An Exploration HDAC8 Missense Swap I94T

Khushmeet Kathuria

Department of Mathematics, University of North Alabama

Abstract

This work is part of a large-scale project with other scientists and undergraduates . The purpose of the large-scale project is to analyze mutations (variations) from several different genes. The gene HDAC8 is explored in this particular project. The variant HDAC8 was assessed through the workflow known as sequence-to-structure-to-function. It was compared to all ClinVar, gnomAD, and TopMed missense variants for the gene HDAC8. All variants were assessed with AlignGVGD, PolyPhen2, Provean, and SIFT. At the beginning of the project, the mutation was considered a variant of uncertain significance. However, during the research process, the variant was classified in the likely disease category. The variant is likely to cause disease and may have moderate impact. A patient with I94T mutation has been presented in this case study. The result of which is Cornelia de Lange syndrome (CdLS)and neuronal disorders. This variant includes a swap in a conserved nucleotide. Conserved sequences, which includes conserved nucleotides, are sequences that stayed preserved in the genome despite normal genetic variations and take longer to mutate than the normal rate. According to the reports of the Slow Homology mutation model of HDAC8 the secondary structure predictions data shows in the Sequence column 19th which is Isoleucine ,the probability of having a strand is 90 percent and in the Sequence first sequence column of Tyrosine the probability of having coil is30 percent and probability of having helix is 60 percent. For the variant 94, I swapped Isoleucine for Tyrosine. This means that the mutation I94T is a hydrophobic acid that is non-polar with the negative charge. This means the mutation I94T would resist water and not dissolve in it.

Introduction

The histone deacetylases (HDACs) enzymes provided crucial role in transcriptional regulation of cells through deacetylation of nuclear histone proteins. HDAC8 enzyme activity signified the importance of HDAC8 isoform in cell proliferation, tumorigenesis, cancer, neuronal disorders, parasitic/viral infections and other epigenetic regulations. Designing of selective HDAC8 inhibitors is a key feature to combat the pathophysiological and diseased conditions involving the HDAC8 activity. This review is concerned about the structural and positional aspects of HDAC8 in the HDAC family. It also covers the contributions of HDAC8 in the pathophysiological conditions, a preliminary discussion about the recent scenario of HDAC8 inhibitors.Cornelia de Lange syndrome (CdLS) is a dominantly inherited congenital malformation disorder caused by mutations in the cohesin-loading protein NIPBL1,2 for nearly 60 of individuals with classical CdLS3-5 and in the core cohesin components SMC1A (5) and SMC3 (<1) for a smaller fraction of probands6,7. In humans, the multi-subunit complex cohesin is comprised of SMC1, SMC3, RAD21 and a STAG protein to form a ring structure proposed to encircle sister chromatids to mediate sister chromatid cohesion (SCC)8 as well as play key roles in gene regulation9. Compiled metrices and statistics comparing these variants to all the known variants listed in ClinVar, TopMed, and GenomAD as well as evolutionary comparison to 225 species has stablished a method for determining the impact of variants as it relates to the other known samples. A 21 codon impact score statistical analysis developed by Dr. Jeremy Prokop can be used to detect and predict individual motif involvement in the development of disease. YASARA software was used to model changes in the protien.

Methodology

Known data was collected from numerous publications about the variant I94T from Uniprot, then data was collect through Cosmic, ClinVar, gnomAD, and TopMED. I-Tasser was used to acquire a pbd file in order to use the YASARA sequencing. YASARA was able to sequence a 3-D animation of HDAC8 I94T, and allowed for the showcase of the wild type and VUS of amino acid 94. The file was sent to Dr. Jared Painter in order to simulate the mds data. A total of 20 nanoseconds of molecular dynamics simulations (mds) were run on a model of the gene HDAC8. All data collected was processed and addressed in various charts. The 21 codon and conservation score obtained from sequenceto-structure-to-function workflow (Prokop et al. 2017).

References

National Center for Biotechnology Information.ClinVar;[VCV000589745.2], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000589745.2 (accessed Sept. 7, 2020).

HDAC8 gene - Genetics Home Reference - NIH. (n.d.). Retrieved September 07, 2020, from https://ghr.nlm.nih.gov/gene/HDAC8.

www.uniprot.com

National Center for Biotechnology Information. ClinVar;[VCV000589745.2], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000589745.2 (accessed Sept. 7, 2020).

HDAC8 gene - Genetics Home Reference - NIH. (n.d.). Retrieved September 07, 2020, from https://www.ghr.nlm.nih.gov/gene/HDAC8. Deardorff, M., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., . . . Shirahige, K. (2012, September 13).

HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Retrieved September 07, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3443318.

Methods

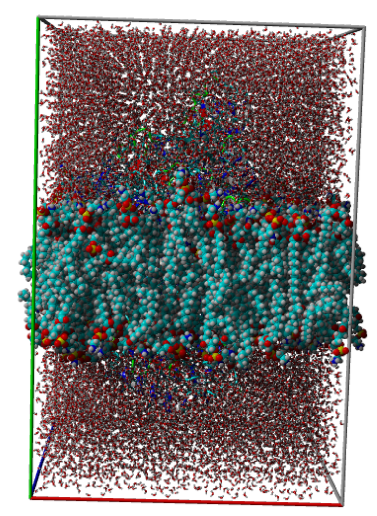

Figure 1:Protein model for HDAC8 I94T generated and embedded into a lipid membrane. HDAC8 I94T variant analysis which is the structural model of HDACS in a lipid membrane looking from the cytoplasmic side. The above shown is a side view of the protein modeled in a typical phospholipid membrane.

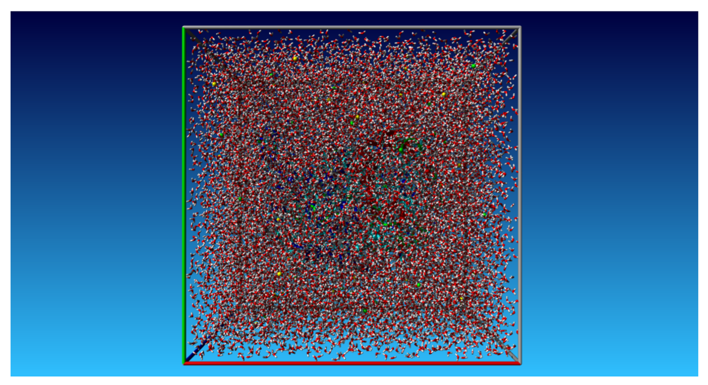

Figure 2:Top view (cross section) of the protein modeled in a typical phospholipid membrane.

Results

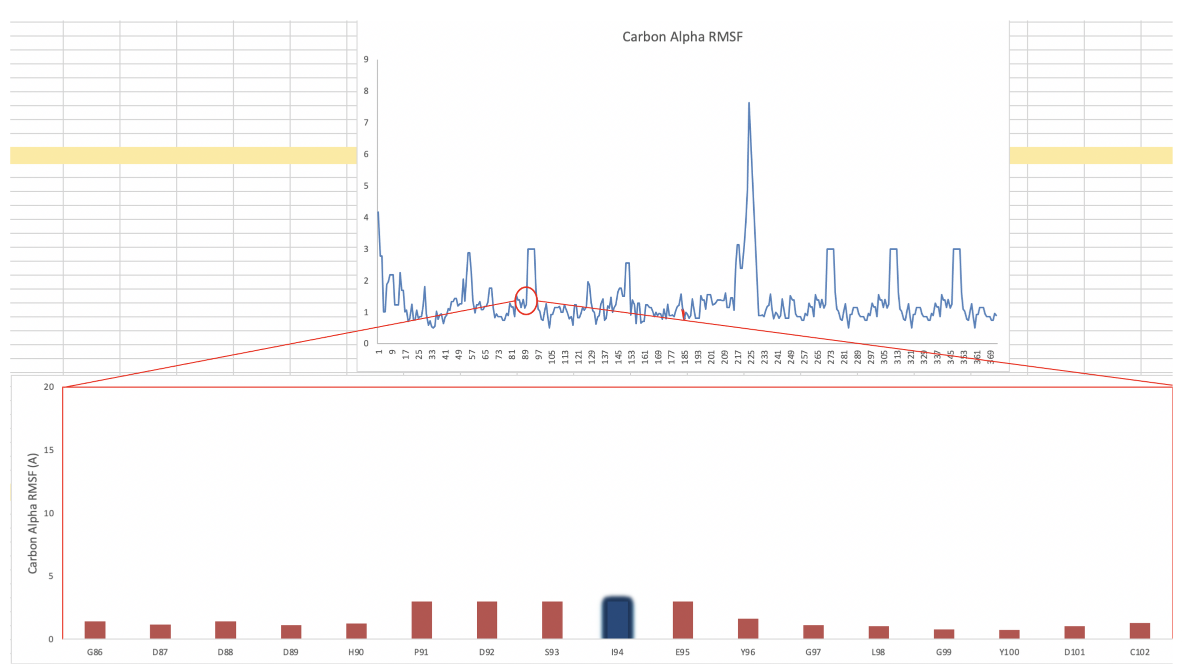

Figure 5:The root mean square fluctuation (RMSF) plot is plotted to estimate the changes in dynamic flexibility of the regions of protein structure. In (Figure 3) all the residues except F209 and M274 looks very stable for RMSF. For Comparison of RMSF, I94(patient) which is present at the surface of the active site has increased flexibility to reduced F209 and M274.

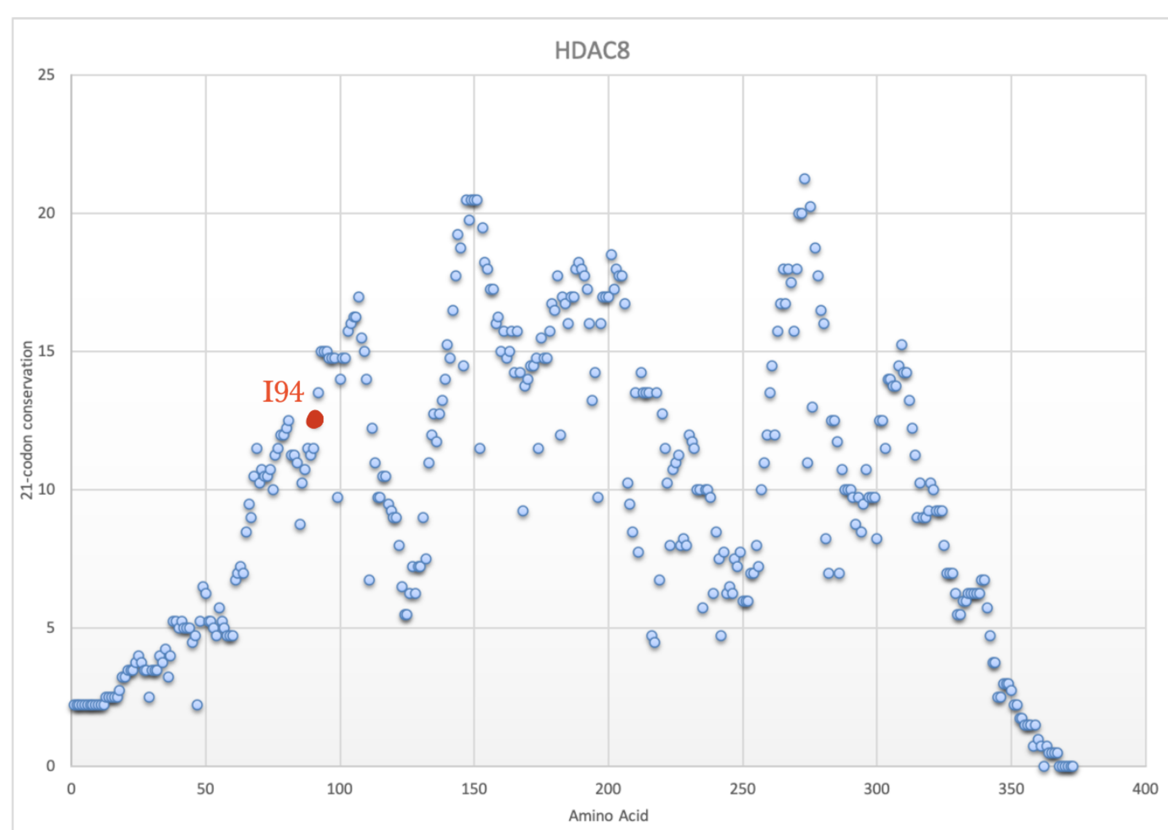

Figure 6:Deep evolutionary analysis using 360 species open reading frame sequences for HDAC8. The plot shows a sliding window calculation for each site, identifying the most selected and conserved linear motifs within the gene. Amino acid 94is identified in red. Based on the analysis of 360 species for HDAC8,it looks like amino acid I94 falls in a conserved region.

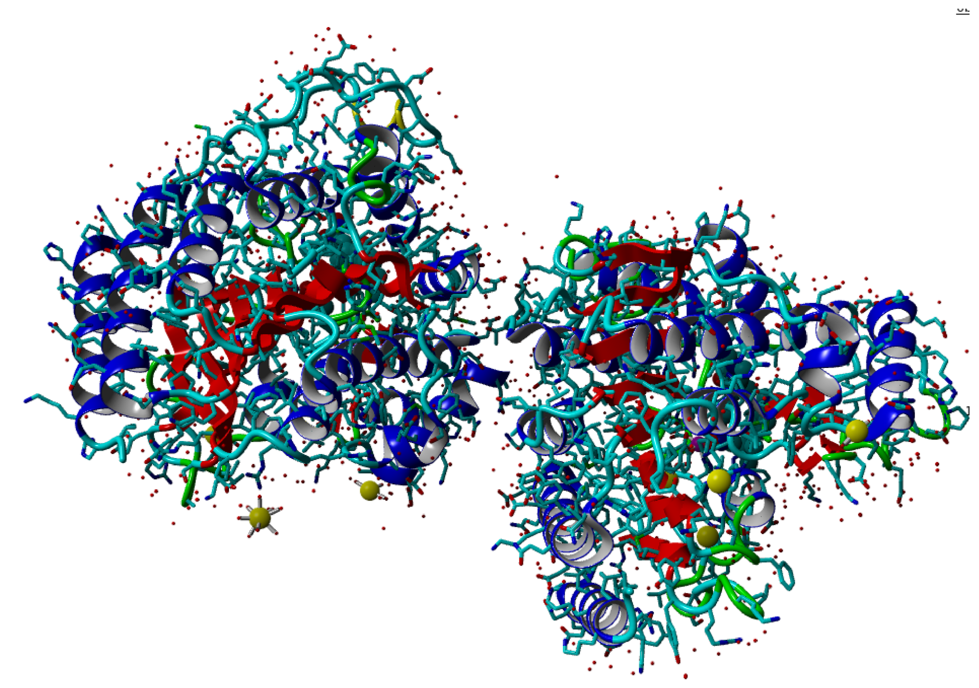

Figure 3:I94T found at base of alpha helix

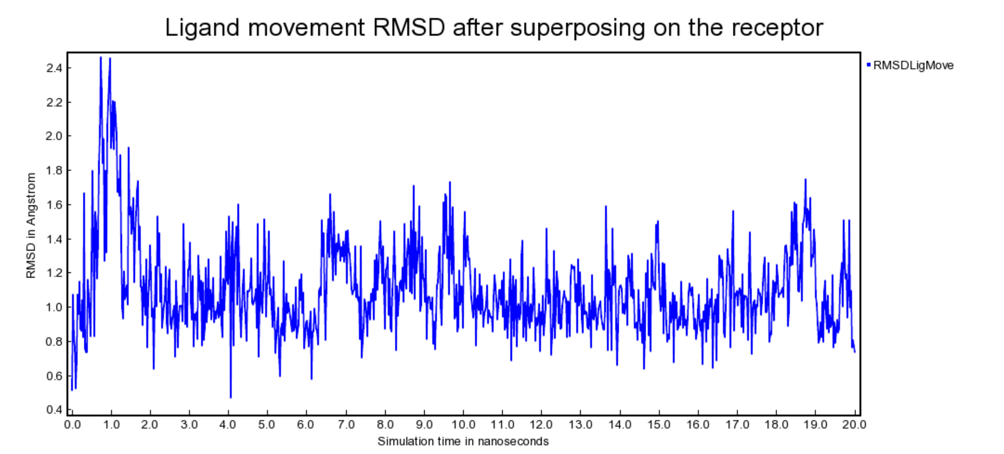

Figure 4:Molecular dynamic simulation of the HDAC8 protein model shown for the carbon alpha root mean squared deviation (RMSD) for each amino acid averaged throughout 20 nanoseconds of simulation. According to the RMSD the region of I94T looks like it is a well packed within the structure based on mds and also shows high stability of the structure.

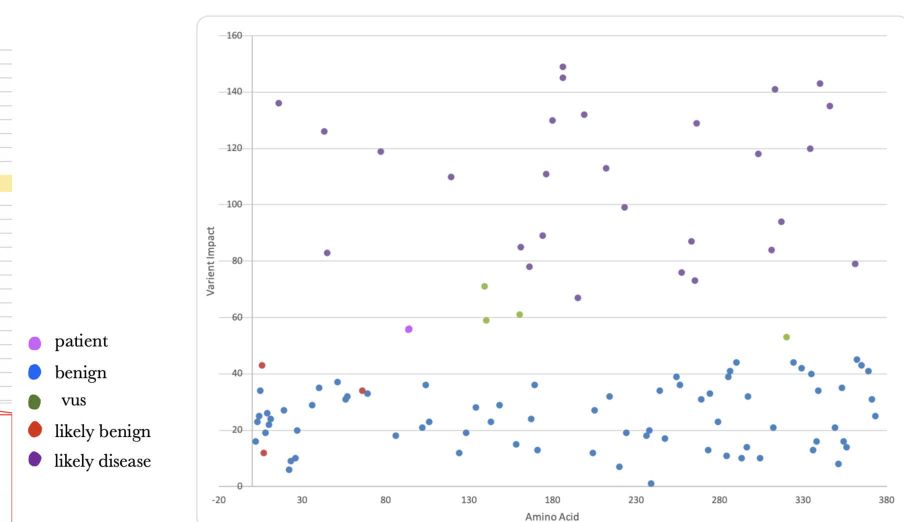

Figure 7:Variant impact scoring for all TopMed / ClinVar (benign-blue, VUS-green, Likely Disease-magenta, likely benign-red), and patient (pink) variants for HDAC8.Variant impact for all gnomAD/TopMed,ClinVar, and patient variants reveals that variant I94T to fall near several other ClinVar variants with higher impact scores.The variants combined impact score is statistically lower than the other categorised ClinVar groups

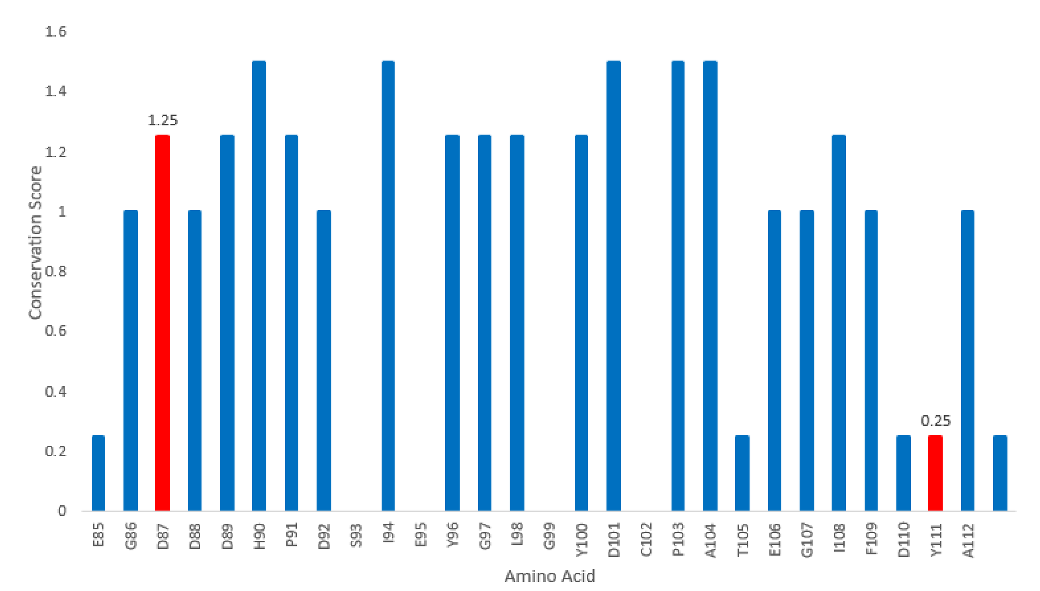

Figure 8:Zoom in view of conservation for amino acid 94 (red) linear motif. The numbers above represent the percent of sequences with synonymous / non-synonymous variants throughout evolution.The patient variants conservation is 1.5 according to the chart, which indicates more likely functional impact of the mutation which is Cornelia de lange syndrome.

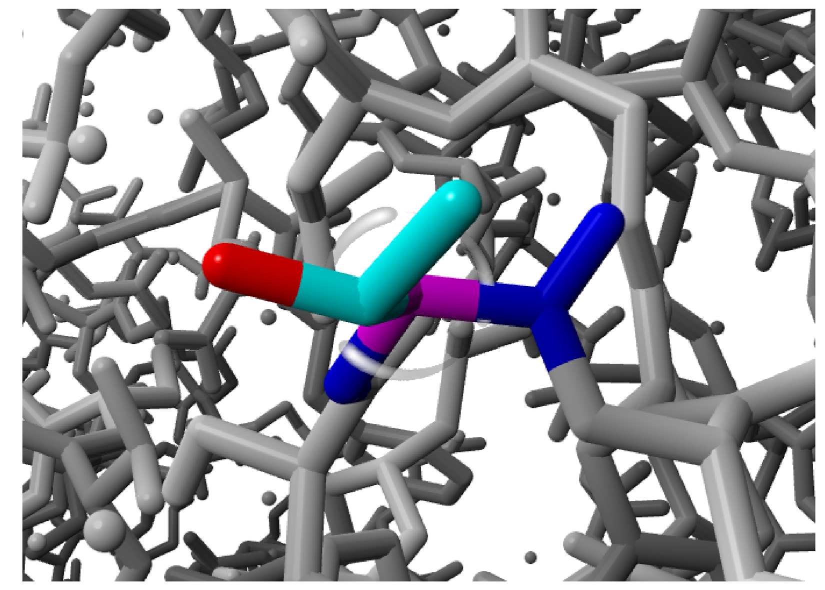

Figure 9:The protein is shown up close with the I94T position marked and zoom in view provided on the right. The model was performed in a membrane matrix but that has been hidden for ease of viewing. This demonstrates that the variant location is close to the end of an alpha helix and is packed into the middle of the protein

Conclusion

In conclusion, we identify a variant in HDAC8, I94T, that associates with Cornelia de Lange syndrome (CdLS) and neuronal disorders.Based on assessments to all known variants in HDAC8 we show that I94T is pretty much in the range for functional impact and structural fold contribution that indicates its critical loss-of-function role in HDAC8. This represents a novel strategy for HDAC8 assessments for VUS that can be used in future screening projects.

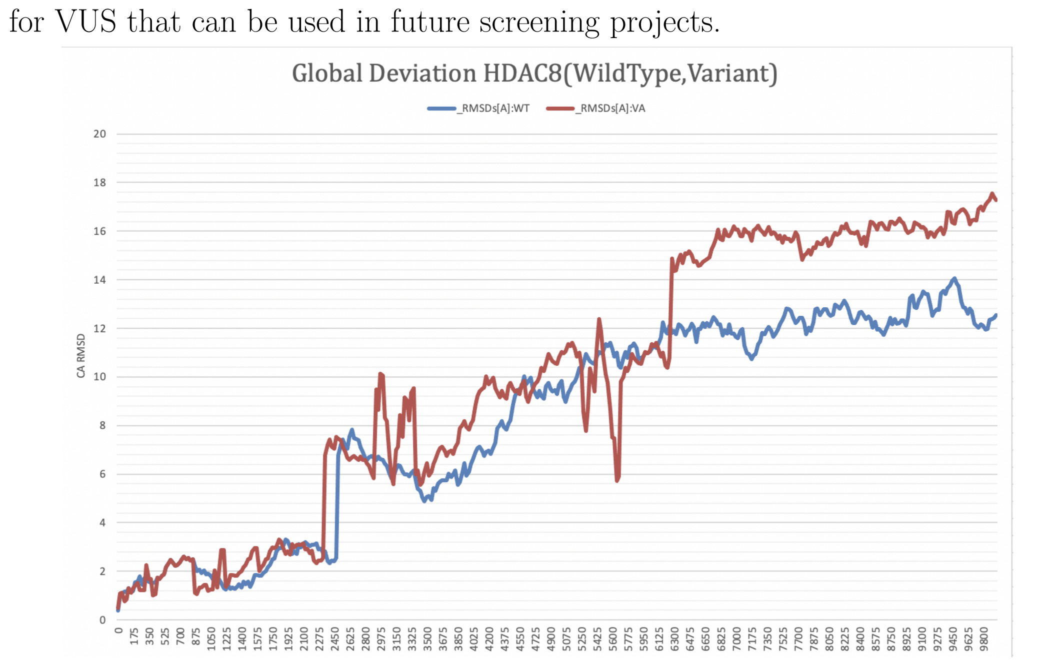

Figure 10:Molecular dynamic simulation of the HDAC8 protein model shown for the carbon alpha root mean squared deviation (RMSD) for each amino acid averaged throughout 20 nanoseconds of simulation. According to the RMSD the region of I94T looks like it is a well packed within the structure based on mds and also shows high stability of the structure.

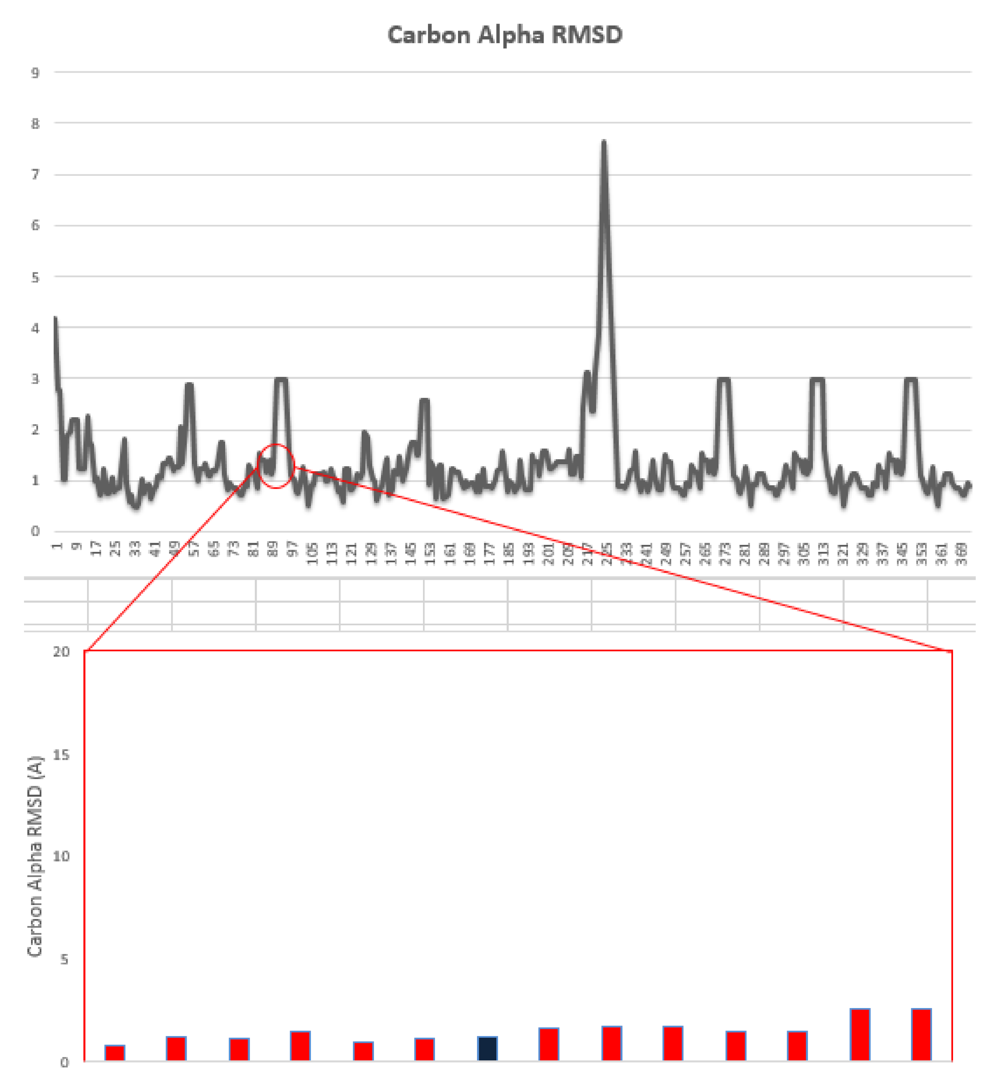

Figure 11:HDAC8 Carbon Alpha RMSD Line (all amino acids) and Bar Graph(amino acids 88-100).

Acknowledgements

Dr.Cynthia Stenger, Dr.Jeremy W Prokop, Dr.Jared Painter, Michele Morris

University of NORTH ALABAMA MATHEMATICS

HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY CODE Characterizing Our DNA Exceptions