**Daniel Alfonsetti**

**UROP Faculty Supervisor: Troy Littleton, M.D., Ph.D.**

**UROP Direct Supervisor: Karen Leopold Cunningham**

**Term: Fall 2018**

**3 Sep. 2018**

**Analysis of polyglutamine tracts in synaptic active zone proteins**

**Project Overview**

*Provide an explanation/background of your UROP project that includes with whom and where you are conducting research.*

A polyglutamine tract (polyQ tract) is a section of a protein that contains several consecutive glutamine monomers. Long polyQ tracts (more than 20 amino acids) are frequent among proteins in the *Drosophilidae* family of flies. In humans, polyQ tracts are the most common homo-amino-acid tract encoded by a single repeated codon (CAG) (Atanesyan, 2012).

It is well known that abnormally long polyglutamine tracts are associated with neurodegenerative diseases. These so-called polyglutamine diseases are caused by expansion in the number of repeats of the CAG codon. Mistakes during DNA replication (such as strand slippage, misalignment, and stalling) can cause CAG repeat expansion. Longer polyQ tracts are generally associated with more severe disease symptoms and earlier disease onset in life (Rinaldi, 2015). The pathogenic polyQ proteins are prone to aggregation, and biophysical studies have shown that inter-protein hydrogen bonding between the glutamines drives this aggregation (Natalello, 2011). There is debate as to whether these aggregations contribute to the pathogenesis or if smaller oligomers are driving the disease. Some research has even suggested that the aggregates confer a protective property to the cells and aids in autophagy pathways (Todd and Lim, 2013*)*.

Prior studies in healthy eukaryotes have shown polyQ tracts to also be enriched in proteins involving regulation of gene expression. These regulatory proteins contain glutamine-rich activation domains involved in protein-protein interactions (Atanesyan, 2012).

Little research has been performed on the function of polyQ tracts outside of neurodegenerative diseases and transcriptional regulation. However, recent preliminary research by Karen Leopold Cunningham (my supervisor) from the Littleton Lab at MIT suggests that polyQ regions may be enriched in synaptic proteins such as RIM (Rab3-interacting molecule) and BRP (brunchpilot). The goal of this project, therefore, is to rigorously determine if synaptic proteins are in fact enriched for polyQ tracts, and if so, to identify their biological function(s). Because the polyQ tracts promote protein-protein interactions in both proteins associated with polyglutamine diseases and proteins involved in gene regulation, we hypothesize that polyQ tracts serve a similar function in synaptic active site proteins. More specifically, we hypothesize that the polyQ tracts help to tether proteins together and confer stability to the active zone scaffold.

**Personal Role & Responsibilities**

*Describe what you are contributing to the project. Be specific about what your personal duties are and what you will be responsible for accomplishing throughout the term.*

While complimentary experimental tests are being performed in parallel by my supervisor, my responsibilities are completely computational in nature. I will be responsible for creating an opensource, reusable, scalable pipeline that can perform, but is not limited to, the following analyses using opensource data from the internet and potentially also from the Littleton Lab:

1. Run a hidden markov model (HMM) to annotate polyQ and non-polyQ regions along a protein’s amino acid sequence. Tune HMM parameters with the Baum-Welch algorithm.
2. Annotate proteins across three (not necessarily independent) categories based on where they are expressed for further statistical analysis: neuronal, synaptic, and nuclear. Use KEGG or GO annotations for this.
3. Identify neural and synaptic proteins that are enriched for polyQ tracts. Are more neural or synaptic proteins enriched than those that are not for each category?
4. Compare the amount of polyQ enrichment for neural and synaptic proteins to the amount of enrichment for non-neuronal/non-nuclear proteins and non-synaptic/non-nuclear proteins, respectively. Use nuclear proteins, many of which are known to have polyQ enriched regions, as a positive control.
5. For proteins that are enriched for polyQ tracts, determine whether their isoforms are also enriched.
6. Determine if/how polyQ enriched protein expression levels change with synapse developmental stages.
7. Perform gene set enrichment analysis to identify categories of genes that are enriched for polyQ tracts.

This pipeline analysis will be performed on both the fly and human genomes. Time permitting, the code may be generalized to analyze other sequence patterns (instead of just polyQ regions).

**Goals**

*Explain what your personal goals for the UROP are, as well as what the overall aim is of the project.*

The overall aim of the project is to determine if glutamine tracts are enriched in synaptic proteins and, if so, to identify their function. My personal goal for this project is to create an easily scalable and generalizable computational pipeline for protein sequence analysis that others both inside and outside the lab can use and build upon after my UROP.

**Personal Statement**

*Briefly state why you are interested in this UROP and explain what you hope to learn from it.*

My long-term career goal is to integrate computer science, neuroscience, and cognitive science in order develop better artificial intelligence. Conversely, I’m interested in applying artificial intelligence and machine learning to computational neuroscience and genomics research. I think this UROP will primarily serve the second goal and will give me more programming and modelling practice in computational biology while simultaneously giving me an opportunity to start learning about neuroscience and the brain.

**References**

1. Atanesyan L, Günther V, Dichtl B, Georgiev O, Schaffner W. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biol Chem*. 2012 Jan;393(1-2):63-70.
2. Rinaldi, C. & Fischbeck, K. H. (2015) Pathological Mechanisms of Polyglutamine Diseases. *Nature Education* 8(4):5.
3. Natalello A, Frana AM, Relini A, Apicella A, Invernizzi G, Casari C, Gliozzi A, Doglia SM, Tortora P, Regonesi ME. A major role for side-chain polyglutamine hydrogen bonding in irreversible ataxin-3 aggregation*. PLoS One*. 2011 Apr 13;6(4):e18789.
4. Todd TW, Lim J. Aggregation Formation in the Polyglutamine Diseases: Protection at a Cost? *Molecules and Cells*. 2013;36(3):185-194.