

# Project Proposal

Group Name: Bioinformatics

Dylan Carpenter, Josh Domal, Hunter Read

## Project Description

In order for chemists, biologists, and other researchers to elucidate the mechanisms of disease and develop medicines to combat illness and improve the lives of people, they must first understand and be able to identify possible causes. In regards to protein mutations, which mutation may have an impact on biological function can be hard to determine, and while wet lab experiments are useful, they are costly and time consuming. Our goal is to create a visualizations from in-silico rigidity analysis which allow a researcher to identify possible mutations that may warrant further investigation.

## Introduction

The data consists of both pdb files and rigidity analysis results. The pdb files describe the protein in various ways, but our interest for this visualization is the sequence of amino acids that make up the protein, which is just a list of abbreviations that represent name of the amino acid. The rigidity analysis data consists of the results regarding the wild type protein, and each possible single-point mutation. The rigidity analysis data contains basic integer data regarding the number of points, number of free points, hinges, bars, bodies determined from the pebble game algorithm, and cluster information. This cluster information can also be used to determine rigidity distance, which describes how destabilizing or stabilizing a mutation is overall.

## Motivation

Our visualization will allow the user to compare and analyze various metrics of protein mutations. This visualization is necessary as it is currently difficult to compare specific amino acid mutations with other mutations and/or the wild type of the protein. Currently, there exists a heat map that shows protein stability across various amino acids but there is no easy way to compare individual mutations via this visualization and determine which amino acids have not been mutated. This heat map also offers no way to investigate the data beyond simply comparing slight color differences. Manually comparing all of these possible mutation across the different proteins incur an exponential amount of time spent from the perspective of the user. With our visualization we look to reduce the time spent on this by providing the user with ways to easily compare mutations with one another. The target user for this system is any researcher (biologists and/or chemists) who are looking to understand the impact of various protein

mutations. This could help in many different disciplines including the development of drugs or the study of a disease.

## Proposed Outcome

Channels: Color, Size, Position

Marks: Area, Points

Visualization Techniques: Heat Map, bar chart

Presentation: Dashboard

## Project Completion Plan

<u>Task</u>	<u>Completion Target</u>
Data Collection	Complete
Data Processing	October 22nd
Prototyping	November 2nd
Implementation	November 30th
Evaluation	December 7th

## Definitions

**Protein:** A string of amino acids that fold up into various shapes and perform biological functions

**Mutation:** In regards to a protein, when the normal amino acid used in construction of a protein is accidentally or intentionally substituted by a different amino acid

**In-silico mutation:** A protein mutation performed on a computer instead of a wet lab

**Wild Type:** The wild (natural) protein as commonly found. A non-mutated protein

**Pebble game:** An algorithm that can be used to determine the flexible and rigid regions of a 2D or 3D graph.