequence, are recognition motifs that appear to target NRXN to the presynaptic region (Fairless et al., 2008). Although the crystal structures of the isolated second, fourth, and sixth LGN domains of αNRXN1 have been solved (Rudenko et al., 1999; Schleier 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 LGN 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 neuromodulator (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 sNRRX1

Protein expression and N-terminal sequencing

Two constructs from the extracellular domain of sNRRX1 were expressed as soluble entities in the cell culture medium of HEK293 G418− cells: one starting at position Leu931 and encompassing sites of alternative splicing #1 and #3 (sNRRX1_1-6) and an N-terminal deletion devoid of the first NLS and EGF domains to yield a protein starting at position Glu280 (sNRRX1_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 homogenous monomeric species. In SDS-PAGE followed by Coomassie blue staining, sNRRX1_1-6 and sNRRX1_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 sNRRX1 protein with its native leader peptide (maptalinggpgdcisslpgwrrgkEFPPG), 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-Pro-Pro-Gly indicated that the mature protein starts at Leu931, consistent with previous findings (Missler et al., 1998) and sequence predictions.

Mass spectrometric analysis of sNRRX1

Both sNRRX1 constructs were expressed in the culture medium of HEK293 G418− cells. These cells lack N-acetylglucosaminyltransferase I (GnT) activity, and consequently glycosylation remains restricted to a homogenous seven-residue oligosaccharide (Reeves et al., 2002), thus simplifying structural analyses. sNRRX1 contains four potential N-linked glycosylation sites at positions N125, N159, N739, and N1223 (Figure 1). Although the peptic 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 578 Da between the two values. Because GnT− 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 oligosaccharide at all four N-linked glycosylation sites. Functionally, sNRRX1_1-6 is fully active in binding neurogin-1 (Figure S1B) (Boucard et al., 2005) and LRRTM2 (de Wit et al., 2009).

Analytical Ultracentrifugation Analyses of the Extracellular Domain of sNRRX1

Sedimentation velocity and equilibrium measurements provided complementary information useful in determining the globularity and oligomerization state of the extracellular domain of sNRRX1 in solution. To determine whether sNRRX1_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 sNRRX1_1-6 has a sedimentation coefficient of 6.36 s (6.28 s, 6.39 s) and a molecular weight of 141 kDa (135.5 kDa, 151.4 kDa), with a frictional ratio of 1.54 (1.51, 1.64), consistent with the monomeric mass of sNRRX1 and indicative of an elongated particle (values in parentheses 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

sNRRX1_1-6

To obtain structural information on the extracellular domain of sNRRX1, the purified protein was negative-stained and imaged using a transmission EM. sNRRX1_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 NLS 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 family 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 sNRRX1_1-6 particles identified the C terminus (LNS6) in the more structured triangular region (Figure 2B), whereas the FLAG Fab-tagged sNRRX1_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 sNRRX1 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.

sNRRX1_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
Structure of the Extracellular Domain of αNRXN1

splice site 1 (Figure 1A), generating a protein beginning at Gly280 (αNRXN1-2-6). Approximately 17,000 particles of αNRXN1-2-6 were analyzed, and 150 class averages were generated. The most common class averages clearly define all five LNS domains in αNRXN1-2-6. 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 shape 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 terminals 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 αNRXN1-1-6 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 analysis and to strengthen our interpretation of the small angle X-ray scattering (SAXS) data, experiments on the truncated sNRXN-2-6 construct were also performed. The forward scattering intensities (I(0)) and radius of gyration (R_g) of each of the sNRXN-1-6 and sNRXN-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 = MW (with c in units of mg/mL) and comparing data with that of tryosyme scattering, estimates of the molecular weight of the scattering particles were 164–174 kDa for sNRXN-1-6, and 109–118 kDa for sNRXN-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 sNRXN-1-6 and sNRXN-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 ~146 Å 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 sNRXN-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 sNRXN 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 sNRXN-1-6 and sNRXN-2-6 can be constructed.

Three-dimensional Reconstruction of the Extracellular Domain of sNRXN

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 sNRXN that closely resemble the shapes obtained with single-particle EM. The crystal structures of LNS 2, 4, and 6 are available (Rüdenko et al., 1999; Scheckler 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 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 (PetroskiN 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 sNRXN-1-6 and sNRXN-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/LNS4 forming the arms (Figure 4), in excellent agreement with the EM images (Figure 2).

Approximately 11% of the mass of sNRXN-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 sNRXN-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 constraints as those used in SASREF against both sNRXN-1-6 and sNRXN-2-6 data sets and 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...
Structure
Structure of the Extracellular Domain of αNRXN1

Figure 4. Rigid Body Modelling 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 excluded in αNRXN 1–6).
(B) Structural ensembles represented as semi-transparent 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 Cα 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 (Lishkaryov et al., 1999). 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; Comolli et al., 2007). The length of this single chain tether may be advantageous to extend the sixth LNS domain so it can approach the center of the synaptic space, permitting association with postsynaptic proteins such as neurophin or LRRTM2 (Comolli et al., 2007; de Wit et al., 2009; Ko et al., 2009). The presence of the stalk domain, however, extends the overall length of sNRXN beyond what a semielongated amino acid sequence would predict.

Sedimentation velocity and equilibrium analyses unambiguously show that the extracellular domain of sNRXN 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 sNRXN, similar to β-neurexin (Comolli et al., 2006), does not self-associate. Thus, unless oligomerization occurs through the short intracellular domain, sNRXN is likely present as a monomer on the cell surface.

Single-particle EM shows that the LNS1-EGF1 pair of sNRXN, 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 sNRXN, 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., 1998; Schoelkler 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 sNRXN1, 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 sNRXN maintains a semi-elongated 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 “snapshot” 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 sNRXN1-6 and sNRXN 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 neurolins bind, is located at the rim of the β sheet sandwich opposite the N and C termini (Rudenko et al., 1999) that reside close together. Consequently when one LNS domain is connected to the next (or 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., 2008). Third, although the extracellular domain of sNRXN, 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 sNRXN, 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 sNRXN, lamininα2, 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., 2008) shows that inter-LNS domain linkers may be extended and flexible. In the case of sNRXN, multiple interacting proteins have been isolated and
Structure of the Extracellular Domain of nNRT1

Characterized In particular, neurexophilin appears to bind to the second LNS domain of nNRT1 (Messier et al., 1998), whereas the neurotrophins and LRRK2 associate with the sixth LNS domain (IchiKerIe et al., 1995; BoUard et al., 2005; De Wil 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 motif, tends to fold back toward LNS4, creating a more compact structure, potentially limiting the accessibility to binding partners (Passer et al., 2009). However, single-particle EM data show that LNS6 retains some flexibility, and the clover-leaf arrangement suggests that the metal binding site of each LNS domain is likely solvent exposed rather than being confined to interdomain stabilization. Considering the decrease in affinity for complex formation at the NIH 3T3 (pH 7.4) and 150 mM NaCl solution applied to the plasmid-coated carbon-coated grids and was negatively stained with 0.75% uranyl acetate, as described elsewhere (Shi et al., 2006), EM images for class averages were collected using a FEI Tecnai-12 RB microscope equipped with an LaB6 filament using low-dose procedures on SO-180 Kfik film at a magnification of 50,000 X 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 nNRT1

Data were collected from the proteins and their soluble forms (ultralow buffers for the nNRT1 and hsp80 deployatory for hsp70) at 20 °C, using an Anton Paar SAXS instrument collection instrument at the University of Utah on 2D position-sensitive image plates (10 mm slit and integration width) as described by Jaffere et al. (2008).

Structure Modelling 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.cub.2010.08.003.

ACKNOWLEDGMENTS

This work was supported by USPHS (grant R01 GM-15830 to P.T.), NIEHS (grant P42ES10327 to P.T.), U.S. Department of Energy (grant DE-FG02- SRC6426 to J.T.), Autism Speaks (grant 2617 to D.C.), and the John Merck Fund and Helman Foundation (support to T.L.). We acknowledge the use of the UCSD Cryo-Electron Microscopy Facility which was supported by NIH grants 1S1RR000116 and GM030305 to Timothy S. Baker and a gift from the Agouron Institute to UCSD. AUC supercomputer analyses were supported by NSF Tarangal allocation TG-MCB070038 (B.D.). Ultrafast development is supported by NIH (R002009 B.D.). Calculations on Lonestar were supported by NSF TeraGrid allocation TG-MCB070038 (B.D.). We thank Dennis Winge (University of Utah, U.T) for quantitative amino acid analysis, and MedOx Genesense, 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 structure 1f89, 1044–1053, August 11, 2010 ©2010 Elsevier Ltd All rights reserved.
helpful discussion on homology modeling, and Greg Fuchs for excellent technical help during the preparation of the cleavable eSIRNAs construct.

Received: March 26, 2010
Revised: June 14, 2010
Accepted: June 17, 2010
Published: August 10, 2010

REFERENCES


Structure
Structure of the Extracellular Domain of αNRXN1


Structure 18, 1044–1053, August 11, 2010 ©2010 Elsevier Ltd All rights reserved 1053
Supplemental Information

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

Davide Comolotti, 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 150 mM NaCl, 10 mM 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 - Weischat 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 7 μM) 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 x 10^4 OD260 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 OD₃₅₀, M⁻¹ cm⁻¹) as implemented in UltraScan. The partial specific volume, v, 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.mayway.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 (FabαNRXN) 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 vacuo, at 110°C. Only reliably quantified amino acids (Ala, Arg, Gin (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,

\[
\frac{4\pi}{\lambda} \sin \theta = \frac{\lambda}{\sin \theta} = \frac{2}{\lambda} \sin \theta
\]

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 Rg and I(0) values were estimated using extrapolations of ln(I(q)) vs q in the linear region qmaxRg ≤ 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 (Rg) and maximum dimension (Dmax) 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:

\[
K = \frac{I(0)}{M_c}
\]

where M 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 (Kingbaum 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; Sheeckler 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 (2J4D) respectively, were used as templates. For EGF1, 2 domains, portions of crystal structures of COX-2 (1CVS, and 1PXX) were used. A portion of prostaglandin H2 synthase-1 (1CQG) was used as model template for EGFR 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 χ² values.

To model such large multi-domain proteins against the SAXS data we defined a number of assumptions as outlined in the Results. Distance constraints were maintained between the domains during multiple SASREF refinements against the 5.86 and 5.7 mg/ml 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 EGFR-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 EGFR-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 was assessed against the desmeared experimental scattering data using the $\chi^2$ discrepancy calculated in CRYSOL (Svergun et al., 1995).

$$
\chi^2 = \frac{(N-1)}{\sum \frac{(I(q) - k\sigma(q))^2}{\sigma(q)}
$$

where $N$ is the number of experimental points, $k$ is a scaling factor. $I(q)$ and $\sigma(q)$ 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 were done at 25°C in 10mM Hepes buffer, pH 7.4, containing 150mM NaCl, 2mM CaCl₂ 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 Ca²⁺ to disrupt the association.

**SUPPLEMENTAL REFERENCES**


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(0)/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

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Sample name</th>
<th>Dmax</th>
<th>Rg (nm)</th>
<th>I(0)/C</th>
<th>Concentration (mg/ml)</th>
<th>I(0)/C/MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>αNRXN−_1-6</td>
<td>170</td>
<td>53.39±0.24</td>
<td>34.75±0.31</td>
<td>5.881</td>
<td>5.93±0.05</td>
<td>0.042</td>
</tr>
<tr>
<td>170</td>
<td>53.04±0.28</td>
<td>27.92±0.27</td>
<td>3.847</td>
<td>6.01±0.06</td>
<td></td>
<td>0.042</td>
</tr>
<tr>
<td>170</td>
<td>53.32±0.30</td>
<td>21.92±0.24</td>
<td>3.88</td>
<td>6.28±0.07</td>
<td></td>
<td>0.044</td>
</tr>
<tr>
<td>170</td>
<td>52.93±0.34</td>
<td>14.11±0.17</td>
<td>2.401</td>
<td>5.88±0.07</td>
<td></td>
<td>0.041</td>
</tr>
<tr>
<td>170</td>
<td>52.63±0.40</td>
<td>6.53±0.12</td>
<td>1.206</td>
<td>5.42±0.10</td>
<td></td>
<td>0.038</td>
</tr>
<tr>
<td>average</td>
<td>170</td>
<td>53.04±0.31</td>
<td></td>
<td></td>
<td></td>
<td>0.042</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Dmax</th>
<th>Rg (nm)</th>
<th>I(0)/C</th>
<th>Concentration (mg/ml)</th>
<th>I(0)/C/MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysozyme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>14.6±0.03</td>
<td>10.37±0.04</td>
<td>20.62</td>
<td>0.50±0.002</td>
<td>0.035</td>
</tr>
<tr>
<td>40</td>
<td>14.61±0.03</td>
<td>7.578±0.03</td>
<td>15.46</td>
<td>0.49±0.002</td>
<td>0.034</td>
</tr>
<tr>
<td>40</td>
<td>14.76±0.04</td>
<td>5.081±0.02</td>
<td>10.31</td>
<td>0.49±0.002</td>
<td>0.034</td>
</tr>
<tr>
<td>40</td>
<td>14.59±0.09</td>
<td>2.888±0.02</td>
<td>5.15</td>
<td>0.56±0.004</td>
<td>0.039</td>
</tr>
<tr>
<td>average</td>
<td>40</td>
<td>14.64±0.05</td>
<td></td>
<td></td>
<td>0.036</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Dmax</th>
<th>Rg (nm)</th>
<th>I(0)/C</th>
<th>Concentration (mg/ml)</th>
<th>I(0)/C/MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>αNRXN−_2-6</td>
<td>140</td>
<td>43.78±0.49</td>
<td>1.45±0.10</td>
<td>5.700</td>
<td>0.26±0.02</td>
</tr>
<tr>
<td>140</td>
<td>42.95±0.59</td>
<td>1.16±0.06</td>
<td>4.560</td>
<td>0.26±0.01</td>
<td>0.022</td>
</tr>
<tr>
<td>150</td>
<td>44.30±0.63</td>
<td>0.90±0.01</td>
<td>3.420</td>
<td>0.26±0.01</td>
<td>0.022</td>
</tr>
<tr>
<td>150</td>
<td>45.37±0.59</td>
<td>0.62±0.08</td>
<td>2.280</td>
<td>0.27±0.01</td>
<td>0.023</td>
</tr>
<tr>
<td>average</td>
<td>145</td>
<td>44.19±0.57</td>
<td></td>
<td></td>
<td>0.022</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Dmax</th>
<th>Rg (nm)</th>
<th>I(0)/C</th>
<th>Concentration (mg/ml)</th>
<th>I(0)/C/MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysozyme</td>
<td>42</td>
<td>14.42±0.03</td>
<td>0.239±0.06</td>
<td>7.39</td>
<td>0.32±0.008</td>
</tr>
</tbody>
</table>
Table S2, related to Figure 4. SASREF calculation parameters

<table>
<thead>
<tr>
<th>LNS1</th>
<th>EGF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>L66</td>
<td>P209</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LNS1</th>
<th>EGF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>C228</td>
<td>Q257</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LNS2</th>
<th>LNS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>E280</td>
<td>V460</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LNS3</th>
<th>LNS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>N466</td>
<td>Q643</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LNS4</th>
<th>LNS5</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>P670</td>
<td>C703</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LNS5</th>
<th>LNS6</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>V709</td>
<td>G889</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LNS5</th>
<th>LNS6</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>A895</td>
<td>C1071</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LNS6</th>
<th>LNS1</th>
</tr>
</thead>
<tbody>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>T1077</td>
<td>C1110</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>LNS1</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>L31</td>
<td>P209</td>
</tr>
<tr>
<td>EGF1</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>C228</td>
<td>Q257</td>
</tr>
<tr>
<td>LNS2</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>E280</td>
<td>V460</td>
</tr>
<tr>
<td>LNS3</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>N466</td>
<td>Q643</td>
</tr>
<tr>
<td>EGF2</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>P670</td>
<td>C703</td>
</tr>
<tr>
<td>LNS4</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>V709</td>
<td>G889</td>
</tr>
<tr>
<td>LNS5</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>G895</td>
<td>C1071</td>
</tr>
<tr>
<td>EGF3</td>
<td></td>
</tr>
<tr>
<td>STARTS</td>
<td>ENDS</td>
</tr>
<tr>
<td>T1077</td>
<td>C1110</td>
</tr>
</tbody>
</table>
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 neurexophilin 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 Ca²⁺-binding sites on the same side of the molecule with similar inter-Ca²⁺ 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 Ca²⁺-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.
Table 4.1: Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th>X-ray source</th>
<th>SSRL BL11-1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal data</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>199.74, 61.24, 155.58</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°)</td>
<td>90.00, 121.21, 90.00</td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Processing software</td>
<td>XDS package</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97945</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30.0-3.02(3.13-3.02)</td>
</tr>
<tr>
<td>$R_{sym}$ (%)</td>
<td>14.0(62.8)</td>
</tr>
<tr>
<td>$I / sI$</td>
<td>8.7(2.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.9(99.5)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>31504</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.4(3.4)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.02(3.10-3.02)</td>
</tr>
<tr>
<td>$R_{work}$ / $R_{free}$ (%)</td>
<td>21.4(31.2)/26.9(39.4)</td>
</tr>
<tr>
<td>No. atoms</td>
<td>7878</td>
</tr>
<tr>
<td>Protein</td>
<td>7795</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>50</td>
</tr>
<tr>
<td>Water</td>
<td>33</td>
</tr>
<tr>
<td>Average $B$ overall (Å²)</td>
<td>47.6</td>
</tr>
<tr>
<td>R.m.s.d. bond length (Å)</td>
<td>0.013</td>
</tr>
<tr>
<td>R.m.s.d. bond angle (°)</td>
<td>1.369</td>
</tr>
<tr>
<td>Ramachandran Plot</td>
<td></td>
</tr>
<tr>
<td>Preferred (%)</td>
<td>93.6</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>6.1</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Figure 4.2: $\alpha$-NRXN_2-6 interfaces 
(a) The $\alpha$-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 $\alpha$-NRXN_2-6 molecule to form the “L” shape. The reference $\alpha$-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 $\alpha$-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 \( \alpha \)-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 Xyl\( \alpha \)1-3Xyl\( \alpha \)1-3Glc\( \beta \)1-O-Ser [29]. In the \( \alpha \)-NRXN_2-6 structure the density accommodates a \( \beta \)-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 Å²) 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 $\beta$-NRXN/NLGN-1 Complex

$\alpha$-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 $\alpha$-NRXN_2-6 structure. To model the complex between $\alpha$-NRXN and NLGN-1, we superimposed the structure of $\alpha$-NRXN_2-6 onto the structures of the $\beta$-NRXN-1 molecules bound to NLGN-1 [21]. The overlaid NLGN molecule interfaces with the $\alpha$-NRXN_2-6 molecule without significant steric obstruction, indicating that NLGN could bind to $\alpha$-NRXN_2-6 in the crystallized conformation (Figure 4.6a). The LNS 6 domain of $\alpha$-NRXN_2-6 shows high overlap at the binding interface to the NLGN molecule, as seen in the $\beta$-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 $\beta$-NRXNs either in complex with NLGN or in their free form (Figure 4.6b&c). The Fourier difference maps for $\alpha$-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 $\alpha$-NRXN with no significant steric obstruction between the $\alpha$-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 $\beta$-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 $\alpha$-NRXN_2-6 overlay with the NLGN:$\beta$-NRXN complex

Overlay of $\alpha$-NRXN_2-6 on the NLGN1:$\beta$-NRXN complex (PDB: 3biw (shown), 3b3q, 2xb6) shows that the $\alpha$-NRXN molecule (gray) can accommodate a NLGN molecule (orange) without significant steric constraint. However, NLGN proximity to the LNS 4 domain in $\alpha$-NRXN indicates a region of the $\alpha$-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 $\alpha$-NRXN_2-5 construct at concentrations approaching 3 $\mu$M (Figure 4.10). Although the specific recognition properties of the individual LNS domains remain to be resolved, the comparative equilibrium dissociation constants between the $\beta$-NRXN and truncated $\alpha$-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 $\alpha$-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 $\alpha$-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

(*) In α-NRXN_1-6, SS#1 E278-Q293 are removed. E258-K277 remain in the construct and represent an alternative splice variant
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 4.9: NRXN-NGLN association at equilibrium measured by SPR
Figure 4.10: SPR analysis of NLGN binding by the α-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 \( \alpha\)-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 \( \alpha\)-NRXN-1 expressed in neurons remains to be determined, we illustrate here the possibility for a new glycosylation pattern of \( \alpha\)-NRXN-1 that is likely to influence its function.

**Disease-Linked Mutations in the \( \alpha\)-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 \( \alpha\)-NRXN-1 leader sequence [2, 5], two in the \( \beta\)-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 $\alpha$-NRXN:NLGN Complex

$\alpha$- and $\beta$-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 $\beta$-NRXNs in their various splice isoforms have been extensively studied using SPR [21, 26, 34, 39, 48]. Information on the $\alpha$-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 $\alpha$-NRXN-1 with NLGN-1 [49]. Pull-down experiments concluded that $\alpha$-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 $\alpha$-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 $\alpha$-NRXN-1 compared to $\beta$-NRXN-1 with NLGN-1 was reported to be $\sim$2-fold less when $\alpha$-NRXN lacks SS#4 and $\sim$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 $\alpha$-NRXN_2-6 structure on the $\beta$-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 $\beta$-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 $\alpha$-NRXN to
accommodate the NLGN molecule, thereby supporting the propensity for \( \alpha \)-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 \( \alpha \)-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 \( \alpha \)-NRXN_2-6 in the conformation shown here. Overall, it remains to be determined if the \( \alpha \)-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 \( \alpha \)-NRXN:NLGN Complex at the Synapse

Both the \( \alpha \)-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 $\alpha$-NRXN:NLGN complex at the synapse.
Overlays of $\alpha$-NRXN_2-6 onto $\beta$-NRXN-1 as bound to NLGN-1 or -4 suggests a possible position for the $\alpha$-NRXN:NLGN complex in the synaptic cleft. The overlaid complex is shown with the C-termini of $\alpha$-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 $\alpha$-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 $\alpha$-NRXN LNS 2-EGF 3 in gray and LNS 6 in blue. Stalk domains for NLGN and $\alpha$-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-acetylglicosaminyltransferase 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% CO₂ 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% NaN₃, 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 chromatography. 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=90.0° β=121.21° γ=90.0°) 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 $R_{\text{work}} = 0.21$ and $R_{\text{free}} = 0.27$ 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 \(\beta\)-NRXN to ~140 kD for \(\alpha\)-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


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 \( \alpha \)-NRXN extracellular region, crystallization of the protein, structure determination of the \( \alpha \)-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\textsuperscript{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\textsuperscript{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 postsynaptic 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


