Drew University

College of Liberal Arts

Modeling and Manipulating Mutant Neurexins in *Caenorhabditis elegans*

A Thesis in Neuroscience

by

Karishma S. Patel

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science

With Specialized Honors in Neuroscience

May 2021

ABSTRACT

To understand mechanisms that produce the human conscious experience, including sensation, perception, cognition, and behavior, researchers have focused on studying the functional unit of the nervous system, the neuron, and the primary location at which neurons communicate with each other, the synapse. Since their discovery, neurexins and neuroligins have been studied as important regulators of the synapse, expressed in species across the eukaryotic evolutionary tree, from *Homo sapiens* – where mutations in each protein have been implicated in human neuropsychological disease – to the nematode *Caenorhabditis elegans*. While expressed in a myriad of subtypes and isoforms in humans, this complexity is vastly reduced in *C. elegans*, giving researchers a simpler context to elucidate function and the consequences of mutant alleles. After reviewing what is known about neurexin and neuroligin structure, function, and binding in humans and *C. elegans*, wild type (WT) human neurexin NRXN1, WT *C. elegans* nrx1, and *C. elegans* ok1649 and ds1 nrx1 alleles were further characterized. The longest human isoform, αNRXN1, and the longest *C. elegans* isoform, αnrx1, both share 28% sequence identity and 47% similarity through BLAST analysis. Protein domain analysis using NCBI's Conserved Domain Search Tool identified 6 functional laminin-neurexin-sex-hormone binding (LNS) sites in the protein sequences of αNRXN1 and αnrx1, one LNS site in βNRXN1, and no LNS sites in γNRXN1 as predicted in the literature; no such sites were recognized in the smaller nrx1 isoform initially described as βnrx1. These findings suggest that the shorter *C. elegans* isoforms may actually be a γ isoforms, due to the lack of extracellular LNS domains, as in γNRXN1. Structural models generated by the C-I-TASSER program of WT NRXN1, nrx1,

and ok1649 and ds1 nrx1 alleles from VC1416 and SG1 strains largely suggest retention of conformation and function, with severe disturbances observed only in the ds1 truncation. Both alleles of nrx1 appear to retain the ability to express functional γnrx1 forms. Therefore, methods to generate three novel nrx1-KO models - total nrx1-KO, an αnrx1- and a γnrx1-KO – using CRISPR-Cas9 technology, microinjection, and selection strategies were developed to further study the roles neurexins may play in the *C. elegans* synapse. Finally, strategies for behavioral characterizations of the three novel *C. elegans* models were discussed. The protein models, proposed neurexin manipulations, and suggested behavioral measures of WT and mutant alleles in *C. elegans* will further current understanding of neurexins and neurexin-neuroligin interactions in the *C. elegans* system. Ultimately, this work may have potential applications in understanding human neuronal communication and disease pathways in neuropsychological conditions.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

aa	amino acids
bp	base-pair
CDS	Conserved Domain Search (an NCBI program)
CGC	Caenorhabditis Genomics Center, University of Minnesota
ChEHD	Cholinesterase Homology Domain
COBALT	Constraint-based Multiple Alignment Tool (an NCBI program)
Cre	Cre Recombinase, excises sequence between two Lox2272 Sites
CysL	Cysteine Loop (protein feature)
D	Dimerization Sequence (protein feature)
EGF	Epidermal Growth Factor-like Domain
Hyg/Hyg^{R}	Hygromycin B (eukaryotic antibiotic) / Hygromycin B Resistance
kb	kilobase-pair
KO	Knock-out
LNS	Laminin/Neurexin Sex-Hormone-Binding-Globulin Domain
nlg1	*C. elegans* neuroligin gene
nlg1	*C. elegans* neuroligin protein
(α/β/γ)nrx1	*C. elegans* (alpha/beta/gamma) neurexin gene

(α/β/γ)nrx1	*C. elegans* (alpha/beta/gamma) neurexin protein
NLGN (1-5)	Human Neuroligin (Types 1-5) Genes
NLGN (1-5)	Human Neuroligin (Types 1-5) Proteins
NLGN4-X	Alternate name for NLGN4
NLGN4-Y	Alternate name for NLGN5
(α/β/γ)NRXN(1-3)	(alpha/beta/gamma) Human Neurexin (Types 1-3) Genes
(α/β/γ)NRXN(1-3)	(alpha/beta/gamma) Human Neurexin (Types 1-3) Proteins
N-gly	N-glycosylated sequence
O-gly	O-glycosylated sequence
O-Link	O-linked sugar modifier (protein feature)
PDZ	(PSD95) - (Dlg1) - (Zo-1) (protein feature)
SEC	self-excising cassette
SS(1-6)	Splice Sites 1-6 in Neurexins
SS(A/B)	Splice Sites SSA or SSB in Neuroligins
Sqt-1	Gene that, when expressed, results in a *C. elegans* roller phenotype
TMR	Transmembrane Region
WT	Wild type

1. INTRODUCTION

Given the complexity of our conscious experience, our ability to sense and perceive the world in minute detail, and to execute a range of responses to our environment, it is no surprise that studying these phenomena poses a formidable challenge. The mind – the seat of our thoughts, experiences, and feelings – is intangible, and is difficult to study directly or quantitatively. Instead, researchers have focused on understanding the nervous system, which, through its physical structure and function, is thought to give rise to the mind.

The fundamental unit of the nervous system is a type of cell called a neuron (**Figure 1**). This cell has the ability to receive chemical signals, either from the environment through a receptor or an upstream partner, at short, outwards-reaching parts of the neuron called dendrites (reviewed in Purves, et al., 2018). Neurons sum up all of the signals they receive at their dendrites and make a decision about whether to fire or not. If enough excitatory inputs are received to outweigh the inhibitory inputs, the neuron propagates an electrical signal through a second type of cellular structure known as an axon, which can carry this signal for a distance depending on the type of neuron. Finally, at the end of the axon, neurons convert the electrical impulse back to a chemical signal, either excitatory or inhibitory depending on the type of neuron. They convey this signal through small molecules known as neurotransmitters, which are released from the axonal membrane, known as the presynaptic membrane. Neurotransmitters drift away from this presynaptic membrane, across a space known as the synapse, before they reach the recipient of this signal, either another downstream partner, or an effector like a muscle or gland.

Neurotransmitters bind to receptors embedded in the postsynaptic membrane to create the excitatory or inhibitory signal meant for the next cell to receive. Thus, neurons are capable of receiving, processing, and relaying a message using chemical and electrical signals.

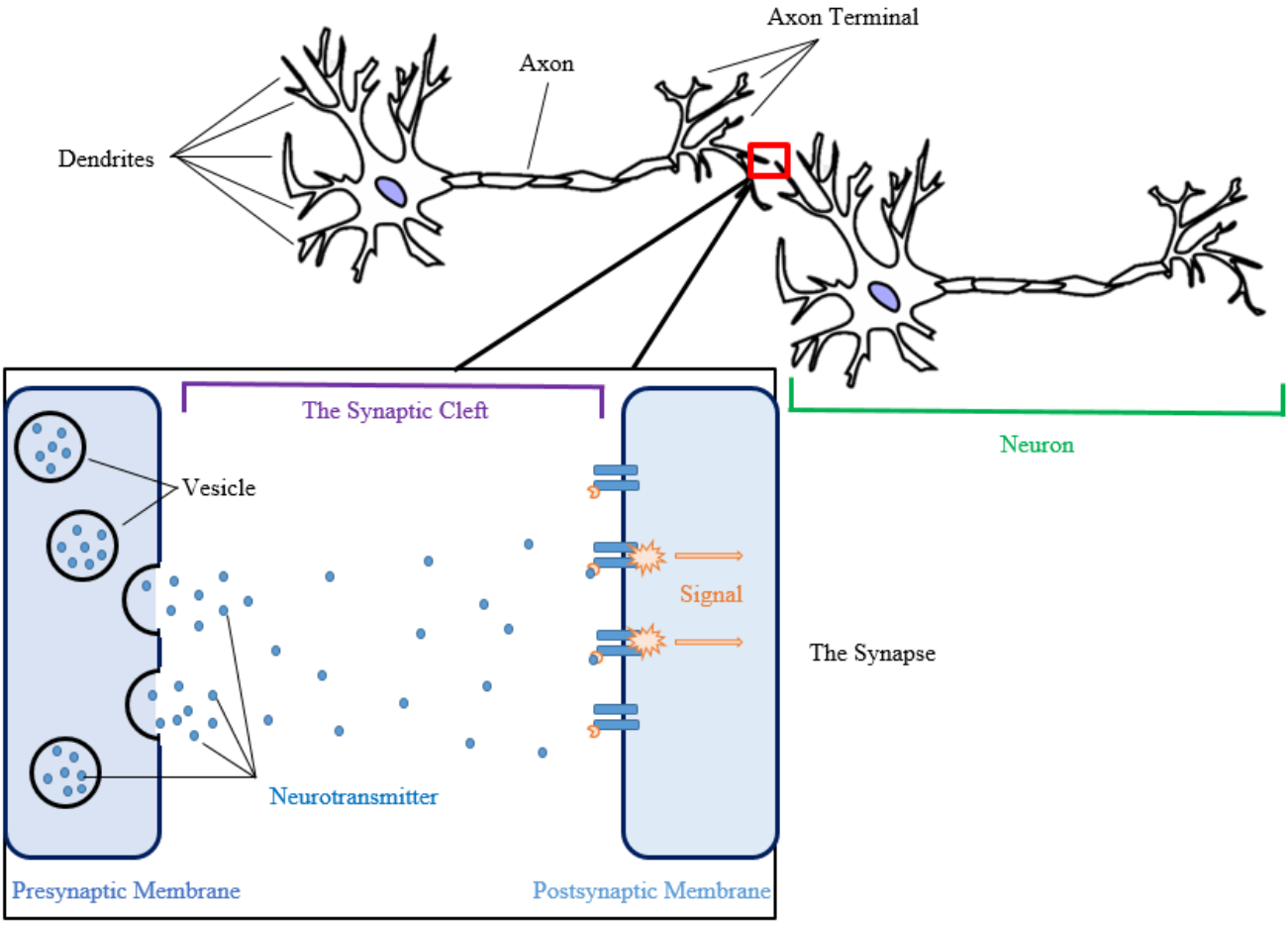


Figure 1. Overview of the Neuron and the Synapse

This illustration represents the structure of canonical neurons, with shorter dendrites that receive a signal, an axon that carries a signal to the vicinity of a neighboring cell, and an axon terminal where chemical messengers called neurotransmitters are released. The synapse, enlarged below the neurons, is the space across which neurotransmitters travel. These neurotransmitters are released from the presynaptic membrane, and are received at the postsynaptic membrane, allowing them to carry the message from one neuron to the next.

While a neuron is a powerful cell on its own, hundreds of billions of neurons (Herculano-Houzel, 2009) come together to form even more powerful networks in organisms (Purves, et al., 2018). The nervous system is generally divided into two major circuits: the peripheral nervous system, and the central nervous system. The neurons of the periphery are dispersed throughout the body and carry both information from receptor cells upwards towards central processing areas, and response commands from central structures downwards towards effectors. In contrast, in the central nervous system, the brain and spinal cord receive information, process it, and send out commands to the rest of the body. It is these central circuits that give rise to things like thoughts, desires, emotions, and behavior through the communication between different types of neurons in various brain regions. Therefore, it is not the neuron itself, but the communication and coordination of whole networks of neurons, that ultimately produce any type of behavioral or cognitive response.

Because even a small difference in how signals are communicated could alter end products as complex as thought or emotion, the synapse – the region at which this communication happens – is tightly regulated (Purves, et al., 2018). Synapses are carefully and regularly pruned to allow only the most productive connections to remain (Purves, et al., 2018). They are patrolled by regulatory enzymes to further modify the neurotransmitter signal. One especially interesting way synapses are regulated, is by structural proteins. Neurexin and neuroligin are a pair of presynaptic and postsynaptic structural proteins respectively that protrude from each membrane and bind to each other in the synaptic cleft. In previous literature, these proteins have been implicated in promoting synapse formation,

stabilization, and function; mutations in either partner have been linked to a myriad of disorders (reviewed in Dean & Dresbach, 2006; reviewed in Sudhof, 2017). Given the key role neurexins appear to play in the creation, operation, and regulation of synapses, further study of neurexins and neurexin mutants may lead to valuable insight into the function of healthy synapses, and the malfunction of diseased synapses.

1.1. Neurexins

In 1992, Ushkaryov and colleagues in the Sudhof lab were searching for the cellular targets of α-latrotoxin, a component of black widow spider venom (Ushkaryov, Petrenko, Geppert, & Sudhof, 1992). Previous electrophysiological work in the field revealed that binding to the α-latrotoxin receptor triggered massive amounts of neurotransmitter release (Nicholls, M, Scott, & Meldolesi, 1982). Others localized this receptor of interest to the presynaptic membrane through visualization of fluorescent antibodies for α-latrotoxin (Valtorta, Madeddu, Meldolesi, & Ceccarelli, 1984). After using peptide sequences of the bovine α-latrotoxin receptor, isolated by affinity chromatography, to create degenerate oligonucleotides, and generate partial cDNA sequences by PCR, researchers used these sequences as probes to isolate the receptor gene from rat brain cDNA libraries (Ushkaryov, Petrenko, Geppert, & Sudhof, 1992). From this search, they identified a set of genes with highly polymorphic, alternatively spliced protein products expressed almost exclusively in the central nervous system that they called neurexins.

Since Ushkaryov and colleagues originally discovered neurexins (1992), additional studies and reviews have focused on further structural characterization, isoforms and alternatively-spliced variants, roles in the synapse, and relevance to human disease

(Ichtchenko, et al., 1995; Ichtchenko, Nguyen, & Sudhof, 1996; reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; Missler, et al., 2003; reviewed in Sudhof, 2008; Zhang, et al., 2010; reviewed in Sudhof, 2017; reviewed in Sudhof, 2018; Trotter, et al., 2019). All neurexins are single-pass membrane proteins generally expressed on the presynaptic membrane. In humans and several other vertebrates, there are three separate neurexin genes, *NRXN1*, *NRXN2,* and *NRXN3,* that produce three types of neurexin proteins: NRXN1, NRXN2, and NRXN3 (reviewed in Sudhof 2018). Each of these types of neurexins has different forms depending on the promoter used in transcription. Transcription of the *NRXN1-3* genes from the furthest-upstream promoter produces a long α form protein (reviewed in Sudhof, 2017). Shorter β forms are transcribed from an internal promoter of those genes (reviewed in Sudhof, 2017). Additionally, a recently discovered third form of NRXN1, but not NRXN2 and NRXN3, called γNRXN1, has been shown to be transcribed from a second internal promoter to produce an even shorter, but still functionally relevant, product (Sterky, et al., 2017) (**Figure 2**). Thus, studies describe NRXN1 in α, β, and γ forms; NRXN2 in α and β forms; and NRXN3 in α and β forms in humans and other mammals.

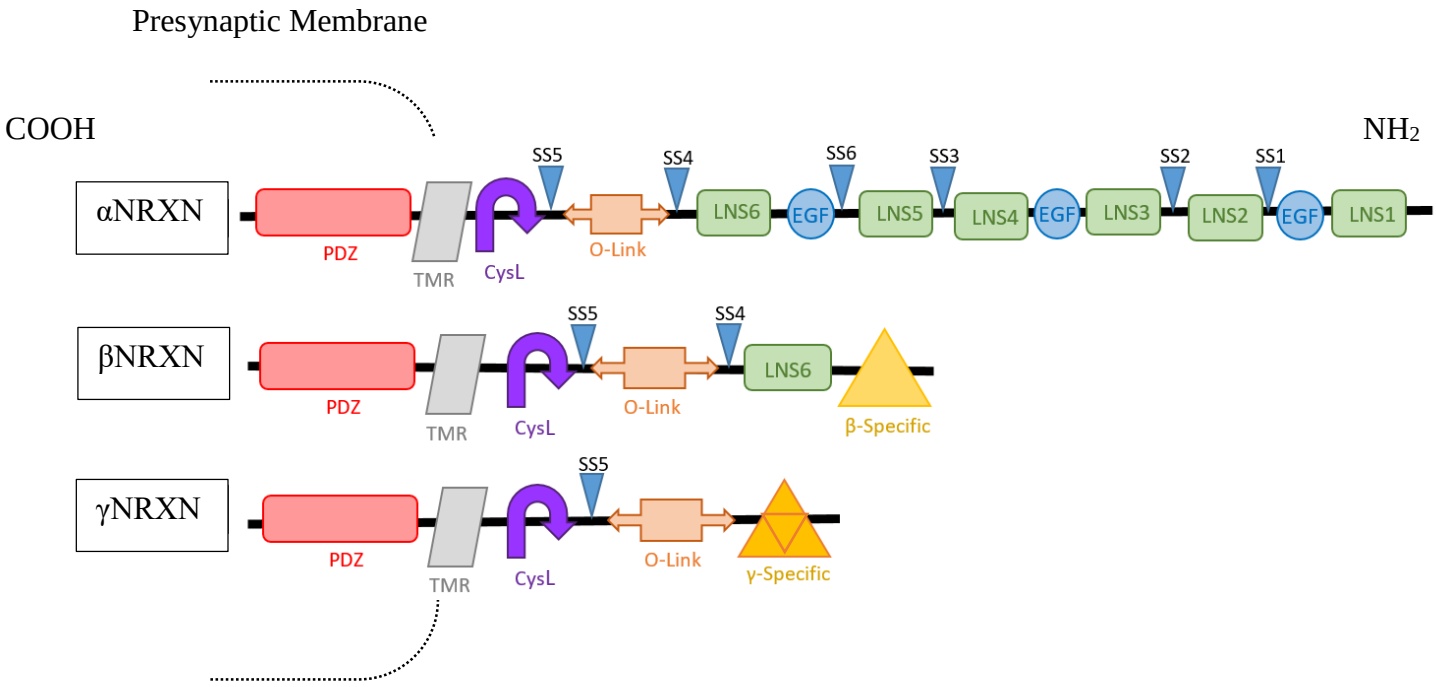


Figure 2. General Domain Schematic of Human NRXN Proteins

Representation of the general domain organizations in human αNRXN1-3, βNRXN1-3, and γNRXN1 proteins. Shown are the six laminin/neurexin/sex-hormone-binding globulin domains (LNS1-6), epidermal growth factor-like domains (EGF), the β-specific sequence (β-Specific) and threonine-rich γ-specific sequence (γ-Specific) regions, O-linked sugar modifier sequence (O-Link), cysteine loop regions (CysL), transmembrane regions (TMR), PDZ-binding domains (PDZ), and alternative-splicing sites (SS1-6). This protein structure is ordered from the C terminus to N terminus, to mimic orientation in the presynaptic membrane.

In humans, αNRXN, βNRXN, and γNRXN contain different protein domains that distinguish each isoform structurally and functionally. αNRXN proteins contain the most domains, with six laminin/neurexin/sex-hormone-binding globulin (LNS1-6) domains, and three interspersed epidermal growth factor (EGF)-like domains (Ushkaryov, Petrenko, Geppert, & Sudhof, 1992; reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; Chen, Jiang, Zhang, Gokce, & Sudhof, 2017; reviewed in Sudhof, 2017) (**Figure 2**). Following these domains are an O-linked sugar modifier sequence (O-Link), a cysteine loop (CysL), a single-pass transmembrane region (TMR) (reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; Chen, Jiang, Zhang, Gokce, & Sudhof, 2017), and finally, an intracellular PDZ-binding domain (PDZ) at the C-terminus (Fairless, et al., 2008) (**Figure 2**). βNRXN proteins, transcribed from an internal promoter, are shorter and contain an alternate starting sequence (β-specific) (reviewed in Ullrich, Ushkaryov, & Sudhof, 1995). In humans and other mammalian neurexins, the β-specific sequence is followed by the same LNS 6, O-Link, CysL, TMR, and C-terminal PDZ-binding domains as αNRXN (reviewed in Ullrich, Ushkaryov, & Sudhof, 1995; reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; reviewed in Sudhof, 2017). Finally, γNRXN1 in humans has been shown to start from a separate promoter, even further downstream on the *NRXN1* gene (Yan, et al., 2015; Sterky, et al., 2017). This form contains the same CysL, TMR, and C-terminal PDZ-binding domains as the other neurexins (Sterky, et al., 2017); however, unlike βNRXN1, γNRXN1 lacks all functional extracellular features like the LNS6 site in βNRXNs (**Figure 2**). Although differing in size, αNRXN1, which is 1477 amino acids (aa) long in humans (UniProt.org, Q9ULB1), has been known to share significant homology with αNRXN2

(1712 aa, 66% sequence identity) (UniProt.org, Q9P2S2), and αNRXN3 (1643 aa, 71% sequence identity) (UniProt.org, Q9Y4C0) (reviewed in Sudhof, 2017). βNRXN1 (472 aa) also shares high homology with βNRXN2 (666 aa, 75% sequence identity) and βNRXN3 (637 aa, 63% sequence identity). Finally, there is significant homology between the α and β isoforms: βNRXN1 residues 88 to 472 share 92% sequence identity with αNRXN1 residues 1123 to 1477; βNRXN2 residues 90 to 666 share 100% sequence identity with αNRXN2 residues 1136 to 1712; and βNRXN3 residues 84 to 637 share 99% sequence identity with αNRXN3 residues 1089 to 1643. As such, α and β NRXN1-3 have all been shown to preserve the general domain organization shown here as well (reviewed in Sudhof, 2017).

Finally, each type of *NRXN* gene can be alternatively spliced (reviewed in Sudhof, 2017). In humans *αNRXN*s have six such splice sites (SS1-SS6) throughout the extracellular domain coding region, *βNRXN*s have two (SS4 and SS5), and *γNRXN*s have one (SS5) (**Figure 2**).

Given the numerous human forms of NRXNs – between NRXN1-3, α and β forms and γNRXN1, as well as splice variants – it is unsurprising that these proteins play a myriad of roles in the synapse. While NRXN1 has been localized across from postsynaptic receptors for excitatory neurotransmitters like glutamate (Trotter, et al., 2019), NRXN2 forms have been associated with receptors for inhibitory neurotransmitters like GABA (Zhang, et al., 2010). Furthermore, alternative splicing, especially at SS4, where a 30 aa exon is included in the SS4+ variant, and excluded in the SS4- variant, has been implicated in allowing presynaptic NRXN proteins to positively or negatively influence responses at

the post-synaptic membrane (Dai, Aoto, & Sudhof, 2019). Even factors including the types of neurons at either end of the synapse, the neurotransmitter released, and the brain region in which those synapses are located, have all been shown to impact NRXN function in those synapses (Chen, Jiang, Zhang, Gokce, & Sudhof, 2017). The complexity of NRXN function in human and mammalian systems only increases when considering post-synaptic binding partners, including neuroligins.

1.2. Neuroligins

One of the major binding partners for neurexin protein is neuroligin. In 1995, Ichtchenko and colleagues created a baculovirus vector of truncated βNRXN1 fused to a histidine tag sequence to facilitate purification from a transfected Sf9 insect cell line (Ichtchenko, et al., 1995). Mixing this tagged protein with solubilized rat brain proteins in the presence of calcium ions (Ca^{2+}), researchers used an affinity column and found that their neurexin protein was bound to a 116 kDa protein, which they named neuroligin-1 (NLGN1). After protein sequencing and cDNA cloning based on this sequence, researchers determined the order of various protein domains, and two alternative splicing sites in the *NLGN* gene (**Figure 3**). Ichtchenko and colleagues experimentally confirmed the presence of several features, including a cholinesterase homology domain (ChEHD); alternative-splicing sites SSA (20 aa) and SSB (9 aa); N- and O-linked glycosylation sites (N-gly and O-gly); and a single-pass transmembrane region (TMR) (1995). It is unclear whether this O-gly in neuroligin proteins are the same one as the O-link sites in neurexin proteins. The dimerization site (D) was first characterized by sedimentation equilibrium experiments (Comoletti, et al., 2006), and multiangle laser light scattering experiments (Arac, et al.,

2007) in subsequent studies. Neuroligin dimerization was later found to be essential to help neurexins and other presynaptic elements localize to the synapse (Shipman & Nicoll, 2012).

Soon after the initial discovery of NLGN1, further scanning of rat brain cDNA libraries with PCR primers constructed from the *NLGN1* sequence revealed the existence of *NLGN2* and *NLGN3* genes and protein products (Ichtchenko, Nguyen, & Sudhof, 1996). Subsequent study of human cDNA libraries revealed human homologs for the three rat *NLGN* genes, as well a novel fourth *NLGN4* (Bolliger, Frei, Winterhalter, & Gloor, 2001), which has since been shown to have two functionally distinct, sex-linked forms: *NLGN4-X* and *NLGN4-Y* (also known as NLGN 5) (Nguyen, et al., 2020). Similar to the different NRXN proteins, NLGN1-5 proteins also show highly conserved domain order and homology (Ichtchenko, Nguyen, & Sudhof, 1996; Bolliger, Frei, Winterhalter, & Gloor, 2001; Nguyen, et al., 2020) (**Figure 3**). Alternative splicing of *NLGN2-4* has also been studied (Ichtchenko, Nguyen, & Sudhof, 1996; Bolliger, Frei, Winterhalter, & Gloor, 2001; Nguyen, et al., 2020). Compared to NLGN1, NLGN2 appears to show 62.8% sequence identity, NLGN3 shows 68.9% sequence identity, and NLGN4 shows 71.2% sequence identity (Bolliger, Frei, Winterhalter, & Gloor, 2001); NLGN4 and NLGN5, the X- and Y-linked forms, are 97% identical to each other (Nguyen, et al., 2020).

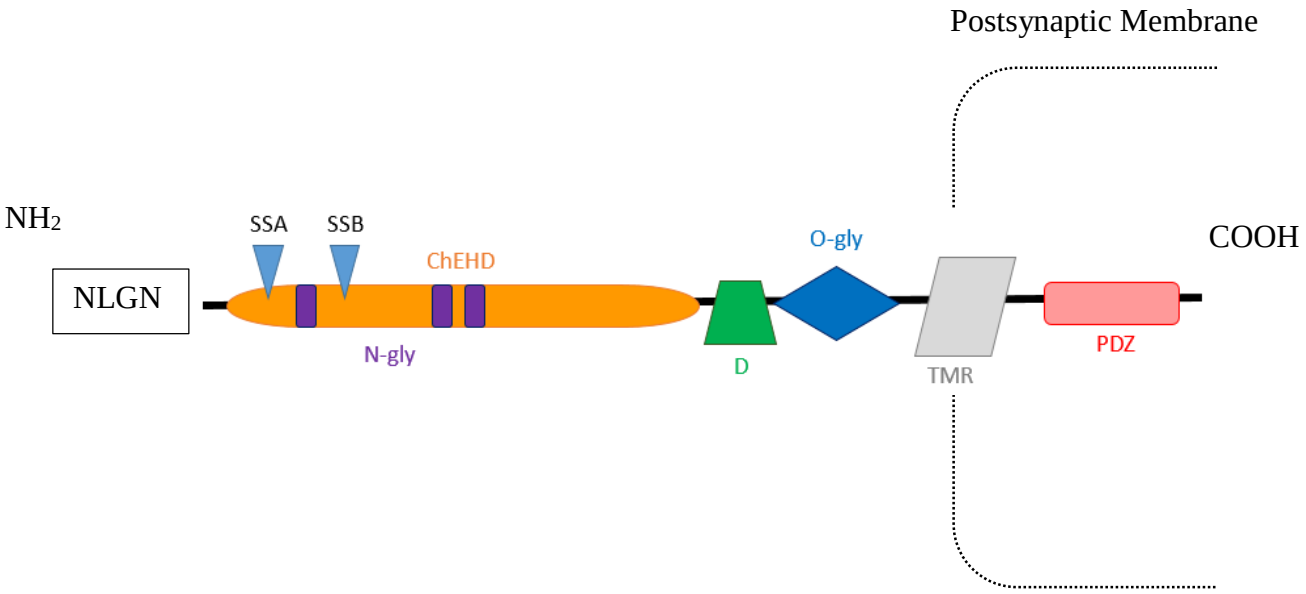


Figure 3. General Domain Schematic of Human NLGN Proteins

Representation of the general domain organization in human NLGN1-5. Shown are the cholinesterase homology domain (ChEHD), N-glycosylation sites (N-gly), dimerization site (D), O-linked glycosylation site (O-gly), transmembrane region (TMR), PDZ-binding domain (PDZ), and alternative-splicing sites (SSA and SSB). This protein organization is shown from the N terminus to the C terminus, corresponding to the orientation of neuroligins in the postsynaptic membrane.

Like neurexins, neuroligins appear to play different roles depending on the type. Studies on NLGN dimerization show that these proteins can exist as NLGN1, NLGN2, and NLGN3 homodimers, or as heterodimers composed of NLGN1-NLGN3, or NLGN2-NLGN3 (Poulopoulos, et al., 2012). Only NLGN dimers traffic to the cell surface; monomers appear to be sequestered to intracellular compartments, a phenomenon that is dependent on hydrophobic sequences in the transmembrane region (Poulopoulos, et al., 2012). Knockout studies in mice have revealed that NLGN1 tends to play critical roles in the postsynaptic membranes of excitatory synapses (Budreck, et al., 2013) and more lasting learning processes like long-term potentiation (Jedlicka, et al., 2015; Jiang, Polepalli, Chen, Zhang, & Sudhof, 2017; Wu, et al., 2019). NLGN2 tends to be localized to the postsynaptic membranes of inhibitory synapses (Varoqueaux, Jamain, & Brose, 2004; Gibson, Huber, & Sudhof, 2009; Liang, et al., 2015). On the other hand, NLGN3 appears to play no role at either type of synapse on its own, but may instead enhance the effects of synapses that already have NLGN1 or NLGN2 (Chen, Jiang, Zhang, Gokce, & Sudhof, 2017; Chanda, Hale, Zhang, Wernig, & Sudhof, 2017). NLGN4/5 are less commonly studied, but are also thought to play a role in inhibitory signaling (Nguyen, et al., 2020). Therefore, just like neuexins, neuroligins come in many forms and perform a variety of functions in the synapse in a form-specific manner.

1.3. Neurexin-Neuroligin Binding

Since their initial discovery, neuroligins have been studied as key binding partners of neurexins. All observed binding has been Ca^{2+}-dependent. When the NLGN1 protein was initially characterized, it appeared only to bind to βNRNXs that lacked the splice insert at SS4 (Ichtchenko, et al., 1995). Subsequent studies with NLGN2 and NLGN3 revealed similar preference for the NRXNβ protein lacking the insert at SS4 (Ichtchenko, Nguyen, & Sudhof, 1996). In all three cases, NLGN alternative splicing did not appear to affect βNRXN binding capabilities (Ichtchenko, et al., 1995; Ichtchenko, Nguyen, & Sudhof, 1996). However, this work failed to test alternative splice variants of NLGN against αNRXN as well; more recent work has demonstrated that αNRXNs, too, are involved in NLGN1-3 binding when the neuroligin lacks a splicing insert at SSB (Boucard, Chubykin, Comoletti, Taylor, & Sudhof, 2005; Comoletti, et al., 2006). Thus, due to structural homology, it seems likely that NLGN1-4 bind NRXN1-3 of both the α and β forms depending on the splicing of NRXNs at SS4 and NLGNs at SSB.

Structural analyses of both NRXN and NLGN proteins have revealed the essential domains for these binding interactions. Using small-angle X-ray scattering and neutron scattering techniques, the structure of NLGN1-βNRXN1 binding was visualized (Comoletti, et al., 2007). From this work, βNRXN1 residues D137 and N238 were proposed as key in binding Ca^{2+} ions, necessary for NRXN-NLGN binding; residues E297 and K306, which flank the SSB splice insert site, are critical for binding of βNRXNs; and that inserts at the SS4 site of NRXNs modulate NRXN-NLGN interactions in a complex, indirect way, as they are removed from the NRXN-NLGN interaction surface (Comoletti,

et al., 2007). Finally, this work proposed a model by which two NRXN monomers interacted symmetrically on either side of the NLGN dimer, spanning the synapse to interact in the middle (Comoletti, et al., 2007) (**Figure 4**).

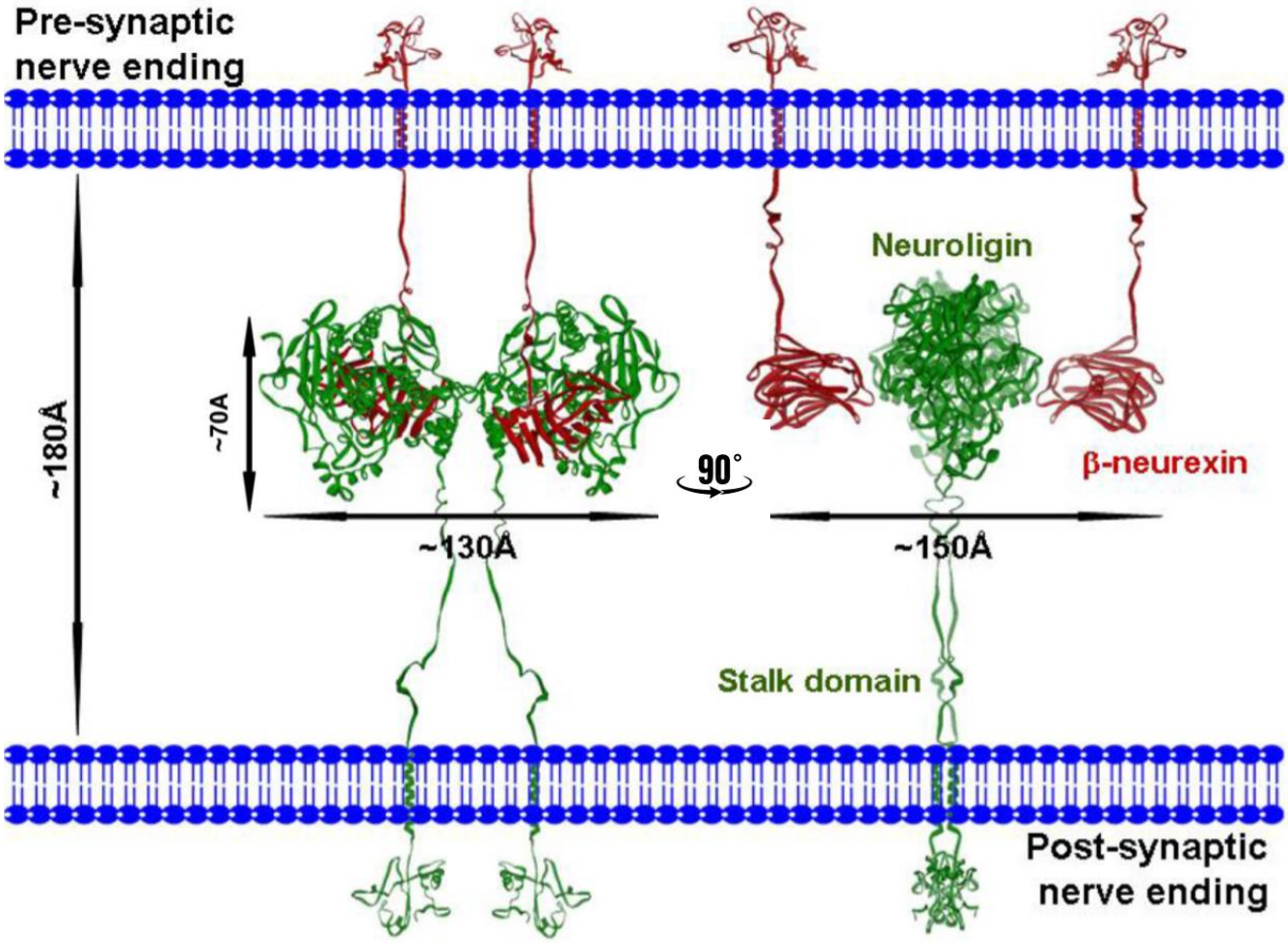


Figure 4. Model of NRXN-NLGN Trans-Synaptic Interaction

NLGN1 in green, and βNRXN1 in red, interact as a heterotetramer in the synaptic cleft, shown from two angles of view. Adapted from Figure 7 in Comoletti et al., 2007.

Structural studies (Comoletti, et al., 2007; Koehnke, et al., 2008) have allowed for important advances in understanding the functions and interactions of neurexin and neuroligin, which can have significant clinical implications. Expression of neurexins and neuroligins has been studied with increasing focus on development of the human brain, as well as the pathophysiology of disease (reviewed in Sudhof, 2008; Harkin, et al., 2017). Researchers have isolated mutations in both *NRXN* and *NLGN* genes associated with a range of neuropsychological diseases – including autism spectrum disorders, schizophrenia, Tourette's syndrome, and intellectual disability – in human patients (Jamain, et al., 2003; Laumonnier, et al., 2004; Zhang, et al., 2009; Gauthier, et al., 2011; Parente, et al., 2016; Huang, et al., 2017; Nakanishi, et al., 2017; Marshall, et al., 2017; Kathuria, et al., 2019). Many, if not all, of the mutations identified in this collection of literature are suspected of having detrimental structural implications for NRXN-NLGN binding in the synapse, disregulating synaptic communication, and resulting in the diseased condition. Therefore, a better understanding of NRXN-NLGN binding, the effect of mutations, and the regulation of the synapse by these proteins is essential to our understanding and, eventually, treatment, of these conditions.

Currently, researchers still have several questions regarding the functions of NRXN, NLGN, and the roles of NRXN-NLGN interactions in the synapse. Comoletti and colleagues (2007) acknowledged that different forms of NRXN and NLGN differ in their splicing, and therefore could exhibit different binding characteristics and structures. Other groups agree, recognizing the challenge of studying wild-type or mutant NRXN and NLGN in mammalian systems, where the effects of cell-type and synapse-type compound the

complexity of having multiple types of NRXN and NLGN interacting (Chen, Jiang, Zhang, Gokce, & Sudhof, 2017; Dai, Aoto, & Sudhof, 2019). Therefore, studying neurexins and neuroligins in a simpler, well-characterized model system may provide a solid foundation of understanding of these proteins, as well as relevant mutations, which can then be applied to mamalian and human systems.

1.4. Caenorhabditis elegans as a Model System for Neurexin-Neuroligin Studies

Caenorhabditis elegans was first proposed as a useful model system for genetic studies because of its small size, simple needs, and short life cycle (Brenner, 1973) and has since been widely used to study genetics in the nervous system (The *C. elegans* Sequencing Consortium, 1998; Hillier, et al., 2005; Felix & Braendle, 2010; Ruszkiewicz, et al., 2018). For the study of neurexins and neuroligins, *C. elegans* offers the additional advantages of having a nervous-system that is completely characterized (White, Southgate, Thomson, & Brenner, 1986; Cook, et al., 2019), as well as expressing only one homologous neurexin gene (*nrx1*) as opposed to the three human proteins (*NRXN1-3*), and one homologous neuroligin gene (*nlg1*) as opposed to the five human proteins (*NLGN1-5*) (Calahorro & Ruiz-Rubio, 2011). Thus, studying nrx1 and nlg1 proteins in *C. elegans* greatly reduces the complexity and difficulty encountered when using mammalian model systems.

C. elegans nrx1 is orthologous to the human NRXN1 protein (Haklai-Topper, et al., 2011; reviewed in Calahorro, 2014; reviewed in Sudhof, 2017). The *nrx1* gene can be transcribed from an upstream promoter to produce the αnrx1 protein, which shares general domain organization, including alternative splice sites, with the human αNRXN proteins

(reviewed in Calahorro, 2014) (**Figure 5a**). This makes the *C. elegans* α isoform similar to the human αNRXN1 protein in terms of domain organization. Additionally, the *C. elegans nrx1* gene can also be transcribed from a downstream promoter to produce a shorter isoform. Several previous studies (Haklai-Topper, et al., 2011; Calahorro & Ruiz-Rubio, 2011; Calahorro & Ruiz-Rubio, 2013; reviewed in Calahorro, 2014) call this shorter isoform βnrx1, and suggest a similar conserved domain and splice site organization, including LNS6 and the remaining C-terminal domains present in human αNRXN (reviewed in Calahorro, 2014) (**Figure 5a**). Experimentally, *C. elegans* nrx1 has been shown to share enough homology with the human proteins that human and rodent neurexins were able to phenotypically rescue the locomotive and sensory behavior of functionally inactive nrx1-mutant *C. elegans* (Calahorro & Ruiz-Rubio, 2013).

Likewise, *C. elegans* nlg1 is orthologous to the human NLGN4-X protein (Hunter, et al., 2010), but has also been compared to NLGN1 and NLGN3 (Calahorro & Ruiz-Rubio, 2012). Although there is evidence of *nlg1* alternatative-splicing in *C. elegans*, changes have been shown to affect intracellular domains rather than functional extracellular regions (Hunter, et al., 2010; Calahorro, 2014; Calahorro, Holden-Dye, & O'Conner, 2015). Finally, the nlg1 protein shares the same general domain organization as human neuroligins (reviewed in Calahorro, 2014; Calahorro, Holden-Dye, & O'Conner, 2015) (**Figure 5b**), and is also similar enough that human and rodent NLGN1 rescued the sensory deficits in functionally inactive nlg1-mutant *C. elegans* (Calahorro & Ruiz-Rubio, 2012).

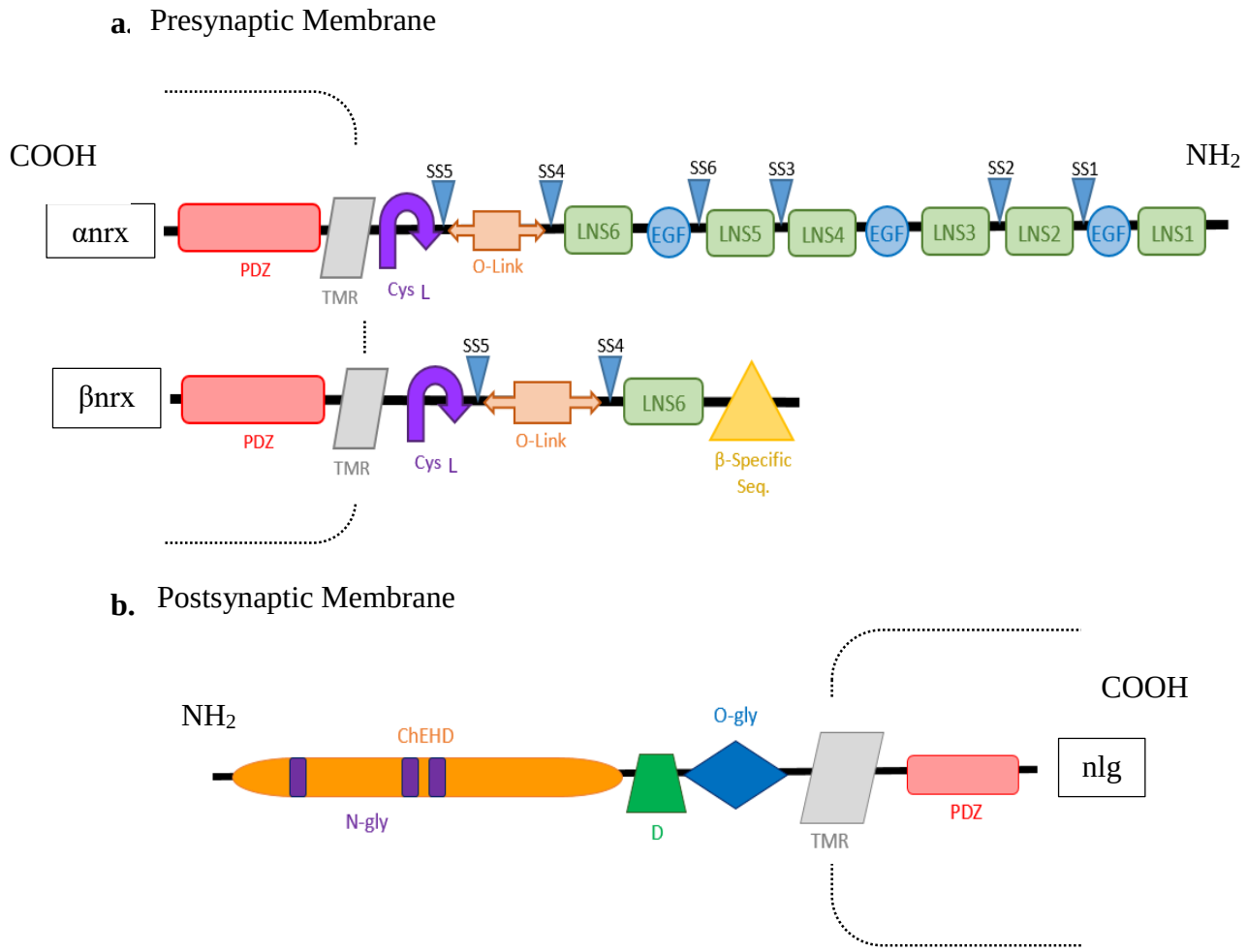


Figure 5. General Domain Schematic of *C. elegans* nrx1 and nlg1 Proteins

a. Representation of the general domain organizations proposed for *C. elegans* αnrx1, and βnrx1 proteins in the presynaptic membrane. Shown are the laminin/neurexin/sex-hormone-binding globulin domains (LNS1-6), epidermal growth factor-like domains (EGF), β-specific sequence (β-Specific), O-linked sugar modifier sequence (O-Link), cysteine loop regions (CysL), transmembrane regions (TMR), PDZ-binding domains (PDZ), and alternative-splicing sites (SS1-6). **b.** Representation of the general domain organization proposed for *C. elegans* nlg1 in the postsynaptic membrane. Shown are the

cholinesterase homology domain (ChEHD), N-glycosylation sites (N-gly), dimerization site (D), O-linked glycosylation site (O-gly), transmembrane region (TMR), PDZ-binding domain (PDZ), and alternative-splicing sites (SSA and SSB).

1.5. VC1416 and SG1 C. elegans Strains

The *C. elegans* model system is far simpler to manipulate than mammalian model systems, has a thoroughly characterized and mapped nervous system (Cook, et al., 2019), and expresses only one ortholog of neurexin and neuroligin. Therefore, researchers have started to characterize WT and mutant nrx1 and nlg1 *C. elegans* (Calahorro, Alejandre, & Ruiz-Rubio, 2009; Haklai-Topper, et al., 2011; CGC, University of Minnesota). While preliminary studies have characterized some nlg1 mutants showing impaired sensory behavor, results from these studies raise some interesting questions about the potential behavioral effects of different nrx1 alleles.

The Caenorhabditis Genetics Center (CGC, University of Minnesota) has two forms of nrx1 prominently discussed in the literature available: the VC1416 strain with the ok1649 allele, and the SG1 strain with the ds1 allele. In VC1416 strains, exon nine is deleted from the WT nrx1 sequence (**Figure 6**).

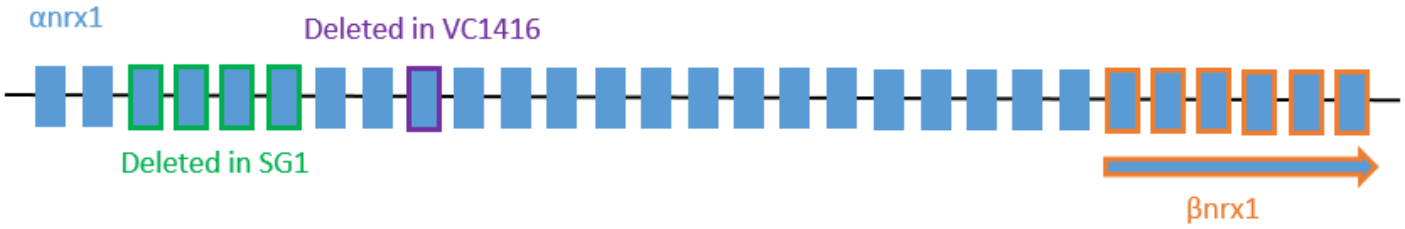


Figure 6. Schematic Representation of the nrx1 Exons

Each of the 29 total exons in αnrx1 is represented by a blue box. Those outlined in purple represent the deletion of exon nine, the ok1649 mutation in the VC1416 mutant strain, while those boxes outlined in green represent the deletion of exons three through six, the ds1 mutation in the SG1 mutant strain. Boxes outlined in orange show the six exons βnrx1 shares with αnrx1 (the seventh β-specific exon from the alternate start is not shown).

Despite resulting in a 62 aa deletion in a functional extracellular region of the only nrx1 present in *C. elegans*, the CGC described the VC1416 strain as having only mild behavioral deficits (CGC, University of Minnesota). Functional studies of this allele also suggested unclear and minor disruptions in *C. elegans* behavior (Calahorro, Alejandre, & Ruiz-Rubio, 2009). The CGC also described the ds1 allele in the SG1 strain, where exons three, four, five, and six are deleted from the WT nrx1 sequence (**Figure 6**). Even with a deletion of 163 aa in a functional region of nrx1, the CGC characterized these worms as having "no gross behavioral abnormalities" and only mildly changed sensitivity to some chemical compounds (CGC, University of Minnesota). In 2011, Haklai-Topper and

colleagues described this allele as well, but also failed to mention any gross phenotypic abnormalities associated with this nrx1 mutant (Haklai-Topper, et al., 2011). Therefore, both VC1416 and SG1 strains unexpectedly showed deficits less severe than anticipated. Given these findings, it may be interesting to further inspect the changes in structure produced by these deletions.

To explore why deficits in VC1416 and SG1 strains appear less severe than expected, we compare the sequence, domain-order, and 3-dimensional structures of WT human and *C. elegans* neurexins, and the VC1416 and SG1 nrx1 alleles. Based on the lack of structural deficits in 3-dimensional protein models, we determine that these mutations may not be sufficient to eliminate all nrx1 activity in *C. elegans*.

2. METHODS

2.1. Sequences and Homology

Human α and βNRXN aa sequences were obtained from the UniProt database (UniProt.org; accession no. Q9ULB1 (αNRXN1), Q9P2S2 (αNRXN2), Q9Y4C0 (αNRXN3), P58400 (βNRXN1), P58401 (βNRXN2), Q9HDB5 (βNRXN3)). The human γNRXN1 sequence was obtained from recent literature (Yan, et al., 2015; Sterky, et al., 2017; Kurshan, et al., 2018). *C. elegans nrx1* unspliced and spliced nucleotide sequences were obtained for the longest isoform available in the WormBase database, isoform k (WormBase.org; accession no. C29A12.4k.1), and aa sequences were also obtained for all thirteen nrx1 isoforms available on WormBase (WormBase.org; accession no.

C29A12.4a.1 - C29A12.4m.1). Sequence identity and similarity were compared between isoform k and the aa sequences of αNRXN1-3 to investigate homology through NCBI's Basic Local Alignment Search Tool (BLAST) (Altschul, Gish, Miller, Myers, & Lipman, 1990). The Constraint-Based Multiple Alignment Tool (COBALT) program (Papadopoulous & Agarwala, 2007) was used to directly compare the aa sequences of nrx1 isoforms a-m to each other to determine the regions of alternate splicing for which experimental expression evidence was reported.

VC1416 and SG1 strains with the ok1649 and ds1 alleles described in the literature (Calahorro, Alejandre, & Ruiz-Rubio, 2009; Haklai-Topper, et al., 2011) were provided by the CGC and mutated nrx1 sequences were confirmed in unpublished work by Stephanie Wang (Bayne Lab). Based on the reported nrx1 exon deletions in the VC1416 and SG1 strains, exon nine was deleted from nrx1 isoform k aa sequence to create the ok1649 allele, and exons three, four, five, and six were deleted from nrx1 isoform k aa sequence to create the ds1 allele. Alterations in splicing due to the deletions were checked using the Department of Bio and Health Informatics NetGene2 Server (Brunak et al., 1991; Hebsgaard et al., 1996). These two nrx1 sequences were also compared to the WT isoforms using COBALT.

2.2. Domain Predictions

Human αNRXN1, βNRXN1 and γNRXN1 aa sequences, along with select *C. elegans* isoform sequences (isoforms e f, k, and m) were entered into NCBI's Conserved Domain Search (CDS) program (Lu, et al., 2020). Results regarding LNS and EGF domain locations were used to inform predictions to localize the remaining domains and features

in α/β/γNRXN1 and nrx1 WT and nrx1 alleles of interest. EGF sites were predicted based on the presence of six cysteine residues within a stretch of 30 to 40 aa (Uniprot.org, KW-0245). O-Link sites were predicted at serine or threonine residues, especially those with a proline residue at +3 or -1 positions (Christlet & K, 2001). CysL sites were predicted as sequences between 5 and 13 aa long between two cysteine residues (Thompson, Lester, & Lummis, 2010). Finally, TMR sequences were predicted as long stretches of primarily hydrophobic residues, and PDZ-binding domain sequences were proposed in previous work (Calahorro, 2014).

2.3. Protein Models

Information about domain locations and boundaries from the nucleotide sequence analysis allowed for the generation of 3-dimensional models of the LNS1-EGF-LNS2-LNS3 region of the αNRXN1, WT nrx1 isoform k, and VC1416 and SG1 nrx1 alleles. Sequences (αNRXN1: 1-658aa; WT nrx1: 1-642aa; VC1416 nrx1: 1-580aa; SG1 truncated nrx1: 1-136aa) were submitted to the online C-I-TASSER program (Zheng, et al., 2021), and resulting PDB files at Rank 1 were downloaded and visualized in the PyMOL program (Schrodinger, LLC).

3. RESULTS

3.1. Sequences and Homology

The longest *C. elegans* nrx1 aa sequence (isoform k) was compared to human αNRXN1, αNRXN2, and αNRXN3 to determine homology using BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990). Of the different human NRXN genes, αNRXN1 shared

the most sequence identity with nrx1, 28% (47% sequence similarity), while αNRXN2 and αNRXN3 each shared around 27% (46% sequence similarity each).

All isoforms of nrx1 available on WormBase, along with the two alleles of interest – VC1416 and SG1 – were compared to each other using the COBALT program (**Figure 7**). The longest isoform, k, was comprised of 1725 aa, while the shortest isoform, e, contained only 137 aa. Isoforms were transcribed from an initial start (a, b, k, l), or from an internal start (b, c, d, f, g, h, i, j, m), except for isoform e, which started even further downstream (**Figure 7**). Exon nine of isoform k was deleted in the submitted sequence to match the ok1649 allele in VC1416 organisms, and exons three, four, five, and six of isoform k were deleted from the submitted sequence to recreate the ds1 allele from SG1 organisms. With no detected effect of the mutations on splicing using NetGene2, the VC1416 strain was expected to express a neurexin with a deletion of 62 aa, but an otherwise-undisturbed product with 1663 aa instead of 1725 aa (**Figure 7**). On the other hand, the larger deletion in the SG1 strain – 163 aa – also shifted the reading frame of the nucleotide sequence and resulted in a nonsense mutation. Therefore, with no change in splicing based on NetGene2 results, it was determined that the SG1 strain would express a truncated product of 136 aa instead of the WT 1725 aa from the initial start (**Figure 7**).

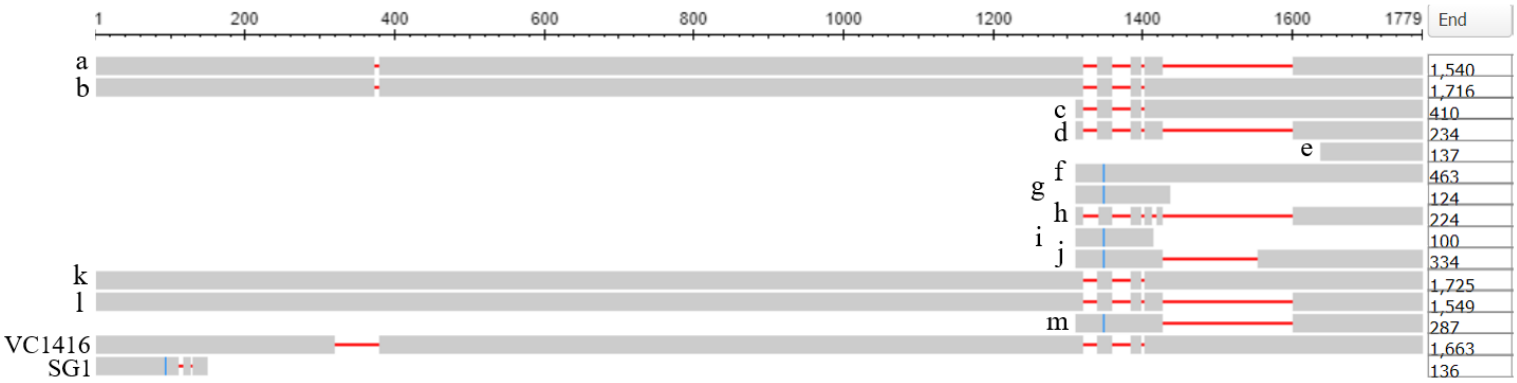


Figure 7. COBALT Analysis of WT Isoforms and nrx1 Alleles of Interest

Nrx1 isoforms (a-m) and alleles of interest are labeled on the left, and total aa length of each isoform is indicated in the column on the right. Grey areas are regions of sequence identity. Blue areas represent regions where the sequence differs. Red areas are regions where there are gaps in the sequence compared to other isoforms.

3.2. Domain Predictions

The location of all previously predicted domains were first mapped onto the aa sequence for WT α/β/γNRXN1 and WT nrx1 isoforms k, f, m, and e, to ultimately identify the domains affected in nrx1 proteins from VC1416 and SG1 strains. For WT αNRXN1, and nrx1 k isoform, the CDS program identified 6 LNS sites, and 1 EGF domain between LNS3 and LNS4 (**Figure 8**). Using literature-derived criteria, two additional EGF sites were predicted in the LNS1-LNS2 and LNS5-LNS6 areas in both the human and *C. elegans* sequences. In the WT βNRXN1 aa sequence, one LNS but no EGF sites were detected by the CDS program. In the WT γNRXN1 aa sequence, there were no LNS or EGF sites

identified by the CDS program. In the WT *C. elegans* m and f isoforms of nrx1, there were no LNS or EGF sites detected by the CDS program. Based on our literature-derived criteria, the remaining O-Link, CysL, TMR, and PDZ-binding regions were predicted and identified in each human and *C. elegans* aa sequence considered, except for isoform e. This isoform, whose transcription started from an even-further downstream promoter, contained only predicted TMR and the PDZ-binding domains based on literature-derived criteria.

With the domains mapped, deletions in the nrx1 alleles of interest were aligned with domains that may be affected. In the VC1416 strain, with the ok1649 allele, the deletion of exon nine was predicted to affect only a portion of LNS2, without interfering with the other domains (**Figure 8**). The SG1 strain, with the ds1 allele, deletions of exons three, four, five, and six were expected to affect LNS1 and all subsequent protein domains (**Figure 8**).

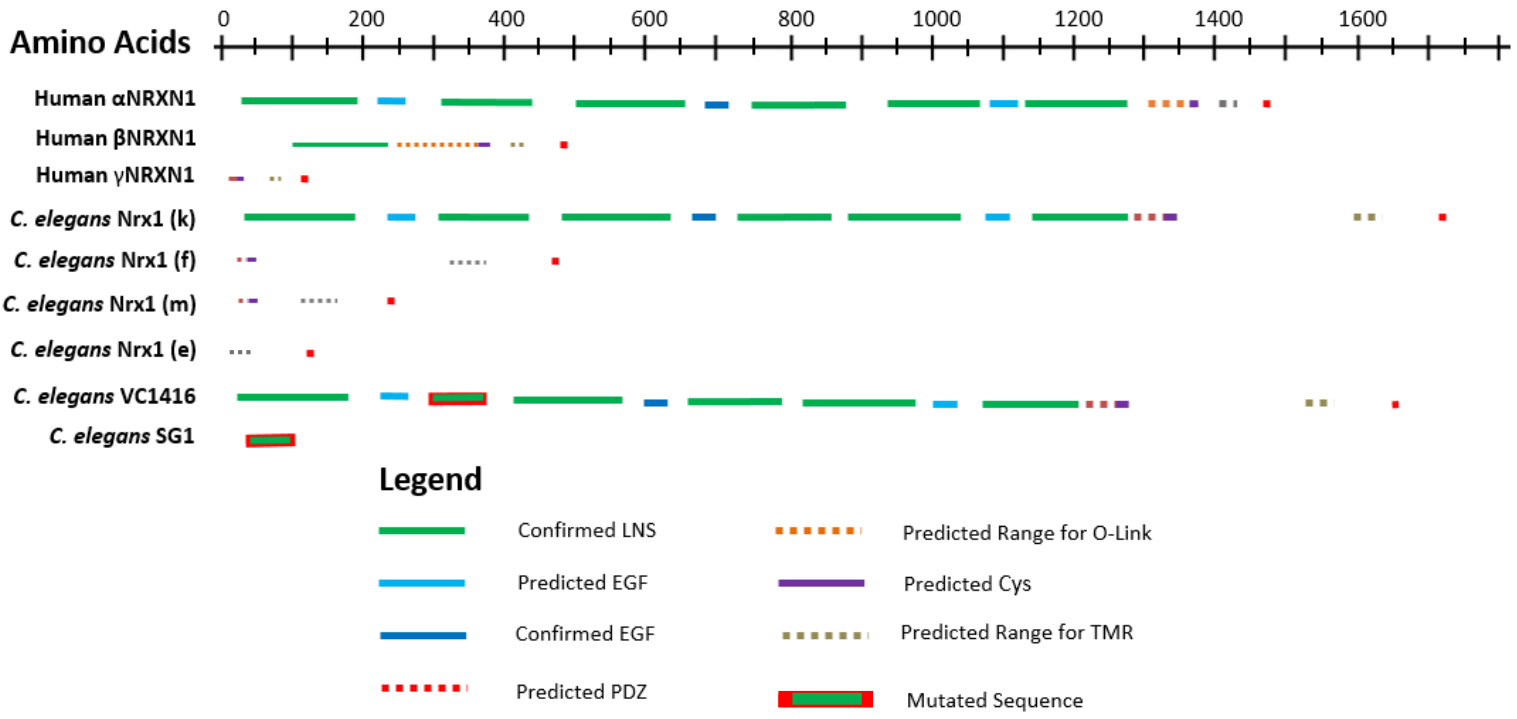


Figure 8. Predicted Locations for Domains in NRXN1 and nrx1 Proteins

Domains found by the CDS program included the confirmed LNS (dark green) and confirmed EGF domains (dark blue). Domain locations and feature ranges predicted based on current literature included the predicted EFG domain (light blue), predicted range for O-Link (orange), predicted CysL (purple), predicted range for TMR (gray), and predicted PDZ-binding domains (red). The domains affected by mutated sequences in VC1416 and SG1 nrx1 proteins are shown in green with a red highlight.

3.3. Protein Models

To gauge potential structural abnormalities in the nrx1 alleles of interest, C-I-TASSER and PyMOL programs were used to generate 3-D models of LNS1-EGF-LNS2-LNS3 sections of human αNRXN1, WT *C. elegans* nrx1 isoform k, and VC1416 and SG1 *C. elegans* strains. Models showed similarity between the structure of αNRXN1 and nrx1 isoform k, as predicted by sequence homology and the domain mapping (**Figure 9a**). Both proteins showed structural sections corresponding to the predicted LNS1, LNS2, EGF, and LNS3 with minor differences throughout each of these domains. These differences included the initial coil region, a slightly shifted LNS1 structure in αNRXN1 compared to nrx1, and a larger coiled structure near the EGF domain in αNRXN1 than nrx1. The two proteins also ended in slightly different locations. Next, structural comparisons were made between WT nrx1 and nrx1 alleles. The ok1649 allele from VC1416 *C. elegans* showed slight differences in the coiled regions at the start of the sequences, but maintained similar LNS1 and EGF domains to the WT (**Figure 9b**). The mutated region in the LNS2 domain differed from the WT, but LNS3 regained similarity, and the two proteins ended in about the same locations. Finally, the ds1 allele from SG1 *C. elegans* showed major differences, not only in the initial coiled regions of the protein, but also in LNS1 (**Figure 9c**). The truncation omitted the EGF, LNS2 and LNS3 domains entirely.

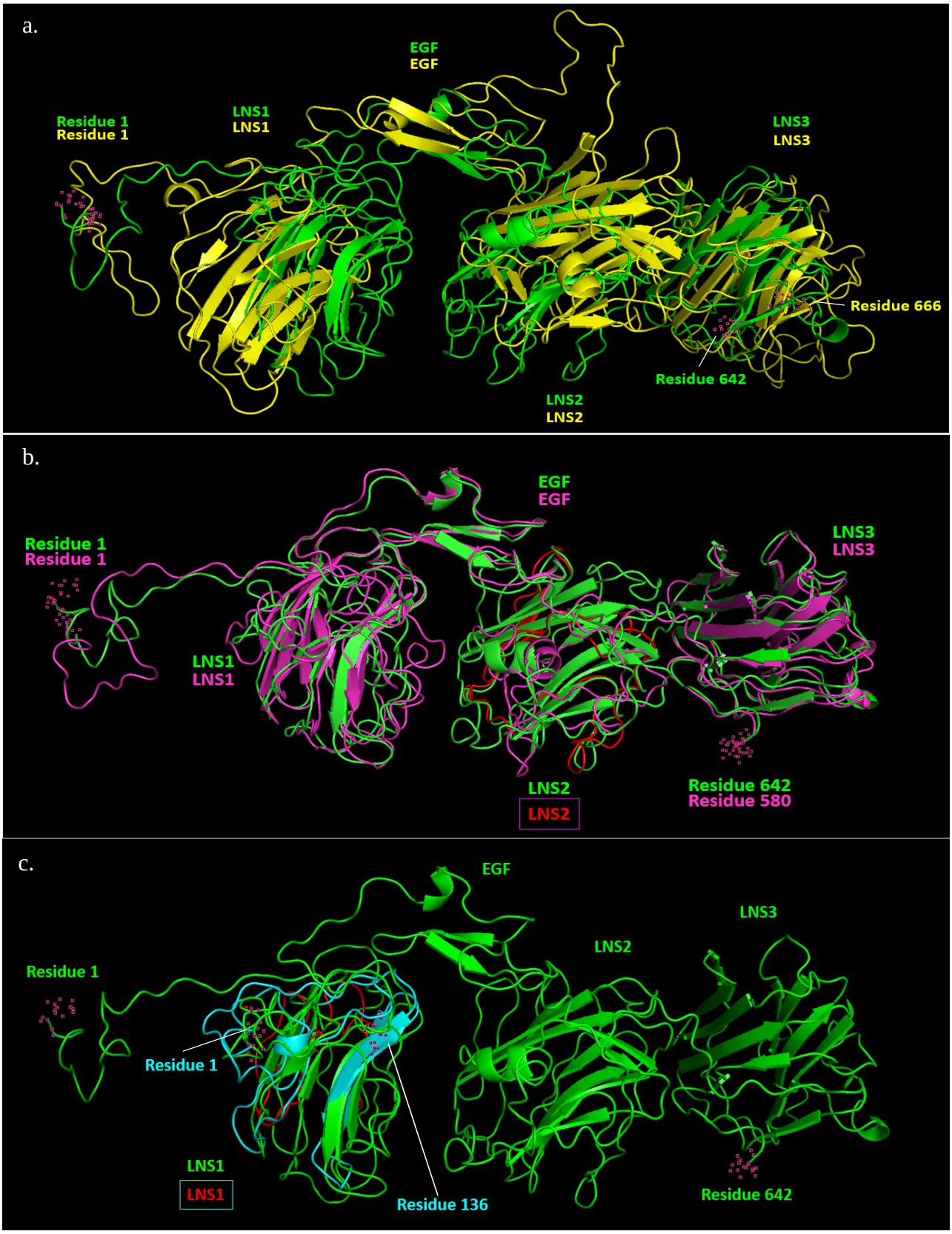

a.
EGF
EGF
Residue 1
Residue 1
LNS1
LNS1
LNS3
LNS3
Residue 666
Residue 642
LNS2
LNS2
b.
EGF
EGF
Residue 1
Residue 1
LNS3
LNS3
LNS1
LNS1
LNS2
LNS2
Residue 642
Residue 580
c.
EGF
LNS3
LNS2
Residue 1
Residue 1
LNS1
LNS1
Residue 136
Residue 642

Figure 9. Structural Comparison of LNS1-EGF-LNS2-LNS3 Domains

C. elegans WT nrx1 (green) compared to (**a.**) the human αNRXN1 (yellow), (**b.**) the ok1649 allele from VC1416 *C. elegans* (magenta), and (**c.**) and the ds1 allele from SG1 *C. elegans* (cyan). Deleted regions are shown in red (**b.** and **c.**).

4. DISCUSSION

4.1. Sequences and Homology

C. elegans have been used extensively in a variety of fields as a simple system in which complex interactions and processes may be modeled. Here, *C. elegans* are proposed as a useful model system to study neurexin, a critical regulator of synaptic formation and function. Current work in the field (Haklai-Topper, et al., 2011; reviewed in Calahorro, 2014; reviewed in Sudhof, 2017) suggests that the *C. elegans nrx1* gene and nrx1 protein are orthologous to the human *NRXN1* gene and NRXN1 protein. Of the thirteen nrx1 proteins reported in WormBase, isoform k (1725 aa) was selected as the most likely to be representative of αnrx1 in *C. elegans*, as it is the longest isoform (**Figure 7**). BLAST analysis of isoform k against αNRXN1, αNRXN2, and αNRXN3 revealed just slightly higher sequence identity and similarity between isoform k and αNRXN1 (28%, 47%), than αNRXN2 and αNRXN3 (27%, 46%), which is supported by the literature (Haklai-Topper, et al., 2011; reviewed in Calahorro, 2014; reviewed in Sudhof, 2017).

In humans, αNRXN1 is transcribed from the initial start site, and is the longest of the three forms. βNRXN1 is transcribed from a start site further downstream, and γNRXN1 is transcribed from the start site furthest downstream. To compare the reported isoforms of

nrx1 in WormBase to the human proteins, all thirteen nrx1 isoforms were analyzed using the COBALT program (**Figure 7**). The graphical output showed three potential start sites for *C. elegans* nrx1 proteins as well. The first start site, furthest upstream, produced isoform k, as well as other variants a, b, and l. These isoforms most likely corresponded to αnrx1 proteins. The majority of the reported nrx1 isoforms – isoforms c, d, f, g, h, i, j, and m – began at the next start site. The longest of these isoforms, isoform f (463 aa) was expected to correspond to βNRXN1. Interestingly, BLAST analysis revealed no significant sequence identity or similarity between isoform f and βNRXN1. In fact, there was no significant sequence identity or similarity found between βNRXN1 and isoforms c, d, f, g, h, i, j, or m based on aa sequence. Fianlly, isoform e, which began at the furtherest downstream start site, was compared against the γNRXN1 sequence. BLAST analysis could not detect any significant sequence identity or similarity between γNRXN1 and isoform e, or between any of the other twelve nrx1 isoforms based solely on aa sequence. Given the previous literature identifying the shorter nrx1 isoform as βnrx1 ((Haklai-Topper, et al., 2011; Calahorro & Ruiz-Rubio, 2011; Calahorro & Ruiz-Rubio, 2013; reviewed in Calahorro, 2014), these results were unexpected. Based on the similarity between αnrx1 and αNRXN1, mutations of αnrx1 are still important to study using a *C. elegans* model. At the same time, these results also suggest that *C. elegans* may have a slightly different set of neurexin isoforms than humans.

The aa sequence for isoform k was changed to recreate nrx1 alleles based on the DNA sequences reported for nrx1 in VC1416 and SG1 strains in the literature (Calahorro, Alejandre, & Ruiz-Rubio, 2009; Haklai-Topper, et al., 2011), commercial databases (CGC,

University of Minnesota), and from in-house sequencing data (Bayne Lab). Based on the deletion of exon nine in the VC1416 strain, only this section of the protein was expected to be missing, and the rest of the protein was expected to be expressed similarly to isoform k (**Figure 7**). Given the deletion of exons three, four, five, and six in nrx1 from the SG1 strain, the protein was expected to have a truncated α form (**Figure 7**). However, the limited characterization of these strains argued that VC1416 and SG1 organisms showed few significant behavioral deficits (Calahorro, Alejandre, & Ruiz-Rubio, 2009; Haklai-Topper, et al., 2011; CGC, University of Minnesota). These reports suggested value in structural characterization of mutant nrx1 prtoeins to predict if the deletions may be in a functionally important area and if these strains could be expected to have siginficant behavioral impairment based on the severity of the structural deformation.

4.2. Domain Predictions

Domain studies were pursued to identify the functionally relevant regions of neurexins as described in the literature (Haklai-Topper, et al., 2011; reviewed in Calahorro, 2014; reviewed in Sudhof, 2017), and to match these regions to those potentially mutated in nrx1 alleles of interest. For these studies, human αNRXN1, βNRXN1, and γNRXN1 were included to further explore the homology between human and *C. elegans* neurexins. Of the *C. elegans* isoforms, isoform k was chosen as a likely candidate for αnrx1; isoform f , the longest isoform starting from the second start site and isoform m, submitted to WormBase as “nrx-1 beta” (Calahorro & Ruiz-Rubio, 2013), were chosen as candidates for βnrx1; and isoform e was chosen for further study as the only isoform starting from the

third start site. Finally, the nrx1 proteins from VC1416 and SG1 strains were included for comparison of which domains were disrupted by the deletions.

Since the early neurexin literature, αNRXNs have been characterized as having six LNS domains, three interspersed EGF domains, an O-Link sequence, a CysL, a TMR, and a C-terminal PDZ-binding domain (Ushkaryov, Petrenko, Geppert, & Sudhof, 1992; reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; Fairless, et al., 2008; Chen, Jiang, Zhang, Gokce, & Sudhof, 2017; reviewed in Sudhof, 2017). The results of the NCBI CDS program identified 6 laminin sites corresponding to the 6 LNS sites discussed by previous literature, and one EGF site between LNS3 and LNS4, as expected (**Figure 8**). There appeared to be substantial regions of sequence without functional importance between LNS1-LNS2 and LNS5-LNS6. Upon further exploration, two additional EGF sites were predicted in these areas based on the characteristic presence of six cysteine residues within a stretch of 30 to 40 aa (Uniprot.org, KW-0245). This concentration of cysteine residues was also present in the EGF site identified by the CDS program, but was not found elsewhere in the αNRXN1 sequence. The predictions for two additional EGF sites in these areas also matched the organization of EGF sites in relation to LNS sites discussed in the literature (Haklai-Topper, et al., 2011; reviewed in Calahorro, 2014; reviewed in Sudhof, 2017). Following LNS6, a region containing S or T candidates for O-Link modifications was identified based on proximity to proline residues as previously suggested (Christlet & K, 2001). The CysL was identified as the only sequence of 5 to 13 aa between two C residues following LNS6 (Thompson, Lester, & Lummis, 2010). The TMR was identified as the longest stretch of hydrophobic residues after LNS6, and the

PDZ-binding domain was located in the sequence as previously identified (Calahorro, 2014). Overall, the domain locations and order of domains predicted for αNRXN1 based on aa sequence closely matched the information reviewed in previous literature (Haklai-Topper, et al., 2011; reviewed in Calahorro, 2014; reviewed in Sudhof, 2017). Based on this framework, the domains and domain order of *C. elegans* αnrx1 protein may be compared as a way to gauge functional similarity between these two proteins.

βNRXNs have been identified as forms of αNRXNs that begin from an internal start site (reviewed in Ullrich, Ushkaryov, & Sudhof, 1995; reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; reviewed in Sudhof, 2017). Characteristically, βNRXN1 features a unique starting sequence as a result of the alternate internal start, and retains only LNS6 along with the remaining C-terminal portions of the protein (reviewed in Ullrich, Ushkaryov, & Sudhof, 1995; reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; reviewed in Sudhof, 2017). In accordance with this previous literature, the CDS program predicted only one LNS domain towards the beginning of the βNRXN1 sequence – leaving space for the β-specific sequence – and identified no EGF sites in the sequence (**Figure 8**). No additional EGF sites were predicted via examination of cysteine residues. Consistent with the hypothesis of a truncated αNRXN1, O-Link, CysL, TMR, and PDZ-binding domains were also predicted in the βNRXN1 sequence based on the criteria discussed above. Comparisons of the known protein domains and domain order between human βNRXN1 and the *C. elegans* short isoforms helped us assess the validity of calling one of these short isoforms βnrx1.

The literature suggests that there are no functional extracellular domains in γNRXN1 proteins, which are transcribed from a third, unique internal start site (Yan, et al., 2015; Sterky, et al., 2017). As expected, the CDS program identified no LNS or EGF sites, in the γNRXN1 aa sequence (**Figure 8**). The criteria above predicted the locations of O-Link, CysL, TMR, and PDZ-binding regions outside of a region at the start of the sequence predicted to be the γ-specific sequence due to the internal start. This human γNRXN1 has not previously been described as having an orthologue in *C. elegans*.

To evaluate relationships between the human and *C. elegans* neurexin proteins, domains and protein features were also mapped onto select WT *C. elegans* nrx1 isofoms. With significant homology to αNRXN1, as well as characterization in the literature (Haklai-Topper, et al., 2011; Calahorro & Ruiz-Rubio, 2011; Calahorro & Ruiz-Rubio, 2013; reviewed in Calahorro, 2014), isoform k was expected to show a similar organization including six LNS domains, three interspersed EGF domains, an O-Link sequence, a CysL, a TMR, and a C-terminal PDZ-binding domain. As expected, the CDS program identified all six LNS sites, as well as one EGF domain (**Figure 8**). Two additional EGF domain were predicted as discussed above, along with O-Link, CysL, TMR, and PDZ-binding domains. Thus, isoform k is considered a close match to human αNRXN1, and may be characterized as a form of αnrx1.

Although isoform k showed significant similarity with αNRXN1, isoforms f and m showed some significant differences from βNRXN1. Isoform f, the longest available isoform whose transcription started from the second internal start, was examined as one potential candidate for βnrx1 in *C. elegans*. While human βNRXN1 contained one LNS

site preceeding the C-terminal elements, the CDS program could not identify any LNS sites in the *C. elegans* isoform f aa sequence (**Figure 8**). Instead, only O-Link, CysL, TMR and PDZ-binding domains were predicted in this sequence, indicating a significant difference between human βNRXN1 and the short *C. elegans* isoform f. Curator comments on WormBase indicated that isoform m may be another βnrx1 candidate. However, similarly to isoform f, there were no LNS or EGF sites identified in the aa sequence – only O-Link, CysL, TMR and PDZ-binding sites. Because the alignment of all thirteen available isoforms suggested that isoforms either began at the initial start or at one of the internal start sites, it is unlikely that any of the other shorter isoforms include uptream sequence corresponding to a LNS site as expected for βnrx1. Instead, based on this absense of LNS6, we determined that short isoforms more closely resemble γNRXN1, which also lacks LNS and EGF domains in the extracellular portion of the protein. Therefore, based on the similarity between isoform k and αNRXN1, and the lack of an LNS site in any short isoforms, we propose that *C. elegans* contain neurexins that correspond to αnrx1 and γnrx1 (**Figure 10**). At this time, there appears to be no expressional evidence to suggest a βnrx1 isoform based on domain analysis.

Finally, analysis of isoform e predicted even fewer elements – only a short TMR and PDZ-binding site – as expected from a protein whose transcription starts even further downstream. evidence for isoform e, a third novel form (**Figure 10**), Evidence for a third, shorter nrx1 isoform had not been previously discussed in the literature. Hovever, based on the elements present in this form, - a short extracellular sequence, the TMR and the PDZ binding domain – it is not expected that this form has many extracellular interactions. The

presence of the PDZ domain on the intracellular side, though, may suggest some ability to interact with intracellular elements, like neutrotransitter receptors or other synaptic regulatory elements.

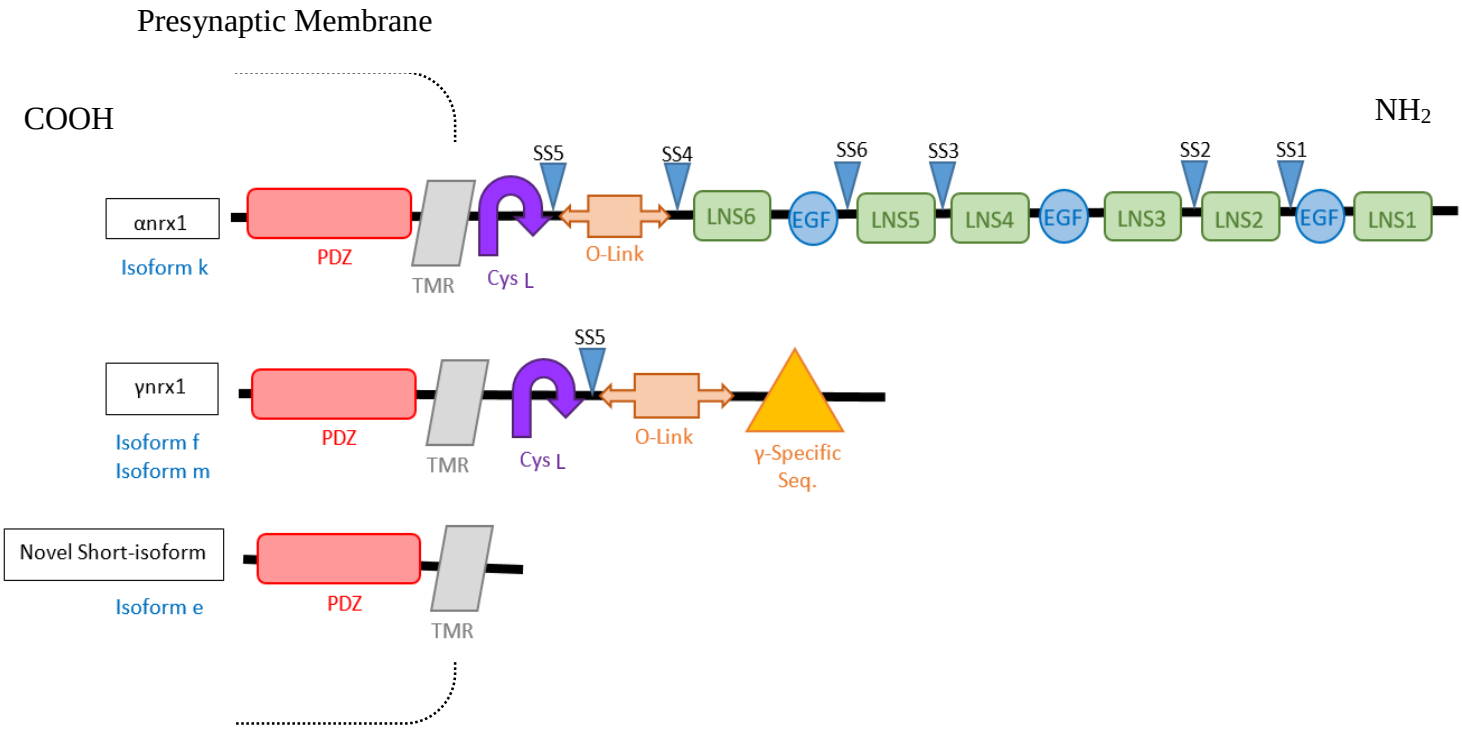


Figure 10. Revised Domain Schematic of *C. elegans* nrx1 Proteins

Representation of the types of nrx1 proteins found in *C. elegans* based on domain mapping studies. αnrx1 proteins like isoform k largely share domain organization with human αNRXN1. On the other hand, shorter isoforms like f and m share more similarity to γNRXN1 – which also lacks extracellular LNS or EGF domains – than to βNRXN1, which contains one LNS site. Therefore, we propose the classification of these isoforms as γnrx1.

One additional short form, isoform e – which was not discussed in the literature – was also identified as starting from a third start site. This form appears to differ from any of the known human NRXN proteins.

Finally, domains of the two VC1416 and SG1 nrx1 proteins were mapped. CDS analysis of the VC1416 nrx1 allele found five complete LNS sites in the same locations as LNS1 and LNS3-6 in the isoform k sequence, with one partial LNS site identified in the location of LNS2 (**Figure 8**). These results were expected, as the deletion of exon nine in the ok1416 allele was expected to disrupt LNS2 and leave the rest of the protein sequence unaffected. In accordance with these predictions, CDS analysis detected the same central EGF site, the two additional EGF sites were predicted manually using the same criteria as above. C-terminal structures – O-Link, CysL, TMR, and PDZ-binding sites – were also predicted as expected, with the characteristic gap between the CysL and TMR that was present in the other nrx1 isoforms studied. Interestingly, given that the rest of the domains appear undisturbed, VC1416 strain may retain the ability to express a largely functional α form. Additionally, because transcription of the γ form begins from an internal start site, VC1416 strains may also retain the ability to express functional γnrx1 proteins. Similar to studies of VC1416 nrx1, CDS analysis of the SG1 nrx1 revealed one partial LNS site (LNS1), but no additional elements were predicted by either the program, or found manually based on literature-derived criteria. These results were also expected, because the ds1 allele's deletion of exons three, four, five, and six was predicted to cause a disruption

of the protein in LNS1 along with a subsequent frameshift that would change the rest of the sequence and lead to a nonsense mutation. While the SG1 strain would therefore not be expected to produce a functional αnrx1 protein, these organisms may, like VC1416, retain the ability to express functional γnrx1 forms from the internal start site.

Overall, studies to map the protein domains according to aa sequence proved valuable in confirming the similarity between αnrx1 and αNRXN1, classifying the *C. elegans* short isoforms more correctly as γnrx1 proteins, uncovering a novel shorter isoform, and most importantly in identifying the functional regions disrupted in the VC1416 and SG1 nrx1 alleles.

4.3. Protein Models

Although the sequence was conserved, the ok1649 allele could still disrupt the folding and structure of the nrx1 protein, resulting in the loss of function. To further study potential disruptions in both VC1416 and SG1 nrx1 protein structure, 3D models were created of the LNS1-EGF-LNS2-LNS3 regions of human α NRXN1, WT *C. elegans* αnrx1, and VC1416 and SG1 *C. elegans* nrx1 proteins (**Figure 9**). As structures were compared, predictions for domain location and order were also confirmed. In the structural protein models (**Figure 9**), three separate LNS domains were identified, with an EGF domain identified between LNS1 and LNS2, just as discussed in the literature (Ushkaryov, Petrenko, Geppert, & Sudhof, 1992; reviewed in Missler, Fernandez-Chacon, & Sudhof, 1998; Chen, Jiang, Zhang, Gokce, & Sudhof, 2017; reviewed in Sudhof, 2017) and suggested by earlier domain mapping results (**Figure 8**).

The structure of WT αnrx1 (isoform k) (642 aa) was compared to αNRXN1 (666 aa) (**Figure 9a**). As the sequence homology and domain similarity suggested, much of the 3D structure is also similar, with some minor differences in the orientation of some LNS sites and in some coil regions. These observed similarities are in accordance with previous experimental work where αNRXN1 was introduced to a *C. elegans* model lacking functional αnrx1 and was able to rescue behavioral abnormalities (Calahorro & Ruiz-Rubio, 2013). These results again reinforce the idea that isoform k, representative of αnrx1 proteins, is structurally – and therefore potentially functionally – to αNRXN1.

Next, the VC1416 strain nrx1 structure (580 aa) was compared to αnrx1 (isoform k) structure (642 aa) (**Figure 9b**). In accordance with the predictions from protein domain mapping, the VC1416 αnrx1 structure shows high degrees of similarity to the WT αnrx1 structure at the start of the sequence, in LNS1 and the EGF domains. As expected, the deletion causes changes in the LNS2 domain, which may be responsible for some slight loss of function. However, the structure once again matches the WT structure in LNS3, and – because there are no other deviations from the WT sequence – for the rest of the protein. Because the VC1416 αnrx1 protein structure is only slightly disturbed in LNS2, and appears to be undisturbed downstream of the deletion, these strains could produce αnrx1 forms that retain the majority of their function. Previous literature also identifies LNS6 as the site responsible for binding and interaction to nlg1 in *C. elegans* (Hu, et al., 2012). This LNS6 site also remains undisturbed in the VC1416 αnrx1 form, suggesting another way this mutant may retain function. Finally, VC1416 worms may also produce functional γnrx1 forms, which have previously been shown to have functional activity of their own in

the synapse, even without interactions with nlg1 (Yan, et al., 2015; Sterky, et al., 2017). Thus, despite their sizeable deletion, it may be possible for VC1416 nrx1 proteins to retain function in the synapse, which could explain the minimal behavioral deficits observed in the literature (Calahorro, Alejandre, & Ruiz-Rubio, 2009; Haklai-Topper, et al., 2011) and by the CGC (CGC, University of Minnesota).

Finally, the SG1 nrx1 protein structure (136 aa) was compared to the WT αnrx1 (isoform k) structure (642 aa) (**Figure 9c**). As expected from the domain mapping studies, SG1 αnrx1 is prematurely truncated in the LNS1 domain, causing even the initial coil portion of the protein to adopt abnormal folding that is different from the WT αnrx1. While it is unlikely that this product retains much function, it is possible that SG1 strains may also continue to produce functional γnrx1 proteins from the internal start site, helping to preserve some behavioral patterns in these *C. elegans* during behavioral characterization in the literature (Calahorro, Alejandre, & Ruiz-Rubio, 2009; Haklai-Topper, et al., 2011) and by the CGC (CGC, University of Minnesota).

In previous work, α and γ neurexin forms have been studied both in mammals and in *C. elegans*. In mice, neurexin proteins were found to be expressed specifically in synapses of the brain (Ushkaryov, Petrenko, Geppert, & Sudhof, 1992). In 2003, Missler and colleagues created mice models to KO all αNRXN proteins, finding that triple-KO mice all died on the first day, double-KO mice all died within the first week, and that single-KO mice showed highly impaired survival (Missler, et al., 2003). Their studies also found the αNRXNs in mammals might be highly related to organization of presynaptic elements, especially those involved in calcium currently related to neurotransmitter release (Missler,

et al., 2003). In *C. elegans*, αnrx1 proteins have been implicated in assembling with neuroligins to organize inhibitory neurotransitter receptors like GABA receptors (Kurshan, et al., 2018). Given this evidence, the mild reported impariments seen in VC1416 and especially SG1 *C. elegans* models – which, according to our protein model evidence, expressed impaired αnrx1 proteins – were unexpected. These results may be further tested by complete knockouts of αnrx1 in *C. elegans* to see the extent of behavioral abnormality.

Simultaneously, Kurshan and colleagues acknowledge that the lack of major defecits seen in loss-of-function studies of αnrx1 specifically may implicate parallel elements that also regulate the organization and function of these synapses (Kurshan, et al., 2018). In *C. elegans*, it may be possible that the γrnrx1 proteins are the dominant species, and carry out the majority of important functions independent of neuroligin binding; this way, even worms that lose αnrx1 may not show significant impairment. In mice, γNRXN1 was highly expressed in the cerebellum, and appeared to have increasing expression with postnatal age (Yan, et al., 2015; Sterky, et al., 2017). KO studies of γnrx1 in *C. elegans* show a substantial decrease in release of neurotransmitters across the synapse in these organisms (Kurshan, et al., 2018). Therefore, it might also be interesting to further study the specific functions of γnrx1 proteins through a γnrx1-KO model in *C. elegans.*

Overall, conclusions from 3D modeling studies support those reached following protein domain mapping. WT αnrx1 (isoform k) shares structural similarities with human αNRXN1 (**Figure 8**, **Figure 9**). Despite deletions in LNS2, VC1416 strain nrx1 proteins have the potential to retain significant function throughout the rest of the αnrx1 form, as well as through the γnrx1 forms. Finally, in spite of a grossly truncated αnrx1 form, SG1

nrx1 proteins may also retain function through the production of γnrx1 forms from the internal start site. These findings corroborate the mildly-deficient characterizations of VC1416 and SG1 strains. Complete removal of the *nrx1* gene, as well as the αnrx1 and γnrx1 forms specificlly, may be needed to produce observable abnormalities in *C. elegans*, to give researchers insight into influence on behavior and function.

6. FUTURE DIRECTIONS

6.1. nrx1-KO in C. elegans

Based on the evidence for mild deficits and possible retention of function in VC1416 and SG1 nrx1 proteins, three KO models were designed to further characterize nrx1 function in *C. elegans*: a total-KO (1), an α-KO (2), and a KO of the shorter nrx1 isoform (3). For each KO model, a separate CRISPR-Cas9 strategy was designed to excise different portions of the *nrx1* gene from the *C. elegans* genome. Three recovery plasmids (pKP1, pKP2, and pKP3) were also designed, containing DNA to replace the portion of the nrx1 gene excised by the CRISPR-Cas9 system. Each recovery plasmid would contain a combination of elements from the pDD287 plasmid and synthetically designed sequences (Syn1, Syn2, and Syn3) inserted into a pUC57 Kan backbone (pKPa, pKPb, and pKPc). These recovery plasmids and KO-models of *C. elegans* may be generated in future work to further study the sensory and behavioral effects of losing nrx1.

Three CRISPR-Cas9 systems were designed to create each of the three nrx1 KO models in *C. elegans*. To generate the total nrx1-KO (total-KO), the Optimized CRISPR

Design Tool (Zhang Lab, 2017) was used to predict guide RNA (gRNA) target sites and generate complimentary gRNA sequences (5':GATAAAGCCAGCATGGCAGG, 3':CTCATGGGCAAAACTTCAAC). The same 5' gRNA sequence was retained for the α-KO model, as the same promoter was targeted. Because the Zhang Lab discontinued the Optimized CRISPR Design Tool, the FlyCRISPR Target Finder program (Gratz, et al., 2014) was used to find a 3' gRNA target sites and generate the complimentary gRNA (3':CAGATTGACAGAATAAATTT). For the third KO-model, the FlyCRISPR Target Finder was also used to find target sequences in αnrx1 introns to disrupted only the internal promoter responsible for producing the γnrx1 isoform for the third KO model (5': ACACACAGACTTGTGGAGCC, 3':ACAGGCCTCTCTGATGACCA), allowing only the longer α form to be expressed as in WT *C. elegans*. The proposed strategy required each of these 5 gRNA sequences to be cloned into pRB1017, gifted by Andrew Fire (Addgene plasmid #59936; Arribere, et al., 2014) at the BasI sites, to prepare three unique sets of 5' and 3' gRNA plasmids for each KO model. The required Cas9 protein would be introduced via the pDD121 plasmid, a gift from Bob Goldstein (Addgene plasmid #91833).

When a set of gRNAs direct the CRISPR-Cas9 machinery to make two double-stranded cuts in the genome, a DNA sequence called a recovery plasmid would be introduced to the cell to be inserted into this space instead of the gene that was previously there. This recovery plasmid would contain elements like flanking regions of homology to encourage homology-directed repair, and other useful gene sequences that may expressed to aid in selection.

The total-KO recovery plasmid, pKP1, was designed to take the place of the entire nrx1 gene in the *C. elegans* genome (**Figure 11)**. The Total-KO recovery plasmid synthesis strategy would involve a combination of the pDD287 (mKate2^SEC^3xMyc) plasmid gifted by Bob Goldstein (Addgene plasmid #70685) and a specially-designed pUC57 Kan (GenScript Cat. No SD1176) plasmid (pKPa) constructed by GenScript. On one hand, pDD287 would be digested with *Stu*I and *Xma*I. This digestion would result in Fragment 1, containing a *Stu*I blunt end; the partial sequence of the mKate2 red fluorescent protein (-ate2; where mK- and –ate2 represent the two fragments of the mKate2 sequence); the self-excising cassette (SEC) sequence, including the Sqt-1 gene, the Cre recombinase, and the Hygromycin B Resistance (Hyg^R) gene, surrounded by Lox2272 sites; and a *Xma*I sticky end (**Figure 11a**).

To generate Fragment 2, the pUC57 Kan plasmid would be digested with *Eco*RV and *Stu*I (**Figure 11b**) to provide the backbone. A specialized synthetic fragment (Syn1) was designed to contain a 500 base pair (bp) 5' homology flank (5'F) needed for homologous recombination, a reconstruction of the red fluorescent protein mKate2 up to the *Stu*I site (mK-), an additional *Xma*I site, and a 500 bp 3' homology flank (3'F) also needed for homologous recombination. This Syn1 sequence would be inserted between the *Eco*RV and *Stu*I sites on pUC57 Kan (**Figure 11c**). In the 5'F and 3'F sequences, the first letter of the PAM sequence originally targeted by the gRNA would be changed to a T nucleotide, to disrupt subsequent targeting by CRISPR-Cas9. Digestion of pKPa with *Stu*I and *Xma*I results in Fragment 2, with ends that match those on Fragment 1 (**Figure 11d**).

Ligation of Fragment 1 with Fragment 2 would result in the complete Total-KO recovery plasmid (pKP1) (**Figure 11e, S4**). This final plasmid would contain the 5'F sequence, the reconstituted mKate2 protein at the *Stu*I site, the SEC sequence surrounded by Lox2272 sites, the reconstituted *Xma*I site, and the 3'F sequence. Given the discussion of neurexin function and importance, (reviewed in Ullrich, Ushkaryov, & Sudhof, 1995; reviewed in Dean & Dresbach, 2006; reviewed in Sudhof, 2008; reviewed in Calahorro, 2014; reviewed in Sudhof, 2017; reviewed in Sudhof, 2018) *C. elegans* lacking the entire nrx1 gene may be expected to show substantial deficits in sensory behaviors.

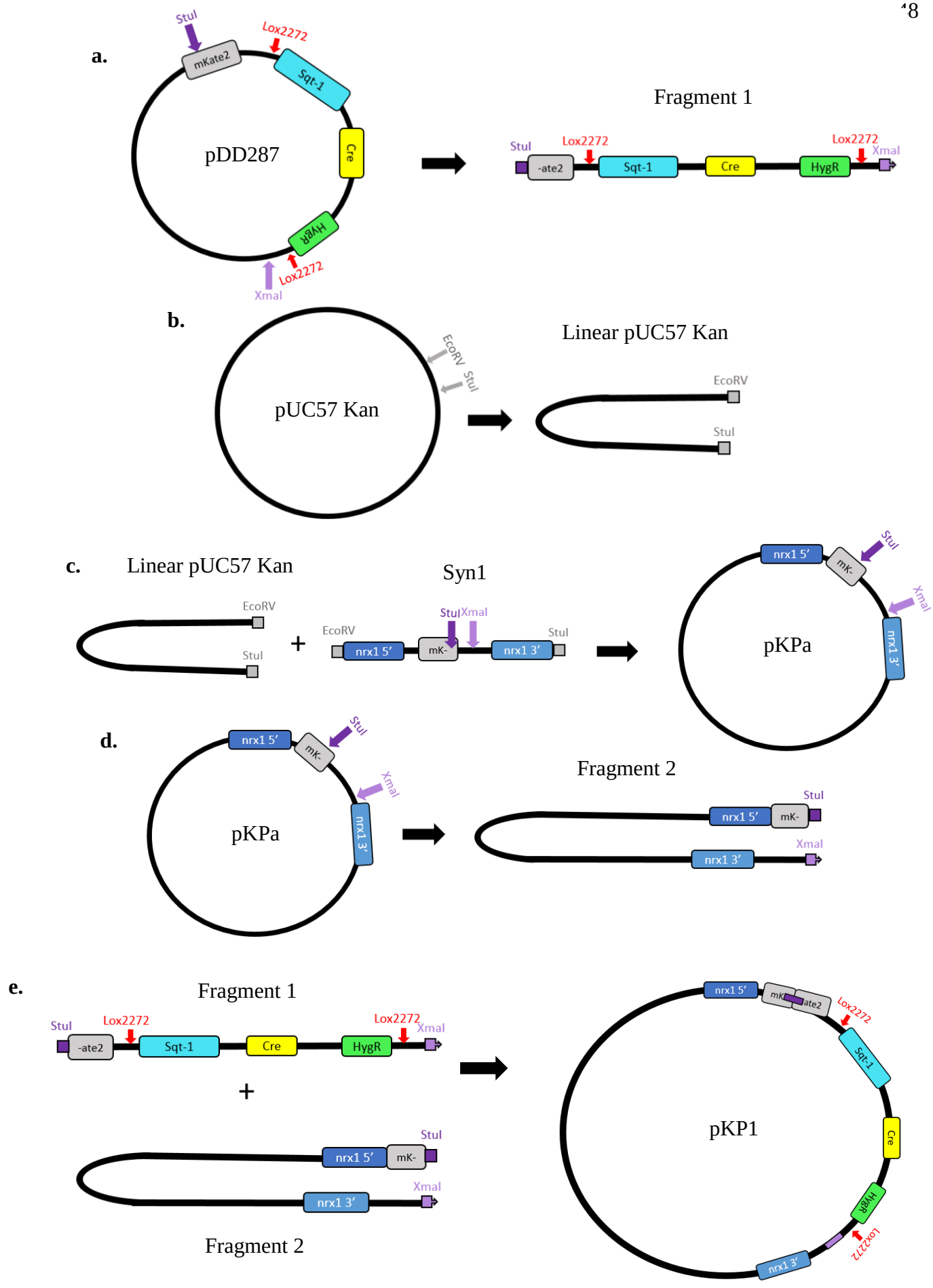

a.
StuI
Lox2272
mKate2
Sqt-1
Cre
HygR
Lox2272
XmaI
pDD287
Fragment 1
StuI
Lox2272
-ate2
Sqt-1
Cre
Lox2272
HygR
XmaI
b.
EcoRV
StuI
pUC57 Kan
Linear pUC57 Kan
EcoRV
StuI
c.
Linear pUC57 Kan
EcoRV
StuI
+
EcoRV
StuI
XmaI
nrx1 5’
mK-
nrx1 3’
StuI
Syn1
nrx1 5’
mK-
StuI
XmaI
nrx1 3’
pKPa
d.
nrx1 5’
mK-
StuI
XmaI
nrx1 3’
pKPa
Fragment 2
StuI
nrx1 5’
mK-
XmaI
nrx1 3’
e.
Fragment 1
StuI
Lox2272
-ate2
Sqt-1
Cre
Lox2272
HygR
XmaI
+
StuI
nrx1 5’
mK-
XmaI
nrx1 3’
Fragment 2
nrx1 5’
mK-
ate2
Lox2272
Sqt-1
Cre
HygR
Lox2272
nrx1 3’
pKP1

Figure 11. Total-KO Recovery Plasmid Synthesis Strategy

a. The pDD287 plasmid would be digested with *Stu*I and *Xma*I to create Fragment 1. During this digest, the mKate2 sequence would be split into pieces, labeled as mK- and –ate2. **b.** pUC57 Kan would be opened by digestion with *Eco*RV and *StuI*. **c.** Linearized pUC57 and Syn1 would be ligated to create pKPa. **d.** pKPa would be opened by digestion with *Stu*I and *Xma*I to create Fragment 2. **e.** Fragments 1 and 2 would be ligated to create pKP1, the total-KO recovery plasmid. Here the mK- and –ate2 sequence fragments would combine to reconstitute the mKate2 sequence in pKP1. pDD287 total size: 10.5 kb. pUC57 total size: 2.6 kb. pKPa total size: 4.3 kb. pKP1 total size: 10.6 kb.

The α-KO recovery plasmid, pKP2, was designed to take the place of the first 22 exons of nrx1, without interfering with the internal start for the transcription of the γnrx1 isoform (**Figure 12**). The α-KO recovery plasmid synthesis strategy involved combining a fragment from pDD287 and a customized synthetic section of the pUC57 Kan plasmid, pKPb. These two components are referred to as Fragment 3 and Fragment 4.

Digestion of pDD287 with *Xma*I and *Sal*I would result in Fragment 3, containing the SEC sequence surrounded by Lox2272 sites along with *Sal*I and *Xma*I ends (**Figure 12a**). The only difference between Fragment 1 and Fragment 3 would be the lack of the mKate2 sequence in Fragment 3 to simplify this synthesis strategy.

To synthesize Fragment 4, the pUC57 Kan plasmid would be digested with *Eco*RV and *Stu*I (**Figure 12b**) to provide the backbone. A different synthetic fragment, Syn2, was designed to contain a similar series of elements. These elements included the same 500 bp 5'F sequence with the modified PAM site as the Total-KO plasmid, a *Sal*I site, a *Xma*I site, and an α-specific 500 bp 3'F sequence with a modified PAM site positioned so as to leave the internal promoter for the shorter isoform undisturbed. Syn2 would then be inserted between the *EcoRV* and *StuI* sites to create the intermediate, pKPb (**Figure 12c, S2**). Digestion of pKPb with *SalI* and *Xma*I would result in Fragment 4, with ends that match those on Fragment 3 (**Figure 12d**).

Ligation of Fragment 3 with Fragment 4 would result in the complete α-KO recovery plasmid (pKP2) (**Figure 12e, S5**). This final pKP2 plasmid would contain the 5'F, the reconstituted *Sal*I site, the SEC sequence surrounded by Lox2272 sites, the reconstituted *Xma*I site, and the specialized 3'F. Previous work (Hu, et al., 2012) suggests that *C. elegans* nrx1-nlg1 interactions may be mediated by the LNS6 domain, present in the αnrx1 forms, but notably absent from the γnrx1 forms. Therefore, deficits observed in these organisms may be attributed not only to a loss of αnrx1, but also to a loss of nrx1-nlg1 interactions in the synapse.

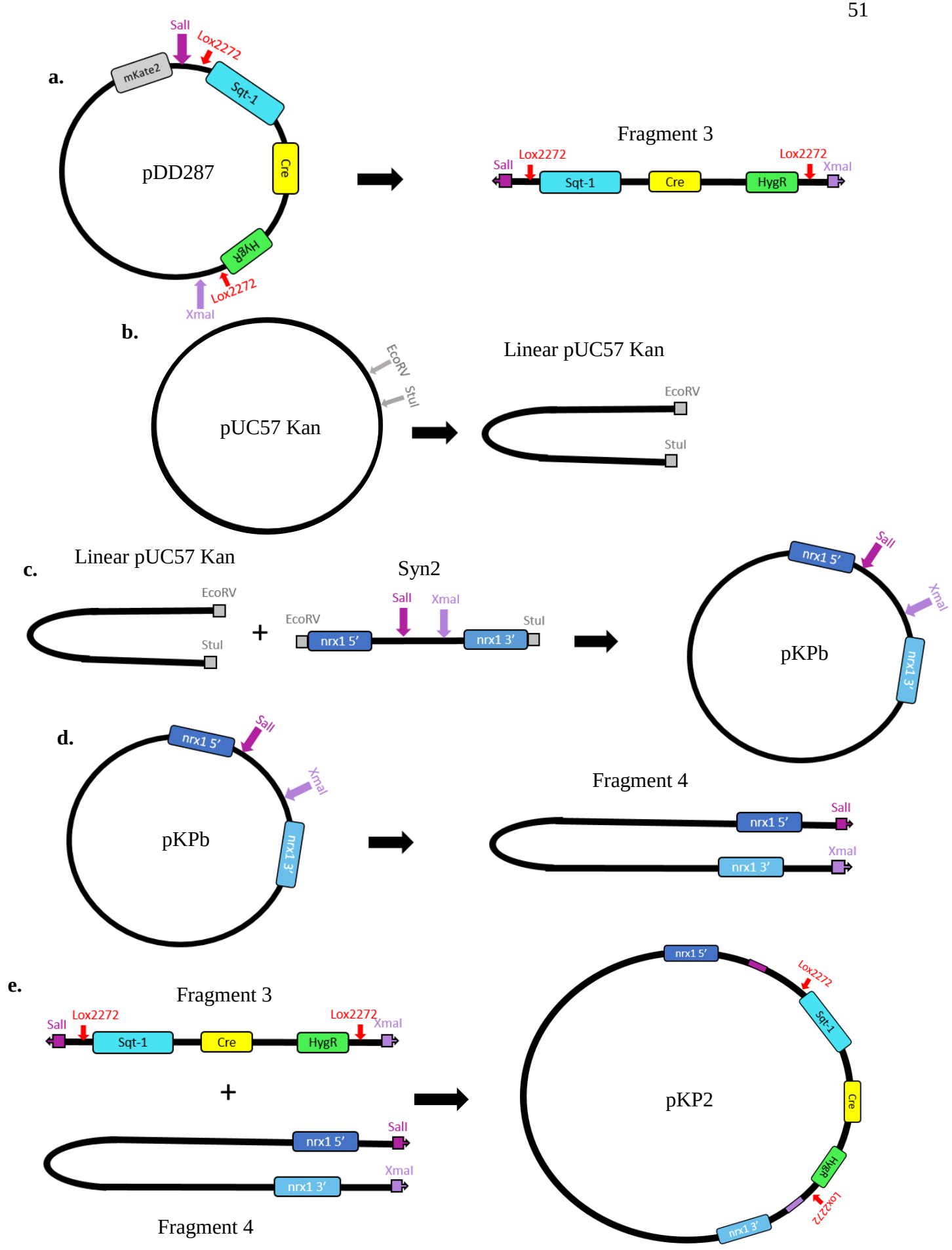

a.
SalI
Lox2272
mKate2
Sqt-1
pDD287
Cre
HygR
Lox2272
XmaI
Fragment 3
SalI
Lox2272
Sqt-1
Cre
HygR
Lox2272
XmaI
b.
EcoRV
StuI
pUC57 Kan
Linear pUC57 Kan
EcoRV
StuI
c.
Linear pUC57 Kan
EcoRV
StuI
+
Syn2
SalI
XmaI
EcoRV
nrx1 5'
nrx1 3'
StuI
pKPb
SalI
nrx1 5'
XmaI
nrx1 3'
d.
nrx1 5'
SalI
XmaI
pKPb
nrx1 3'
Fragment 4
SalI
nrx1 5'
XmaI
nrx1 3'
e.
Fragment 3
SalI
Lox2272
Sqt-1
Cre
HygR
Lox2272
XmaI
+
SalI
nrx1 5'
XmaI
nrx1 3'
Fragment 4
nrx1 5'
Lox2272
Sqt-1
pKP2
Cre
HygR
Lox2272
nrx1 3'

Figure 12. α-KO Recovery Plasmid Synthesis Strategy

a. The pDD287 plasmid would be digested with *Sal*I and *Xma*I to create Fragment 3. **b.** pUC57 Kan would be opened by digestion with *Eco*RV and *Stu*I. **c.** Linearized pUC57 and Syn2 would be ligated to create pKPb. **d.** pKPb would be opened by digestion with *Sal*I and *XmaI* to create Fragment 4. **e.** Fragments 3 and 4 would be ligated to create pKP2, the α-KO recovery plasmid. pDD287 total size: 10.5 kb. pUC57 total size: 2.6 kb. pKPb total size: 3.6 kb. pKP2 total size: 9.7 kb.

Finally, the short-isoform-KO recovery plasmid was designed to take the place of only the internal γnrx1 start site, without interfering with any of the exons expressed in αnrx1 (**Figure 13**). Synthesis of the short-isoform-KO recovery plasmid would involve combining the same fragment from pDD287 (Fragment 3) and a third customized section of the pUC57 Kan plasmid (Fragment 5).

Digestion of pDD287 with *Xma*I and *Sal*I would result in Fragment 3, as used in the α-KO recovery plasmid (**Figure 13a**). This fragment would contain the SEC sequence surrounded by Lox2272 sites along with *Sal*I and *Xma*I ends, and again, lacks the mKate2 sequence for simplicity.

To synthesize Fragment 5, the pUC57 Kan plasmid would be digested with *Eco*RV and *Stu*I (**Figure 13b**) to provide a backbone. A third synthetic fragment, Syn3, was designed to contain an internal-promoter-specific 500 bp 5'F sequence for homologous

recombiation, a *Sal*I site, a *Xma*I site, and an internal-promoter-specific 500 bp 3'F sequence for homologous recombination. Again, the first nucleotide of the PAM sequence in both the 5'F and 3'F sequences would be changed to a T to prevent subsequent CRISPR-Cas9 activity. The location of the gRNA sites excises the short-isoform-specific exon in regions that are introns in the αnrx1 sequence; this section is not built back in the recovery plasmid in order to disrupt the transcription of this short-isoform. Syn3 would then be ligated between the *Eco*RV and *Stu*I sites of pUC57 Kan to form an intermediate plasmid, pKPc (**Figure 13c, S3**. Digestion of pKPc with *Sal*I and *Xma*I would result in Fragment 5, with ends that match those on Fragment 3 (**Figure 13d**).

Ligation of Fragment 3 with Fragment 5 would result in the complete short-isoform-KO recovery plasmid (pKP3) (**Figure 13e, S6**). This final pKP3 plasmid would contain the unique 5'F sequence, the reconstituted *SalI* site, the SEC sequence surrounded by Lox2272 sites, the reconstituted *Xma*I site, and the unique 3'F sequence. γnrx1, previously discussed as βnrx1, has not been extensively characterized in the literature in *C. elegans*. However, based on the characterization of γNRXN1 (Yan, et al., 2015; Sterky, et al., 2017; Kurshan, et al., 2018), which concluded that the γ form had activity independent of other NRXN1 isoforms, moderate sensory deficits may also be expected in these organisms.

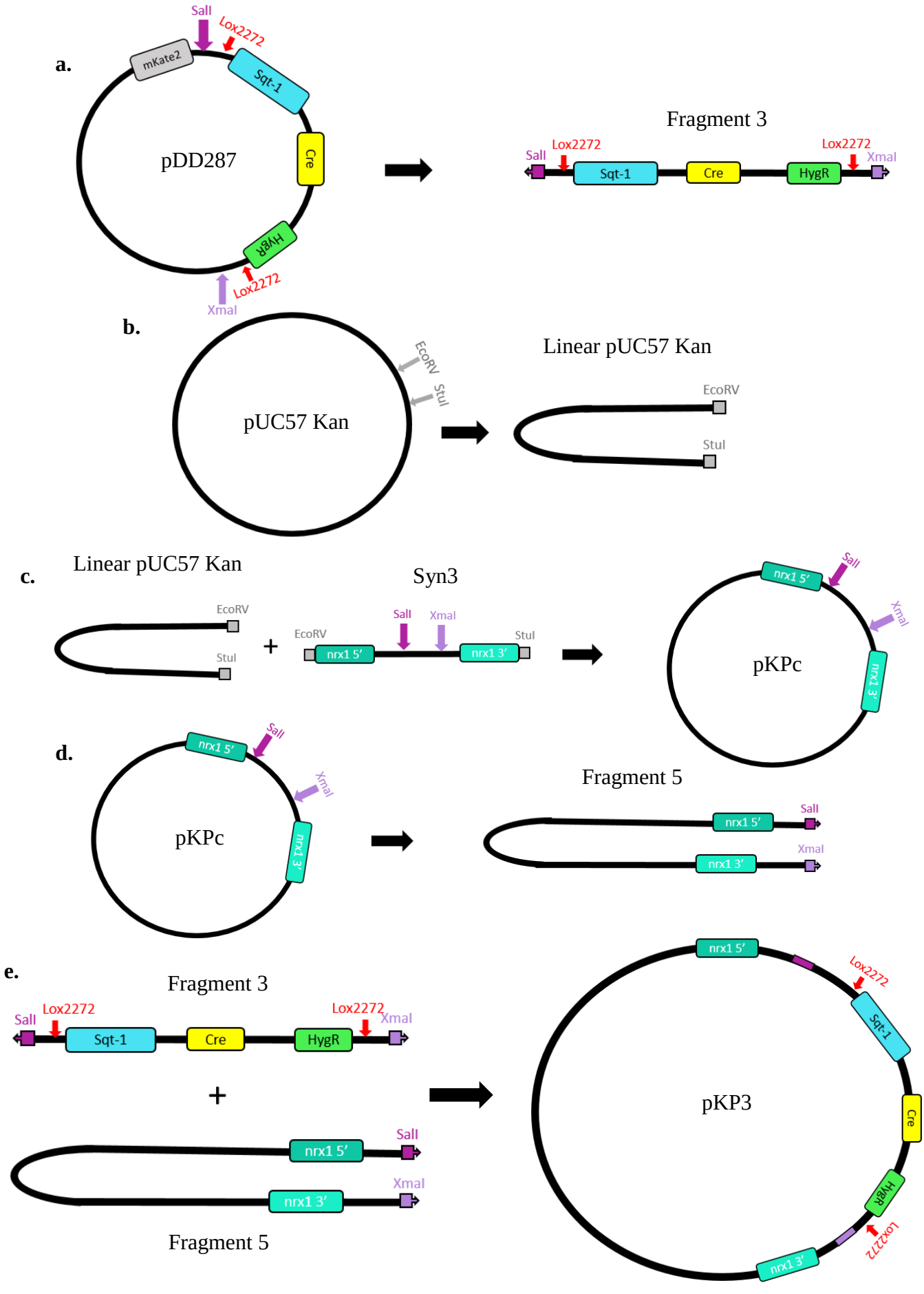

a.
SalI
Lox2272
mKate2
Sqt-1
Cre
HygR
Lox2272
XmaI
pDD287
Fragment 3
SalI
Lox2272
Sqt-1
Cre
HygR
Lox2272
XmaI
b.
EcoRV
StuI
pUC57 Kan
Linear pUC57 Kan
EcoRV
StuI
c.
Linear pUC57 Kan
Syn3
EcoRV
StuI
EcoRV
SalI
XmaI
StuI
nrx1 5'
nrx1 3'
pKPc
SalI
XmaI
nrx1 3'
d.
nrx1 5'
SalI
XmaI
nrx1 3'
pKPc
Fragment 5
SalI
nrx1 5'
XmaI
nrx1 3'
e.
Fragment 3
SalI
Lox2272
Sqt-1
Cre
HygR
Lox2272
XmaI
+
SalI
nrx1 5'
XmaI
nrx1 3'
Fragment 5
nrx1 5'
Lox2272
Sqt-1
Cre
HygR
Lox2272
nrx1 3'
pKP3

Figure 13. Short-isoform-KO Recovery Plasmid Synthesis Strategy

a. The pDD287 plasmid was digested with *Sal*I and *Xma*I to create Fragment 3. **b.** pUC57 Kan was opened by digestion with *Eco*RV and *Stu*I. **c.** Linearized pUC57 and Syn3 were ligated to create pKPc. **d.** pKPc was opened by digestion with *Sal*I and *Xma*I to create Fragment 5. **e.** Fragments 3 and 5 were ligated to create pKP3, the short-isoform-KO recovery plasmid. pDD287 total size: 10.5 kb. pUC57 total size: 2.6 kb. pKPc total size: 3.6 kb. pKP1 total size: 9.7 kb.

6.2. Microinjection and Confirmation Strategies

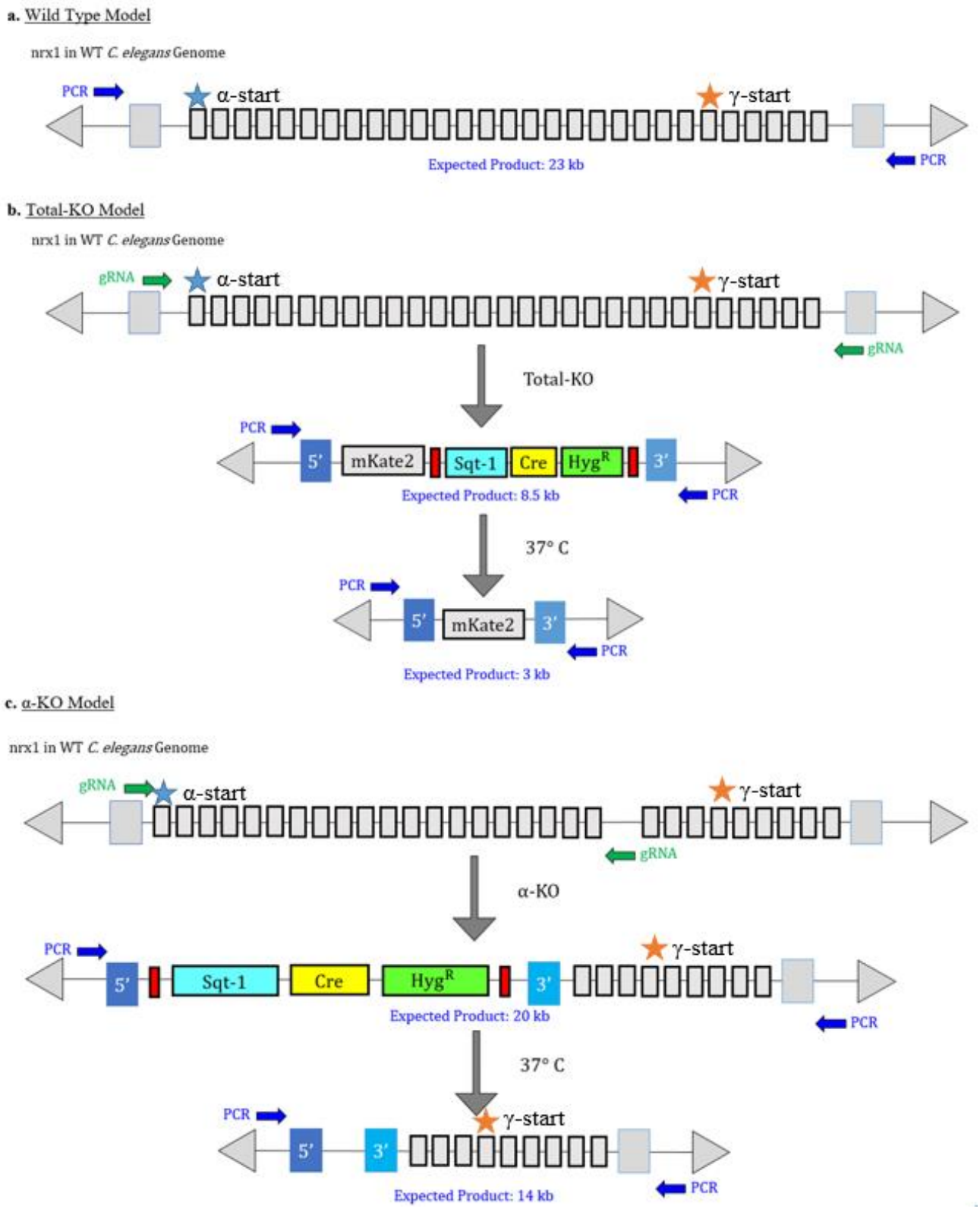
Once the CRISPR-Cas9 gRNAs and three recovery plasmids are synthesized, the next step would be to create the three nrx1-KO strains of *C. elegans*. To accomplish this, the following plasmids would be delivered: the two pRB1017 plasmids containing the relevant 5' and 3' gRNAs, the pDD121 plasmid with the Cas9 protein expression cassette, the appropriate recovery plasmid (pKP1, pKP2, or pKP3), and an additional GFP plasmid, pDD04neo pymo2::gfp, useful for subsequently proposed selection strategies.

These plasmids would be introduced into the *C. elegans* organism via microinjection (Berkowitz, Knight, Caldwell, & Caldwell, 2008). In this strategy, hermaphrodite organisms in the L3 developmental stage would be picked, immobilized onto a 3% agarose pad and paralyzed with hydrocarbon oil. A specialized needle would be used to inject the plasmids into the ovaries. The injected worms would be returned to a

fresh plate, and freed from paralysis with M9 buffer. Organisms that survive the microinjection would be expected to resume normal movement after the M9 wash and lay eggs, which will develop into *C. elegans* carrying the delivered plasmids.

A selection strategy was then developed to distinguish *C. elegans* that were successfully transformed following the microinjection. Firstly, transformed organisms would acquire the Hyg^R gene from the recovery plasmid. This selectable marker would allow them to grow on Hyg^+ plates, while organisms lacking the Hyg^R gene are expected to die on these plates. Successful transformants would also acquire and express the Sqt-1 reporter gene, which results in *C. elegans* that display a roller phenotype that is visually distinct from the WT conformation due to a mutated form of collagen (Kramer et al., 1988). Finally, organisms that express green fluorescence from the pDD04neo pymo2::gfp plasmid would be excluded. This plasmid, which would lack the homology flanks at the 5' and 3' ends, should have no way to enter the chromosome. As such, expression of GFP may indicate expression of extrachromosomal arrays, in place of recovery plasmids correctly integrated into the genome. Therefore, worms that express both Hyg^R and the roller phenotype, and do not express GFP, would be selected for subsequent experiments. To minimize changes to the genome with recovery plasmid integration, the Hyg^R and Sqt-1 sequences may be excised from the genome once this stage of selection is complete (**Figure 14**). This excision will remove interference from the Sqt-1 phenotype in subsequent behavioral analysis. These elements, along with the Cre recombinase sequence, were bookended by two Lox2272 sites. Organisms would be grown at 37° C to induce heat shock and expression of Cre, resulting in a protein product that cleaves the sequence at

both LoxP sites. During double stranded break repair by non-homologous end joining, the final genomic sequence would not include the SEC sequence (Sqt-1- Cre-Hyg^{R}).



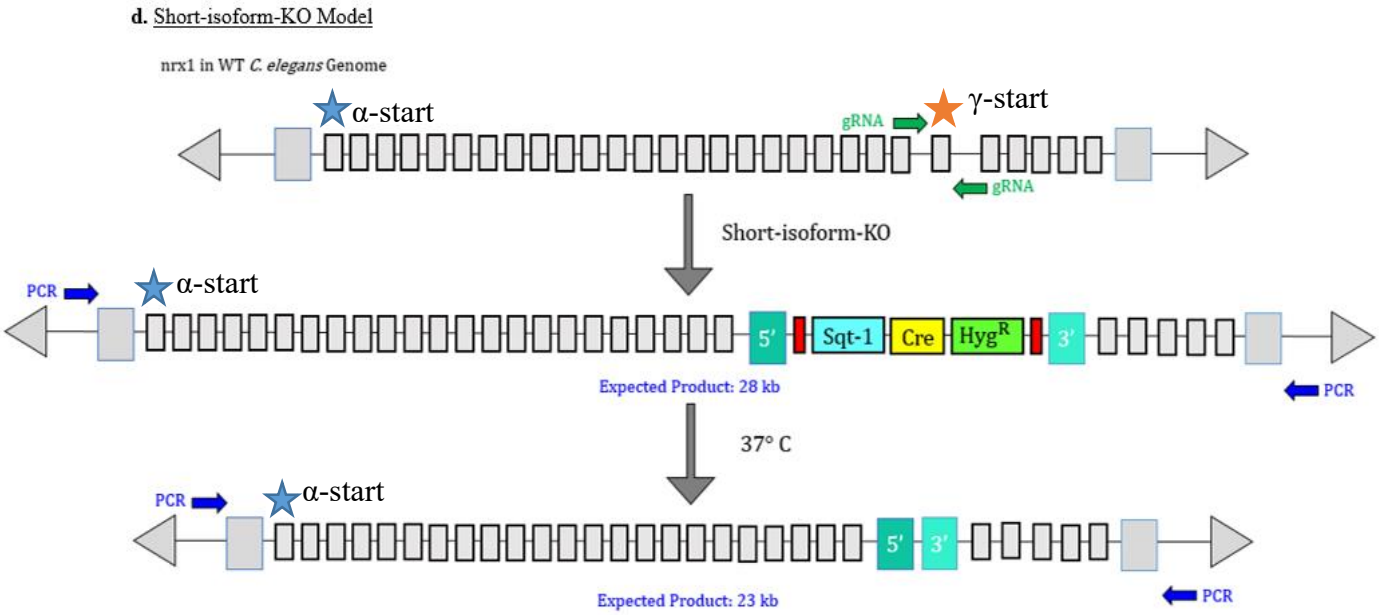


Figure 14. Recovery Plasmid Insertion and SEC Excision

a. The WT nrx1 gene in *C. elegans* would be amplified by PCR and considered for size comparison with products from mutant models. **b.** The Total-KO Model would be expected to replace all 29 *C. elegans* nrx1 exons (gray boxes) with the sequences contained in pKP1. Following growth in 37° C, the SEC sequence will be excised. **c.** The α-KO Model would be expected to replace the vast majority of exons in the *αnrx1* gene with the sequences contained in pKP2, leaving the γ-start site undisturbed. The SEC sequence would be lost after growth in 37° C. **d.** The Short-isoform-KO Model would be expected to only delete the internal γ start site, leaving the α start site and full exon sequence undisturbed. The product would contain sequences from pKP3, and would be longer than the WT *nrx1* gene. Following growth at 37° C, the SEC sequence would be lost. These diagrams only represent the exons and general strategy, and are not to scale.

To confirm the sequence at the junctions of recovery plasmid insertion, as well as successful excision of the SEC sequence, forward and reverse PCR primers were designed to amplify these segments for further analysis (**Figure 14**). The forward primer was located upstream of the pKP1 5'F sequence, and reverse primer was located downstream of the pKP1 3'F sequence to capture the entire insertion sequence regardless of which KO model was tested. Both the forward primer (5' – CCACTACCGATCGGTAGTGC – 3' on the plus strand) and the reverse primer (5' – CTTTACCTCGGCAATTTCG – 3' on the minue strand) were checked for off-target binding, self-complimentarity, and secondary-structure formation using the NCBI PrimerBLAST program (Ye, et al., 2012).

Using these primers to perform PCR, successful insertion of each recovery plasmid and successful SEC excision would produce an amplicon of easily distinguishable size (**Figure 14** and **Table 1**), with the exception of the WT genomic sequence compared to the Short-isoform-KO following SEC excision, which would both result in a 23 kb amplicon. To distinguish between these possibilities, these fragments may be digested with NaeI, which would result in 18 kb and 5 kb fragments in the short-isoform-KO-SEC-excised sequence, but would not make any cuts in the wild type sequence, resulting in one 23 kb fragment. As a final confirmation of the insertion and excision junctions, fragments at these boundaries may be sequenced.

Table 1. Experimental Condition and Expected Amplicon Size following PCR

Condition	Expected Amplicon Size (kb)
Wild Type	23
pKP1 Insertion	8.5
pKP1 SEC Excision	3
pKP2 Insertion	20
pKP2 SEC Excision	14
pKP3 Insertion	28.5
pKP3 SEC Excision	23

Amplicons shown in red cannot be distinguished, and must therefore by digested by NaeI

An overview of the experimental design proposed is provided in **Figure 15**. This proposal includes design and inclusion of the appropriate plasmids (**Figure 15a**), microinjection (**Figure 15b**), and selection strategies (**Figure 15c** and **Figure 15d**) to confirm successfully edited organisms. Finally, following the implementation of the strategies designed and outlined here, next steps may include behavioral characterization of the three model systems generated (**Figure 15e**).

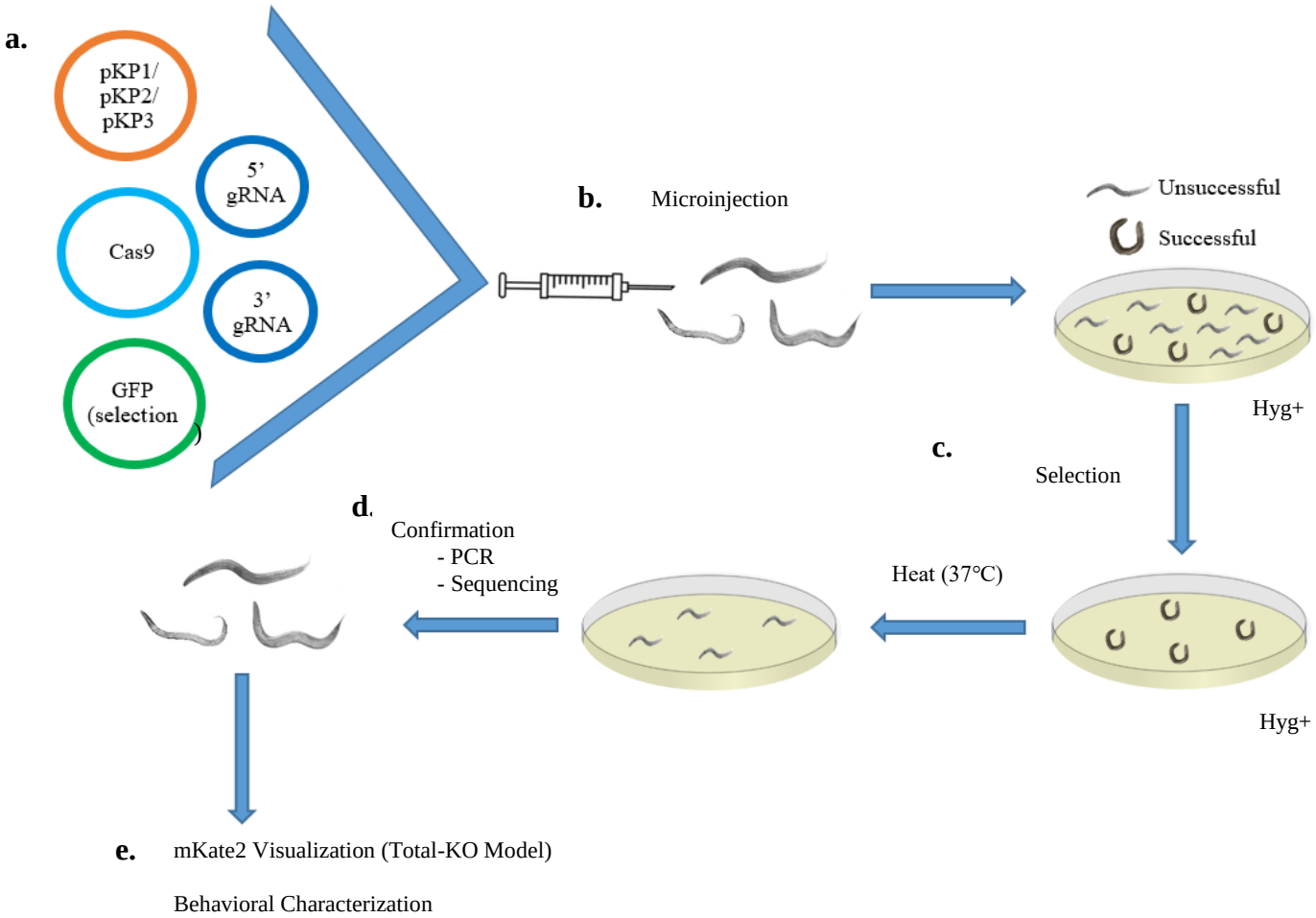


Figure 15. Summary of Experimental Design to Generate Three KO Models

The following strategy was designed to generate Total-KO, α-KO, and γ-KO *C. elegans* models. **a.** One of three recovery plasmids (pKP1, pKP2, or pKP3), two guide RNA (gRNA) plasmids, the Cas9 expression cassette, and a GFP plasmid would be introduced to *C. elegans*. **b.** Microinjection would deliver the plasmids to the organisms. **c.** Selection strategies, including Hyg^R and Sqt-1 phenotypes, would be used to identify successful transformants before the SEC sequence is excised. **d.** Correct insertions would be confirmed with PCR and sequencing techniques. **e.** KO-Models may subsequently be characterized through behavioral testing, and expression at the nrx1 promoter may be visualized using mKate2 fluorescence.

Behavioral studies have been previously used to test the effects of genetic disruption of both neurexins and neuroligins in *C. elegans* (Calahorro, Alejandre, & Ruiz-Rubio, 2009; Calahorro & Ruiz-Rubio, 2011; Haklai-Topper, et al., 2011; Calahorro & Ruiz-Rubio, 2012; Calahorro & Ruiz-Rubio, 2013; DeFronzo, 2020). Characterizations of sensory behavior such as thermotaxis or chemotaxis may be pursued, alongside examinations of basic motor behavior. To further specify which tests may be most relevant, pKP1 would include the mKate2 red fluorescent protein. When this recovery plasmid is delivered to total-nrx1 KO organisms, expression of mKate2 would allow for the visualization of neurons where the nrx1 gene is typically expressed. Referencing the recently available complete *C. elegans* connectome (Cook, et al., 2019) would allow for the identification of the specific pathways and synapses that rely on nrx1, and may suggest which behaviors are expected to be particularly disrupted in KO models.

Ultimately, the creation and characterization of these three nrx1-KO models will allow for further study of mutant *C. elegans*, and will advance current understanding of nrx1 importance and function in the synapse.

7. CONCLUSIONS

Since their initial identification, neurexins have been implicated as critical synaptic proteins in a variety of organisms. By studying the nucleotide sequences, the domain organization, and the 3D protein structure of NRXN and nrx1, *C. elegans* nrx1 isoform k was confirmed to be similar to the human αNRXNs and *C. elegans* nrx1 isoforms f and m

were confirmed to be similar to human γNRXN1. Furthermore, based on these 3D structural studies, we proposed explanations for the observed lack of behavioral deficits in VC1416 and SG1 *C. elegans*. To further study the effects of losing nrx1 proteins, we proposed and designed three separate CRISPR-Cas9 strategies to KO expression of the total nrx1 protein, the α-nrx1 protein, and the γnrx1 protein. Finally, we proposed microinjection, selection, and behavioral testing strategies to fully study these three models. Ultimately, an improved understanding of nrx1 function in *C. elegans* may help uncover a key to decoding the mechanisms behind our own synapses, both in health and disease.

SUPPLEMENTAL MATERIALS

S1. Syn1 and pKPa Sequence: https://docs.google.com/document/d/1Dl8d-8WrtVjDDVvowjCb-f20if5y3Mp3R3qaksctsP8/edit?usp=sharing

S2. Syn2 and pKPb Sequence: https://docs.google.com/document/d/1oC7KGSUcMfv_szVYAvkFyE4IVYqgiA9HNwC_JbO1O_Y/edit?usp=sharing

S3. Syn3 and pKPc Sequence: https://docs.google.com/document/d/1q7DLLhGxnHx0k3Qk4uNo7ddpevILCBy9z2l7k1zSw8Q/edit?usp=sharing

S4. pKP1 Sequence:

https://docs.google.com/document/d/1uU9okkhgSVpKNzSXR4sc1dabcRj-J0A8Wq9xplUR44s/edit?usp=sharing

S5. pKP2 Sequence:

https://docs.google.com/document/d/1sF1DGGIwU5KdT7oED4WnAGQkR6eM7V49KsZ0xWtrJbM/edit?usp=sharing

S6. pKP3 Sequence:

https://docs.google.com/document/d/1rQSn6cd8BIHlZx_T0UivVgKhpKdHvbvjcC4PTy-Gq0I/edit?usp=sharing

ACKNOWLEDGEMENTS

We would like to extend our sincere thanks to the following people and organizations for supporting this research: The University of Minnesota's Caenorhabditis Genomics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440), for providing the VC1416 and SG1 strains of *C. elegans*; Andrew Fire, for gifting Addgene plasmid #59936; Bob Goldstein, for gifting Addgene plasmid #91833 and Addgene plasmid #70685; and Denis Dupuy, for gifting Addgene plasmid #26393.

REFERENCES

Altschul, S., Gish, W., Miller, W., Myers, E., & Lipman, D. (1990). Basic local alignment search tool (BLAST). *Journal of Molecular Biology, 215*(3), 403-410.

Arac, D., Boucard, A., Ozkan, E., Strop, P., Newell, E., & Sudhof, T. B. (2007). Structures of neuroligin-1 and teh neuroligin-1/neurexin1β complex reveal specific protein-protein and protein-Ca2+ interactions. *Neuron, 56*, 992-1003.

Arribere, J., Bell, R., Fu, B., Artiles, K., Hartman, P., & Fire, A. (2014). Efficient marker-free recovery of custom genetic modiciations with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics, 198*(3), 837-846.

Berkowitz, L., Knight, A., Caldwell, G., & Caldwell, K. (2008). Generation of stable transgenic C. elegans using microinjection. *Journal of Visualized Experiments, 18*, e833.

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

Boucard, A., Chubykin, A., Comoletti, D., Taylor, P., & Sudhof, T. (2005). A splie code for trans-synaptic cell adhesion mediated by binding of neuroligin1 to α- and β-neurexins. *Neuron, 48*, 229-236.

Brenner, S. (1973). The genetics of Caenorhabditis elegans. *Genetics, 77*, 71-94.

Brunak, S., Engelbrecht, J., Knudsen, S. (1991). Prediction of human mRNA donor and acceptor sites from the DNA sequence. *Journal of Molecular Biology, 220*, 49-65.

Budreck, E., Kwon, O., Jung, J., Baudouin, S., Thommen, A., Kim, H., . . . Kim, J. (2013). Neuroligin-1 controls synaptic abundance of NMDA-type glutamate receptors through. *PNAS, 110*(2), 725-730.

Calahorro, F. (2014). Conserved and divergent processing of neuroligin and neurexin genes: from the nematode C. elegans to human. *Invertebrate Neuroscience, 14*, 79-90.

Calahorro, F., & Ruiz-Rubio, M. (2011). Caenorhabditis elegans as an experimental tool for the study of complex neurological diseases: Parkinson's disease, Alzheimer's disease, and autism spectrum disorder. *Invertebrate Neuroscience, 11*, 73-83.

Calahorro, F., & Ruiz-Rubio, M. (2012). Functional phenotypic rescue of Caenorhabditis elegans neuroligin-deficient mutants by the human and rat NLGN1 genes. *PLoS ONE, 7*(6), e39277.

Calahorro, F., & Ruiz-Rubio, M. (2013). Human alpha- and beta- NRXN1 isoforms rescue behavioral impairments of Caenorhabditis elegans neurexin-deficient mutants. *Genes Brain and Behavior, 12*, 453-464.

Calahorro, F., Alejandre, E., & Ruiz-Rubio, M. (2009). Osmotic avoidance in Caenorhabditis elegans: synaptic function of two genes, orthologues of human NRXN1 and NLGN1, as candidates for autism. *Journal of Visualized Experiments, 34*, e1616.

Calahorro, F., Holden-Dye, L., & O'Conner, V. (2015). Analysis of splice variants for the C. elegans orthologue of human neuroligin reveals a developmentally regulated transcript. *Gene Expression Patterns, 17*, 69-78.

Chanda, S., Hale, W., Zhang, B., Wernig, M., & Sudhof, T. (2017). Unique versus redundant functions of neuroligin genes in shaping excitatory and inhibitory synapse properties. *Journal of Neuroscience, 37*(29), 6816-6836.

Chen, L., Jiang, M., Zhang, B., Gokce, O., & Sudhof, T. (2017). Conditional deletion of all neurexins defines diversity of essential synaptic organizer functions for neurexins. *Neuron, 94*, 611-625.

Christlet, T., & K, V. (2001). Database analysis of O-glycosylation sites in proteins. *Biophysical Journal, 80*(2), 952-960.

Comoletti, D., Flynn, R., Boucard, A., Demeler, B., Schirf, V., Shi, J., . . . Taylor, P. (2006). Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for b-neurexins. *Biochemistry, 45*(42), 12816-12827.

Comoletti, D., Grishaev, A., Whitten, A., Tsigelny, I., Taylor, P., & Trewhella, J. (2007). Synaptic arrangement of the neuroligin/Bneuresin complex revealed by x-ray and neutron scattering. *Structure, 15*(6), 693-705.

Cook, S., Jarrell, T., Brittin, C., Wang, Y., Bloniarz, A., Yakovlev, M., . . . Emmons, S. (2019). Whole-animal connectomes of both Caenorhabditis elegans sexes. *Nature, 571*, 63-71.

Dai, J., Aoto, J., & Sudhof, T. (2019). Alternative splicing of presynaptic neurexins differentially controls postsynaptic NMDA and AMPA receptor responses. *Neuron, 102*, 993-1008.

Dean, C., & Dresbach, T. (2006). Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. *Trends in Neuroscience, 29*(1), 21-29.

DeFronzo, S. (2020). Worms on the brain: the role of autism-related genes nlg1 and nrx1 in thermotactic and chemotactic behavior of C. elegans. (Undergraduate Honors Thesis) Drew Unviersity, Madison, NJ.

Fairless, R., Masius, H., Rohlmann, A., Heupel, K., Ahmand, M., Reissner, C., . . . Missler, M. (2008). Polarized Targeting of neurexins to synapses is regulated by their C-terminal sequences. *Journal of Neuroscience, 28*(48), 12969-12981.

Felix, M., & Braendle, C. (2010). The natural history of Caenorhabditis elegans. *Current Biology, 20*(22), R956-R969.

Gauthier, J., Siddiqui, T., Huashan, P., Yokomaku, D., Hamdan, F., Campagne, N., . . . Rouleau, G. (2011). Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Human Genetics, 130*(4), 563-573.

Gibson, J., Huber, K., & Sudhof, T. (2009). Neuroligin-2 deletion selectively decreases inhibitory synaptic transmission originating from fast-spiking but not from somatostatin-positive interneurons. *Journal of Neuroscience, 29*(44), 13883-13897.

Giordano-Santini, R., Milstein, S., Svrzikapa, N., Tu, D., Johnsen, R. B., Vidal, M., & Dupuy, D. (2010). An antibiotic selection marker for nematode transgenesis. *Nature Methods, 9*(7).

Goldstein, B. (n.d.). FP-SEC vectors for homologous recombination in C. elegans. *Unpublished*.

Goldstein, B. (n.d.). SapTrap/SEC plasmids. *Unpublished*.

Gratz, S., Ukken, F., Rubinstein, C. T., Donohue, L., Cummings, A., & O'Connor-Giles, K. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics, 196*, 961-971.

Haklai-Topper, L., Soutschek, J., Sabanay, H., Scheel, J., Hobert, O., & Peles, E. (2011). The neurexin superfamily of Caenorhabditis elegans. *Gene Expression Patterns, 11*, 144-150.

Harkin, L., Lindsay, S., Xu, Y., Alzu'bi, A., Ferrara, A., Gullon, E., . . . Clowry, G. (2017). Neurexins 1-3 each have a distinct pattern of expression in the early developing human cerebral cortex. *Cerebral Cortex, 27*, 216-232.

Hebsgaard, SM., Korning, PG., Tolstrup, N., Engelbrecht, J., Rouze, P., Brunak, S. (1996). Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. *Nucleic Acids Research, 24*(17),3439-3452.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontires in Human Neuroscience, 3*, Article 31. doi:doi: 10.3389/neuro.09.031.2009

Hillier, L., Coulson, A., Murray, J., Bao, Z., Sulston, J., & Waterson, R. (2005). Genomics in C. elegans: So many genes, such a little worm. *Genome Research, 15*, 1651-1660.

Hu, Z., Hom, S., Kudze, T., Tong, X., Choi, S., Aramuni, G., . . . Kaplan, J. (2012). Neurexin and neuroligin mediate retrograde synaptic inhibitoon in C. elegans. *Science, 337*(6097), 980-984.

Huang, A., Yu, D., Davis, L., Sul, J., Tsetsos, F., Ramensky, V., . . . Chen, J. e. (2017). Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. *Neuron, 94*, 1101-1111.

Hunter, J., Millen, G., McManus, J., Heatherly, M., Duke, A., & Rand, J. (2010). Neuroligin-deficient mutants of C. elegans have sensory processing deficits and are hypersensitive to oxidative stress and mercury toxicity. *Disease Models and Mechanisms, 3*, 366-376.

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

Ichtchenko, K., Nguyen, T., & Sudhof, T. (1996). Structures, alternative splicing, and neurexin binding of multiple neuroligins. *Journal of Bioligical Chemistry, 27*, 2676-2682.

Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I., . . . Bourgeron, T. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genetics, 34*, 27-29.

Jedlicka, P., Cnencak, M., Kreuger, D., Jungenitz, T., Brose, N., & Schwarzacher, S. (2015). Neuroligin-1 regulates excitatory synaptic transmission, LTP and EPSP-spike coupling in the dentate gyrus in vivo. *Brain Structure and Function, 220*(1), 47-58.

Jiang, M., Polepalli, J., Chen, L., Zhang, B., & Sudhof, T. (2017). Conditional ablation of neuroligin-1 in CA1 pyramidal neurons blocks LTP by a cell-autonomous NMDA receptor-independent mechanism. *Molecular Psychiatry, 22*, 375-383.

Kathuria, A., Lopez-Lengowski, K., Warmuff, B., McPhie, D., Cohen, B., & Karmacharya, R. (2019). Synaptic deficits in iPSC-derived cortical interneurons in schizophrenia are mediated by NLGN2 and rescued by N-acetylcysteine. *Trasnlational Psychiatry, 9*, 321.

Koehnke, J., Jin, X., Burdreck, E., Rosy, S., Scheiffele, P., Honig, B., & Shapiro, L. (2008). Crystal structure of the extracellular cholinesterase-like domain from neuroligin-2. *Proceedings of the National Academy of Sciences, 105*(6), 1873-1878.

Kramer, JM., Johnson, JJ., Edgar, RS., Basch, C., Roberts, S. (1988). The sqt-1 gene of *C. elegnas* encodes a collagen critical for organismal morphogenesis. *Cell, 55*(4), 555-565.

Kurshan, P., Merrill, S., Dong, Y., Ding, C., Hammarlund, M., Bai, J., . . . Shen, K. (2018). γ-neurexin and frizzled mediate parallel synapse assembly pathways antagonized by receptor endocytosis. *Neuron, 100*, 150-166.

Laumonnier, F., Bonnet-Brilhault, F., Gomot, M., Blanc, R., David, A., Moisard, M., . . . Briault, S. (2004). X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *American Journal of Human Genetics, 74*(3), 552-557.

Liang, J., Xu, W., Hsu, Y., Yee, A., Chen, L., & Sudhof, T. (2015). Conditional neuroligin-2 knockout in adult medial prefrontal cortex links chronic changes in synaptic inhibition to cognitive. *Molecular Psychiatry* , 850-859.

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M., Geer, R., NR, G., . . . Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research, 48*(D1), D256-D268.

Marshall, C., Howrigan, D., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D., . . . al., e. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics, 49*(1), 27-35.

Missler, M., Fernandez-Chacon, R., & Sudhof, T. (1998). The making of neurexins. *Journal of Neurochemistry, 71*(4), 1339-1347.

Missler, M., Zhang, W., Rohlmann, A., Kattenstroth, G., Hammer, R., Gottmann, K., & Sudhof, T. (2003). a-Neurexin couple Ca2+ channels to synaptic vesicle exocytosis. *Nature, 423*, 939-948.

Nakanishi, M., Nomura, J., Ji, X., Tamada, K., Arai, T., Takahashi, E., . . . Takumi, T. (2017). Functional significance of rare neuroligin 1 variants found in autism. *PLoS Genetics, 13*(8), e1006940.

Nguyen, T., Wu, K., Pandey, S., Lehr, A., Li, Y., Bemben, M., . . . Roche, K. (2020). A cluster of autism-associated variants on X-linked NLGN4X functionally resemble NLGN4Y. *Neuron, 106*, 1-10.

Nicholls, D., M, R., Scott, I., & Meldolesi, J. (1982). α -latrotoxin of black widow spider venom depolarizes the plasma membrane, induces massive calcium influx, and stimulates transmitter release in guinea pig brain synaptosomes. *PNAS, 79*(24), 7924-7928.

Papadopoulous, J., & Agarwala, R. (2007). COBALT: constraint-based alignment rool for multiple protein sequences. *Bioinformatics, 23*(9), 1073-1079.

Parente, D., Garriga, C., Baskin, B., Douglas, G., Cho, M. A., & Shinawi, M. (2016). Neuroligin 2 nonsense variant associated with anxiety, autism, intellectual disability, hyperphagia, and obesity. *American Journal of Medical Genetics, Part A 9999A*, 1-4.

Poulopoulos, A., Soykan, T., Tuffy, L., Hammer, M., Varoqueaux, F., & Brose, N. (2012). Homodimerization and isoform-specific heterodimerization of neuroligins. *Biochemical Journal, 446*, 321-330.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A., Mooney, R., . . . White, L. (Eds.). (2018). *Neuroscience* (Sixth Edition ed.). New York, NY, Unites States of America: Oxford University Press.

Ruszkiewicz, J., Pinkas, A., Miah, M., Weitz, R., Lawes, M., Akinyemi, A., . . . Aschner, M. (2018). C. elegans as a model in developmental neurotoxicology. *Toxicology and Applied Pharamcology, 354*, 126-135.

Schrodinger, LLC. (n.d.). The PyMOL Molecular Graphics System. *Version 2.0*.

Shipman, S., & Nicoll, R. (2012). Dimerization of posysynaptic neuroligin drives synaptic assembly via transsynaptic clustering of neurexin. *PNAS, 109*(47), 19432-19437.

Sterky, F., Trotter, J., Lee, S., Recktenwald, C., Du, X., Zhou, P., . . . Sudhof, T. (2017). Carbonic anhydrase-related protein CA10 is an evolutionarilly conserved pan-neurexin ligand. *PNAS, 114*(14), E1253-E1262.

Sudhof, T. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. *Nature, 455*, 903-911.

Sudhof, T. (2017). Synaptic neurexin complexes: a molecular code for the logic of neural circuits. *Cell, 171*(4), 745-769.

Sudhof, T. (2018). Towards an understanding of synapse formation. *Neuron, 100*, 276-293.

The C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science, 282*, 2012-2018.

Thompson, A., Lester, H., & Lummis, S. (2010). The strucutral basis of function in Cys-Loop Receptors. *Quarterly Reviews of Biophysics, 43*(4), 449-499.

Tong, X., Hu, Z., Liu, Y., Anderson, D., & Kaplan, J. (2015). 2015. *eLife, 4*, e09648.

Trotter, J., Hao, J., Maxeiner, S., Tsetsenis, T., Liu, Z., Zhuang, X., & Sudhof, T. (2019). Synaptic neurexin-1 assembles into dynamically regulated active zone nanoclusters. *Journal of Cell Biology, 218*(8), 2677-2698.

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

University of Minnesota. (n.d.). Carnorhabditis Genetics Center.

Ushkaryov, Y., Petrenko, A., Geppert, M., & Sudhof, T. (1992). Neurexins: synaptic cell surface proteins related to the a-Latrotoxin receptor and laminin. *Science, 257*(5066), 50-56.

Valtorta, F., Madeddu, L., Meldolesi, J., & Ceccarelli, B. (1984). Specific localization of the α-latrotoxin receptor in the nerve terminal plasma membrane. *Journal of Cell Biology, 99*(1), 124-132.

Varoqueaux, F., Jamain, S., & Brose, N. (2004). Neuroligin 2 is exclusively localized to inhibitory synapses. *European Journal of Cell Biology, 83*, 449-456.

White, J., Southgate, E., Thomson, N., & Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philosophical Transactuons of the Royal Society of London, Biological Sciences, 314*(1165), 1-340.

Wu, X., Morishita, W., Riley, A., Hale, W., Sudhof, T., & Malenka, R. (2019). Neuroligin-1 signaling controls LTP and NMDA receptors by distinct molecular pathways. *Neuron, 102*, 621-635.

Yan, Q., Weyn-Vanhentenryck, S., Wu, J., Sloan, S., Zhang, Y., Chen, K., . . . Zhang, C. (2015). Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proceedings of the National Academy of Sciences, 112*, 3445-3450.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics, 13*(134).

Zhang Lab. (2017). Optimized CRISPR Design. Retrieved February 22, 2018, from www.crispr.mit.edu

Zhang, C., Atasoy, D., Arac, D., Yang, X., Fucillo, M., Robison, A., . . . Sudhof, T. (2010). Neurexins physically and functionally react with GABAa receptors. *Neuron, 66*, 403-416.

Zhang, C., Milunsky, J., Newton, S., Ko, J., Zhao, G., Maher, T., . . . Sudhof, T. (2009). A neuroligin-4 missense mutation associated with autism impairs neuroligin-4 folding and endoplasmic reticulum export. *Journal of Neuroscience, 29*, 10843-10854.

Zheng, W., Zhang, C., Li, Y., Pearce, R., Bell, E., & Zhang, Y. (2021). Folding non-homology proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. Retrieved from https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/