Applying Atomistic Modeling to Predict NLGN3 Isoform Binding to Neurexin 1- Beta

N. Doan[a], T. A. Deisher[a], S. Solares[b], Z. Talebizadeh[c]

[a]Sound Choice Pharmaceutical Institute, Seattle, WA; [b] University of Maryland, College Park, MD; [c]Children's Mercy Hospital and University of Missouri-Kansas City, Kansas City, MO

Introduction

Autism spectrum disorders have been associated with over 300 genetic mutations. As a polygenic disease, ASDs require at least one additional insult in addition to the underlying genetic susceptibility in order for manifestation of the disease phenotype (Caldwell, 2010). For example, Bradley et. al. (2010) have identified large de novo deletions in NRXN1 (neurexin 1) exons associated with ADHD probands. Notably, many studies, including Anello 2009, demonstrate that autism prevalence is four or more times higher in males than in females. Hence, X-chromosome genes are hypothesized to be especially important in the etiology of the disease (Smith, Anne, & Flodman, 2009).

Among all autism-associated genes located on the X-chromosome, neuroligin 3 (NLGN3) has been well studied in relation to ASDs. Additionally, it is believed that "alterations in neuronal circuitry and/or neuronal signaling are responsible for the behavioral and cognitive aberrations in autism patients" (Chih et al., 2004). As a post-synaptic protein, NLGN3 binds with Neurexins, pre-synaptic proteins that trigger post-synaptic differentiation through their ligands the neuroligins, which in turn trigger pre-synaptic differentiation. Thus, their binding impact nerve cell to nerve cell communication. A point mutation in NLGN3 (R451C), found in two brothers with ASD, leads to endoplasmic reticulum protein retention and reduced neurexin binding of the mutated NLGN3 that is surface expressed (Chih et al., 2004). However, population screening indicates that NLGN and NRXN mutations are rare among ASD in general. In 2006, Talebizadeh et al. identified the expression of two NLGN3 isoforms in ASD lymphoblastoid samples from autistic subjects. One of the two isoforms is a truncated NLGN3 isoform missing exon 7 using predicted three dimensional structure for the esterase-like domains of the NLGN3. To date, 6 isoforms have now been isolated. However, predicted structural modeling of various isoform expression patterns on neurexin binding and synapse formation has not been reported.

In this study, the tertiary structures of various NLGN3 isoforms are predicted using threading modeling. These structures are then equilibrated and optimized by applying atomistic in order to gain deeper understanding of the interaction between neurexin 1-beta and various isoforms of NLGN3 as well as create a fine protein structures for NLGN3 isoforms and NRX1β.

	NLGN3	Coding Exon	NRXN beta 1
Ca 2+	Gly 360	7	Asn 238
	Gln 359	7	Asp 137
			Val 154
			Ile 236
Direct Interaction	Gln 359	7	
	Gly 360		
	Glu 361		
	Phe 362		
	Leu 363		
	Asn 364		
	Tyr 463		
	Gly 464		
	His 268/Glu 270	6	
Salt Bridge	Glu 361/Glu 270	6	Arg 109
	Asp 351	7	Arg 232
Polar Interaction	Arg 561	8	Asn 103

Table 1. Expected NLGN3 and NRX1B contacts based on published crystal structure.
In NLGN3, exon 6, 7 and 8 are observed to have direct, salt bridge and polar interactions with NRX1β. In particular, exon 7 is involved in Ca2+ binding, a crucial factor in neuroligin – neurexin interaction.

Methods and Results

Six NLGN3 splice variants extracted from lymphoblastoid cells from autistic subjects were sequenced to obtain full protein sequences. Each variant's predicted protein structure was generated by I-TASSER web server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). NRX1β protein structure was extracted from a published structure of NLGN4X and NRX1β from NCBI.

In order to determine potential residues involved in binding between all six variants of NLGN3 and NRX1β, the published crystal structure of NLGN4X, which is highly homologous to NLGN3, from Protein Data Bank was used in addition to a published study by Fabrichny et al. (2007). Full length NLGN3 crystal structure was not available from the PDB.

NAMD, a scalable Molecular Dynamics protocol, was used to optimize and equilibrate the 3 D structure of each NLGN3 variant prior to docking with NRX1β. Docking software, Rosetta Dock Protein-Protein Server (http://rosettadock.graylab.jhu.edu/), was then applied to dock or combine each variant of NLGN3 with NRX1β. The docking results were subsequently refined and equilibrated by using NAMD, through minimization and dynamic modeling processes.

Isoform	Length	Exon loss
V1	848	wildtype
V2	828	3, part of 4
V3	808	3, 4
Ve345	758	3, 4, 5
Ve7	344	7, 8
Vi2	218	3, 4, 5, 6, 7, 8

Table 2- Characteristics of each NLGN3 isoforms including their amino acid length and exon loss

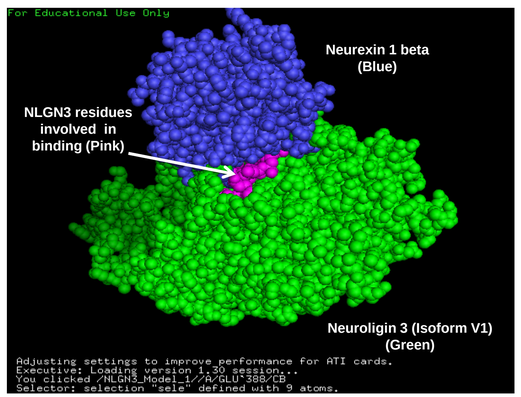

Figure 1- A protein structure of NLGN3 isoform V1 (green) docking with NRX1β (blue) analyzed by PyMol. Residues of NLGN3 isoform V1 involved in direct binding with NRX1β colored in pink include Gly 360, Gln 359, Glu 361, Phe 362, Leu 363, Asn 364, Tyr 463, Gly 464, His 268, Glu 270, Asp 351 and Arg 561.

I. I-TASSER

For each NLGN3, the top five structures are predicted by I-TASSER server. The top ranked structure of each NLGN3 isoforms was highly similar to the published neuroligin 4 on X-chromosome (NGLN4X) named 2WQZ missing exons 1 and 2.

II. ROSETTA

Table 3	Predicted direct contact residues &				Ca binding
2WQZ	359-364	463-464	265-266	270	359-360
V1	393-398	497-498	300-301	304	393-394
V2	373-378	477-478	280-281	284	373-374
V3	353-358	457-458	260-261	264	353-354
V345	303-308	407-408	210-211	214	303-304
V7			300-301	304	
Vi2	There are no significant residues for binding				

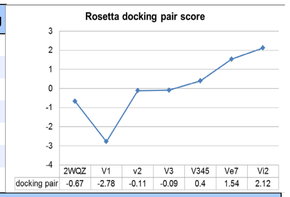

Table 4	Calculated Rosetta binding residues									
2WQZ	383	385	387-388	391-394	396-397	430-434	436			
V1	178	181-183	185-193	281						
V2	165-172	250	253	254						
V3	159	226	229	239	233	238-241	261-268	604-612	614	617-620
V345	176	179-181	183	187-191	211	214-216	218	305-307	309	
V7	238	249-251	253-255	258-263	266-267	270-271	300-303	306	327	344
Vi2	2	8	13-14	17-18	21-22	24-25	28-29	47	161-164	166-170

Table 3 **(*upper table*)** shows the residues of each NLGN3 isoforms that are predicted to be involved in neuroligin – neurexin binding based on the crystal structure in Table 1.
Table 4 **(*below table*)** shows the Rosetta output of calculated residues of NLGN3 isoforms docking with NRX1β.
Graph 1 ***(top left)*** shows the docking pair energy which is the score of interacting residue – residue pairing between 2 individual proteins during docking. In the table, the first isoform V1, containing all the exons, has the lowest docking pair energy; while the other isoforms, missing 2 exons or more, require more energy to dock with NRX1β.

III. NAMD

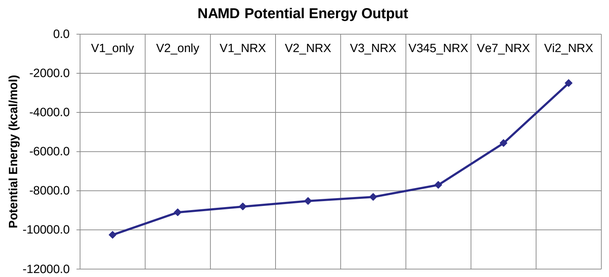

Table 5- NAMD Potential Energy Output for NLGN3 V1, V2 isoforms and NLGN3-NRX complex
The potential energy from NAMD was calculated using the molecular mechanics force field and the change in solvation energy and the CHARMM22/CMAP force field.

	V1_only	V2_only	V1_NRX	V2_NRX	V3_NRX	V345_NRX	Ve7_NRX	Vi2_NRX
Potential Energy	-10255.2	-9104.9	-8809.0	-8525.0	-8320.2	-7705.1	-5569.9	-2502.0
Ratio to V1_only	1	0.888	0.859	0.831	0.811	0.751	0.543	0.244
percentage change from V1_only		11%	14%	17%	19%	25%	46%	76%
percentage difference from V1_only		12%	15%	18%	21%	28%	59%	122%
Ratio to V1_NRX			1.000	0.968	0.945	0.875	0.632	0.284
percentage change from V1_NRX				3%	6%	13%	37%	72%
percentage difference from V1_NRX				3%	6%	13%	45%	112%

Discussion

With a comprehensive threading protocol from I-TASSER, the tertiary structures for all NLGN3 isoforms have been predicted for docking.

The docking pair energy calculated by Rosetta, representing the energy needed to dock residues, demonstrates a correlation between the number of missing exons and unfavorable NLGN3 – NRX1β binding energies. These results show that the missing exons in NLGN3 isoforms could lead to unfavorable binding energy during the docking process between NLGN3 and NRX1β.

These docking structures are later refined through molecular dynamics equilibration by NAMD software. According to Potential Energy output, V1 and V2 isoforms are ranked the highest with only 3% difference in energy between them. The other isoforms' structures have 6% to 112% difference in energy. The ranking shows that V1 and V2 have a higher capability of predicting the correct conformation of intermediate folding of the predicted docking structures than other isoforms. Especially, the percentage difference jumps up from 13% to 45% and even 112% for Ve7 and Vi2 isoforms, respectively, which are missing exon 7. Therefore, NAMD outputs are consistent with result from the study of Talebizadeh et al. (2006) on the significance of exon 7 in NLGN3- NRX1β binding.

Unfortunately, the Rosetta web-server output residues of NLGN3 isoforms docking with NRX1β are different than those residues from the study of Fabrichny et al. (2007). It is due to the limitation of the number of input residues and fixed docking protocol. We suggest writing our own protocol for Rosetta that is more applicable to the project.

Summary

1. I-TASSER is a comprehensive web-server tool to generate three-dimensional structures. Rosetta and NAMD software are able to create docking NLGN3 and NRX1β complex and then refine those structures.
2. Rosetta result show the unfavorable binding between NLGN3 and NRX1β when missing exons, especially exon 7. NAMD energy output shows that wild type V1 and V2 have more capacity of predicting the correct conformation.

References

1. Bradley, W. E., Raelson, J. V., Dubois, D. Y., Godin, E., Fournier, H., Prive, C., et al. (2010). Hotspots of Large Rare Deletions in the Human Genome. PLOS , 5 (2), 1-7.
2. Ezzeldin, Hussein M., Jeffery B. Klauda, and Santiago D. Solares. "Molecular Mechanics- and Dynamics-Based Modeling of the Major Gas Vesicle Protein, GvpA." Journal of Structural Biology, 2011.
3. Fabrichny, Igor P., Philippe Leone, Gerlind Sulzenbacher, Davide Comoletti, and Meghan T. Miller. "Structural Analysis of the Synaptic Protein Neuroligin and Its b-Neurexin Complex:Determinants for Folding and Cell Adhesion." Neuron , 2007.
4. Lyskov S., Gray J.J. "The RosettaDock server for local protein-protein docking" Nucleic Acids Research 36 (Web Server Issue), W233-W238 (2008).
5. Phillips, JC, et al. "Scalable Molecular Dynamics with NAMD." Journal of Computational Chemistry, 2005.
6. Smith, M., Anne, S. M., & Flodman, P. (2009). Nuclear and Mitochondrial Genome Defects in Autisms. The Year in Human and Medical Genetics , 102-132.
7. Z Talebizadeh, Lam, D Y, M F Theodoro, D C Bittel, G H Lushington, and M G Butler. "Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism." J Med Genet., 2006.
8. Yang Zhang. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, vol 9, 40 (2008)

Conflict of Interest: None
Acknowledgments: This work was funded by the M.J. Murdock Charitable Trust and private donations.