September 2021

# A computational study of genotype-phenotype mutation patterns

Kamaludin Dingle
*Gulf University for Science and Technology*

Omar Tawfik
*Gulf University for Science and Technology*, xbox.ps3.psp.ps2@gmail.com

Ahmed Aldabagh
*Gulf University for Science and Technology*, gust0019831@gust.edu.kw

# A Computational Study of Genotype-Phenotype Mutation Patterns

Ahmed Aldabagh and Omar Abdulazim
Faculty Supervisor: Prof. Kamaludin Dingle

## Abstract

Understanding properties of genotype-phenotype maps is important for understanding biology and evolution. In this project we make a computational study of the statistical effects of genetic mutations, in particular computing the probabilities of each phenotype transitioning to any other phenotype. We also investigate the importance of the local phenotypic environment of a single genotype, and its role in determining mutation transition probabilities. We use HP protein folding, RNA structure, and a simplified GRN matrix model to study these questions.

## Overview

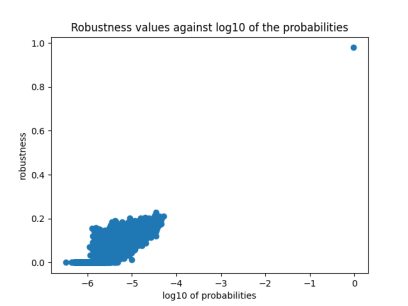
When we say that a protein or RNA mutated, we mean that a change happened to its structure by changing one allele or more. This change can lead to a decrease or complete loss in its expression. However, it may not affect the structure at all in some cases.
We will be studying these mutations in details for both of Protein and RNA structures.

We will be using Protein of length 20, and RNA of length 12 for all of the testings.

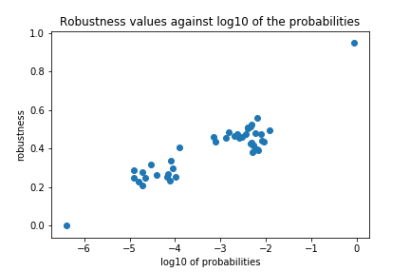
Some Protein sequences give us nothing, we call this the “NO SHAPE” and we represent it by X.

## How does the robustness correlate with the probabilities of structures?

**Protein:**
The robustness, which is the probability of the same protein structure mutating to itself again, was calculated then was plotted against the probabilities of the structures.

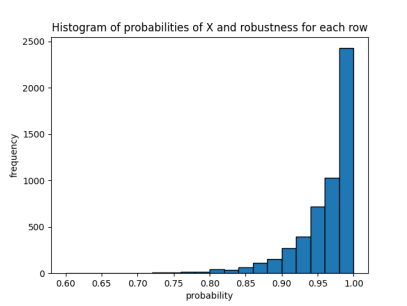



**RNA:**
We wanted to do the same for the RNA to confirm our predictions before proceeding.

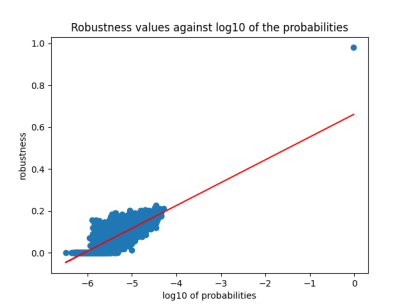



## Since the robustness and the probability of X are high, can this lead to a formula to predict the mutation of a structure?

**Protein:**
Our analysis showed that the probabilities of robustness and X are dominating. And this can be clearly seen in the below histogram.

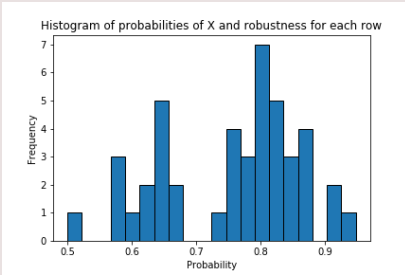



Both probabilities added together are most of the times almost 1. So, the probability of a structure mutation can be represented as P(S) = R + P(S -> X) = 1, where S is the structure and R is the robustness.




After fitting, we came up with this equation:
R = 0.11 log10(P) + 0.66, where log10(P) is the log10 of the probability of a structure.
From the equation above we can calculate the robustness of any protein structure of length 20, then we can substitute the R into the second equation so:
P(S -> X) = 1 – R

**RNA**:




In RNA, the spread is more distributed so we can’t use the same method here.

## Future Work

1. Come up with a formula to calculate the probability for RNA.
2. Find the correlation of the probabilities of individual sequences mutating against probabilities of their structures mutating.
3. Experimenting with other genotype-phenotype maps (e.g. GRN and Polyomino).