AWSEM Studies of Cyclophillin A Evolving Pseudogenes

Michael Allison, University of Houston
Hana Jaafari, Rice University

Motivation

Noncoding regions of the genome have been a recent subject of interest for their genome regulation roles. Pseudogenes in particular, give us a glimpse at former coding genes that were under selection pressure, and now are neutrally evolving[1]. By looking at multiple pseudogenes of a parent gene we can see snapshots of evolution with dying genes and genes that are regaining function[2]. We can study using AWSEM and DCA the energetic landscapes of pseudogenes of various levels of activity for a select parent protein that has never been done in literature.

What are Pseudogenes?

- Pseudogenes are defined as nonfunctional relatives of their protein coding parent genes[3].
- Pseudogenes were once functional genetic elements that lost their "functionality" due to mutations and are free to acquire mutations with little or no selective pressure.

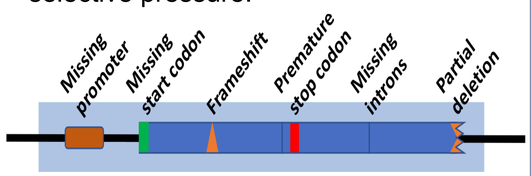

Figure 1. Potential Defects of Pseudogenes[4]

- Subsequent research has revealed that pseudogenes play essential roles in gene regulation and the development of diseases like cancer.
- We expect pseudogenes to be more energetically unstable than their parent coding genes.

How do we study the effects of devolving/evolving pseudogenes energetically?

- 5 pseudogene candidates for the cyclophillin A parent ENSP00000419425 demonstrating various signs of devolution or evolution—evidence of transcription, polymerase II and transcription factor binding sites, and are in active chromatin regions—were identified from PsiDr.
- AWSEM Suite was chosen to predict the pseudogene protein structures since it was best able to reproduce the parent protein structure.
- AWSEM Hamiltonian function: $V_{total} = V_{backbone} + V_{contact} + V_{fragmem} + V_{hydrogen} + V_{template} + V_{core}$
- DCA Hamiltonian function: $E = \sum_{i<j} e_{ij} + \sum_i h_i$

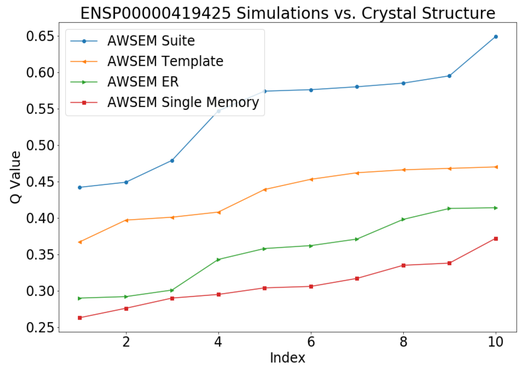

Figure 2. Q values of protein structure determination of various AWSEM versions

Results

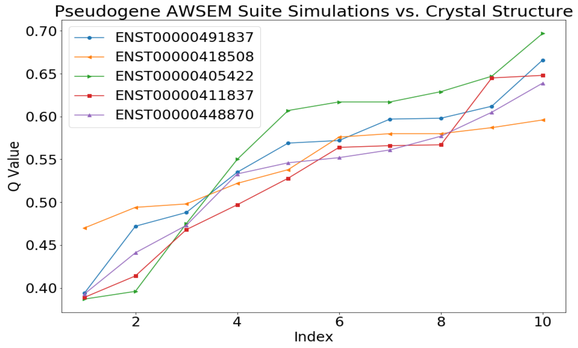

Figure 4. Q value of different Pseudogene AWSEM simulations. The Q values are all relatively high and show preservation of parent gene structure regardless of neutral selection.

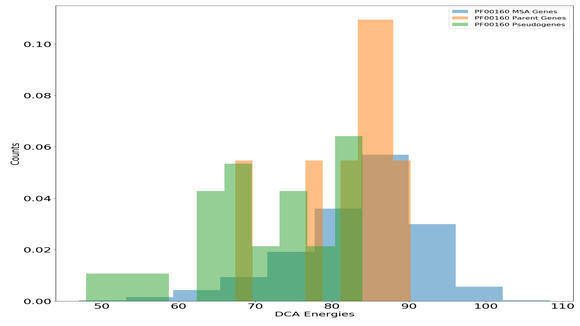

Figure 5. DCA histogram reveals pseudogenes (green) have lower evolutionary energy than MSA (blue) and parent genes (orange)

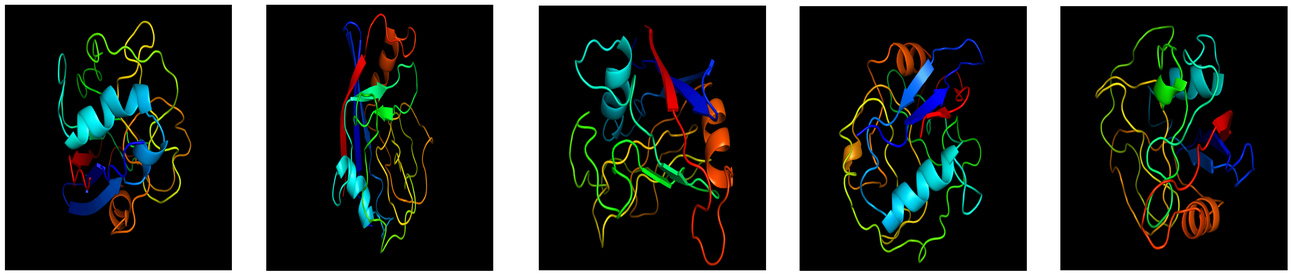

Figure 6. Representative structures of the 5 pseudogenes: ENST00000405422, ENST00000411837, ENST00000418508, ENST00000448870, ENST00000491837

Cyclophillin A

- Cyclophilin A is a member of the immunophilin family—a family of cellular proteins that posses binding abilities to immunosuppressive drugs[5].
- Cyclophilin A plays critical roles in the development of several human diseases such as cardiovascular disease, cancers, viral infections etc[6].

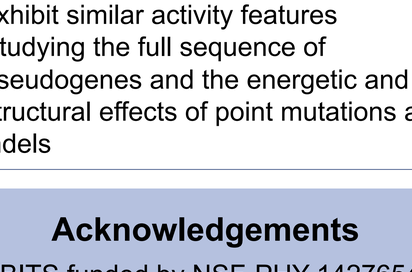

Figure 7. Crystal Structure of a Cyclophilin A parent protein (PDBID: 6U5E)

Takeaways from Research

- Pseudogenes were observed to have surprisingly similar physical energies compared to their parent crystal structure despite mutations and varying activity levels. Point mutations appear to not significantly alter the overall configuration.
- The evolutionary energies are expectedly lower between pseudogenes and parents.
- Dying genes still look similar to parents even under neutral selection
- Resurrected pseudogenes are closer energetically to proteins and potentially regained some evolutionary pressure.

What about the future?

- Looking at different protein families that exhibit similar activity features
- Studying the full sequence of pseudogenes and the energetic and structural effects of point mutations and indels

Contact Information

Michael Allison
University of Houston, Biology Major
Email: michaelallisonk@gmail.com
Phone: (817) - 891 - 8715

References

1. ZhaoLei Zhang and Mark Gerstein, *Science Direct*, Current Opinion in Genetics & Development, **Volume 14**, Pages 328-335 (2004)
2. Baikang Pei et al., *Genome Biology*, **Volume 13**, R51 (2012)
3. Wen Li et al., *Genet Genomics*, **Volume 40**, pages 171-177 (2013)
4. Rosie Redfield, *Wikipedia, The Free Encyclopedia* (2020)
5. Amaravadhi Harikishore and *Ho Sup Yoon*, Curr Mol Pharmacol, **Volume 9**, pages 37-47 (2015)
6. Pasquale Nigro et al., *Cell Death Dis*, **Volume 4**, e888 (2013)
7. Shikai Jin et al., *Journal of Chemical Theory and Computation*, **Volume 16**, pages 3977-3988 (2020)

Acknowledgements

ORBITS funded by NSF-PHY-1427654
NOTS funded by NSF-CNS-1338099