Drew University

College of Liberal Arts

Modeling neuroligin based autism spectrum disorders in *Caenorhabditis elegans*

A Thesis in Neuroscience

By

Stephanie H. Wang

Submitted in Partial Fulfillment

Of the Requirements for the Degree of

Bachelor of Science

With Specialized Honors in Neuroscience

May 2022

ABSTRACT:

Autism Spectrum Disorder (ASD) is disorder associated with difficulties in social skills, communication, sensory information, and repetitive behaviors. A small subset of patients have mutations in the pre and post synaptic transmembrane synaptic adhesion molecules, neurexin (NRX) and neuroligin (NLG). In humans, there are five forms of NLGN proteins and three forms of NRXN. Genetically liked autism may be associated with mutations in *NLGN*3 or *NLGN*4.

Caenorhabditis elegans (*C. elegans*) are widely used as model organisms, due to their relatively short life cycle of approximately three days. They have a total of 302 neurons and have homologues to *NLGN* and *NRXN*. The nine different isoforms of worm neuroligin do not currently have predicted crystal structures. Utilizing the online resources SWISS Model, the Zhanglab I-TASSER program and PyMol, the structures were predicted. Exonal differences were also determined; isoform E contains all sixteen exons. Seven of nine isoforms are missing exon 14. Two other isoforms begin at the second start found within exon 8. The protein structure of the nlg-1 knockout strain VC228, was also predicted. It was hypothesized that it is unable to form dimers; nor is it capable of binding to nrx-1 following computational modeling.

To better understand the role of neuroligin in *C. elegans*, a complete knockout of the gene by CRISPR/Cas9 was designed. A rescue null knockout plasmid was constructed utilizing PCR and Gibson assemblies. Sanger sequencing determined multiple mutations within coding regions of the final construct (pSW13): thus, a restriction digest method was utilized to complete the construction. Components were isolated from pSW13 and pDD282 (commercial plasmid used as part of pSW13 construction) a combinations of different restriction enzymes. Site-directed mutagenesis was used to correct one PCR mutation. Additional rounds of restriction fragment replacement were used to correct remaining mutations. Completion of the rescue

plasmid is ongoing. Once transgenic *C. elegans* are created, the knockout mutant can be used to explore functions of individual isoforms, mutations, and even human variations of neuroligin.

As such computational modeling of the C. elegans nlg-1 protein showed that the wild type isoforms should all be capable of forming dimers, however the knockout strain, VC228 cannot, thereby making its less drastic phenotype even more confusing. Thus, creating a complete knockout will allow for a greater understanding of the protein.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

aa	Amino acids
bp	base-pair
CGC	Caenorhabditis Genomics Center, University of Minnesota
Cre	Cre Recombinase, removes sequence between two lox-p Sites found in self-excising cassette
EGF	Epidermal growth factor like domain
gRNA	Guide RNA
Hyg/HygR	Hygromycin B (eukaryotic antibiotic) / Hygromycin B Resistance
Kb	Kilobase-pair
LNS domain	laminin A/Neurexin sex hormone binding domain
NLG	Non species specific neuroligin gene
NLG	Non species specific neuroligin protein
nlg-1	*C. elegans* neuroligin gene
nlg-1	*C. elegans* neuroligin protein
NLGN(1,2,3,4X,4Y)	Five human neuroligin genes
NLGN(1,2,3,4X,4Y)	Five human neuroligin proteins
NRX	Non species specific neurexin gene
NRX	Non species specific neurexin protein
nrx-1	*C. elegans* neurexin gene
nrx-1	*C. elegans* neurexin protein
NRXN	Human neurexin gene
NRXN	Human neurexin protein
N-gly	N-glycosylated protein region

O-gly	O-glycosylated protein region
PDZ	(PSD95) - (Dlg1) - (Zo-1), protein domain
SEC	Self-excising cassette
Sqt-1	Gene for the *C. elegans* roller phenotype
WT	Wild type

INTRODUCTION:

Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that in 2020 diagnosed in 1 in 54 children in the United States (CDC 2020a). Since 2000, the number of children diagnosed with ASD has continued to rise in the United States. Such an increase in prevalence in the US may be due to increased and better screening protocols, early intervention, and a larger general understanding of ASD in the public (Doherty et al. 2021). Children may be diagnosed with ASD as young as 18 months and generally before the age of two years. However, children with mild ASD may be diagnosed much later (Hyman et al. 2020). Prior to 2013, ASD was comprised of many subtypes according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-4). The fifth edition of the DSM combines the previous subtypes into a singular syndrome-based disorder.

ASD usually presents as a variety of abnormal social, communicational, and developmental behaviors (CDC 2020a, Niedźwiecka and Pisula 2022, Doherty et al, 2021). Some examples of abnormal social behaviors include the refusal to engage in eye contact, and the preference to be alone. Some will prefer to not engage in physical contact while others may not understand personal space. Examples of communicational abnormalities include delayed speech, repeating words or phrases or reversing pronouns. Some individuals may not understand jokes or sarcasm. Some others may also be completely nonverbal, or they may speak very little. Furthermore, ASD patients may exhibit temper tantrums, aggression, or strange emotional reactions to different situations. Some patients may also present with over or under stimulation in different situations, leading to difficulties or a diminished quality of life. There are many more

symptoms not discussed however, each patient may present with different combinations of these symptoms.

As stated above, diagnosing ASD changed in 2013, as prior to this the DSM4 had three different classifications (NIH). These included Autistic Disorder, Asperger's Syndrome and Pervasive developmental disorder not otherwise specified (PDD-NOS). Following the revision of the DSM these three classifications became autism spectrum disorder. Catching and diagnosis of ASD early is crucial; this allows for early intervention and access to resources at a much younger age. Diagnosing ASD can be different depending on the age of diagnosis. For young children it is a two-step process which include routine screenings during normal physicals with their primary care doctor. If the doctor notices developmental delays or if the child is at a higher risk, the doctor will move to step two where more evaluation of the child takes place. During this step doctors work with other professionals to evaluate cognition levels, language abilities, and the skill level of daily activities. Doctors may also perform hearing and blood tests to determine if there are any other disorders the child may experience compounded with ASD.

Diagnosing ASD can also occur in older children, adolescents, and adults. In school age children, it is traditionally teachers and school staff who notice the signs of ASD and perform initial screening (Duesenberg and Burns 2022). The parents may then be asked to bring their child to their primary care doctor for further evaluation and testing. In adults, procuring an ASD diagnosis can be more difficult; there currently exists no definitive diagnostic criteria for adults (Hayes et al. 2022, Huang et al. 2020). Prior to testing, adults must first identify the symptoms of ASD in themselves and come to an understanding of what a diagnosis may mean for their life, all before mentioning it a doctor. Finding professionals who would be willing to diagnose adults is another challenge many may face. When a professional is found, a developmental history of the

patients as well as interviews with family members will also be needed to make the correct diagnosis.

Currently, there are a multitude of different treatment approaches for autistic children and their families ranging from medications to different behavioral/occupational therapy. For instance, medication can be used to treat some symptoms a child may experience. If a child is co-diagnosed with ADHD, medications could be used to help the hyperactivity, aggression, and attention challenges (Carlson et al. 2006, Mayes et al. 2020). Medications may also be used to treat seizures, gastrointestinal issues, and other associated medical conditions. More non-medical approaches also exist such as teaching the child how to eventually lead independent lives (Hymen et al. 2020). Usually, these treatments are adult lead and can focus on developing the child's communication and social skills. One example is applied behavior analysis (ABA), where children are rewarded for correctly performing a task. However, this therapy's efficacy has been called into question in recent years (Sandoval-Norton et al. 2021). There are other therapies, such as floor time, which critics of ABA have been suggesting is more beneficial to autistic children; however, more research is needed to fully determine which therapy system is most beneficial. Children may also see a speech language pathologist to help teach better communication techniques through sounds, gestures, and words (Santhanam and Bellon-Harn 2022). Other more nontraditional therapies also exist, such as essential oils, crystals, and hugging. However, these treatment routes do not have any clinical evidence to support their use. New and innovative methods such as transcranial magnetic stimulation are being developed adding to the arsenal of tools available to parents and caregivers, to improve the quality of life for each individual (Santhanam and Bellon-Harn 2022). There are more treatment options for ASD patients and their families that are not mentioned, but each individual case is unique and treatment/therapy

methods that worked in one case may not work for another. Thus, it is important that implementing different treatments is done on a case-by-case basis.

Controversies in Study

In the past few years, the study of autism has become increasingly controversial among the autism community. Within the community, there are those who strongly believe that autism is simply a case of neurodiversity. According to the Merriam-Webster dictionary, neurodiversity is defined as "the concept that differences in brain functioning within the human population are normal and that brain functioning that is not neurotypical should not be stigmatized" (2021). With a rise in use of the internet and an increased abillity to connect with other autistic individuals, there has been an increase in the way information relating to ASD has been distributed, spreading the aims of the self-advocates in and out of the autistic community. They believe that autism, among other disorders, is just a version of diversity, just like hair, eye, and skin color (Leadbitter et al. 2021). They also champion equal rights for and the decrease of discrimination against those with disabilities. Of course, there are many who advocate against the concept of neurodiversity. These people are often parents or caretakers of children who require more care. They believe that their children are being left out of the conversation (Hughes 2021). These high care children may not hold the capability to advocate for themselves, if they can, it may not be to the quality or the extent that many of their low care counterparts possess.

Some in the neurodiversity movement do not view scientific autism research in a positive light (Grinker 2015). Many believe the end goal of autism related medicine and research is to find a cure. The self-advocates within the community do not believe there is anything that must be fixed; it is a part of who they are. Thus, they may view any science relating to autism or other neurodevelopmental disorders as a personal attack. However, within this neurodiversity debate is

a belief that the medical model of autism is wrong, to the point where they do not even want to understand the underlying causes.

Genetic Causes of Autism Spectrum Disorder

Currently there exists no clear, concreate cause of autism spectrum disorder, however, multiple different factors have been suggested. Epigenetics, older parental age both maternal and paternal, history of autoimmune disorders, infection during pregnancy, low birth weight, and genetics have all been correlated to ASD (Hodges et al. 2020). The credibility and actual causation of study results in these fields, have been mixed. Despite the uncertainty, the genetic aspects of ASD have been one of the largest areas of study; there have been approximately 900 different genes that have been linked to ASD as of 2021 (Huang et al 2022). With cheaper and more accessible gene testing technologies, such as Next Generation Sequencing, the number could continue to rise. Of these 900 gene, 26 have been well characterized as potential risk factors for ASD. Two of these genes, *NLGN*3 and *NLGN4*X are part of a family of genes known as neuroligins. Prior research has linked mutations in the neuroligin (*NLG,* non-species specific) genes, subsequent proteins misfolding and alterations to ASD phenotypes (Jamain et al. 2003, Ulbrich et al. 2016).

Neuroligins and Neurexins

Neuroligins (NLG) are transmembrane proteins that participate in cell adhesion of neurons when interacting with their cross-synapse binding partner, neurexins (NRX). The presynaptic NRX interacts with the postsynaptic NLG during normal function (Südhof 2008). The main role of NRX and NLG is that they are involved in synaptogenesis, maintenance, and the maturation of new synapses (Varoqueaux et al. 2006). The current body of work that characterizes neuroligins and neurexins has greatly increased since neurexins were first

discovered in 1992 by Ushkaryov and colleagues as they searched for a receptor for α-latrotoxin from the black widow spider (1992). They discovered that *NRX* had two genes (I. and II.) and each had two different sized forms, the alpha (α) and a beta (β) due to different promoters. Furthermore, they also determined that each gene and form were able to undergo different alternative splicing events resulting in at least 100 different isoforms of neurexins. It was later determined in 1994 that there were three *NRX* genes and a total of six mammalian forms, Iα-IIIβ (Ushkaryov et al. 1994).

Approximately three years following the initial discovery and characterization of NRX, Neuroligin 1 was identified in rats by Ichtchenko and colleagues as the cross-synapse binding partner for β- neurexins (1995). This paper was one of the first to suggest that the interaction between neurexin and neuroligin function as synapse adhesion molecules due to the tightness of the interaction. In 2001, Bolliger and colleagues were able to identify and characterize four human neuroligin genes using reverse transcription PCR and human brain tissue samples (2001). They utilized the three known rat neuroligins to find the three human orthologs. They are further responsible for the discovery of a fourth human neuroligin and proposed a fifth. Over 20 years later, the information regarding neuroligins and their roles in the central nervous system continues to improve and grow. It is now known that there are five neuroligin genes in humans, *NLGN*1, *NLGN*2, *NLGN*3, *NLGN*4X and *NLGN*4Y.

Neurexin and Neuroligin Domains

The structures and domains of both human neurexin and neuroligin have been characterized over the years. Both are single transmembrane proteins that include an intracellular transmembrane PDZ ((PSD95) - (Dlg1) - (Zo-1)) domain on the C terminal end of the protein (**Figure 1)** (Lisé and Husseini 2006, Dean and Dresbach 2006, Chen et al. 2008, Lee and Zheng

2010, Miller et al. 2011, Bang and Owczarek 2013). However, the PDZ domain found in neuroligin is a type 1 PDZ recognition site where the one found in neurexins is a type 2 PDZ interaction site. Both proteins also contain an O-linked glycosylation (O-Glyc) domain just next to the extracellular side of the transmembrane region. In the case of neurexin, there are two distinct forms that each of the three neurexins can take: α-neurexin and the β-neurexins. In both neurexin forms there is at least one LNS domain on the extracellular side of the N-terminus of the protein. The LNS domain contains sequences relating to laminin A, neurexin and a sex hormone binding protein. The two neurexins differ however in that α- neurexin contains 6 LNS domains that are separated by epidermal growth factor (EGF)-like sequences. Beta-neurexins only contain 1 LNS domain and have a unique sequence on the N terminal side of the LNS domain. Alpha-neurexins also contain five to six alternative splice sites; however, the beta form only contains splice sites 4 and 5.

Neuroligins in comparison, to neurexins, are much simpler in that they are a single transmembrane protein with a cholinesterase-like domain **(Figure 1)** (Lisé and Husseini 2006, Dean and Dresbach 2006, Chen et al. 2008, Miller et al. 2011, Bang and Owczarek 2013). Neuroligins are missing one residue which makes the cholinesterase-like domain enzymatically inactive when compared to a normal cholinesterase domain. They also contain an oligomerization site in addition to the other similar domains it shares with neurexins. Neuroligins also contain three N-glycosylation (N-gly) sites and two EF-hand motifs (not pictured) which are included to aid in calcium ion (Ca^{2+}) binding (Calahorro 2014, Lisé and Husseini 2006). Furthermore, there are two alternative splice sites found within the cholinesterase-like domain.

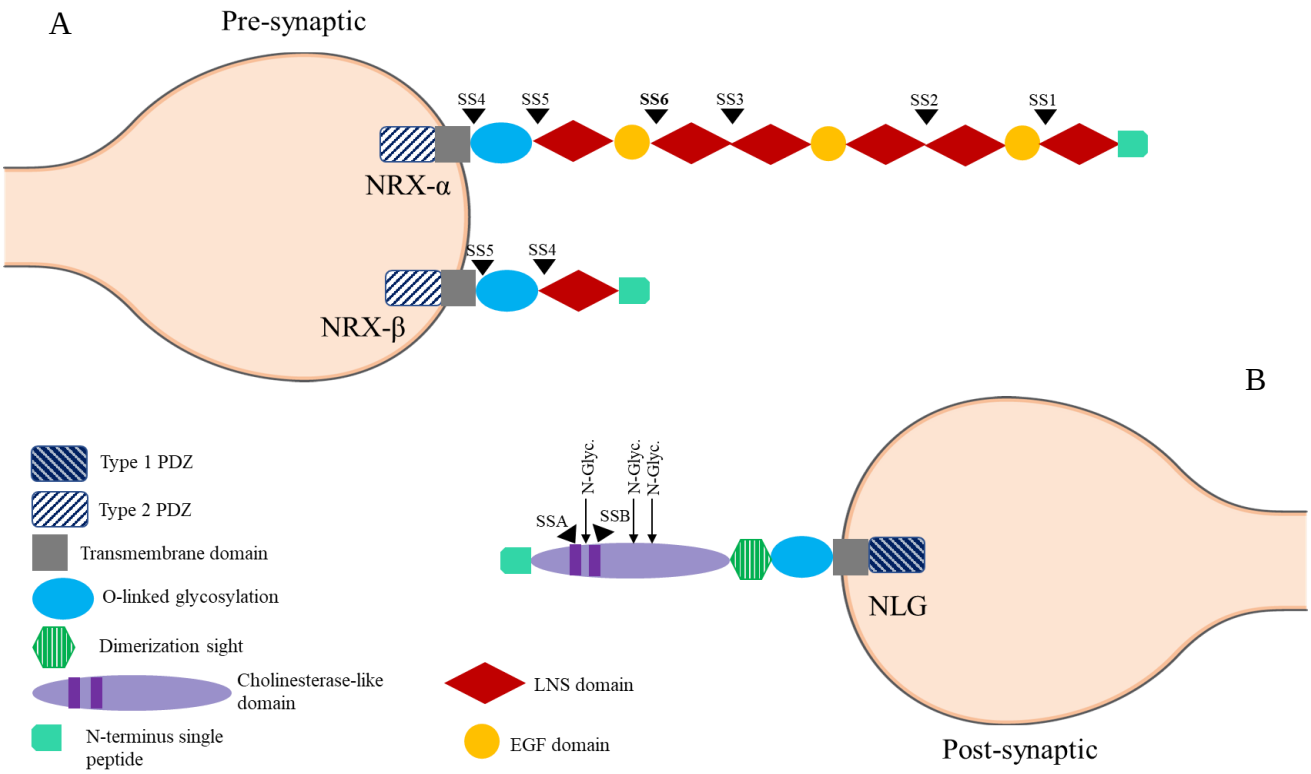


Figure 1. Visualization of Neurexin and Neuroligin domains. (Adapted from Calahorro 2014). A) Visualization of Neurexin (NRX) domain structures for both alpha and beta forms. Six alternative splice sites (SS1-6) can be found in neurexin alpha forms and only two can be found in in the beta form. B) Visual representation of Neuroligin (NLG) domains. Includes the three N glycosylation (N-Glyc.) sites and the two alternative splice sites (SSA and SSB). Dimerization site is also noted.

Human Neuroligin

In humans there are five NLGN genes; two, NLGN3 and NLGN4 are located on the X chromosome (NLGN4X) these are of particular interest because mutations in these genes have been linked previously to cases autism spectrum disorders. (Südhof 2008). NLGN4X has a chromosomal complement involving the Y chromosome, often called NLGN5 or NLGN4Y. Prior research has discovered that NLGN1, NLGN3, and NLGN4X/NLGN5 are very similar to each other, and that NLGN2 is the most different. Evolutionarily, NLGN4 (both X and Y versions) are the least evolutionarily conserved among different species (Marro et al. 2019, Nguyen et al. 2020). As such, NLGN4Y animal models are more difficult to create, leading to limited characterization and other research.

The variety of NLGNs differ depending on the type of neuron, specifically whether the neuron is inhibitory or excitatory. It has been seen that NLGN3 is present in both inhibitory and excitatory neurons (Budreck and Scheiffele 2007). NLGN1 and NLGN4 are found in excitatory (glutamatergic) synapses (Song et al. 1999, Marro et al. 2019). Whereas NLGN-2 is found in inhibitory (GABAergic) synapses (Varoqueaux et al. 2004). NLGN3 can be found in both glutaminergic (excititory) and GABAergic (inhibitory) synapses. Each synapse can also contain multiple different of NLGN proteins at a time (Budreck and Scheiffele 2007). For example, either a combination of NLGN1 and NLGN3 or NLGN2 and NLGN3 can be found in the same synapse. However, NLGN1 will not be found in the same synapse as NLGN2.

Each of the four neuroligin proteins are responsible for different roles within the growth and maturation of synapses. One example is that the NLGN1 protein was determined to play a role in the formation of dendritic spines of neurons (Nakanishi et al. 2017). In a paper by Nakanishi and colleagues they studied four separate mice NLG1 point mutations associated with

autism. Each mice point mutation had corresponding human variant. In models containing these mutations they found that in each instance the number of dendritic spines was greatly reduced from that of the wild type.

As mentioned above NLGN2 is located within the inhibitory synapses and functions to recruit other proteins such as gephyrin to the cell membrane surface (Poulopoulos et al. 2009, Choii and Ko 2015). Research utilizing HEK293 cells and mice cerebellar granule cells has indicated that NLGN2 is critical to synaptogenesis and maturation of inhibitory synapses (Fu and Vicini 2009). In the research completed by Fu and Vicini they created a culture system where synapses could form (2009). In this culture system, expression of only NLGN-2 allowed for the formation of inhibitory synapses. In addition, NGLN2 recruits and activates additional proteins to the membrane surface (Papadopoulos and Soykan 2011, Poulopoulos et al. 2009). NLGN2 activates the molecule collybistin, a GDP/GTP-exchange factor, that helps to recruit gephyrin to the synapse plasma membrane (Papadopoulos and Soykan 2011, Poulopoulos et al. 2009). Gephyrin is responsible for the clustering γ-aminobutyric acid type A ($GABA_A$) receptors as well as acting as providing posttranslational modifications and binding for other proteins such as $GABA_A$ receptor-associated protein (Choii and Ko 2015). As such NLGN2 is necessary for inhibitory synapse formation and maturation.

Recent research has identified that NLGN3 is active in processes involved with growth and complexity of the dendrites in excitatory neurons (Etherton et al. 2011a, Xu et al. 2019). The, *NLGN*3, arginine451cysteine (R451C) mutation was shown to increase the complexity of dendrites in the hippocampus (Etherton et al. 2011a). This mutation also increased the synaptic transmission seen with both the AMAP and NMDA receptors indicating that NLGN3 may be involved in regulating the signal transmission. In research utilizing another mutation, the *NLGN*3

arginine704cysteine (R704C) it was seen that there was an increase in the amount of AMPA receptors but not in the amount of NMDA receptors (Etherton et al. 2011b). Thus, further solidifying that NLGN3 may play a role in the recruitment and retention of excititory receptors.

Unlike the other NLGNs, NLGN4 is not well conserved among different species. To combat this challenge in research, researchers may use human neuronal stem cells to determine overall changes. For example, to elucidate the function of NLGN4X, Shi and colleagues identified and measured the knocked down of the gene (Shi et al. 2013). They noticed that during a six-week period, the neuron with their *NLG*4X knocked down had less cell growth. This included fewer numbers of axons, neurites and neuron differentiation. Thus, this suggests that NLGN4 is vital to growth and the differentiation of neurons. NLGN4Y is even more of an interesting case as there is only a 3% difference from the sequence of NLGN4X sequences, yet there are drastic changes in the cellular processing (Nguyen et al. 2020). The biggest change is that NLG4Y is unable to induce synapses due to the lack of cellular transport to bring the protein to the cell surface. This is believed to be due to an amino acid (aa) change from serine to proline change at position 93 in the extracellular domain.

With five different genes for neuroligins found in humans, the possibility for different mutations in genes and subsequent proteins to contribute to ASD phenotypes, is expansive. Thus, understanding the function of each protein is vitally important.

Cross-synapse binding

Binding of neurexins and neuroligins can be influenced by multiple different factors. These include alternative splicing, the presence of calcium ions (Ca^{2+}) and the form of neurexin and mutations (Chen et al. 2008, Miller et al. 2011, Liu et al. 2013, Tsetsenis et al. 2014). Specifically in species with more than one neuroligin, the combination of neurexin and

neuroligin forms may influence binding. Despite this, there are critical components that are required for the cross-synapse binding (binding of proteins that hold two synapses togeter) to occur. Neurexins and neuroligin have been shown to bind in a hydrophobic calcium dependent manner (Rissner et al. 2008, Miller et al. 2011). Specifically, the hydrophobic Ca2+ binding pocket can be found in the 2nd, 3rd, and 6th LNS domains of α-neurexins and the sixth LNS domain of β-neurexin (Miller et al. 2011). Another critical component is that neuroligins must form dimers to reach the plasma membrane found in the synapse (Poulopoulos et al. 2012). However, these dimers have been shown to exist as both homodimers and heterodimers within the nervous system. Some dimerization sites have been determined though both computer modeling and mutagenesis work. In one example, the three human NLGN3, homodimerization sites were determined (Dean et al. 2003, Ko et al. 2009, Shipman and Nicoll 2012). The first dimerization site was determined to be located at Phenylalanine (F) 437, Methionine (M) 438, and Tryptophan (W) 442, where the second dimerization site is located at Histidine (H) 606 and Leucine (L) 607 (Dean et al. 2003, Ko et al. 2009, Shipman and Nicoll 2012). The third dimerization site is located at Lysine (K) 600 and Lysine (K) 606 ((Dean et al. 2003, Ko et al. 2009, Shipman and Nicoll 2012).

Alterations to other, less critical, regions allow for variation in the binding of NRXs and NLGs. As mentioned above, both neurexins and neuroligins each have multiple alternative splice sites, resulting in potentially thousands of different combinations, each possibly affecting cross-synapse binding differently (Boucard et al. 2005). For example, Boucard and colleagues found that synapse size and density differed when alternatively spliced neuroligin was bound to the two forms of neurexin (2005). In another example, altering the specific neuroligin, also affected the binding affinity and tightness (Koehnke et al. 2010). In a paper by Koehnke and colleagues they

determined that NLG3 had a weaker interaction with β-NRX as compared to both NLG1 and NLG2. Furthermore, they found that the insertion and retention of the β-NRX alternative splice site 4 either weakened or strengthened the interaction between the NLG and the three different NRX translated proteins. Thus, indicating that the form and alternative splicing do play a critical role in NRX/NLG interactions.

Neuroligin: an Underlying Cause of Autism Spectrum Disorder

In many cases of ASD the cause is unknown, however for a small subset of individuals their ASD is caused by genetic mutations (Ulbrich et al. 2016, CDC 2020b). Mutations of neuroligin (*NLGN*) and neurexin (*NRXN*) genes have been linked to ASD. Most commonly it is mutations in *NLGN*3 and *NLNG*4 that have been linked to ASD. Two specific mutations include the *NLGN*3 mutations of glycine to arginine at position 221 (G221R) and arginine to cysteine at position 451 (R451C) (Ulbrich et al. 2016,). The R451C mutation has been characterized to a much larger extent than the G221R mutation. In both mutations, the quaternary structure of NLGN is altered, leading to difficulties in cellular trafficking. Research has shown that the R451C mutation causes an activation of the unfolded protein response due to local protein misfolding causing a total loss of approximately 90 percent of NLGN3 (Ulbrich et al. 2016, Shen et al. 2014). Previous research has shown that individuals with the R451C mutation, had some mutant protein at the cell surface; however, the rest of these proteins were degraded in the endoplasmic reticulum. The G221R mutation has been shown to increase the degree of misfolding of the α/β-hydrolase fold domain due to its centralized location in the domain (De Jaco et al. 2010).

Recent research utilizing mice has identified that mutations in *NLGN*3 (R451C mutation) cause growth and complexity in the dendrites (Xu et al. 2019). They suggest that NLGN3 relates

to the Akt/mTOR signaling pathway, a pathway related to cellular division. The mutation up-regulates the pathway and increases the dendritic growth. This could potentially be a mechanism of ASD. Furthermore, research with R451C transgenic mice indicates that there appears to be deficiencies in social behavior (Etherton et al. 2011a). Interestingly the mice seemed to show an increased ability to learn in a Morris Water Maze environment. However, they noticed when they bred a second generation of R451C mutant mice, different phenotypes presented themselves. This could further suggest that like humans, mice with ASD-like symptoms can all be different from each other and need individualized treatment approaches.

Model System

Caenorhabditis elegans (C. elegans) are a species of nematodes that since the 1960s have been widely used to model different diseases (Corsi et al. 2015). *C. elegans* have a relatively short life cycle of approximately three days and a life span of about 10 days. This allows for generational studies to occur relatively quickly. In addition, most are self-fertilizing hermaphrodites allowing much easier breeding and repopulation of a specific strain. Most importantly however is that *C. elegans* only have 302 neurons. This makes studying their nervous system and neurons much less complicated compared to, say, a mouse nervous system with millions of neurons (Corsi et al. 2015).

Neuroligins and C. elegans

C. elegans have one neuroligin gene (*nlg*-1) which is an ortholog of the five human *NLGN* genes (Calahorro 2014). This provides another strong reason to use *C. elegans* as a model system. Despite having only one neuroligin gene, there are nine separate known and characterized isoforms of the nlg-1 protein (Wormbase.org). However, it has been suggested that there may be a many as 24 different isoforms for the single neuroligin gene (Hunter et al. 2010).

Of the nine confirmed isoforms, the longest contains 847 amino acids (aa) and 16 exon regions. Each of the nine isoforms are the result of alternative splicing, isoforms A-I. It has been suggested that the different isoforms of nlg-1 may compensate for the lack of multiple neuroligins genes found in humans and other organisms (Calahorro et al. 2015). Thus, introducing different interactions and functions that otherwise may be unseen.

Currently there are neuroligin knockout strains of *C. elegans* available from the Caenorhabditis Genomics Center (CGC) at the University of Minnesota. One example is the strain VC228, genotype nlg-1(ok259), which is a model system designed to mirror the ASD like symptom and genetics (*C. elegans* Deletion Mutant Consortium 2012). However, prior work demonstrated that these worms phenotypically appear to resemble the wild type (WT) worms even with a 2341 base pair (bp) deletion and a 334 base pair insertion in their *nlg*-1 gene (Hunter et al. 2010, DeFronzo 2020). This deletion led to the loss of exons eight through thirteen. Thus, despite containing such a large mutation, it was surprising that the worms do not show a greater behavioral deficit. Perhaps the knockout mutation of VC228 was not drastic enough to display the desired phenotype.

In the current body of literature, it is still unclear how the different isoforms of nlg-1 affect the functionality of the protein. There is also limited knowledge on how these nine different isoforms may affect the overall folding of subsequent proteins. Thus, we aimed to computationally model the crystalline protein structure of the isoforms. Further we aimed to create a complete knockout of the *C. elegans* neuroligin gene that expresses a green fluorescent protein from the *nlg*-1 promoter.

Creating a complete neuroligin knockout strain of *C. elegans* would further improve the knowledge surrounding the function of the protein at the synapse. Moreover, the knockout strain

can be further modified to express individual isoforms and even a “humanized” protein containing known ASD associated proteins. To accomplish this, a plasmid can be constructed to contain the 5 prime and 3 prime flanks of the neuroligin gene annealed together with a self-excising cassette and a green fluorescent protein (GFP) reporter **(Figure 2)**. The self-e excising cassette contains a temperature induced Cre which will facilitate self-removal, the Sqt-1 gene which will alter behavioral phenotype, and a hygromycin resistance gene (Dickinson et al. 2015). The Sqt-1 gene and the hygromycin resistance are selection markers for the transgenic animals. Following the successful microinjection of the CRISPR/Cas9 system, *C. elegans* would be screened before the self-excising cassette was removed leaving the complete knockout with the GFP transcriptional maker behind.

To better understand the role of neuroligins in both *C. elegans* and its role in autism spectrum disorder, computation modeling and the creation of a null knockout mutant were initiated. Through the computational modeling we hoped to gain insight into the reason the existing *nlg*-1 mutant strain, VC228, did not show many phenotypic changes. We then worked on creating the null knockout mutant to fully ensure no function of the protein was present; the null knockout strain will allow for future selective expression of different isoforms and mutations associated with ASD.

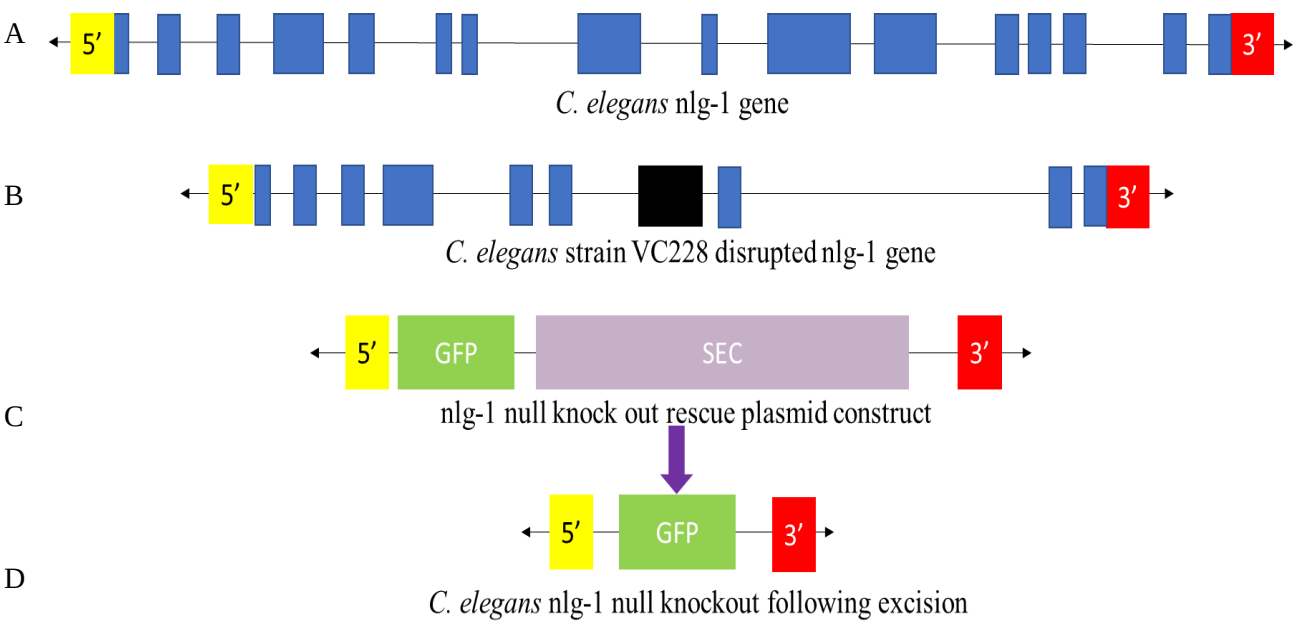


Figure 2. Design schematic for designing a complete null knockout. (Blue box: exons, black box: VC228 insert) **A)** Full exon structure of the *C. elegans* nlg-1 gene. **B)** Exon structure for the disrupted VC228 disrupted nlg-1 gene. **C)** Structure of the nlg-1 null knockout rescue plasmid. Contains green fluorescent protein and a self-excising cassette. **D)** Depiction of the nlg-1 null gene following excision of self-excising cassette.

METHODS

Computational Modeling

Utilizing different computer software (NCBI's BLAST, SWISS-Model and C-I-TASSER, and PyMol) the difference in isoforms and the crystalline structure and of the nine known *C. elegans* neuroligin proteins were predicted. First differences between the nine different isoforms were determined by aligning each of the smaller protein sequences against the full *nlg*-1 gene which is present in isoform E. The remaining isoform information can be found in Table 1. Each of the isoform sequences were analyzed with NCBI's nucleotide BLAST software. The differences were noted and compared against the known location of each of the introns and exons. This information was also available from WormBase.org.

Table 1 *Caenorhabditis elegans* nlg-1 Isoforms

Isoform	Coding Sequence Name (WormBase.org)	Coding Sequence length (base pairs)	Protein Length (amino acids)
Isoform A	C40C9.5a	2397	798
Isoform B	C40C9.5b	2388	795
Isoform C	C40C9.5c	2385	794
Isoform D	C40C9.5d	2292	763
Isoform E	**C40C9.5e**	**2544**	**847**
Isoform F	C40C9.5f	2397	798
Isoform G	C40C9.5g	2460	801
Isoform H	C40C9.5h	1269	422
Isoform I	C40C9.5i	1278	425

Bolded Isoform E indicates largest final protein. Table adapted from WormBase.org

Once the differences were known each protein sequence was placed through the programs SWISS-Model and C-I-TASSER to generate unsubmitted Protein Data Bank (PDB) files (Guex et al. 2009, Bertoni et al. 2017, Bienert et al. 2017, Studer et al. 2020, Waterhouse et al. 2018, Zheng, et al. 2021). The structures of the different isoforms would be predicted based upon using the crystal structure of human NLGN4X (PDB 3BE8) as a reference (Fabrichny et al. 2007). The Program PyMOL (Schrodinger, LLC) was used to help visualize the predicted structures of each isoform. Each visualized isoform was also aligned with isoform E, thus allowing for visualization of the difference present. Furthermore, the crystal structure of the knockout nlg-1 protein from the *C. elegans* strain VC228 was also predicted also utilizing SWISS-Model, C-I-TASSER and PyMOL.

Predicting* nlg-1 *dimerization and neurexin binding sites

From prior work completed by Ko and colleagues and Dean and colleagues, the dimerization location for human neuroligin 3 was known (Dean et al. 2003 and Ko et al. 2009). For NLGN3 there are three dimerization sites. The first contains three residues at phenylalanine (F) 437, methionine (M) 438, and tryptophan (W) 442 (Ko et al. 2009). The second and third dimerization sites only contain two residues; for dimerization site 2, the residues are histidine (H) 606 and leucine (L) 607, and for dimerization three, the residues are lysine (K) 600 and valine (V) 601 (Dean et al. 2003). To find the predicted corresponding *C. elegans* residues, the translated NLGN3 sequence (NP_061850.2) was compared to the translated *nlg*-1 sequence (NP_872254.1) utilizing NCBI's protein BLAST. The location of the NLGN3 residues were identified on the corresponding *C. elegans* protein sequence. From there the sequence was compared to the known sequence of each exon.

A similar process was used to find the predicted nlg-1 and nrx-1 binding sites. Studies modeling three of the human neuroligins indicated each had its own binding sites (Uchigashima 2021, Yoshida et al. 2021). The binding sites for NLGN1 and NRXN were found at D387, E397 and N400; the binding sites for NLGN3 and NXRN are found at K609, R611 and R613. Furthermore, for NLGN4X and NRXN the binding sites were found at D351, E361, and N364. Each of the three translated *NLGN* sequences (NLGN1: XP_016861382.1/ NLGN3: NP_851820.1/NLGN4X: NP_001269074.1) were compared to the translated *nlg*-1 sequence (NP_872254.1) utilizing NCBI's protein BLAST. Each of their residues were found and the corresponding *C. elegans* regions was noted and again found in the reverse translated protein sequence. The partial base pair sequence was found within the known sequence of each exon.

Neuroligin Complete Knockout:

To better understand the function of neuroligins, a complete knockout model was designed. This model would use a CRISPR/Cas9 system to incorporate the transgenic DNA into the *C. elegans* genome (Jiang et al. 2013, Doudna and Charpentier 2014, Dickinsonet al. 2015, Singh et al 2017, Zerbini et al 2017). To utilize the CRISPR system the 3' guide RNA (gRNA) and the rescue plasmid were created. The plasmid carrying the 5' guide RNA was previously made by Stefanie DeFronzo and named pSD1. Following the completion of all components, the rescue plasmid, gRNA plasmids and the Cas9 plasmids would be microinjected into WT *C. elegans* for transfection. After the null model is created, others can be made to express specific isoforms or mutation, further characterizing the role of neuroligin in synapses and in autism spectrum disorders.

The Neuroligin 3 prime gRNA

The gRNA will be used to specifically locate the location within the *C. elegans* genome where the targeted gene editing is to occur. The nlg-1 3’ guide RNA (gRNA: pSW19) was created by isolating the plasmid pRB1017, an empty gRNA plasmid obtained from Addgene, using the miniprep plasmid purification kit from Invitrogen (Arribere et al. 2014). The plasmid pRB1017 was digested (protocol in Appendix 1) with *Bsa*I to obtain the empty backbone into which the gRNA sequence was inserted (Arribere et al. 2014). The oligonucleotides 3’g1 (5' TCTTGAAGTTGGAGATCATGAGCGG 3') and 3’g2 (5' AAACCCGCTCATGATCTCCAACTTC 3') were annealed and ligated to pRB1017 utilizing the New England BioLabs (NEB) Quick Ligase kit before being transformed into NEB 10-Beta competent *E. coli* cells. Colony polymerase chain reactions (PCR protocol; see Appendix 1) was performed with the primers M13F1 (5' TGTAAAACGACGGCCAGT 3') and M13R1 (5' CAGGAAACAGCTATGAC 3') and the NEB LongAmp Hot Start Taq 2X Master Mix. The PCR product was purified following the NEB Monarch PCR and DNA Clean Up Kit protocol. Sanger sequencing reaction was completed by using the Applied Biosystem’s BigDye Terminator v3.1 Cycle Sequencing Kit with the primer M13R1. An ethanol precipitation was completed to clean the reaction before being loaded into the sequencing machine. The samples were analyzed for the correct sequence.

nlg-1 Null Knockout

Creating pSW13

The CRISPR/Cas9 system contains a rescue plasmid which will replace the genomic DNA with the alternate sequence that will be incorporated into the wild type animals (Jiang et al. 2013, Doudna and Charpentier 2014, Dickinson et al. 2015, Singh et al 2017, Zerbini et al 2017).

This plasmid was constructed to include a green fluorescent protein and a self-excising cassette, used for introducing and removal of selection markers (Dickinson et al. 2015). Following the completion of this plasmid, additional modifications can be introduced to selectively express specific isoforms and mutations; to better characterize function, location, and potential pathogenesis of a mutation. The *nlg*-1 null knockout rescue plasmid was made by linearizing the plasmid pDD282 (obtained from AddGene, pDD282 contains the self-excising cassette) with *Sal*I and the plasmid pJR33 (backbone containing the 5' and 3'flanks of the *C. elegans nlg*-1 gene made in the Bayne laboratory by Janaya Reeves, class of 2019) with *Eco*RV **(Figure 3)**. Linearization occurred to reduce the possibility of super coiled DNA interference. The restriction sites, for this and subsequent restriction mapping, were predicted utilizing the online program RestrictionMapper (Blaiklock). Each digest was run on an 0.8% agarose gel to ensure linearization before PCR reactions were run to isolate the desired Gibson Assembly fragment. All project specific sequencing primers were designed with NCBI's online Primer BLAST software (Ye et al. 2012). Gibson Assembly primers were designed utilizing NEBuilder Assembly tool (New England BioLabs Inc). pJR33 underwent PCR with the forward primer pJR3B_fwd (5' ataagcgtgaCTCCAACTTCTTTCGTTTC 3'), the reverse primer (5' ctttactcatGCCTGTTCACTTCCAAATTC 3') and the NEB LongAmp Hot Start Taq 2X Master Mix. pDD282 underwent PCR with the forward primer gfpsec_fwd (5' gtgaacaggcATGAGTAAAGGAGAAGAATTGTTC 3'), the reverse primer gfpsec_rev (5' gaagttggagTCACGCTTATCGTCATCG 3') and the NEB LongAmp Hot Start Taq 2X Master Mix. A Gibson assembly (see appendix 1) was completed to anneal the two PCR products together (**Figure 3)**. The assembly product was transformed into NEB 10-Beta competent *E. coli* cells following manufacturers protocol. A mini prep following the Invitrogen

PureLink Quick Plasmid Miniprep Kit was completed to isolate plasmid DNA, which was used to complete an identifying digest using *Puv*II and 0.8 % gel. The expected band sizes were 4766, 3140, 2364, and 384 base pairs (bp) due to the number of restriction enzyme cut sites.

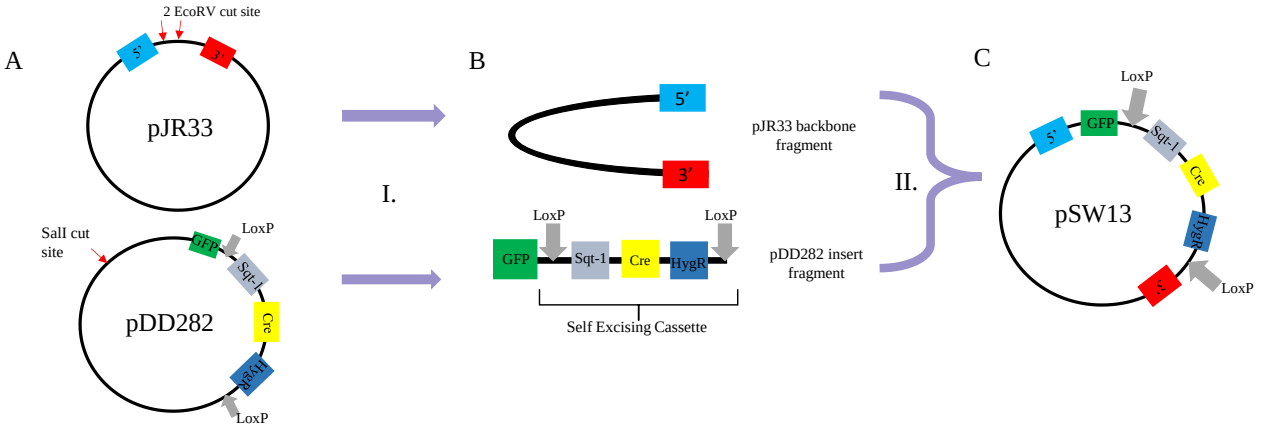


Figure 3. Creating pSW13 through Gibson Assembly. (5': 5 prime flank, 3': 3'flank, GFP: green fluorescent protein, Sqt-1: gene for visualizing phenotypic changes in transgenic animals, Cre: Cre recombinase, HygR: Hygromycin resistance). A) The starting plasmids and their contents needed for creating the null knockout rescue plasmid. I.) Actions taken to isolate plasmid fragments. Each plasmid was linearized with their respected restriction enzymes and underwent PCR to obtain each fragment. B) Fragments isolated from each starting plasmid and its final construct important components. II.) Gibson Assemblies were performed to anneal the fragments together. C. The final null knockout rescue plasmid was generated and given the designation pSW13.

Promising colonies were further sequenced using Sanger sequencing in reactions with the primer 282F (5' AATCAGATCCGGCTCCCTG 3') or JRA2F (5' CTTCTACCCGCCCTTCTTGTCTTC 3'). Following the initial identification sequencing, a full plasmid sequence was completed utilizing 24 overlapping sequencing primers. A full list of pSW13 sequencing primes can be found in Table 2. These sequencing primers were also designed using NCBI's Primer BLAST software (Ye et al. 2012). The resulting sequence confirmed that the pSW13 plasmid contains the pJR33 backbone and pDD282 insert was designated pSW13. After receiving the raw sequencing data, files were compiled to read the full

transcript. The intended sequence was compared to the experimental sequencing data using NCBI's Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990).

Table 2. pSW13 sequencing primers

Primer Name	Sequence 5' to 3'
JRA1F	GGCCAAGCTGGGATAAACAC
JRA2F	CTTCTACCCGCCCTTCTTGTCTTC
JRA3F	TTCCAACACATTTCAGACTCAAGC
JRB1F	AGCACATGTTTGAAACAAATCCCG
JRB2F	AAGGGCAACTATGGGCTCCTTGAC
JRB3F	ACCCTGCGAGAGACCATCAAGTTC
JRB4F	AATGGCAAGACCTGGAGCAC
SWF2	CCCCACTCACACTCCACTACTACAAGGAGAATCTGTACTTTC
NLG1	CACGACGTTGTAAAACGACG
NLG2	CTTCTACCCGCCCTTCTTGTCTTC
NLG3	ACGGAAAGCTCACCCTCAAG
NLG4	TTTCAGATCCGTCACAACATCG
NLG5	TTTCTCTTTCCGTCTCCCAATTCG
NLG6	AAAGGAGACGAATCCGGTAGG
NLG7	TGGACCAACCTGCAACTGCAATG
NLG8	ACCAGACGGAAGACCAGGAAAG
NLG9	GCGCCAGTTTCAGAATCCAAG
NLG10	CGATGTGGATGAGCATTCTTC
NLG11	ACTGAGAGAAGCATCTTCCAG
NLG12	TACCACTATTTCCGTCCAGC
NLG13	AGGGCGAAGAATCTCGTGCTTTC
NLG14	ACGCGGATTTCGGCTCCAACAATG
NLG15	GGGTCCAATTACTCTTCAACATC
NLG16	GGGACACACCTCTTAGATTGAAAG

Sequencing primers were designed in the NCBI Primer BLAST program (Ye et al. 2012).

Creating pSW1312

pSW1312 was constructed to rectify mutations in pSW13 by ligating two restriction enzyme digest fragments; both fragments were generated by useing *Sac*II and *Bgl*II. pSw13 was digested into two fragments of 6768 and 3886 bps and pDD282 was digested into two fragments

of 6601 and 3874 bps (**Figure 4)**. The large backbone fragment was collected from pSW13, and the small self-excising cassette (SEC) fragment was collected from pDD282. These two fragments were gel purified using the NEB Monarch DNA Gel Extraction kit. The two purified fragments were joined together using the NEB Quick Ligation kit before being transformed into 10-Beta competent *E. coli* cells. The transformants were sequenced using NLG sequencing primers 1-16 (Table 2). After sequencing, this plasmid also underwent BLAST analysis like that of pSW13. This new plasmid was given the designation of pSW1312.

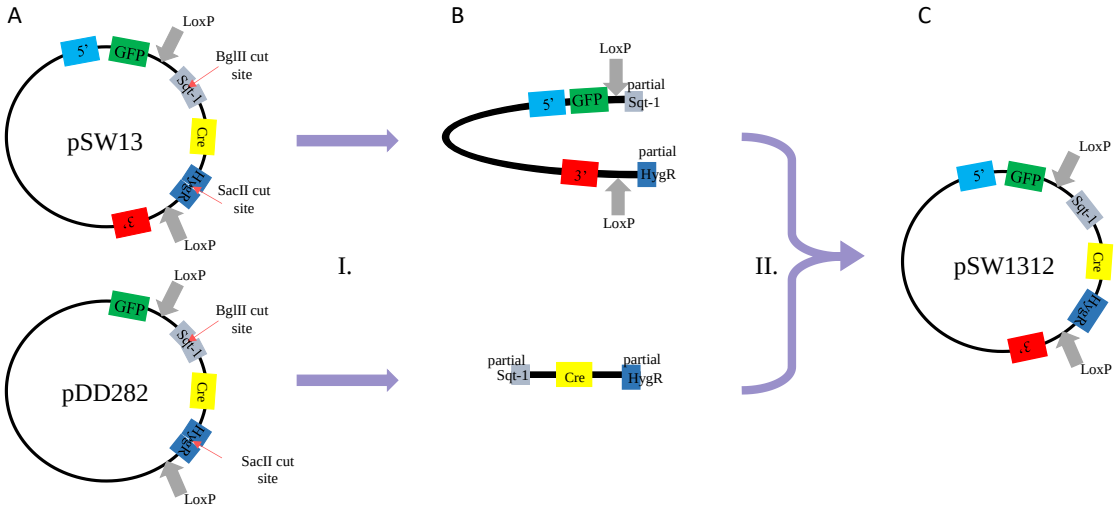


Figure 4. Creating pSW1312 through restriction enzyme digest. (5': 5 prime flank, 3': 3'flank, GFP: green fluorescent protein, Sqt-1: gene for visualizing phenotypic changes in transgenic animals, Cre: Cre recombinase, HygR: Hygromycin resistance). A) These are the starting plasmids from which pSW1312 was made. Important regions needed for pSW1312 are shown along with *Bgl*II and *Sac*II cut sites. I.) Each starting plasmid underwent restriction enzyme digest with *Bgl*II and *Sac*II to retrieve needed fragments. B) Visualization of the critical regions within each of the fragments. Within the self-excising region, the location of digest sites cut within marker sequences. II.) Fragments underwent ligation with NEB's Quick Ligase to generate the latest iteration of the null knockout rescue plasmid. C) Visualization of the final pSW1312 plasmid following ligation.

Creating pSW13126

pSW13126 was generated to correct a point mutation found in pSW1312 utilizing site-directed mutagenesis; and was completed using the NEB Q5 site-directed mutagenesis kit. Mutagenesis primers were designed in NEBaseChanger (New England BioLabs inc): GFPF (5' GTTCTCCGTCTCCGGAGAGG3') and GFPR primer (5' TTGTGTCCGTTGACGTCTCC 3'). Clones were digested with *Hind*III for identification of SEC insert; fragment were expected to have band sizes of 3032, 2973, 2246,1514, and 889 bp each die to the number of *Hind*III cut sites in the pSW13126 sequence. This was followed by Sanger sequencing with the JRA2F primer (5' CTTCTACCCGCCCTTCTTGTCTTC 3') to ensure that the site-directed mutagenesis occurred. Potential clones identified with the SEC insert were fully sequenced with NGL sequencing primers 1-16 (Table 1). This plasmid was designated pSW13126.

Creating pSW13126X in progress

pSW13126X was designed to fix additional mutations found within the pSW13126 plasmid. For this construction, both pSW13126 and pDD282 were digested with *BamH*I. Digestion of pSW13126 was predicted to produce two fragments with the length of 9298 and 1356 bp long. Digestion of pDD282 was predicted to also have two fragments of 9119 and 1356 bp (**Figure 5)**. Fragments were run on an 0.8% agarose gel. The large fragment from pSW13126 was isolated and the small fragment from pDD282 was also collected. The fragments underwent gel purification following a protocol identified as "freeze and squeeze" (see Appendix 1) (Patharkar 2019). The fragments were ligated together using the NEB Quick ligation kit before being transformed into 10-Beta competent *E. coli* cells. Clones were digested with *BamH*I to ensure that the fragments were present before sequencing; the same fragment size found prior in pSW13126 would be expected here.

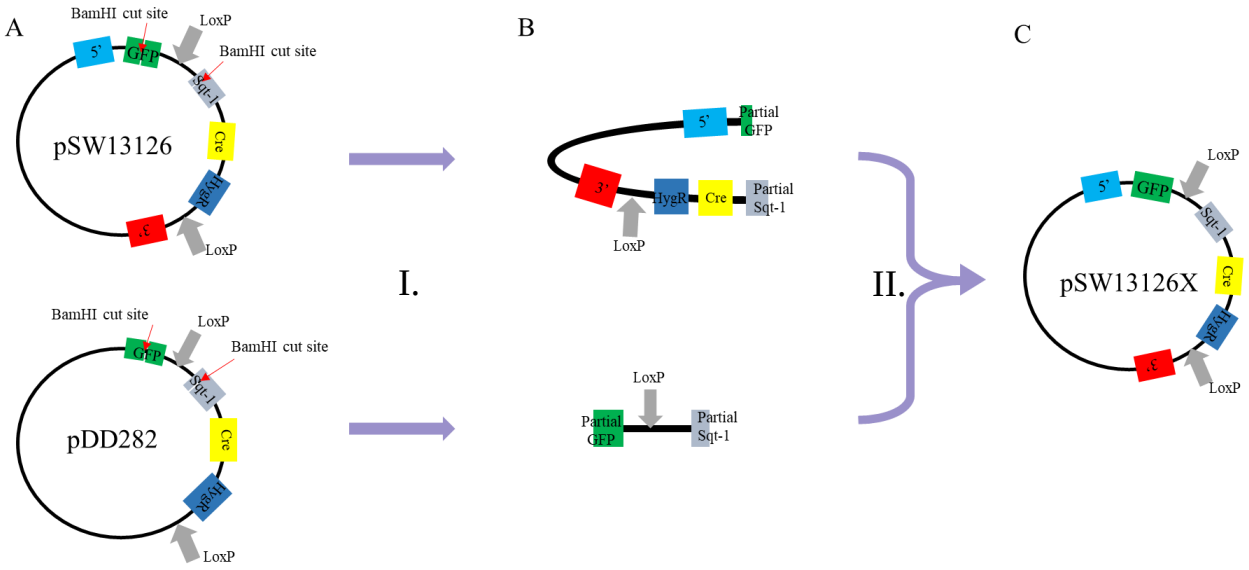


Figure 5. Creating pSW13126X through restriction enzyme digest. (5': 5 prime flank, 3': 3'flank, GFP: green fluorescent protein, Sqt-1: gene for visualizing phenotypic changes in transgenic animals, Cre: Cre recombinase, HygR: hygromycin resistance). A) starting products for pSW13126X, are depicted as well as key components found within each plasmid and approximations to the location of each *BamH*I site. I.) Each plasmid underwent *BamH*I digest and gel purification. B) Visualization of the fragments and their components following restriction enzyme digest. II.) The fragments underwent ligation to form the final construct for psSW13126X. C) Final intended construct for pSW13126X and its components.

RESULTS

Computational Modeling

C. elegans neuroligin 1 modeled isoforms

Utilizing NCBI's BLAST software and information obtained from WormBase, the differences between the nine different *C. elegans* neuroligin isoforms were determined. All isoforms were compared to the full transcript of nlg-1 found in isoform E with 847 aa (**Figure 6)**. Isoform A is missing three amino acids, an arginine, phenylalanine, and another arginine (RFR), in exon 4 and is also missing exon 14. Isoform B is also missing the three amino acids (RFR) from exon 4, as well as exon 14. However, isoform B also lacks three amino acids, glycine, asparagine, and serine (GNS) at the very end of exon 15 and beginning of exon 16. Isoform C was discovered to only be missing exons 15 and 16. Isoform D was also determined to be missing the three, RFR, amino acids from exon 4 as well as exons 13 and 14. Isoform F is missing the three GNS amino acids from the end of exon 15 and beginning of exon16. It was further determined that Isoform G is only missing exon 14. Isoforms H and I start at the second start codon found within exon 8, thus, both are missing the portion of exon 8 prior to the start codon and the first 7 axons **(Figure 6)**. Both isoforms are also missing exon 14. Isoform H differs from Isoform I in that the three amino acids GNS are missing from the end of exon 15 and beginning of exon 16. Isoform I begins at the second start codon and is missing exon 14. Thus, these two isoforms were expected to have different predicted crystal structures compared to the other isoforms which began coding at exon 1.

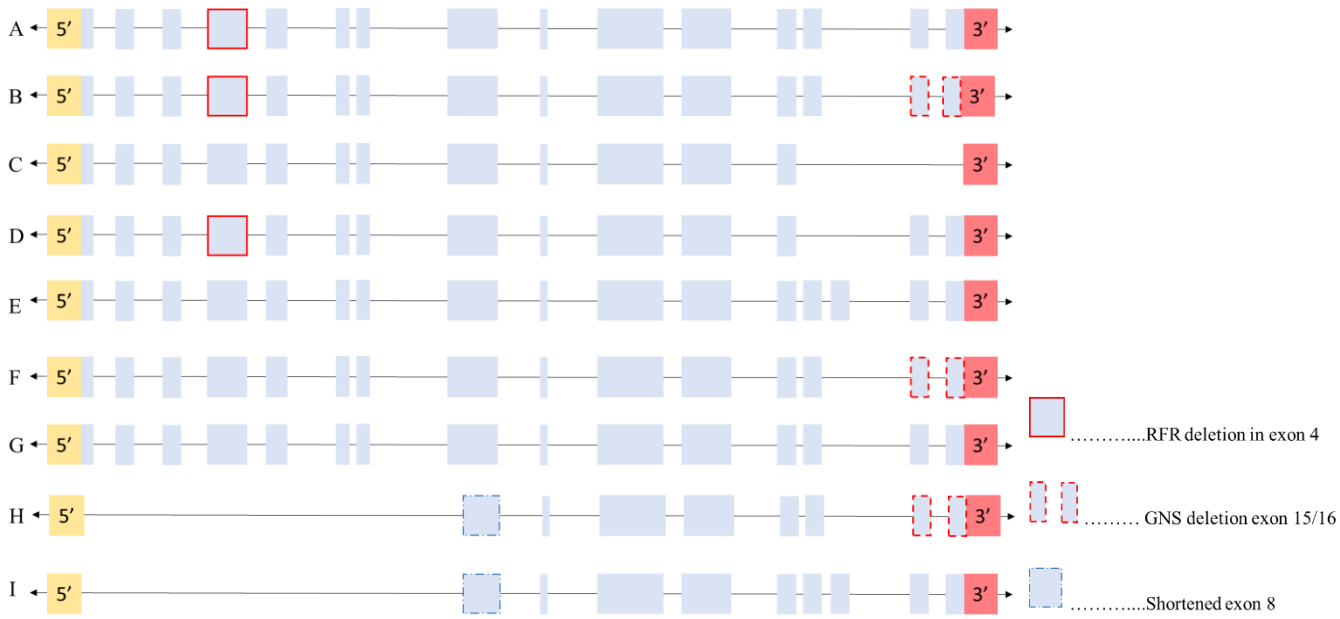


Figure 6. Visual representation of nlg-1 isoforms. (Light blue box: exon, RFR deletion in exon 4: solid red outline, GNS deletion in exons 15/16: dashed red outline, shortened exon 8: dashed blue outline) There are nine different isoforms of nlg-1, two of which start at the second start codon found within exon 8. If a particular exon is missing from the sequence, there will be no representative box in the known position. Each isoform is represented with their letter designation on the left followed by the exon structure on the right. Isoform E is the largest isoform that contains no deletions and is what all other isoforms were compared to.

Visualization of nlg-1 Protein Isoform Structures

Utilizing C-I-TASSER, the predicted structures (PDB file) of the different nlg-1 isoforms was generated and visualized in PyMol (see Appendix 3 for the prediction files). These predicted crystal structures allowed for a visual determination on potentially how each of the different isoforms altered the structure of nlg-1. In isoforms H and I, the isoforms starting at the second start codon, the protein is much smaller but still capable of forming dimers, despite missing large sections of the N-terminus region of the full protein **(**See **Figure 7** for isoform H predicted structure**).** In panels B and C of Figure 7, the pink- and orange-colored section of Isoform E indicate each of Isoform H's monomers. The overlapping of isoform E and H allowed for a visual representation of the similarity of both the proteins and the dimerization sites.

In isoforms with a arginine, phenylalanine, arginine deletion in exon 4, such as in isoform B, it was hypothesized that the deletion would cause a β-sheet to turn into α-helix (red circle in **Figure 8**). This change can be seen in all three-isoform (A, B, and D) containing the three amino acid deletion but not seen in the others. This structural change should not be the result of any protein sequence change after the second start codon. This region of the visualized protein is not found within the two isoforms that begin their coding there **(Figure 7)**. Structural changes caused by the exon deletion were much more difficult to determine. No visual difference could be detected that were unique to those of isoforms without the exon deletion. However, the three amino acid deletion from the end of exon 15 and beginning of 16 in isoforms B, F, and H did affect how a section of loop interacted with itself changing the conformation (**Figure 9)**. It was determined that all isoforms are capable of forming dimers with other neuroligin proteins. The structures of all nine isoforms independently and compared to isoform E can be found in Appendix 3.

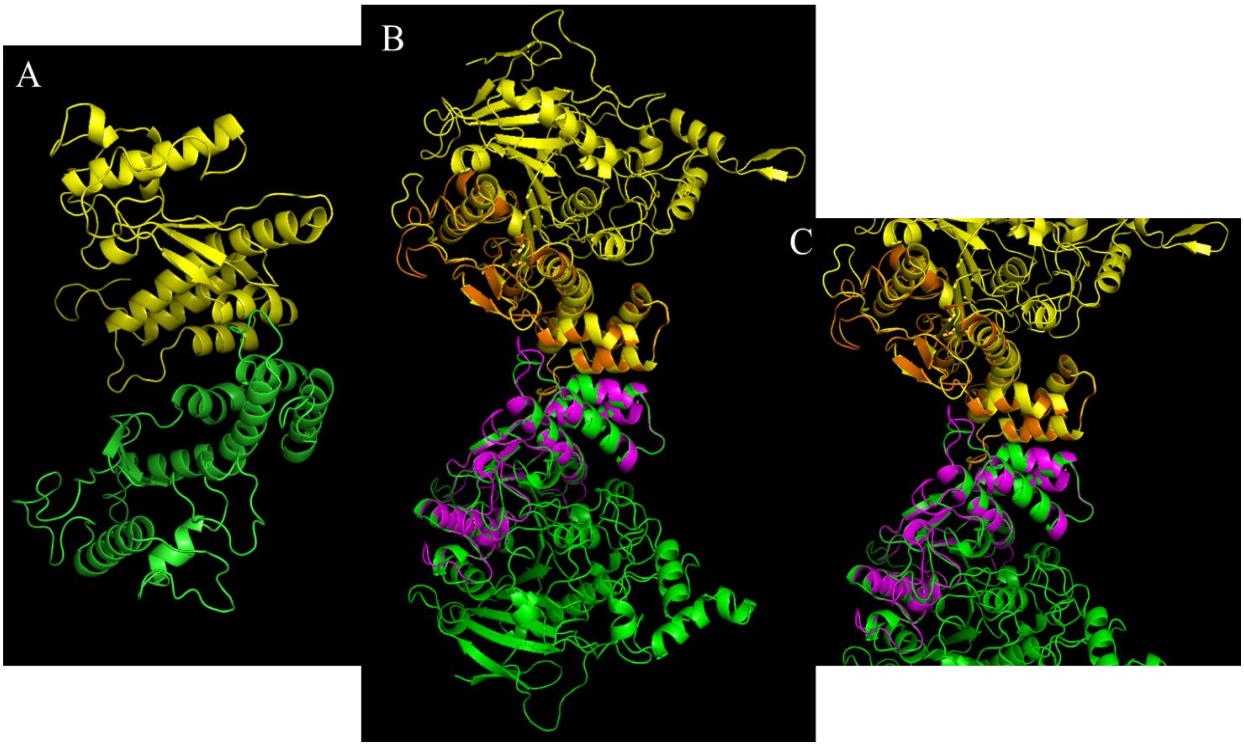


Figure 7. Comparison of the predicted structure of nlg-1 isoform H with isoform E. (all proteins visualized in PyMol) Isoform H starts at the second start codon as such is expected to be smaller in size to the other isoforms. While isoform I is not pictured in this figure, both share the common similarity of starting at the second start codon found within this eighth exon. **A)** The predicted dimer formed by translating nlg-1 isoform H without alignment to full protein. **B)** Isoform H (pink and orange) aligned with Isoform E (yellow and green). The predicted isoform H size is much smaller than isoform E but can still form dimers in the same fashion as isoform E. **C)** This a zoomed in view of the aligned isoform H and isoform E focusing on the regions translated from Isoform H. This panel does not focus on the large overall picture of isoform E.

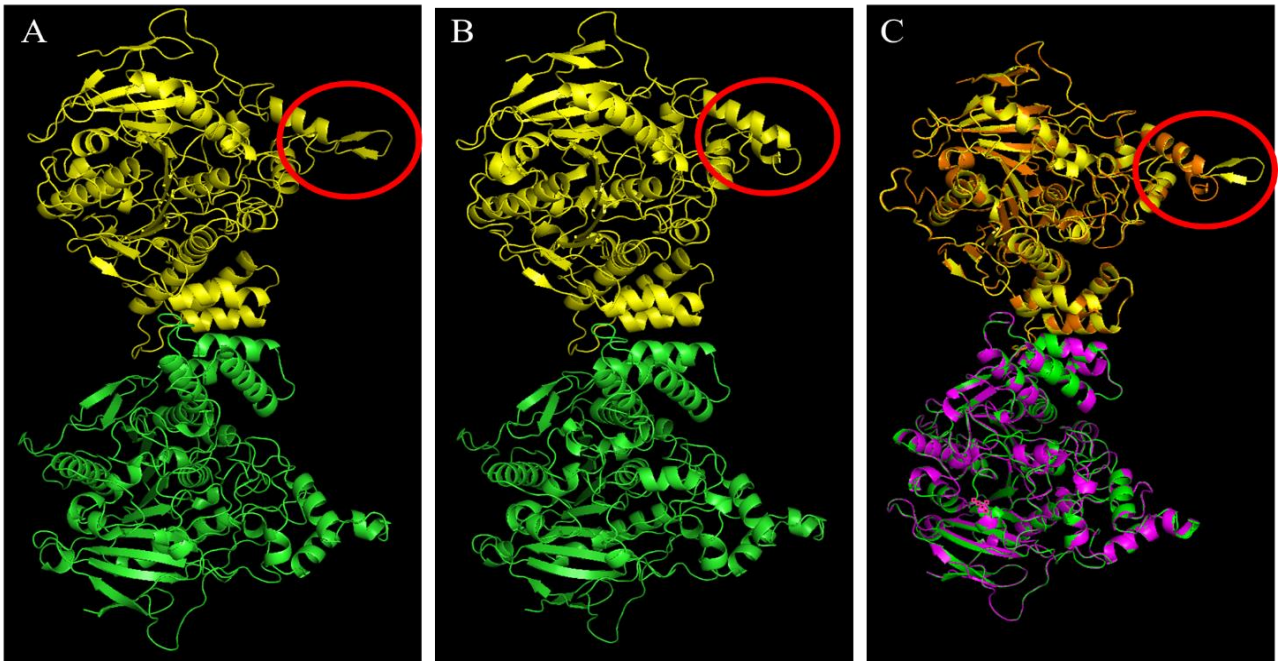


Figure 8. Hypothesized alterations in protein structure due to the arginine, phenylalanine, arginine deletion. (Red circle indicates area of interest, all structures visualized in PyMol). **A)** Visualization of predicted dimer that nlg-1 forms when translated from isoform E. The red circle highlights the initial β-sheet conformation. **B)** Visualization of predicted structure for dimer formed by nlg-1 isoform B. Red circle highlights the conformation change from β-sheet to α-helix in the same location as in panel A. **C)** An overlayed alignment of Isoforms E (green and yellow) and B (orange and pink). In the red circle the conformation change can be see clearly changing from a β-sheet to an α-helix.

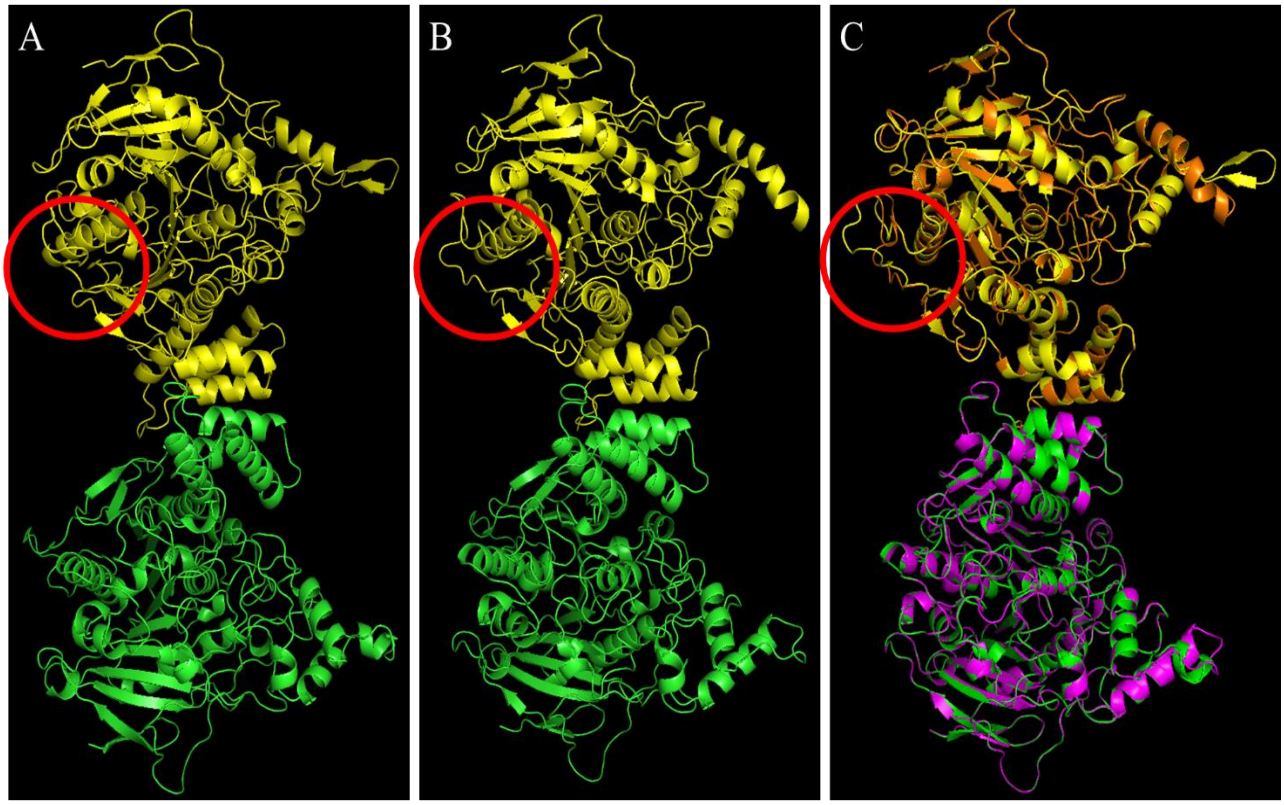


Figure 9. Predicted structural change due to the glycine, asparagine, serine deletion at the end of exon 15. A) Visualization of predicted dimer that nlg-1 forms when translated from isoform E. The red circle highlights the initial loop conformation. **B)** Visualization of predicted structure for dimer formed by nlg-1 isoform F. Red circle indicates highlights the conformation of loop in the same location as in panel A. **C)** An overlayed alignment of Isoforms E (green and yellow) and F (orange and pink). In the red circle, the conformation change of loop can be seen.

VC228 isoforms

Prior sequencing in the Bayne laboratory (completed by Stefanie DeFronzo; class of 2020) provided the full sequence of VC228 (data not shown). This sequence was compared to the wildtype (WT) nlg-1 protein sequence. VC228 is missing exons 8-13 and a 334 base pair insert exists in its place. VC228 does not appear capable of forming dimers **(Figure 10)**. It also appears smaller than the other isoforms. This mutated protein does include the β-sheet to α-helix change, like some of the wild type isoforms. Further computational modeling information can also be obtained in Appendix 3.

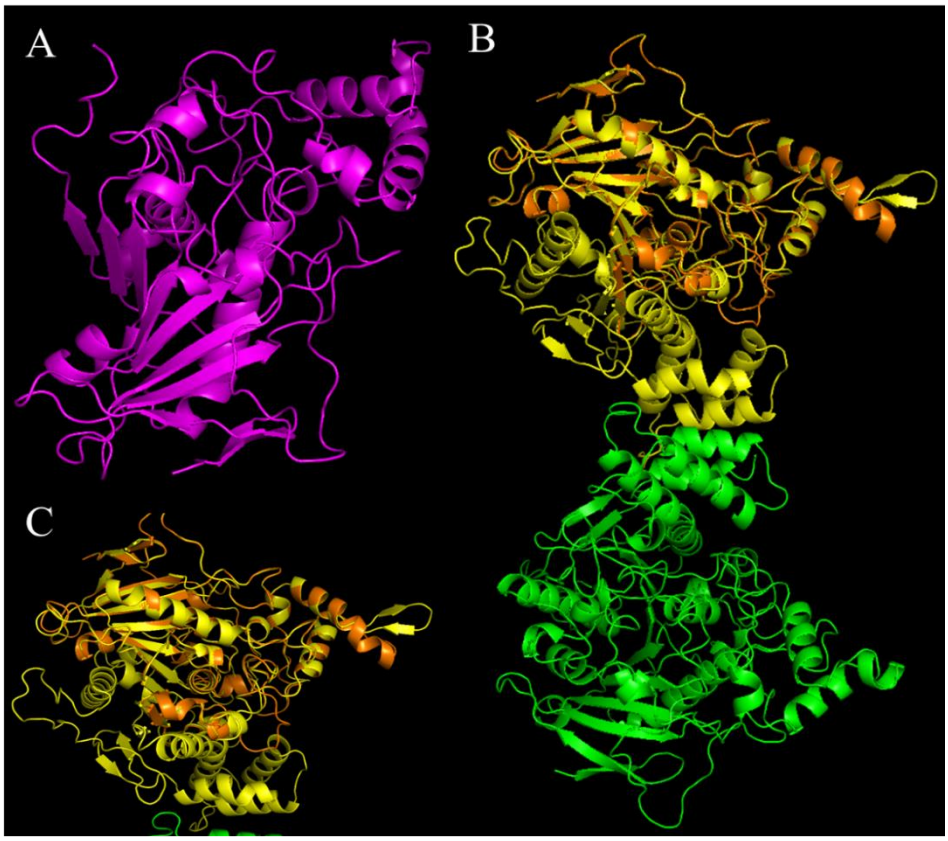


Figure 10. Predicted structure of the nlg-1 from the strain VC228. A) Predicted structure of a monomer of VC228, nlg-1 protein (pink). **B)** The predicted VC228 monomer compared to the predicted structure of wildtype nlg-1 isoform E. The alignment is shown in orange and yellow for increased contrast between the two proteins. **C)** A magnified view of the VC228 monomer compared to a monomer of the wild type nlg-1 isoform E.

Predicted nlg*-1 dimerization and neurexin binding sites*

Identification of the predicted neuroligin dimerization site, utilizing research of human neuroligin, showed the exon location of each of each residue within the *C. elegans* protein. As mentioned in the methods section there are three NLGN3 dimerization sites, found by Dean and colleagues and Ko and colleagues. The corresponding *C. elegans* residues for the first site were determined to be located at positions asparagine (N) 414, glutamic acid (E) 415 and tryptophan (W9) 419. The residues for the nlg-1 dimerization site two are located at asparagine (N) 594 and phenylalanine (F) 595. The third dimerization site is located at positions glutamine (Q) 588 and valine (V) 589. It was further determined that the first dimerization site is located within exon 8 of nlg-1 and both sites two and three are in exon 11. Further confirmation of these predicted sites is need. However, this serves as an initial proof of concept that if the dimerization sites of NLGN3 and nlg-1 are conserved, they should behave within the cell in similar fashions.

The NLGN/NRXN human binding site was also used as a reference to determine the potential C. *elegans* nlg-1/nrx-1 binding site. Currently, the NRX/NLG binding sites have been identified for NLGN1, NLGN3 and NLGN4X. All were used to predict the binding sites as it is unclear which NLGN, nlg-1 would behave more like in at different synapses. For NLGN1 the binding sites are located at aspartic acid (D) 387, glutamic acid (E)97, and asparagine (N) 400 (Yoshida et al. 2021, Uchigashima 2021). The binding sites for NLGN3 are located at lysine (K) 609, arginine (R) 611 and arginine (R) 613. Finally, the binding sites for NLGN4X are located at aspartic acid (D) 351, glutamic acid (E) 361 and asparagine (N) 364. The corresponding *C. elegans* residues for each NLGN were located by using NCBI protein BLAST. The corresponding residues by using NCBI protein BLAST for NLGN1 are located at lysine (K) 344,

leucine (L) 356, and glutamic acid (E) 359, where the associated residues for NLGN3 are found at arginine (R)577, arginine (R) 579 and lysine (K) 615. Finally, the NLGN4X corresponding residues are found at asparagine (N) 343, lysine (L)356 and glutamic acid (E) 359. From *C. elegans* residues, the binding sites associated with NLGN1 and NLGN4X are located within the eighth exon of nlg-1 and the associated binding site for NLGN3 can be found in within exons 10 and 11. While both the nrx-1 binding and dimerization sights of nlg-1 are predictions, they provide the abillity to speculate the overall effect mutations may have on these behaviors.

Complete nlg-1 Knockout

Neuroligin 3 Prime gRNA Plasmid

Following the computational modeling, creation of the *nlg*-1 knockout commenced with the creation of the 3' guide RNA and the rescue plasmid. All plasmids underwent sanger sequencing to confirm identity and sequence. The two gRNA oligonucleotides were annealed together and ligated into a *Bsa*I digested pRB1017 (confirmed through an agarose gel; not shown). The product was transformed into NEB 10-Beta competent *E. coli* cells as described in methods. The colony PCR and subsequent agarose gel electrophoresis identified sample 3 and 6 as potential candidates due to the slightly larger band size (**Figure 11)**. Sample 3 was identified to contain the correct sequence, after confirmational sequencing and analysis with the primer M13R1. However, sample 6 was determined to not contain the correct sequence and was discarded. The third sample plasmid was given the designation pSW19.

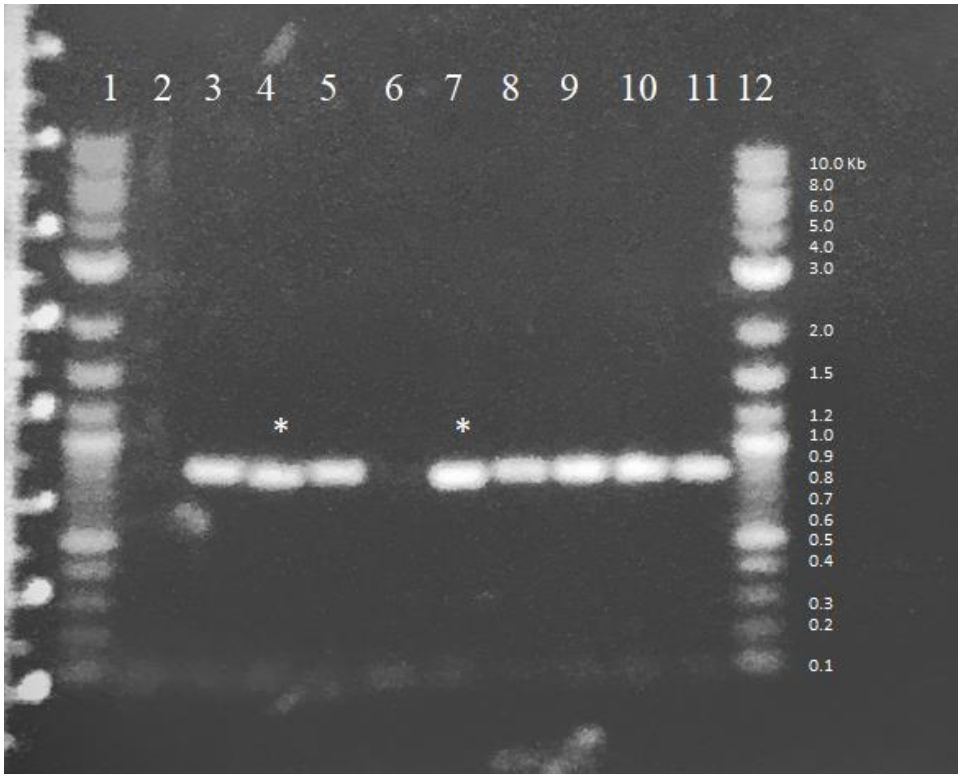


Figure 11. Screening for potential 3' gRNA candidates. (Lane 1: New England BioLabs 1kb Plus DNA Ladder, Lane 2: Clone 1, Lane 3: Clone 2, Lane 4: Clone 3, Lane 5 Clone 4, Lane 6: Clone 5, Lane 7: Clone 6, Lane 8: Clone 7, Lane 9: Clone 8, Lane 10: Clone 9, Lane 11: Clone 10, Lane 12: New England BioLabs 1kb Plus DNA Ladder) * indicates colonies chosen to be sequenced for presence of 3' gRNA.

pSW13

This plasmid (pSW13) was designed to function as the rescue plasmid in the CRISPR/Cas9 system to make a knockout of the *C. elegans nlg*-1. pSW13 was created through PCR and Gibson assembly from the starting plasmids, pJR33 and pDD282. pJR33 contains the 5' and 3 prime flank regions of the *nlg*-1 gene. These flanking regions improve the efficiency of homologous recombination to making the complete null knockout. pDD282 contains the green fluorescent protein and the self-excising cassette, loxP sites, sqt-1, Cre and hygromycin (an eukaryote antibiotic) resistance (HygR). Once each of the PCR fragments were identified through an agarose gel, they were annealed using a Gibson Assembly and transformed into 10-

Beta competent *E. coli* cells. Liquid cultures of the transformation bacteria were used to isolate the plasmid DNA, which were subsequently digested with *Pvu*II **(Figure 12)**. Only one of the 16 prepared samples showed, promise of being the correct plasmid; the promising sample had four fragments at approximate sizes of 4766, 3140, 2364, and 384 base pairs as predicted. These sizes were expected due to the location of cut sites within the sequence. This information will determine if the SEC insert was successfully ligated to the backbone.

The plasmid was sequenced utilizing a series of 24 overlapping primers. Analysis of the sequencing data demonstrated that the Gibson assembly and transformation had indeed worked. Upon closer inspection, there appeared to be many point mutations throughout the self-excising cassette, backbone, and green fluorescent protein (GFP) regions. There was also a missense mutation within the GFP region (data not shown). This mutation causes a serine to proline change at position 961. Moreover, there was two other mutations that were introduced that further caused concern. There is missense mutation at position 6313 of pSW13, due to the change of an adenine to a guanine, that caused a histidine to become an arginine within the self-excising cassette region. Thus, further work was needed to correct these mutations.

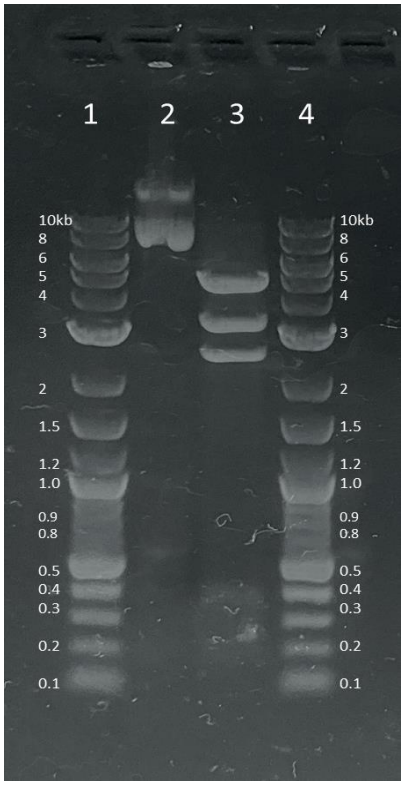


Figure 12. Identification of pSW13 cloneuseing *Puv*II. (Lane 1: New England BioLabs 1kb Plus DNA Ladder, Lane 2: uncut pSW13, Lane 3: digested pSW13, Lane 4: New England BioLabs 1kb Plus DNA Ladder. Expected fragment sizes 4766, 3140, 2364, and 384 base pairs.

pSW1312

Due to the mutations found in the sequencing data from pSW13, subsequent plasmids were created to rectify many of the errors. Following the assembly of the two restriction enzyme digest fragments (fragments made with *Bgl*II and *Sac*II, see methods) and verification through sequencing; the data showed a less mutated self-excising cassette region (data not shown). Mutations were still, however, found throughout the plasmid. Despite the self-excising cassette region containing no new mutations, the serine to proline mutation at position 961 mutation of the GFP region remained. This mutation would need to be remedied to ensure proper function of the GFP post translation.

pSW13126

Site-directed mutagenesis was performed, as described in methods, on pSW1312 to return the proline at position 961 back to a serine. Multiple clones were grown, and DNA was isolated before being digested with *Hind*III as an initial verification step. The digestion (agarose gel not shown) showed that all samples except for sample 3 contained the correct fragments (see methods for band sizing). Following the initial screening, clones were sequenced utilizing the JRA2F primer, clone 6 and clone 7 contained the expected base change pair (**Figure 13)**. Clone 6 was chosen due to the higher concentration of DNA in the sample following sequencing. This sequencing showed a less mutated sequence within the self-excising cassette as compared to pSW13. Despite this there still were four point mutations and two single base pair insertions within the non-coding region between the first loxP site and the start of the Sqt-1 coding region. Thus, these mutations would not affect the overall formation of the protein. Moreover, there was a mutation was a tyrosine to cystine mutation at position 4310. This mutation, found within the coding region for green fluorescent protein, could prevent proper function of the protein. Thus, another round of cleanup was determined to be needed.

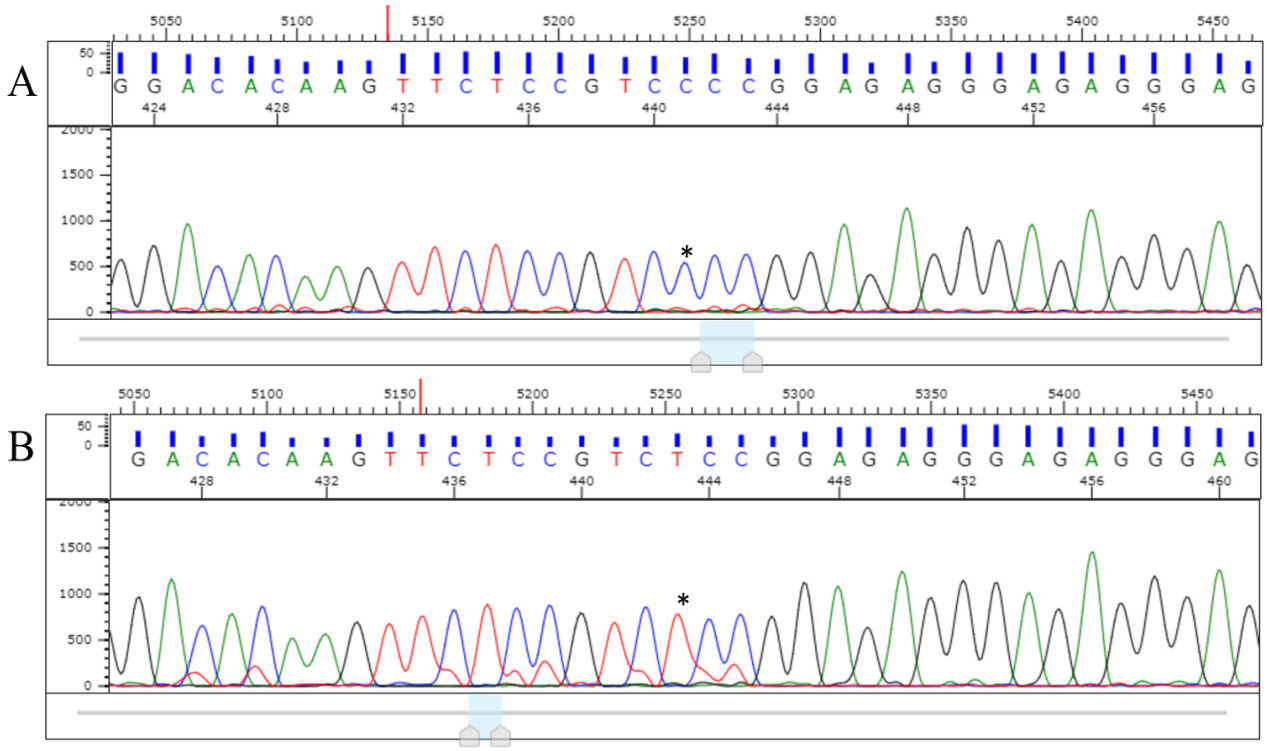


Figure 13. Identification of successfully completed mutagenesis creating pSW13126. (* indicates target for mutagenesis) A) sequence for pSW1312 with presence of the in-frame serine to proline mutation. (proline: CCC) B) Sequence for pSW13126 demonstrating successful mutagenesis of the in-frame proline (CCC) mutation back to a serine (TCC).

pSW13126X

Construction of this plasmid this plasmid is still progress; thus, it does not have a unique designation, but rather a temporary identifier of pSW13126X. This plasmid was created to further remove mutations from the self-excising cassette and the green fluorescent protein. Thus, pSW13126 was digested with *BamH*I to fix the tyrosine to cystine change at position 4310 mutation along with the other six mutations. The *BamH*I fragment from pDD282 (contains the unmutated sequence) replaced the mutated *BamH*I fragment from pSW13126. The pSW13126 plasmid was successfully digested with *BamH*I and the 9298 bp fragment was isolated through the "Freeze and Squeeze" agarose gel purification (gel not shown). The pDD282 1356 bp fragment caused challenges. This fragment had been difficult to isolate through restriction digest. The small plasmid fragments also did not appear well on the 1% gel. After multiple trials in creating this plasmid through restriction enzyme digest and ligation failed to produce colonies. Additional trials demonstrated that higher concentrations of DNA may be needed for distinctive bright bands and a higher gel purification concentration. In instances of no colonies growing, the concentration of the smaller pDD282 fragment was low following isolation, thus there was an even smaller amount of DNA isolated after being purified through the agarose gel (quantified through NanoDrop). At higher concentrations, an agarose gel showed bright correct bands for both pSW13126 and pDD282 **(Figure 15)**. The concentration of the pDD282 samples is 340 ng/µL. For pSW13126 the concentration is at 220 ng/µL. The pDD282 samples were at a much higher concentration compared to previous attempts where it was frequently under 100 ng/µL. This last plasmid purification trial was the stopping point; thus, any future work should continue from these samples.

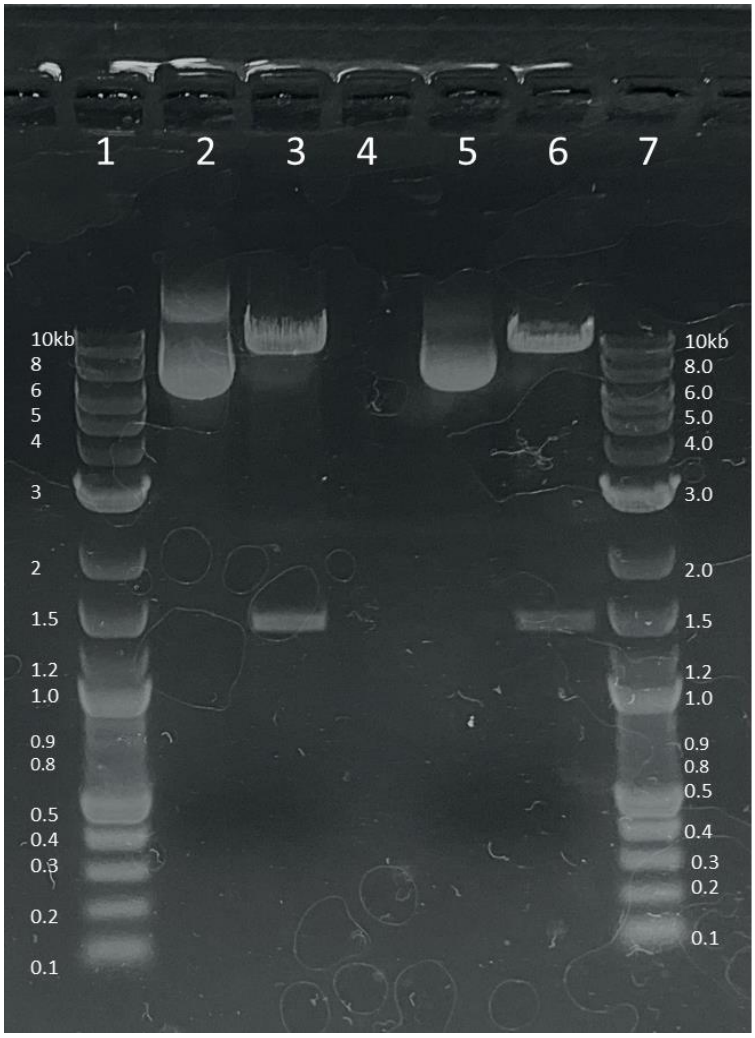


Figure 14. Restriction enzyme digest of pDD282 and pSW13126 with *BamH*I. Lane 1 New England BioLabs 1kb Plus DNA Ladder, Lane 2: Undigested pDD282, Lane 3: Digested pDD2828, Lane 4: Blank, Lane 5: Undigested pSW13126, Lane 6: Digested pSW13126. (1% agarose gel)

DISCUSSION

Autism spectrum disorder have been linked in the past to mutations within neuroligin genes. *C. elegans* and humans both have neuroligin genes that has been very well conserved. As such, a *C. elegans* model system in which to study human ASD. The model system could also be used to study each of the different isoforms of wild type *nlg*-1, this would help to better understand the function and location of each. To build this model, computational modeling of the individual nlg-1 isoforms was completed along with the determination of potential dimerization and neurexin binding sites. This allowed for visualization of current proteins (wild type and mutated) and understanding of how they interact with themselves and neurexins. A CRISPR/Cas9 system was used to build a complete knockout of the *C. elegans nlg*-1 gene. A green fluorescent protein (GFP) was included to allow for visualization of which neurons expressed nlg-1 in its synapses. The GFP also may determine if nlg-1 was expressed in any other location inside of a neuron.

Computational Modeling

Following the prediction of each *C. elegans* nlg-1 isoform structure, further analysis of the protein sequence was completed. It was determined from the structure that each of the nine isoforms should be capable of forming dimers, and thus, able to be trafficked to the plasma membrane of the synapse where interactions between neuroligin and neurexin take place **(Figures 7, 8, 9, and Appendix 3)**. The sequence of the protein further supports this analysis as, the predicted three dimerization (based on human NLGN3 dimerization sites) sites are located withing exons 8 and 11 (Ko et al. 2009 and Dean et al. 2003). Each of the nine isoforms, including the two shortened forms, contain exons 8 and 11. Utilizing similar techniques, the potential binding sites of neuroligin and neurexin were also predicted. The binding sites from

NLGN1, NLGN3 and NLGN4X were used to obtain corresponding regions of nlg-1. These corresponding regions were found within most of the nlg-1 isoforms except for the shortened isoform forms H and I. The binding sites for NLGN1/NRXN and NLGN4X/NRXN correspond to residues within the beginning of exon 8, prior to the start codon used in isoforms H and I. However, the binding sites of NLGN3 and NRXN correspond to regions exon 10 and 11within nlg-1. Therefore, binding would still occur if the binding region of nlg-1 is most like that of NLGN4X or NLGN1 in most instances. The binding sites for the NLGN2 was not used in these predictions as this information was not determined in referenced literature.

The neuroligin protein found in the strain VC228, when modeled, did not appear to capable of forming dimers **(Figure 10)**. Therefore, the binding and dimerization sites for VC228 were also analyzed. Due to the known deletion of exon 8 through 13, the dimerization sites for VC228 are not present. The 334 bp insertion does not appear to influence the dimerization of VC228 as no dimerization sites appear within the insertion (Hunter et al. 2012). Further investigation would be required to determine full effects of this insertion. The lack of dimerization theoretically would prevent the VC228 produced protein from being brought to the plasma membrane of the synapse therefor, there would be no interactions between neuroligin and neurexin. However, due to the limited deficits found within this strain, the lack of dimerization is peculiar and requires further study. Perhaps the addition of a GFP tag will provide insight into where this protein is ultimately trafficked to.

Due to the uncertain effect of the 334 bp insertion, it was further thought that perhaps VC228 still contained the neuroligin and neurexin binding sites. Analysis showed VC228 also does not contain the corresponding NLGN/NRXN binding site regions when compared to the wild type. As mentioned above the VC228 gene lacks exons 8-13; the NLGN1/NRXN and the

NLGN4X/NRXN corresponding binding sites are located within the missing exon 8. The NLGN3/NRXN corresponding binding site is also located within deleted exons, specifically exons 10 and 11. The corresponding regions also did not appear within the 334 base pair insertion; thus, it can be expected that there would be little unintended interaction. The lack of all binding sites further adds to the confusion relating to the limited deficit phenotype despite altered structure. The knowledge that despite the altered protein, the limited behavioral changes remain, provide a compelling reason to create a null knockout mutation.

Complete *nlg*-1 Knockout Model

The complete *nlg*-1 knockout model has yet to be completed due to multiple mutations found within the plasmid. Multiple iterations of the rescue plasmid construction have been needed; each contained corrections for the different mutations found in the SEC and GFP found during the previous attempts (**Figure 15)**. There still are mutations to be corrected; however, the number is much smaller, these were fixed alongside the larger amino acid mutations. Multiple correction methods were used, including site-directed mutagenesis, and restriction enzyme fragment replacements and ligations.

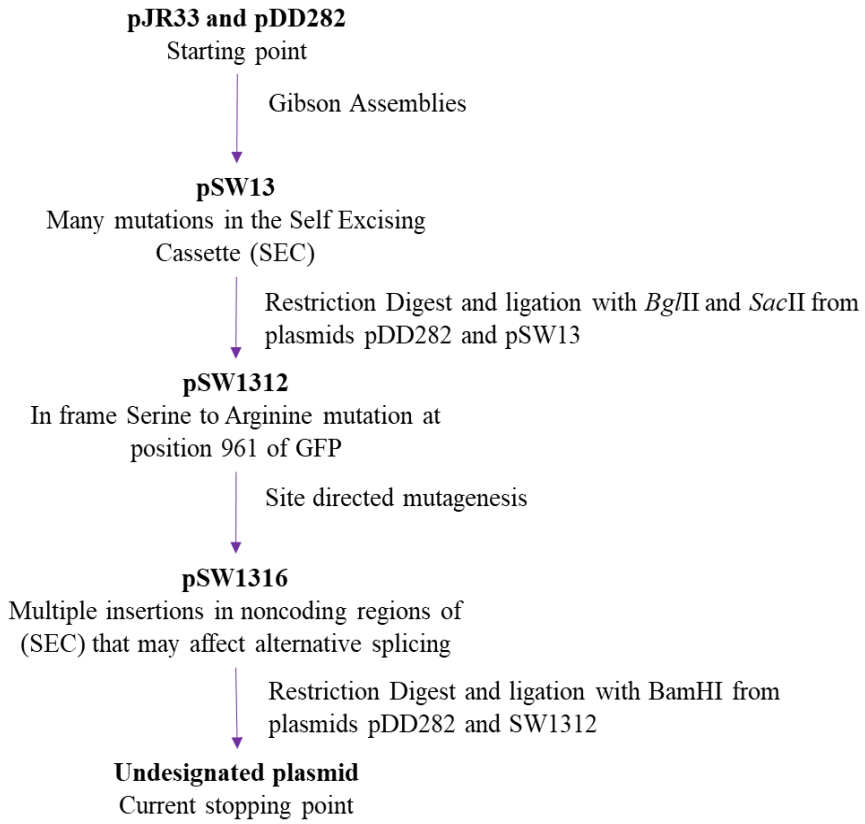


Figure 15. Flowchart of creating the nlg-1 knockout rescue plasmid. Described above is a general depiction of the overall plan followed in creating the null knockout rescue plasmid along with the unique challenges found within each step.

Following the initial completion of the pSW13 plasmid through PCR and Gibson assemblies, the Sanger sequencing data showed many different mutations (data not shown). Each of the following mutations was corrected in a later iteration of the rescue plasmid. One of the most problematic mutations that was fixed, was the in-frame serine to proline change at base pair position 961 of pSW13. Followed by a second, which was the missense mutation, of a histidine to arginine mutation at base pair position 6313, this change caused concern due to it unknown effects on protein folding. Another example of a mutation found was the point mutation within a splice site at base pair position 4264; a cytosine was mutated to a thymine. The number of

mutations within the self-excising cassette region resulted in developing a new protocol to helpfully minimize the mutations introduced.

To reduce the mutations in the rescue plasmid sequence, we decided that the plasmid was to be remade minimizing the number of PCR steps used. A restriction enzyme digest and ligation method were developed and used in following iterations of the null knockout rescue plasmids. The first attempt utilizing this technique involved the restriction enzymes *Bgl*II and *Sac*II and was named pSW1312 **(Figure 4)**. This plasmid contained a substantially cleaner self-excising cassette region, and both the point and missense mutations were both remedied. However, the plasmid still contained the serine to proline mutation at position 961from pSW13. As a result, the serine to proline mutation at position 961 was also further remedied by site-directed mutagenesis utilizing the NEB Q5 site-directed mutagenesis kit. The mutagenesis was successful in returning the proline to the rightful serine **(Figure 13)**. This iteration was named pSW13126. However, following sequencing, a tyrosine to cystine mutation at position 4310 was discovered. Thus, further modifications are needed.

The newest iteration is focused on removing the tyrosine to cystine mutation at position 4310. Due to its location within a coding region of GFP, we decided that it was of important that the sequence be free of mutations to ensure proper function of the fluorescence marker **(Figure 5)**. This iteration utilized *BamH*I as the restriction enzyme, there were two cut sites, which removed the need for a second restriction enzyme. Multiple attempts had been made to correctly make this plasmid; however, each of these trials failed. In some trials, no colonies would grow, in others, there would be colonies, however when analyzed though a gel, these colonies did not contain the new insert. Often, they were either empty backbone or nothing would show up on a gel. Multiple different routes were taken in attempts to solve the problem, including increasing

the concentration of insert used, utilizing different competent cells, and isolating plasmid DNA from new colonies grown from frozen stock. Prior to the stopping point it was decided to attempt to isolate high concentration DNA to work with from both pDD282 and pSW13126. This was completed by purifying larger quantities of liquid culture in one purification column. After multiple attempts, isolating DNA from colonies grown from frozen stock granted DNA of high concentrations that had not been previously seen, and when ran on an agarose gel the bands were super bright and had the correct fragment sizes **(Figure 14)**. This indicats that future attempts should be completed with the newest DNA isolation or with high, greater than 150 ng/μL concentration DNA. Creation of a successful pSW13126X will allow for the completion of the components of the CRISPR/Cas9 system as the two gRNAs are already made. The system, with the rescue plasmid would be microinjected into wild type *C. elegans* to produce the new neuroligin null model system.

Possible Reasons for Errors in pSW13 and pSW13126X

As mentioned above, sequencing pSW13, created using Gibson Assembly, revealed multiple mutations throughout the plasmid. After analysis of the sequence and steps utilized to create the plasmid, we decided that the mutations were introduced during the multiple PCR steps. In the following attempts that do not use PCR, no new mutations were introduced.

In pSW13126X, there are multiple possible reasons for the challenges seen in creating this plasmid. First, the pDD282 plasmid is unstable due to the temperature induced Cre and lox-p sites present within the sequence (Dickinson et al 2015). Thus, the depositing laboratory suggest that it should be grown at 30° C to maintain stability. Furthermore, the stock of pDD282 obtained from AddGene is only stable for approximately two weeks without being placed in the -

80° C freezer. Therefore, one should return to the original frozen stock plasmid anytime there may be concerns relating to growth quality or presence of the SEC insert.

Another issue that was faced in the construction of pSW13126X came from only utilizing *BamH*I as it created two identical ends, in which can be combined with the backbone in either orientation. Thus, a screening method prior to sequencing is needed. This method is a work in progress; however, it was decided that screening will either involve colony PCR with carefully selected primers designed to determine the orientation of the insert. Another challenge that was faced was when transformations were completed, to insert the plasmid DNA into competent cells, often no colonies would grow. It was suspected that there was not a large enough concentration of insert DNA from pDD282 that was used in the annealing reaction. Thus, not enough plasmids were created and used in the transformation despite multiple different concentration being attempted. This would explain why no cells grew as they were plated on ampicillin plates and the plasmids contained the ampicillin resistance gene. It was also suspected that the competent cells were just old and lost their effectiveness; however, when others in the lab, used the same supply of competent cells, their experiments worked, thus, proving this theory unlikely. Changing the type of competent cells was also tested to no avail. Following attempts with these modifications, it was determined that the likely cause of the issues was a combination of the instability of pDD282 and low concentrations of pDD282 insert DNA.

Future Directions

Despite the many attempts at creating a mutation-free *nlg*-1 complete null knockout, no iteration was completely mutation free. There were many instances where the desired GFP and self-excising cassette insert were not found when colonies were grown and purified from older existing cultures. Also, the purified plasmid DNA did not appear stable when frozen at -20°C,

samples that had the insert when initially purified (confirmed through gel electrophoresis; not shown), did not contain the insert weeks later. To increase the probability of success during the next trial, the pDD282 plasmid should be streaked from frozen stock (-80°C) every time it is to be used.

In the latest iteration pSW13126X, there were difficulties in the transformation and uptake of the plasmid into the competent *E. coli* cells (data not shown). One potential way to improve the transformation is to adjust the concentration of vector used in the annealing step within and outside of the kit's manufacturers recommended range. Since the insert is smaller (approximately 15%) compared to the backbone, a lower concentration would be needed. In the current concentrations there are far more molecules of insert than vector per nanogram, as such the number of annealed plasmids is limited by the concentration of vector. To achieve a higher concentration of DNA overall, the starting concentration must be higher. Some of the DNA is lost during both the restriction digest and gel purification steps. Thus, more colonies should be grown up overnight at 28°C from a frozen stock plate; multiples of these samples (maximum of 9 mL liquid culture) should be combined and purified together to increase the concentration isolated following a miniprep. Furthermore, during the gel purification step, an attempt should be made to recover the largest amount of the band with as little extra agarose gel as possible. This will increase the concentration of gel purified DNA as there would be less liquid for the DNA to be soluble in. The isolated gel piece should also be "mushed" and pulverized as best as possible before being added to the filter pipet tip, possibly increasing and amount of DNA fragments that can be recovered. When enough insert is isolated, the ligation step should include the addition of phosphatase of the vector to prevent the large ends of the backbone from re-ligating together without the presence of an insert. A restriction enzyme digest with *Hind*III would be completed

to distinguish between the two different orientations due to a change in the location of the restriction sites. If the insert is not reversed, there will be five bands with the lengths of 3032, 2973, 2246,1514, 889 bps. If the insert was reversed there will only be 3 bands with the lengths of 4648, 3032, 2973 bps.

Following the completion of the null knockout plasmid the next step would be to microinject the *C. elegans* with the CRISPR/Cas9 system plasmids. The *C. elegans* would be initially screened for the roller (sqt-1) phenotype, a *C. elegans* behavior where they move in circle instead of a straight line, and for the hygromycin resistance as this indicates successful incorporation of the self-excising cassette (Kramer et al. 1988, Park and Kramer 1994). Finally, the temperature at which the successful transgenic animals are growing would be raised to remove the self-existing cassette, thus the hygromycin resistance and the roller phenotype would no longer be present in these animals. Allowing for animals to only express the GFP and making this a full knockout. Fluorescent and behavioral testing would follow. Fluorescent testing would allow for identification of neurons and synapse which express neuroligin in the worm body (Calahorro and Ruiz-Rubio 2012). The behavior testing will identify possible phenotypic deficits associated with the missing neuroligin protein; some behavioral characteristics that may be tested include chemotaxis and thermotaxis.

Another direction to continue in the future would be to create a *C. elegans* model that expressed the human *NLGN*3 as mutations in this gene have been associated with ASD In the past (Jamain et al. 2003, Ulbrich et al 2016). Once the humanized model system is completed, site-directed mutagenesis would be completed to create a *C. elegans* model of the R451C and G221R mutations. The former mutations have been associated with autism in the literature; the latter is a mutation often used as a positive control of protein misfolding, but not associated with

autism (Jamain et al. 2003 and Ulbrich et al. 2016). To accomplish this, a synthetic gene component containing artificial *C. elegans* introns with the beginning of the human neuroligin gene. Introns were chosen to be included as in *C. elegans* express transgenic genes better if there are introns present (Mary Colasanto, InVivo Biosystems personal communication). This would be ligated to a neuroligin cDNA to create a full human gene that can be expressed within *C. elegans*. Creating these mutant model systems would allow for further identification and understanding of the phenotypic implications surrounding autism associated mutations.

Conclusions

Predicting the protein structure of nlg-1 has allowed for valuable understanding into possible reasons for the strains VC228's nlg-1 lower deficit phenotype. Through this analysis, we determined that VC228 is unable to form dimers nor is it predicted that this nlg-1 will bind with nrx-1. Despite this, question still remain, how does this strain not exhibit a more severe behavior deficit. Thus, the goal was to create a complete null knockout mutant strain to avoid confounding variables. Despite the challenges seen in each attempt, valuable information regarding future experimental processes were obtained. With each new iteration, the null knockout plasmid continues to move forward towards microinjection and the creation of transgenic animals.

Research into autism spectrum disorders has become a controversial topic within the last ten years as the medical model of disabilities has been rejected (Pellicano and den Houting 2022). However, with such limited knowledge regarding the role of genetics, environment, and other potential causes, work completed in this field would provide scientific and peer reviewed reasoning, to potential help individuals, parents, and caregivers understand that their choices and undermine the harmful misinformation that is so often seen among the public. Research may also

provide insight into helpful, long-term treatments to improve the quality of life of autistic individuals without ever curing the autism. The research completed in this study does not seek to cure autism through the use of CRISPR/Cas9 or another method, but rather to understand the effect and function of an autism associated protein.

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.

Arribere JA, Bell RT, Fu BXH, Artiles KL, Hartman PS, Fire AZ. 2014. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. Genetics. 198(3):837–846. doi:10.1534/genetics.114.169730. http://dx.doi.org/10.1534/genetics.114.169730.

Bang ML, Owczarek S. 2013. A matter of balance: role of neurexin and neuroligin at the synapse. Neurochem Res. 38(6):1174–1189. doi:10.1007/s11064-013-1029-9. http://dx.doi.org/10.1007/s11064-013-1029-9.

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. 2017. The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res. 45(D1):D313–D319. doi:10.1093/nar/gkw1132. http://dx.doi.org/10.1093/nar/gkw1132.

Bolliger MF, Frei K, Winterhalter KH, Gloor SM. 2001. Identification of a novel neuroligin in humans which binds to PSD-95 and has a widespread expression. Biochem J. 356(Pt 2):581–588. doi:10.1042/0264-6021:3560581. http://dx.doi.org/10.1042/0264-6021:3560581.

Boucard AA, Chubykin AA, Comoletti D, Taylor P, Südhof TC. 2005. A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. Neuron. 48(2):229–236. doi:10.1016/j.neuron.2005.08.026. http://dx.doi.org/10.1016/j.neuron.2005.08.026.

Budreck EC, Scheiffele P. 2007. Neuroligin-3 is a neuronal adhesion protein at GABAergic and glutamatergic synapses: Synaptic localization of neuroligin-3. Eur J Neurosci. 26(7):1738–1748.

Calahorro F. 2014. Conserved and divergent processing of neuroligin and neurexin genes: from the nematode C. elegans to human. Invert Neurosci. 14(2):79–90. doi:10.1007/s10158-014-0173-5. http://dx.doi.org/10.1007/s10158-014-0173-5.

Calahorro F, Holden-Dye L, O'Connor V. 2015. Analysis of splice variants for the C. elegans orthologue of human neuroligin reveals a developmentally regulated transcript. Gene Expr Patterns. 17(2):69–78. doi:10.1016/j.gep.2015.02.002. http://dx.doi.org/10.1016/j.gep.2015.02.002.

Carlson JS, Brinkman T, Majewicz-Hefley A. 2006. Medication treatment outcomes for school-aged children diagnosed with autism. Calif School Psychol. 11(1):21–30. doi:10.1007/bf03341112. http://dx.doi.org/10.1007/bf03341112.

CDCa. 2020 Sep 25. Data & statistics on autism spectrum disorder. Cdc.gov. [accessed 2020 Oct 13].https://www.cdc.gov/ncbddd/autism/data.html.

CDCb. Signs and symptoms of autism spectrum disorders. Cdc.gov. 2020 Aug 12 [accessed 2020 Oct 13].https://www.cdc.gov/ncbddd/autism/signs.html

C. elegans Deletion Mutant Consortium. 2012. large-scale screening for targeted knockouts in the Caenorhabditis elegans genome. G3 (Bethesda). 2(11):1415–1425. doi:10.1534/g3.112.003830. http://dx.doi.org/10.1534/g3.112.003830.

Chen X, Liu H, Shim AHR, Focia PJ, He X. 2008. Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions. Nat Struct Mol Biol. 15(1):50–56. doi:10.1038/nsmb1350. http://dx.doi.org/10.1038/nsmb1350.

Choii G, Ko J. 2015. Gephyrin: a central GABAergic synapse organizer. Exp Mol Med. 47(4):e158. doi:10.1038/emm.2015.5. http://dx.doi.org/10.1038/emm.2015.5.

Corsi AK, Wightman B, Chalfie M. 2015. A Transparent Window into Biology: A Primer on Caenorhabditis elegans. Genetics. 200(2):387–407.

Dean C, Scholl FG, Choih J, DeMaria S, Berger J, Isacoff E, Scheiffele P. 2003. Neurexin mediates the assembly of presynaptic terminals. Nat Neurosci. 6(7):708–716. doi:10.1038/nn1074. http://dx.doi.org/10.1038/nn1074.

Dean C, Dresbach T. 2006. Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. Trends Neurosci. 29(1):21–29. doi:10.1016/j.tins.2005.11.003. http://dx.doi.org/10.1016/j.tins.2005.11.003.

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 HonorsThesis) Drew Unviersity, Madison, NJ.

De Jaco A, Lin MZ, Dubi N, Comoletti D, Miller MT, Camp S, Ellisman M, Butko MT, Tsien RY, Taylor P. 2010. Neuroligin trafficking deficiencies arising from mutations in the alpha/beta-hydrolase fold protein family. J Biol Chem. 285(37):28674–28682.

Dickinson DJ, Pani AM, Heppert JK, Higgins CD, Goldstein B. 2015. Streamlined genome\ engineering with a self-excising drug selection cassette. Genetics. 200(4):1035–1049. doi:10.1534/genetics.115.178335. http://dx.doi.org/10.1534/genetics.115.178335.

Doherty M, Haydon C, Davidson IA. 2021. Recognising autism in healthcare. Br J Hosp Med (Lond). 82(12):1–7. doi:10.12968/hmed.2021.0313. http://dx.doi.org/10.12968/hmed.2021.0313.

Doudna JA, Charpentier E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. 346(6213):1258096. doi:10.1126/science.1258096. http://dx.doi.org/10.1126/science.1258096.

Duesenberg MD, Burns MK. 2022. Autism spectrum disorder identification in schools: Impact of criteria, assessments, and student data for identification decisions. Psychol Sch. 59(4):845–865. doi:10.1002/pits.22649. http://dx.doi.org/10.1002/pits.22649.

Etherton M, Földy C, Sharma M, Tabuchi K, Liu X, Shamloo M, Malenka RC, Südhof TC. 2011a. Autism linked neuroligin-3 R451C mutation differentially alters hippocampal and cortical synaptic function. Proc Natl Acad Sci U S A. 108(33):13764–13769.

Etherton MR, Tabuchi K, Sharma M, Ko J, Südhof TC. 2011b. An autism-associated point mutation in the neuroligin cytoplasmic tail selectively impairs AMPA receptor-mediated synaptic transmission in hippocampus: Autism-associated neuroligin mutation. EMBO J. 30(14):2908–2919. doi:10.1038/emboj.2011.182. http://dx.doi.org/10.1038/emboj.2011.182.

Fabrichny IP, Leone P, Sulzenbacher G, Comoletti D, Miller MT, Taylor P, Bourne Y, Marchot P. 2007. Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion. Neuron. 56(6):979–991. doi:10.1016/j.neuron.2007.11.013. http://dx.doi.org/10.1016/j.neuron.2007.11.013.

Fu Z, Vicini S. 2009. Neuroligin-2 accelerates GABAergic synapse maturation in cerebellar granule cells. Mol Cell Neurosci. 42(1):45–55. doi:10.1016/j.mcn.2009.05.004. http://dx.doi.org/10.1016/j.mcn.2009.05.004.

Grinker RR. 2015. Reframing the science and anthropology of autism. Cult Med Psychiatry. 39(2):345–350. doi:10.1007/s11013-015-9444-9. http://dx.doi.org/10.1007/s11013-015-9444-9.

Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis. 30 Suppl 1(S1):S162-73. doi:10.1002/elps.200900140. http://dx.doi.org/10.1002/elps.200900140.

Hayes J, Ford T, McCabe R, Russell G. 2022. Autism diagnosis as a social process. Autism. 26(2):488–498. doi:10.1177/13623613211030392. http://dx.doi.org/10.1177/13623613211030392.

Hodges H, Fealko C, Soares N. 2020. Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. Transl Pediatr. 9(Suppl 1):S55–S65. doi:10.21037/tp.2019.09.09. http://dx.doi.org/10.21037/tp.2019.09.09.

Huang J, Liu J, Tian R, Liu K, Zhuang P, Sherman HT, Budjan C, Fong M, Jeong M-S, Kong X-J. 2022. A next generation sequencing-based protocol for screening of variants of concern in autism spectrum disorder. Cells. 11(1):10. doi:10.3390/cells11010010. http://dx.doi.org/10.3390/cells11010010.

Huang Y, Arnold SR, Foley K-R, Trollor JN. 2020. Diagnosis of autism in adulthood: A scoping review. Autism. 24(6):1311–1327. doi:10.1177/1362361320903128. http://dx.doi.org/10.1177/1362361320903128.

Hughes JA. 2021. Does the heterogeneity of autism undermine the neurodiversity paradigm? Bioethics. 35(1):47–60. doi:10.1111/bioe.12780. http://dx.doi.org/10.1111/bioe.12780.

Hunter JW, Mullen GP, McManus JR, Heatherly JM, Duke A, Rand JB. 2010. Neuroligin deficient mutants of C. elegans have sensory processing deficits and are hypersensitive to oxidative stress and mercury toxicity. Dis Model Mech. 3(5–6):366–376. doi:10.1242/dmm.003442. http://dx.doi.org/10.1242/dmm.003442.

Hyman SL, Levy SE, Myers SM, council on children with disabilities, Section on developmental and behavioral pediatrics. 2020. Identification, evaluation, and management of children with autism spectrum disorder. Pediatrics. 145(1):e20193447.

Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Missler M, Moomaw C, Südhof TC. 1995. Neuroligin 1: A splice site-specific ligand for β-neurexins. Cell. 81(3):435–443. doi:10.1016/0092-8674(95)90396-8. http://dx.doi.org/10.1016/0092-8674(95)90396-8.

Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, et al. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet. 34(1):27–29. doi:10.1038/ng1136. http://dx.doi.org/10.1038/ng1136.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol. 31(3):233–239. doi:10.1038/nbt.2508. http://dx.doi.org/10.1038/nbt.2508.

Ko J, Zhang C, Arac D, Boucard AA, Brunger AT, Südhof TC. 2009. Neuroligin-1 performs neurexin-dependent and neurexin-independent functions in synapse validation. EMBO J. 28(20):3244–3255. doi:10.1038/emboj.2009.249. http://dx.doi.org/10.1038/emboj.2009.249.

Koehnke J, Katsamba PS, Ahlsen G, Bahna F, Vendome J, Honig B, Shapiro L, Jin X. 2010. Splice form dependence of beta-neurexin/neuroligin binding interactions. Neuron. 67(1):61–74. doi:10.1016/j.neuron.2010.06.001. http://dx.doi.org/10.1016/j.neuron.2010.06.001.

Kramer JM, Johnson JJ, Edgar RS, Basch C, Roberts S. 1988. The sqt-1 gene of C. elegans encodes a collagen critical for organismal morphogenesis. Cell. 55(4):555–565. doi:10.1016/0092-8674(88)90214-0. http://dx.doi.org/10.1016/0092-8674(88)90214-0.

Leadbitter K, Buckle KL, Ellis C, Dekker M. 2021. Autistic self-advocacy and the neurodiversity movement: Implications for autism early intervention research and practice. Front Psychol. 12:635690. doi:10.3389/fpsyg.2021.635690. http://dx.doi.org/10.3389/fpsyg.2021.635690.

Lee H-J, Zheng JJ. 2010. PDZ domains and their binding partners: structure, specificity, and modification. Cell Commun Signal. 8(1):8. doi:10.1186/1478-811X-8-8. http://dx.doi.org/10.1186/1478-811X-8-8.

Lisé M-F, El-Husseini A. 2006. The neuroligin and neurexin families: from structure to function at the synapse. Cell Mol Life Sci. 63(16):1833–1849. doi:10.1007/s00018-006-6061-3. http://dx.doi.org/10.1007/s00018-006-6061-3.

Marro SG, Chanda S, Yang N, Janas JA, Valperga G, Trotter J, Zhou B, Merrill S, Yousif I, Shelby H, et al.2019. Neuroligin-4 regulates excitatory synaptic transmission in human neurons. Neuron. 103(4):617-626.e6.

Mayes SD, Waxmonsky JG, Baweja R, Mattison RE, Memon H, Klein M, Hameed U, Waschbusch D. 2020. Symptom scores and medication treatment patterns in children with ADHD versus autism. Psychiatry Res. 288(112937):112937. doi:10.1016/j.psychres.2020.112937. http://dx.doi.org/10.1016/j.psychres.2020.112937.

Merriam-Webster. 2021. Neurodiversity Definition. Merriam-Webster Dictionary. [accessed 2021 Dec 13]. https://www.merriam-webster.com/dictionary/neurodiversity

Miller MT, Mileni M, Comoletti D, Stevens RC, Harel M, Taylor P. 2011. The crystal structure of the α-neurexin-1 extracellular region reveals a hinge point for mediating synaptic adhesion and function. Structure. 19(6):767–778. doi:10.1016/j.str.2011.03.011. http://dx.doi.org/10.1016/j.str.2011.03.011.

Nakanishi M, Nomura J, Ji X, Tamada K, Arai T, Takahashi E, Bućan M, Takumi T. 2017. Functional significance of rare neuroligin 1 variants found in autism. PLoS Genet. 13(8):e1006940. doi:10.1371/journal.pgen.1006940. http://dx.doi.org/10.1371/journal.pgen.1006940.

Nguyen TA, Wu K, Pandey S, Lehr AW, Li Y, Bemben MA, Badger JD 2nd, Lauzon JL, Wang T, Zaghloul KA, et al. 2020. A cluster of autism-associated variants on X-linked NLGN4X functionally resemble NLGN4Y. Neuron. 106(5):759-768.e7. doi:10.1016/j.neuron.2020.03.008. http://dx.doi.org/10.1016/j.neuron.2020.03.008.

nlg-1 (gene) - WormBase : Nematode Information Resource. Wormbase.org. [accessed 2022 Jan 3]. https://wormbase.org/species/c_elegans/gene/WBGene00006412.

Niedźwiecka A, Pisula E. 2022. Symptoms of autism spectrum disorders measured by the qualitative checklist for autism in toddlers in a large sample of Polish toddlers. Int J Environ Res Public Health. 19(5):3072. doi:10.3390/ijerph19053072. http://dx.doi.org/10.3390/ijerph19053072.

Nih.gov. Autism spectrum disorder. [accessed 2020 Oct 14]. https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml.

Park Y-S, Kramer JM. 1994. The C. elegans sqt-1 and rol-6 collagen genes are coordinately expressed during development, but not at all stages that display mutant phenotypes. Dev Biol. 163(1):112–124. doi:10.1006/dbio.1994.1127. http://dx.doi.org/10.1006/dbio.1994.1127.

Patel, KS. (2021). Modeling and manipulating mutant neurexins in *Caenorhabditis elegans* (Undergraduate Honors Thesis) Drew University, Madison, NJ.

Patharkar R. 2019. DNA Gel Purification - Freeze & Squeeze Method. https://www.youtube.com/watch?v=b3wfhofZJSI.

Papadopoulos T, Soykan T. 2011. The role of collybistin in gephyrin clustering at inhibitory synapses: facts and open questions. Front Cell Neurosci. 5:11. doi:10.3389/fncel.2011.00011. http://dx.doi.org/10.3389/fncel.2011.00011.

Pellicano E, den Houting J. 2022. Annual Research Review: Shifting from "normal science" to neurodiversity in autism science. J Child Psychol Psychiatry. 63(4):381–396. doi:10.1111/jcpp.13534. http://dx.doi.org/10.1111/jcpp.13534.

Poulopoulos A, Aramuni G, Meyer G, Soykan T, Hoon M, Papadopoulos T, Zhang M, Paarmann I, Fuchs C, Harvey K, et al. 2009. Neuroligin 2 drives postsynaptic assembly at perisomatic inhibitory synapses through gephyrin and collybistin. Neuron. 63(5):628–642. doi:10.1016/j.neuron.2009.08.023. http://dx.doi.org/10.1016/j.neuron.2009.08.023.

Poulopoulos A, Soykan T, Tuffy LP, Hammer M, Varoqueaux F, Brose N. 2012. Homodimerization and isoform-specific heterodimerization of neuroligins. Biochem J. 446(2):321–330. doi:10.1042/BJ20120808. http://dx.doi.org/10.1042/BJ20120808.

Sandoval-Norton AH, Shkedy G, Shkedy D. 2021. Long-term ABA therapy is abusive: A response to gorycki, Ruppel, and Zane. Adv Neurodev Disord. 5(2):126–134. doi:10.1007/s41252-021-00201-1. http://dx.doi.org/10.1007/s41252-021-00201-1.

Santhanam SP, Bellon-Harn ML. 2022. Speech-language pathologist's role in understanding and promoting self-advocacy in autistic adults. Am J Speech Lang Pathol. 31(2):649–663. doi:10.1044/2021_AJSLP-21-00223. http://dx.doi.org/10.1044/2021_AJSLP-21-00223.

Shen C, Huo L-R, Zhao X-L, Wang P-R, Zhong N. 2015. Novel interactive partners of neuroligin 3: new aspects for pathogenesis of autism. J Mol Neurosci. 56(1):89–101.

Shi L, Chang X, Zhang P, Coba MP, Lu W, Wang K. 2013. The functional genetic link of NLGN4X knockdown and neurodevelopment in neural stem cells. Hum Mol Genet. 22(18):3749–3760. doi:10.1093/hmg/ddt226. http://dx.doi.org/10.1093/hmg/ddt226.

Shipman SL, Nicoll RA. 2012. Dimerization of postsynaptic neuroligin drives synaptic assembly via transsynaptic clustering of neurexin. Proc Natl Acad Sci U S A. 109(47):19432–19437. doi:10.1073/pnas.1217633109. http://dx.doi.org/10.1073/pnas.1217633109.

Singh V, Braddick D, Dhar PK. 2017. Exploring the potential of genome editing CRISPR-Cas9 technology. Gene. 599:1–18. doi:10.1016/j.gene.2016.11.008. http://dx.doi.org/10.1016/j.gene.2016.11.008.

Song, J.-Y., Ichtchenko, K., Südhof, T. C. & Brose, N. 1999. Neuroligin 1 is a postsynaptic cell Adhesion molecule of excitatory synapses. Proc. Natl Acad. Sci. USA 96: 1100–1125.

Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. 2020. QMEANDisCo-distance constraints applied on model quality estimation. Bioinformatics. 36(6):1765–1771. doi:10.1093/bioinformatics/btz828. http://dx.doi.org/10.1093/bioinformatics/btz828.

Südhof TC. 2008. Neuroligins and neurexins link synaptic function to cognitive disease. Nature. 455(7215):903–911.

Tsetsenis T, Boucard AA, Araç D, Brunger AT, Südhof TC. 2014. Direct visualization of trans-synaptic neurexin-neuroligin interactions during synapse formation. J Neurosci. 34(45):15083–15096. doi:10.1523/JNEUROSCI.0348-14.2014. http://dx.doi.org/10.1523/JNEUROSCI.0348-14.2014.

Uchigashima M, Cheung A, Futai K. 2021. Neuroligin-3: A circuit-specific synapse organizer that shapes normal function and autism spectrum disorder-associated dysfunction. Front Mol Neurosci. 14:749164. doi:10.3389/fnmol.2021.749164. http://dx.doi.org/10.3389/fnmol.2021.749164.

Ulbrich L, Favaloro FL, Trobiani L, Marchetti V, Patel V, Pascucci T, Comoletti D, Marciniak SJ, De Jaco A. 2016. Autism-associated R451C mutation in neuroligin3 leads to activation of the unfolded protein response in a PC12 Tet-On inducible system. Biochem J. 473(4):423–434.

Ushkaryov YA, Hata Y, Ichtchenko K, Moomaw C, Afendis S, Slaughter CA, Südhof TC. 1994. Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences in receptor-like neuronal cell-surface proteins. J Biol Chem. 269(16):11987–11992. doi:10.1016/s0021-9258(17)32671-6. http://dx.doi.org/10.1016/s0021-9258(17)32671-6.

Ushkaryov YA, Petrenko AG, Geppert M, Südhof TC. 1992. Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. Science. 257(5066):50–56. doi:10.1126/science.1621094. http://dx.doi.org/10.1126/science.1621094.

Varoqueaux F, Aramuni G, Rawson RL, Mohrmann R, Missler M, Gottmann K, Zhang W, Südhof TC, Brose N. 2006. Neuroligins determine synapse maturation and function. Neuron. 51(6):741–754.

Varoqueaux F, Jamain S, Brose N. 2004. Neuroligin 2 is exclusively localized to inhibitory synapses. Eur J Cell Biol. 83(9):449–456.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1):W296–W303. doi:10.1093/nar/gky427. http://dx.doi.org/10.1093/nar/gky427.

Xu Jing, Du Y-L, Xu J-W, Hu X-G, Gu L-F, Li X-M, Hu P-H, Liao T-L, Xia Q-Q, Sun Q, et al. 2019. Neuroligin 3regulates dendritic outgrowth by modulating Akt/mTOR signaling. Front Cell Neurosci. 13:518.

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).

Yoshida T, Yamagata A, Imai A, Kim J, Izumi H, Nakashima S, Shiroshima T, Maeda A, Iwasawa-Okamoto S, Azechi K, et al. 2021. Canonical versus non-canonical

transsynaptic signaling of neuroligin 3 tunes development of sociality in mice. Nat Commun. 12(1):1848. doi:10.1038/s41467-021-22059-6. http://dx.doi.org/10.1038/s41467-021-22059-6.

Zerbini F, Zanella I, Fraccascia D, König E, Irene C, Frattini LF, Tomasi M, Fantappiè L, Ganfini L, Caproni E, et al. 2017. Large scale validation of an efficient CRISPR/Cas-based multi gene editing protocol in Escherichia coli. Microb Cell Fact. 16(1):68. doi:10.1186/s12934-017-0681-1. http://dx.doi.org/10.1186/s12934-017-0681-1.

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

Appendix 1. Non-Kit Protocols

The following are protocols for procedures used not obtained from commercially available kits:

PCR Set Up

- 25 µL New England BioLabs LongAmp Hot Start Taq 2X Master Mix
- 2.5 µL 5 µM Forward primer
- 2.5 µL 5 µM Reverse primer
- 1 µL DNA
- 19 µL Nuclease free water

Follow appropriate annealing temperature for primers (NEB TM calculator)

NEB Thermocycling program: (final holding temperature 4°C)

Initial denaturization	94°C	30 seconds

30 cycles	94°C	15-30 seconds
	45-65 °C (annealing temperature)	60 seconds
	65°C	50 seconds/Kilo base (kb)
Final extension	65°C	10 minutes

Restriction Enzyme Digest

- 20 µL Reaction
    - 2 µL NEB CutSmart Buffer (1x) or another appropriate buffer
    - 5 µL mini prep DNA
    - 2 µL Restriction enzyme
        - was reduced to 1 µL of each enzyme if more than one restriction enzyme was needed
    - 11 µL Nuclease free water

May scale this reaction depending on quantity of digested DNA needed

Gel Sample Preparation:

- 12 µL Uncut sample:
    - 5 µL DNA sample
    - 5 µL Tris-HCLl or TE or nuclease free water
    - 2 µL 6X Loading dye
- 24µL Digest sample
    - 20 µL Digest sample

- 4 μL 6X Loading dye

Agarose Gel:

1. Make 800 mL of 1X Tris-acetate-EDTA (TAE) (80 mL 10X TAE and 720 mL distilled water)
2. Weigh out appropriate amount of agarose powder
    a. Measurements for 100 ml gel apparatus:
        i. 0.8% gel= 0.8 gram (g) agarose powder
        ii. 1X gel= 1.0 g agarose powder
3. Measure out appropriate amounts of 1X TAE
    a. Small gel uses 50 ml of 1X TAE
    b. Large gel uses 100 mL 1X TAE
4. Combine and heat until the solution is clear with no granules remaining
5. Let cool until tolerable to hold without gloves or heat protection
6. Add 2 μL, 10 mg/ml ethidium bromide (EtBr) for every 100 mL of agarose gel
7. Pour into pre prepared gel casting apparatus and wait for it to solidify
8. Remove gel from casting apparatus and place into running apparatus, cover with enough 1X TAE that the tops of wells can no longer be seen
9. Add EtBr (10 mg/mL) to buffer in the same 2 μL/100 mL ratio
10. Load samples w/ 5 μL NEB 1kb Plus DNA ladder on the outside lanes of sample on gel
11. Run gel 80 volts/2 amps/ 120 min (adjust time as needed)

"Freeze and Squeeze" Agarose Gel DNA Purification (Patharkar 2019)

- Run desired DNA on an agarose gel
- Cut a large 1 mL pipet tip with a filter to fit in a 1.5 mL microcentrifuge tube and place inside
- Cut the desired fragment out of gel with as little excess gel as possible
- Place the fragment within the cut pipette tip, chopping the gel fragment into small pieces
- Place the sample in -80°C freezer for 5 minutes
- Following freezing, immediately spin in microcentrifuge for 3 minutes at maximum speed
- Purified DNA will be in the flow through; discard filter.

Gibson Assembly

- 2x Annealing Buffer (makes 10 milliliters)
    - 200 μL 1M Tris
    - 160 μL 125mM EDTA
    - 58.44 mg NaCl
    - Nuclease free water to 10 mL
- Assembly Reaction
    - 10 μL 2X annealing buffer
    - 2 μL small fragment
    - 1 μL large fragment
    - 7 μL nuclease free water
    - Incubate 15 minutes at 50°C

Ethanol Precipitation:

Section 1:

- 7 µL PCR sequencing reaction
- 1.5 µL Exonu*
- 1.5 µL Alkphos*
    1. Incubate 15 minutes at 37°C
    2. Incubate 15 minutes at 80°C

*Sold together as ExProStar by Cytiva

Section 2:

1. Transfer 8 µL into 1.5 mL microcentrifuge tube
2. Add
    a. 1 µL of 20µg/mL glycogen
    b. 2.25 µL 125 mM EDTA, pH 8.0
    c. 27 µL 95% EtOH
3. Vortex to mix
4. Incubate at room temperature for 20 minutes
5. Spin in 18 °C microcentrifuge for 20 minutes
6. Carefully aspirate off supernatant with pipette
7. Add 250 µL 70% EtOH vortex 20 seconds
8. Spin in 18 °C microcentrifuge for 10 minutes
9. Carefully aspirate off supernatant with pipette
10. Repeat steps 7-9

11. Place samples in speed vac to dry for 15 minutes under both pressure and temperature (caps open)
12. Add 20 µL HiDI (highly deionized) formamide and vortex for 30-60 seconds
13. Transpher to sequencer tubes and load into sequencer

Appendix 2. Sequences for pSW13

Introns

Introns within sequences

5'F

GFP

SEC

LoxP

Sqt-1

Cre

Hygr

LoxP

3"F

CTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCAT
AGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACA
TACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG
AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG
GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGA
CTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTC
AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCAT
CACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACT
ATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG
GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCG

GTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTT
CAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
CCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTG
GTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC
CGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTT
TTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGAT
TTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAA
TTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGAT
CTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA
CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAG
CCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGC
CTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTC
GCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGT
GGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAA
CGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTT
AGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTG
TTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGC
CATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA
TACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCA
TTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGT
TGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGC
AAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA
CTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAA
TAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAG
AAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGA
GGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGAC
ACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCG
GGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGT
CGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
CACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA
CCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGG
GCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGG
ATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTC
ACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGG
CCAAGCTGGGATAAACACACTCGAAAGTTGAAAATTGAAAATTGTTTT
AATTTTTCTCTTGAAAGAAATCTGATAATTTTTTAATTCAATTTCTCCA
ATTGTTTTGGTTTTTTTCACTTGAGCTTTGTAGACATATATGTGTCAAC
ATCAAATGTTTGTTACATATTTCAAAAGCCAACTTAATAAAACTAAAA

AAAAACCGAAAATGAAAACGATCTATTAATGTTTTTTTTGCAGAATTTT
TCTTTCAGAATGCAAAAAATTATTTGAACGTGTCAGATTCAAATTAAG
ATGCTTATATTCAACTTCAATCAAAAAAGCTTAACATTTTAACAAAAC
CTTTTTATATTCAATATTATTTTCCTACTTCCTATCACATTTCAAAATG
TCCATCTCATTGCAACTCGACAAGTTTTGTGCAGATTGAGACAATTGA
CCGGACAACAATTTCCTATTGCTGGACCCCCTGTTGCCCCCCAAATGA
TGACATTACCTCTTTTCCTCCCTTCTACCCGCCCTTCTTGTCTTCTTCTT
CTTCTTTTTCGGGATTTTTCTCTACAACGATCATCATTTTTTGTCATGC
AGCTAGGTTGTTGCATCTGCTTCCAAGATAACAAACCTAATTTTTGTC
ATGCGAAACAATTTACTCGTTTCATGTTTTTTTTCCTTTTGACTTCTTC
AAAACTATCTTATGCTTTTATTTTATAGTCGTATATTATTTATTTATGA
CTCAACTCAAAATATGACAATTTTAATGATGCCGATTACTTCTAAACT
TATTGCCAAAATTATTATATGTGTGTTTTCAGATACTTTCCATTCGAAA
AACAACGAAAATGGCGGAAAGCAGAGAATCATGCGAATTTGGAAGTG
AACAGGCATGAGTAAAGGAGAAGAATTGTTCACTGGAGTTGTCCCAA
TCCTCGTCGAGCTCGACGGAGACGTCAACGGACACAAGTTCTCCGTCT
CCGGAGAGGGAGAGGGAGACGCCACCTACGGAAAGCTCACCCTCAAG
TTCATCTGCACCACCGGAAAGCTCCCAGTCCCATGGCCAACCCTCGTC
ACCACCTTCTGCTACGGAGTCCAATGCTTCTCCCGTTACCCAGACCAC
ATGAAGCGTCACGACTTCTTCAAGTCCGCCATGCCAGAGGGATACGTC
CAAGAGCGTACCATCTTCTTCAAGgtaagtttaaacatatatatactaactactgattattta
aattttcaggaCGACGGAAACTACAAGACCCGTGCCGAGGTCAAGTTCGAG
GGAGACACCCTCGTCAACCGTATCGAGCTCAAGgtaagtttaaacagttcggtact
aactaaccatacatatttaaattttcagGGAATCGACTTCAAGGAGGACGGAAACATC
CTCGGACACAAGCTCGAGTACAACTACAACTCCCACAACGTCTACATC
ATGGCCGACAAGCAAAAGAACGGAATCAAGGTCAACTTCAAGgtaagttt
aaacatgattttactaactaactaatctgatttaaattttcagATCCGTCACAACATCGAGGAC
GGATCCGTCCAACTCGCCGACCACTACCAACAAAACACCCCAATCGG
AGACGGACCAGTCCTCCTCCCAGACAACCACTACCTCTCCACCCAATC
CGCCCTCTCCAAGGACCCAAACGAGAAGCGTGACCACATGGTCCTCCT
CGAGTTCGTCACCGCCGCCGGAATCACCCACGGAATGGACGAGCTCTA
CAAGGAGAATCTGTACTTTCAATCCGGAAAGGTAAGTTTAAAATAACT
TCGTATAGCATACATTATACGAAGTTATTTTCAGGGAGAGCAAAAGCT
TATCTCCGAGGAGGACCTCTAAGGATGATCGACGCCAACGTCGTTGAA
TTTTCAAATTTTAAATACTGAATATTTGTTTTTTTTCCTATTATTTATTT
ATTCTCTTTGTGTTTTTTTTTCTTGCTTTCTAAAAAATTAATTCAATCCA
AATCTAAACATTTTTTTTTCTCTTTCCGTCTCCCAATTCGTATTCCGCTC
CTCTCATCTGAACACAATGTGCAAGTTTATTTATCTTCTCGCTTTCATT
TCATTAGGACGTGGGGGGAATTGGTGGAAGGGGGAAACACACAAAAG
GATGATGGAAATGAAATAAGGACACACAATATGCAACAACATTCAAT
TCAGAAATATGGAGGAAGGTTTAAAAGAAAACATAAAAATATATAGA
GGAGGAAGGAAACTAGCTGCCCTCGATTGTCTATCTTCCATTCTCACC
AAGTTTCAACAACTTTATTCAGTCATTCTTCTCTGGTTCATTTCCCAAC
CAATCTTCCCGAATAAAGCCCATCTGTCCTTTCGGCCACATTCCTCACC
TACATCCGTACTCCTTATCTCCCGTTCTTATCCAGTCTTTTATCGTCTCT
GCGTCTCGCACACTGTCTCGTCGGTAATGTAAAAAAAAAACAAAAAA

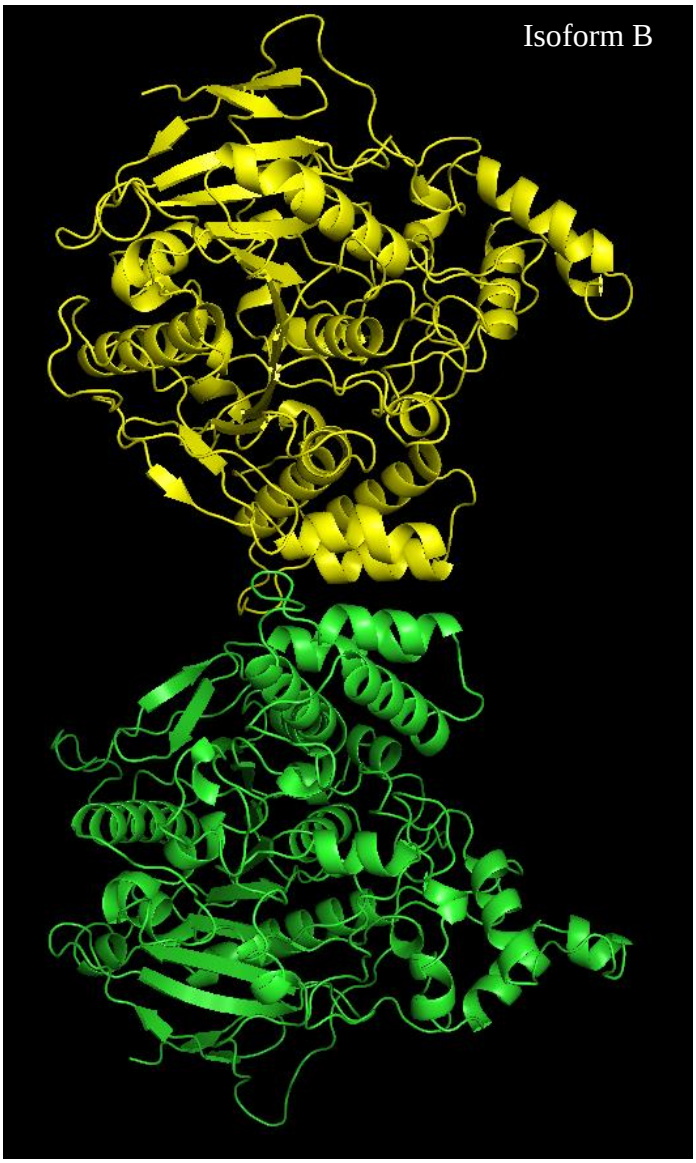
AACTTGAAAACTTGCGGAAAGATTTCCTAATAAAAGGAGACGAATCC
GGTAGGAAATCGACAAGTCTCTGCTTCTCGAAGACTAGGTGAGTTACA
TTTCTTTGGCCGAATTATTTTGAAATCCTAAATTTGATAATTTCAGAGA
TGTCTGTAAAACTTGCGTGTTATGTGACGGCTTCGGTCACTGTCGCCA
CTCTTATGGTATGCTTCATGACCATGTCAACCATCTACTCGGAGGTTG
ATGGATTCAGAGAGAAGCTTGATACCGAGATGAATGTGTTCAGAgtgag
ttaaattatattttgaattttaaataatttttaattttctagCAATCTACCAATGGATTGTGGAA
GGACATAGTTGTCATCGGAAGATCTAGCAAGCGTGTCCGTTGTCAATA
TGAAGAGACCAACGCTACCCCAACTCCACATGCTGATGGATCCCCATC
TGCTCCACCAGGTCAACCACCAGCAGTTCCACCAGTCTTCAACCAGCC
AAAGACTCCAAATGGAGCCAATGGAAATGGACCAACCTGCAACTGCA
ATGCTGATAACAAGTGCCCAGCTGGACCATCCGGACCAAAGGGAGTT
CCAGGAGTTCCAGGACTCGACGGAGTTCCAGGACTTGACGGTGTTCCA
GGAGTTGGAGCTGATGATATCGCTCCACAACGCGAGTCTGTCGGATGC
TTCACTTGCCCACAAGGACCAGTTGGACCACCAGGAGCTCTTGGAAGA
CCAGGACCACGTGGACTTCCAGGACCAAGAGGACAAAATGGAAACCC
AGGAAGAGATGGACAACCAGGACATCCAGGAGAGCAAGGATCATCCG
GTCAAATCGGAAAGATCGGAGAGCCAGGACCACCAGGAGAGAAGGGA
CGCGACGCCGAGCATCCAATCGGAAGACCAGGACCAAAGGGACCAAG
AGGAGATCAAGGACCAACAGGACCAGCTGGACAGAACGGTCTTCACG
GACCACCAGGAGAGCCAGGAACCGTTGGACCAGAAGGACCATCTGGA
AAGCAAGGACGTCAAGGACCAGACGGAACCCAGGGAGAGACTGGACC
AGACGGAAGACCAGGAAAGGATGCCGAGTACTGCCAGTGCCCAGACA
AGTCTCCACCATCAGAGGCTGTCAACGCCAACCGTGGATACAGAAAT
ATCTAAATTGTTGGTGTTTTCTAATAAAATATTTGAGATTCAGTGAGA
CTTTTTTCTTGGCGGCACAATAAAGTTTCCTCTTTTTGTTGGGTCACTC
AACGAACACAATACAAATATTTGAGCGCTCGCATTTTTCTCAACCTTT
CGGGTTTAAAGCACTTCAAAATAGAATTGAAGAGAGATTGTGAAAGT
ACATCAGACAGAATCACATTTTGATAGCATTCACTTCACTCAGATGCA
AACTTTTCAGTGACTATAAAAAGTCATAAAATAATAACATTAAAAAAG
CAAATAAATTAGCGAGAGAATTTTTTGACAAAAAGAAAGAAGAGTGA
TAGAGAAGAAGGGAATGCTTGAAAGGATTTTGCATTTATCTTAATCAC
CGTCTTCGAGAAGACGAACCATGGCGCCAGTTTCAGAATCCAAGTTTC
TAATATAGTTCATCACAATGTTGACGTTGGTCCAACCTCCGGCTTGCA
TAATCTCTGGGATGGAAACTCCAGCGCGAGCCATGTCACGGGCAGCTC
CGACACGGGCGGAATGTCCAGACCAAGCCAAGTAACGTTGTCCGGAA
TCATCCTTGGCTCCGTAGATGAGACGGTGGGTAGCTTCGAAGATTcctga
aaatttattttcagcttattcttagctaatatgttgtttaaacttaCCTTCAAGAGCTCGAGTCGA
GAGTTGTGATGTTGCACTCGGGGCAGCAACTCCGTTCTTGCGAACACG
ACAGAACAGGTAGTTGTTTGGATCATCGGCGACTCCGGAAACCGAAAT
CCAGCGCTCAACGAGCTTGGTAACTCCAAGGGAGAGCGCTTTTTCGAC
TCCGGCAGTCGAGACATCACCCAGCGTTTTTGTACGTCCGATGTGGAT
GAGCATTCTTCCTCCGTCTGTGCGACTAATATCCTTGACGCGAATGCG
GGCAATTTCGGCGATACGGAGAAGTGTGTTGTAAGCAATTCCGAGGA
AGGCGAGGTTTCGGATGTCCTGGCAACGATCGGAATTTTCCATAAGAG
AACGGACCTGATCGAAGTCGGTtctgaaaatttcgtcaagtcttaggattgaatacattaat

atttaaacttaCCTTTCGAAGGCGAGGGCTTGCTTGGCTCTCTCTCCAGCATC
AACATTTTCCTTGCGGATACGTCTCATTACGAGAGATACGGCATTAGA
ATCTGATGGGCGTGGAAGTCCAGATCGGCGATGGAGCATATTGAGCTG
TCCGAGATGTTGTTGGATCGTCTTCACCGCCAATCCACGAGCTTGGAG
GTAGAGCAAGTAGTCACGAACGTCTTCTGGCTCGGCTGGGAACCACTT
GCGGTTGTTAAGCTTGCACCAGGCAGCCCAAGAACGGCACACTGAGA
GAAGCATCTTCCAGGTGTGCTCGGAGAAGGCTTGTCTGTCACGGAACA
TATCCATGAGGTTCTTGCGAACTTCATCACTTGTCGCGTCAACTGGAA
GTGCTGGAAGATTTTGATGAACAGTGAGAAGATTTGACATATTTTCGA
AGTTTTTTAGATGCACTAGAACAAAGCGTGTTGGCTTCCTCTGAGCCC
GCTTTCCTTATATACCCGCATTCTGCAGCCTTACAGAATGTTCTAGAA
GGTCCTAGATGCATTCGTTTGAAAATACTCCCGGTGGGTGCAAAGAGA
CGCAGACGGAAAATGTATCTGGGTCTCTTTATTGTGTACACTACTTTTC
CATGTACCGAATGTGAGTCGCCCTCCTTTTGCAACAAGCAGCTCGAAT
GTTCTAGAAAAAGGTGGAAAATAGTATAAATACCGTTGAAAATAAAT
ACCGAACAACATTTGCTCTAATTGTGAAATTAGAAATCTTCAAACTAT
AATCATGTCACTTTACCACTATTTCCGTCCAGCTCAACGTTCCGTTTTT
GGTGATCATTTTTGCTTTCGTCGTAAATCTACACACGCGTCTCTTCCGT
GCGAGAGTCCAAGCCAGCAGCCAAATTCGTTGACTGAGTATTCAACGT
TTATACGTTGTCGGCAACGAGAAATAGGAAAATGCATCGGGAAATGTT
CTTTTTTCGATTTTTTCCAAGGTTTTGACAAATTTTACCACGAATTTTG
CTATGTTTTCAATTAAAAAATATGTTATTCAACTGTTTCTATGAGGAA
AATAAGGCTTTGCATGTAATTTTCTTATTCAGCATAATTTTTAATTAAT
TTGAATTTTCTGTCCTAACGTTTATTTTGTTTTCTTGGTTATGACTGATC
TGAAATTAATTTTTGAATTTTAAGGTAATATGTCAGGCGGTGCCGCAA
GTTTGTACAAAAAAGCAGGCTCCATGAAAAAGCCTGAACTCACCGCG
ACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGAC
CTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGAT
GTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGT
TTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCC
CGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCT
ATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTG
AAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGATG
CGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCG
GACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTCATATGCG
CGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACA
CCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGG
CCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCGGCT
CCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACT
GGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCT
TCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACT
TCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCG
TATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGAC
GGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATC
GTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCCCGCAG
AAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAG

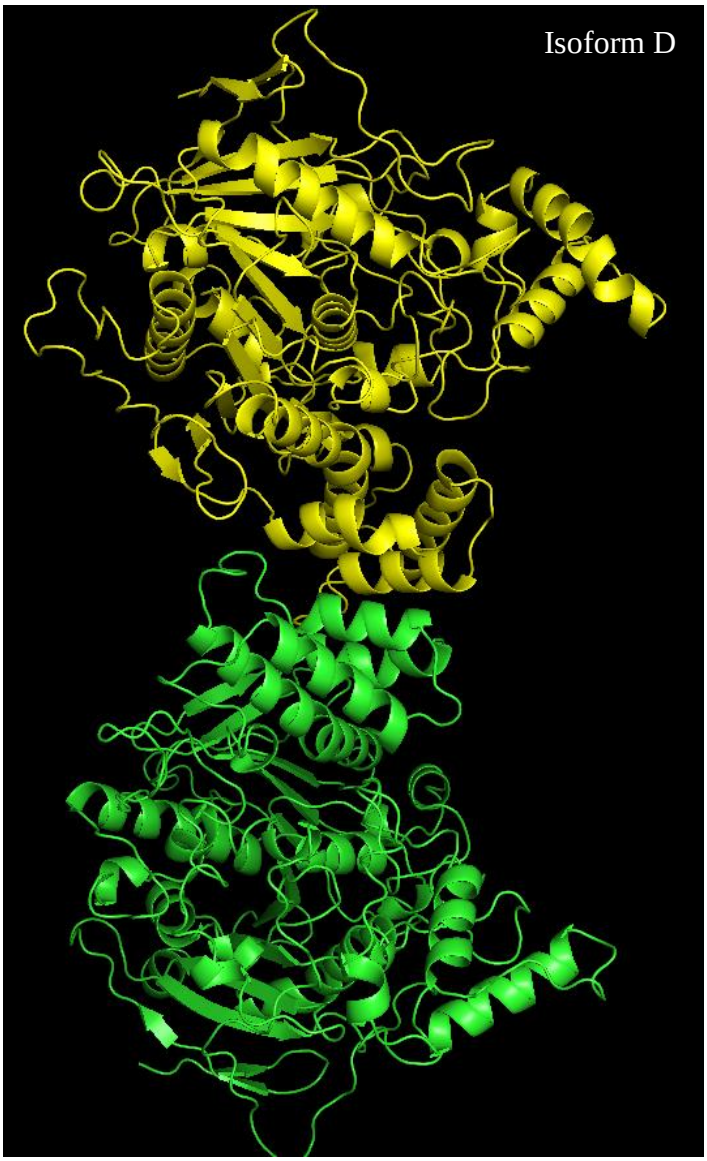
TGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAGACCC
AGCTTTCTTGTACAAAGTGGGTCCAATTACTCTTCAACATCCCTACAT
GCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAA
CTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAA
TGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAAAAGTCG
AAAAAAATTGTGCTCCCTCCCCCCATTAATAATAATTCTATCCCAAAA
TCTACACAATGTTCTGTGTACACTTCTTATGTTTTTTACTTCTGATAAA
TTTTTTTGAAACATCATAGAAAAAACCGCACACAAAATACCTTATCAT
ATGTTACGTTTCAGTTTATGACCGCAATTTTTAATAACTTCGTATAGCA
TACATTATACGAAGTTATTTTCAGGGAGCCGGATCTGATTATAAAGAC
GATGACGATAAGCGTGACTCCAACTTCTTTCGTTTCCCTCCTGAAATG
TACACATCCATTGTTGCGGTGTTTTGTACATATTTTCCTTTTGCAAAGT
TTTCTTATTTTCTTTTTTTTATCATTGCTGTTTGTCATAATAAAATCCGAG
TATACTATTTTCGGCATGTATGCTTATAATTTCCAAATTCAAAGGAACTG
GAAACAATAAAAATTGGGGACACACCTCTTAGATTGAAAGATATGTC
ATTTTGTAAATTGAGATGTTTTTAGCTAACAAGAAATAATTTTTCCAAT
CTTACAAAGAATCTGACAAAAGTTCTTCGGCCTTTCATCGTGTTTCATC
TAGAAACGTGGAAGTTTATTTTAATCATCTAGATTTTTTTTTAAAATCT
AAATCTCTTTTCAGATAAAGTGATTTTTTATCAGTTAATATATATATAT
GTTAATAGATAAGCCGTAATAATCATAAAATATCACTTATTGCTTGAT
ACACCCAAAATTACCAAATAGGCCAATTTTCACGTTCACATTTTGCAT
TATATGGCAGCTTCCTTAAGTCTAAG

Appendix 3. Additional Protein Structures

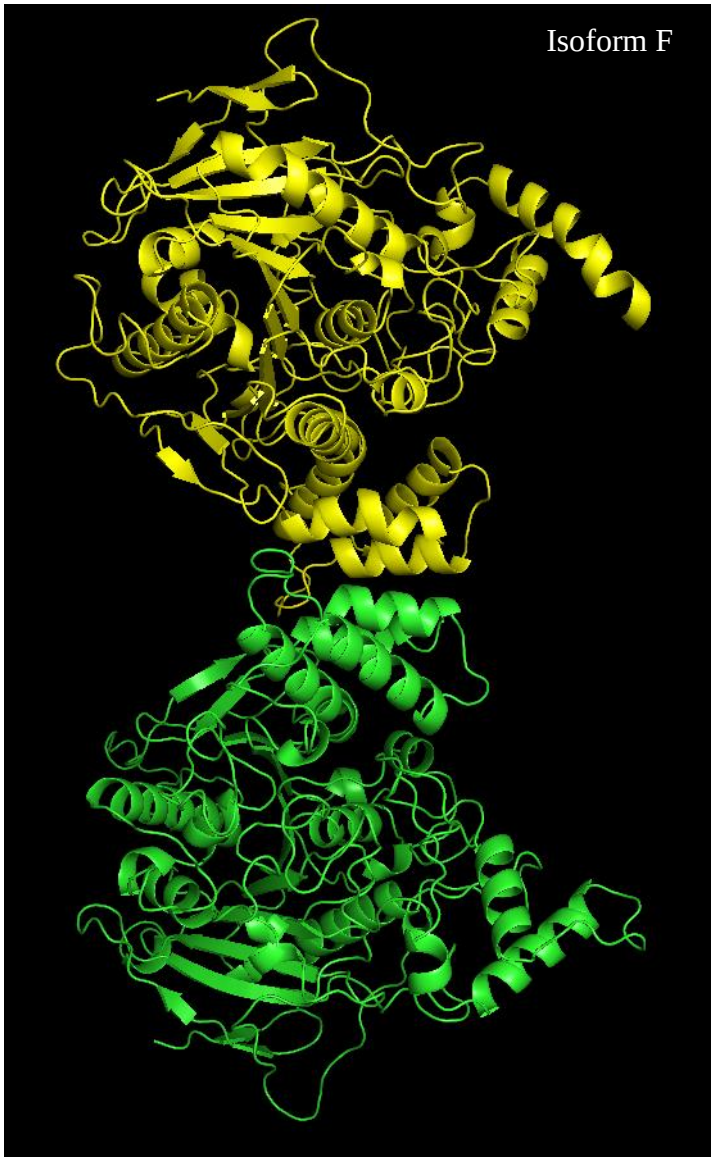
Individual Isoforms

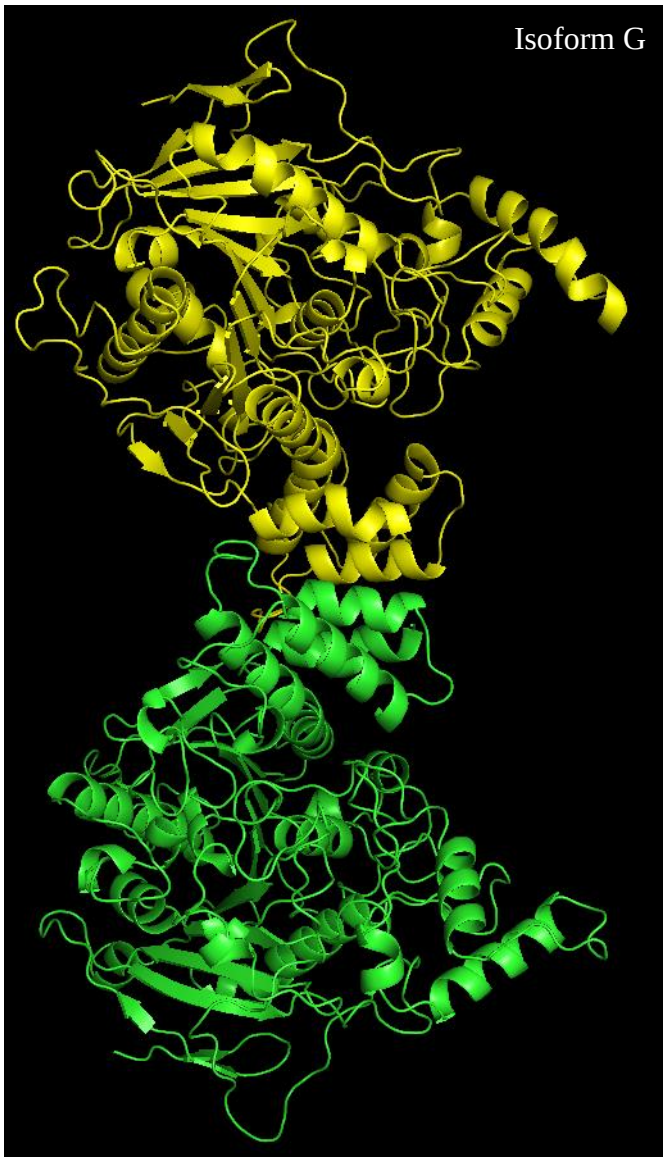


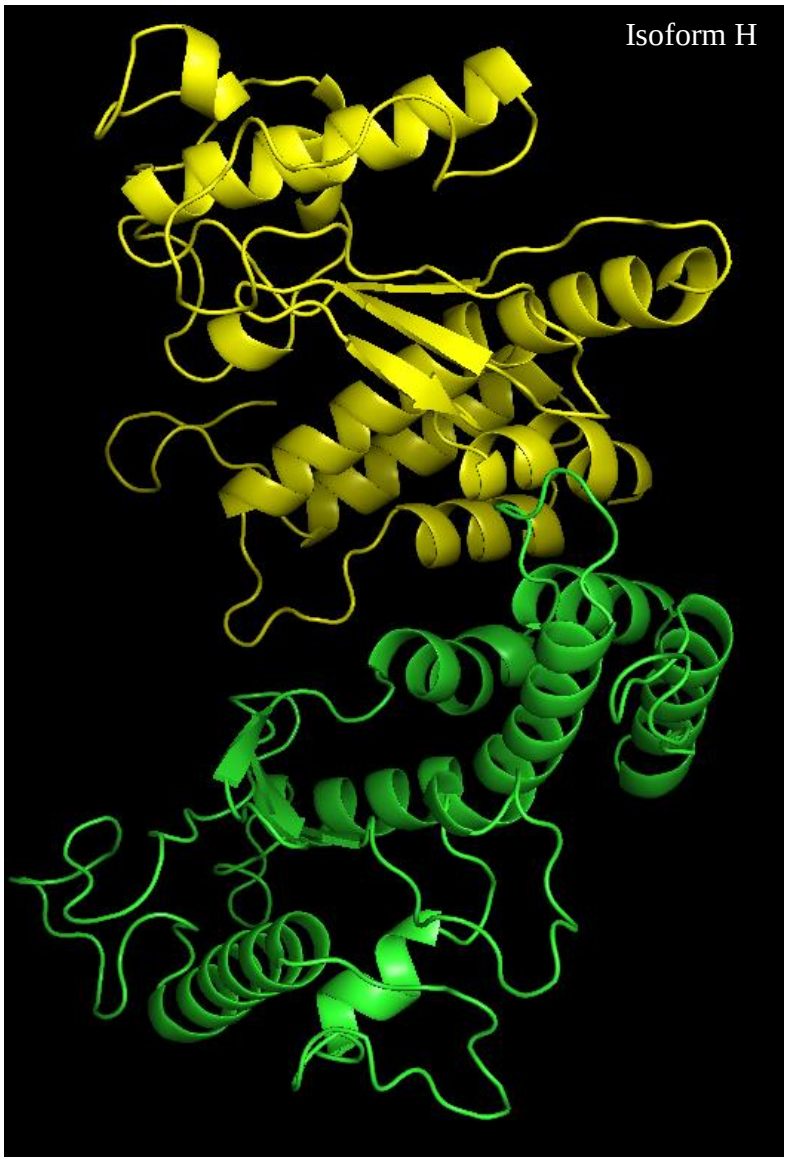

Isoform C


Isoform D


Isoform E


Isoform F

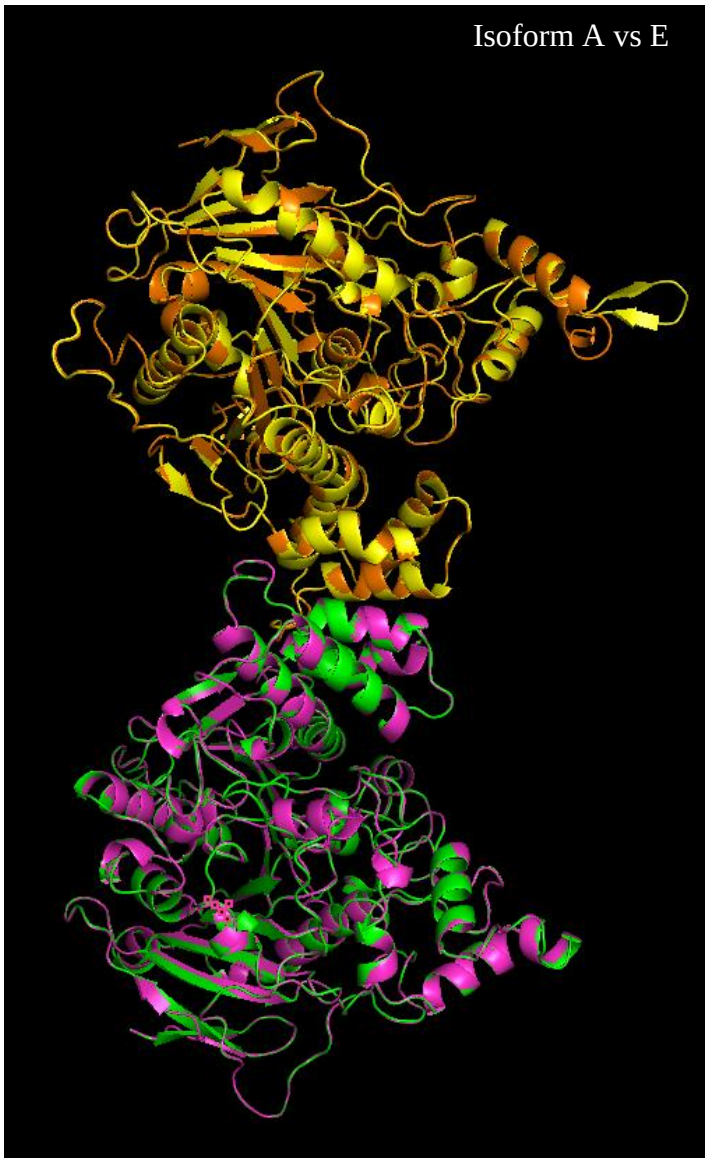

Isoform G

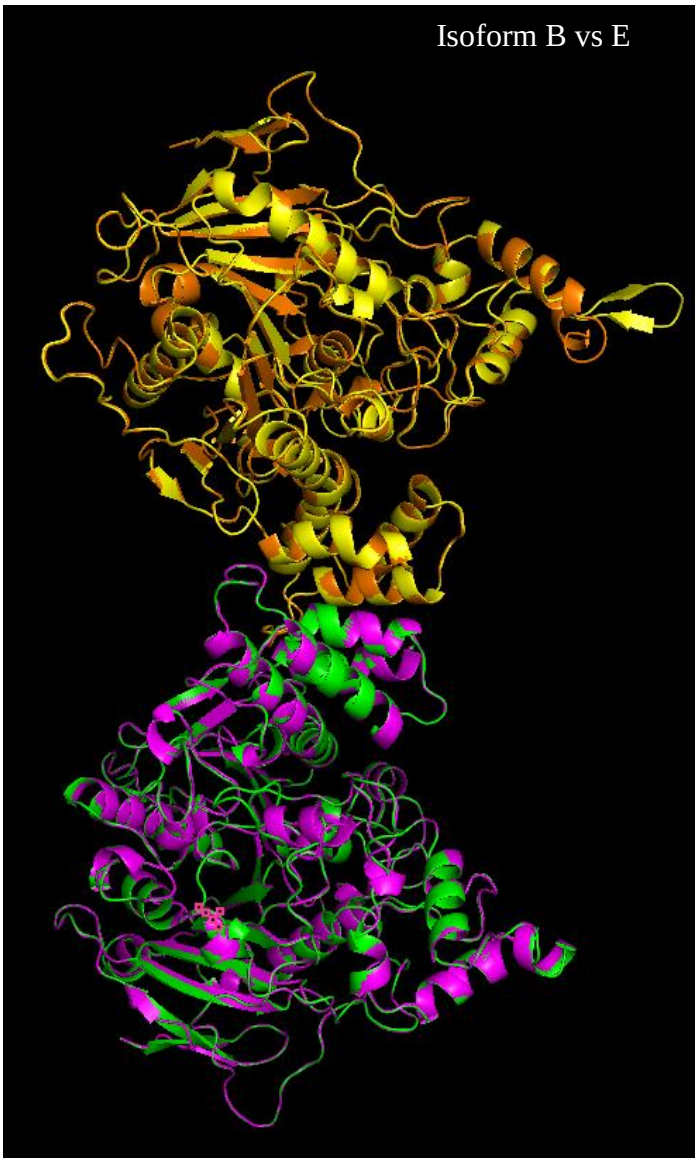

Isoform H

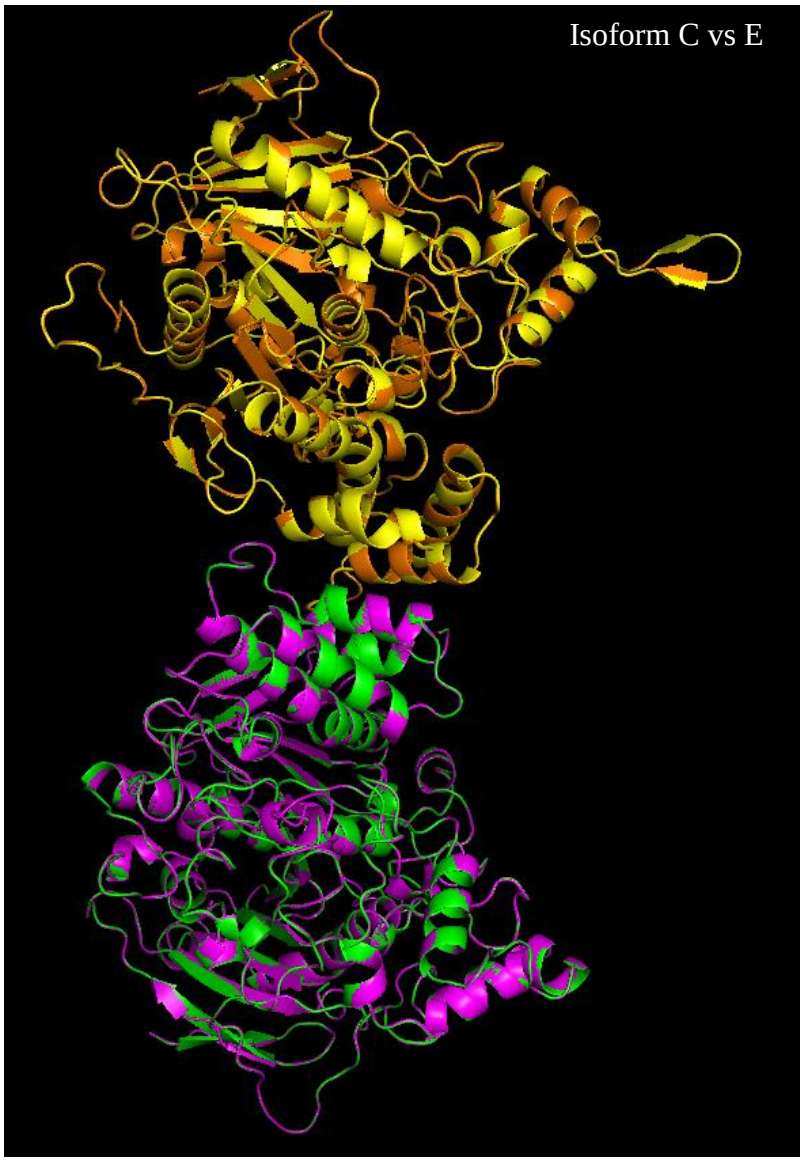

Isoform I

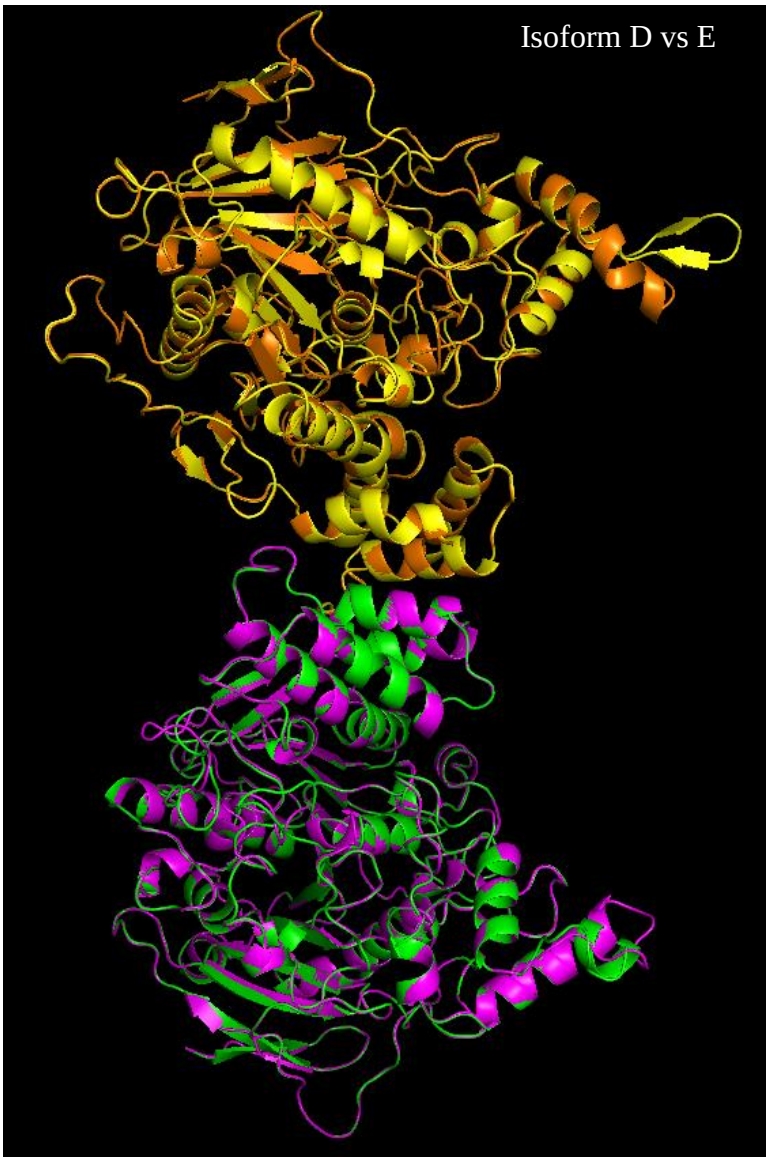
Alignment of each isoform with isoform E

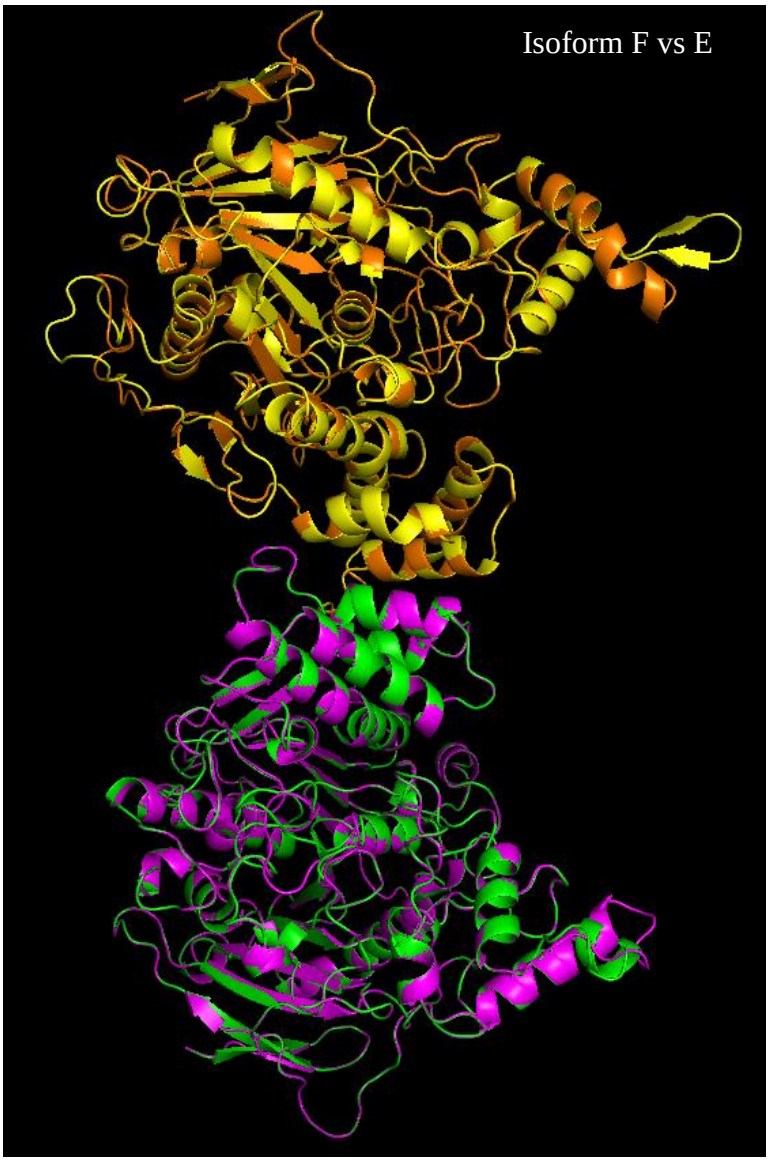
Isoform E: Yellow and Green Compared Isoforms: Pink and Orange

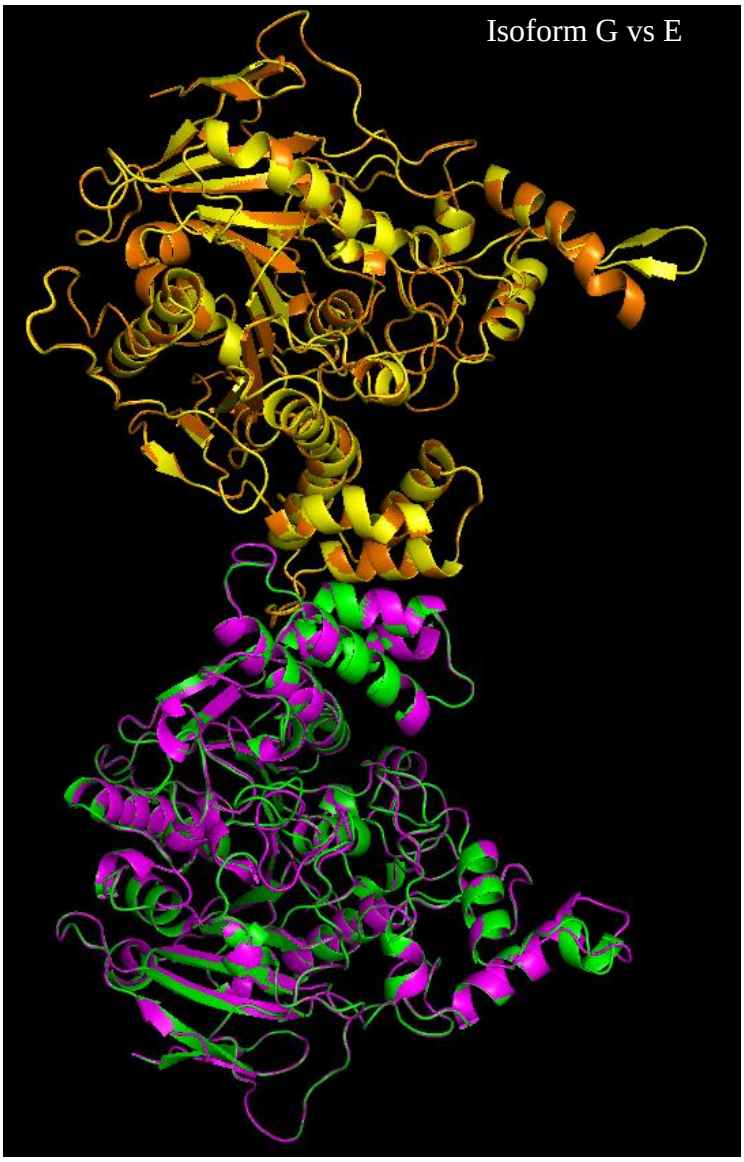


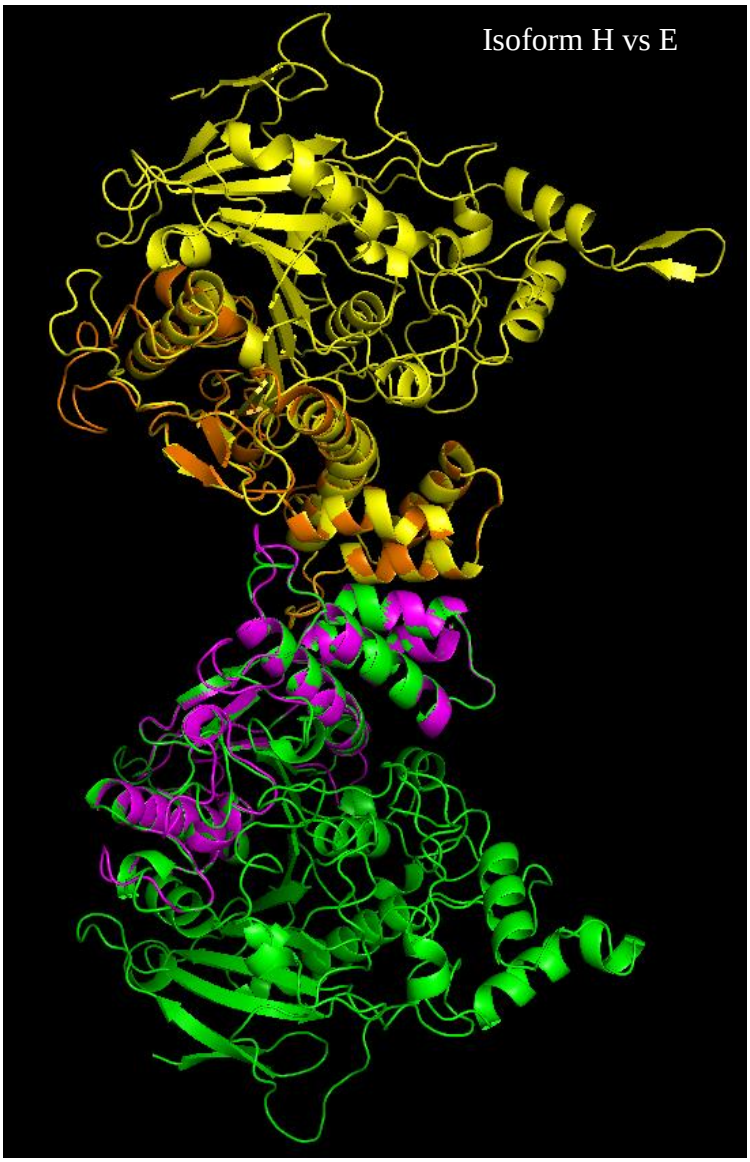


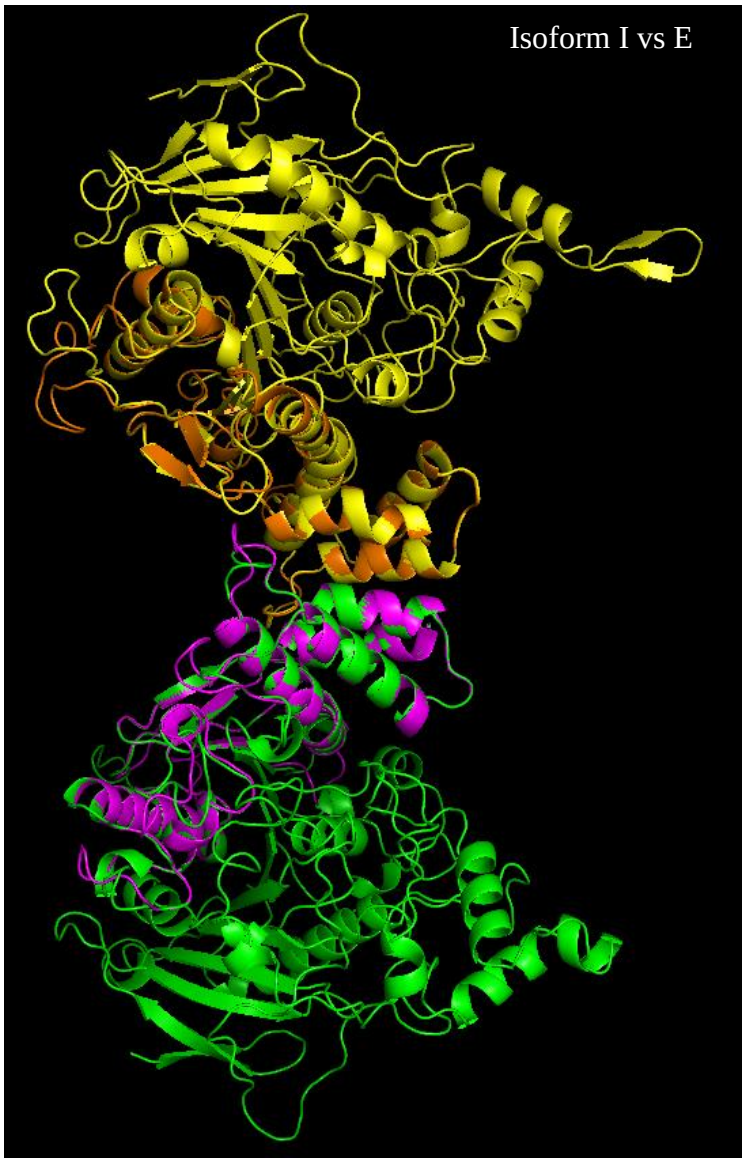

Isoform C vs E


Isoform D vs E


Isoform F vs E


Isoform G vs E


Isoform H vs E


Isoform I vs E

VC228


VC228 alignment with
WT nlg-1 isoform E
VC228: Orange
WT, nlg-1 Yellow