TABLE OF CONTENTS

ABSTRACT

Proteomics is the study of protein that includes protein function and structure. One of the main objectives of this field is to explore the 3D structure of proteins. This application utilizes the main objective of proteomics. The user can explore a specific set of genes involved in cancer. It allows users to access information about that specific biomolecule, and run basic analysis and visualize its models.

Specific Objectives

- Users can access pre-populated data such as features, gene ontology, cross-references on certain genes that are involved in cancer
- User can select certain aspects of the protein to explore, such as looking at spatial distribution of the protein chains, a basic computation of the molecular distances, and 3D structure of the protein

Technical Requirements

- Python 3
- BioPython
- UniProt REST API
- Requests module
- Pandas
- Matplotlib
- PyMol
- D3.js(Optional)

DESIGN

From development perspective, initial data model would be created using UniProt's API, all the genes of interest will be pulled in from there, an initial analysis related to features, GO, and other information will be performed and these would be stored in the MySQL database. From the user's perspective when they click on a protein name, the user should expect all this information to show up on the main page.

Once the user access the basic protein information, they will see more options to explore the structure of the protein further. From development perspective this would be achieved by utilizing the Protein Data Bank(PDB), a repository with structural data of proteins. Once the user choses a protein to explore a pre-written PDB parser script will be executed. There are different options for creating this script

1. Writing the PDB parser from scratch
2. Using BioPython's PDBParse or MMCIIF Parser

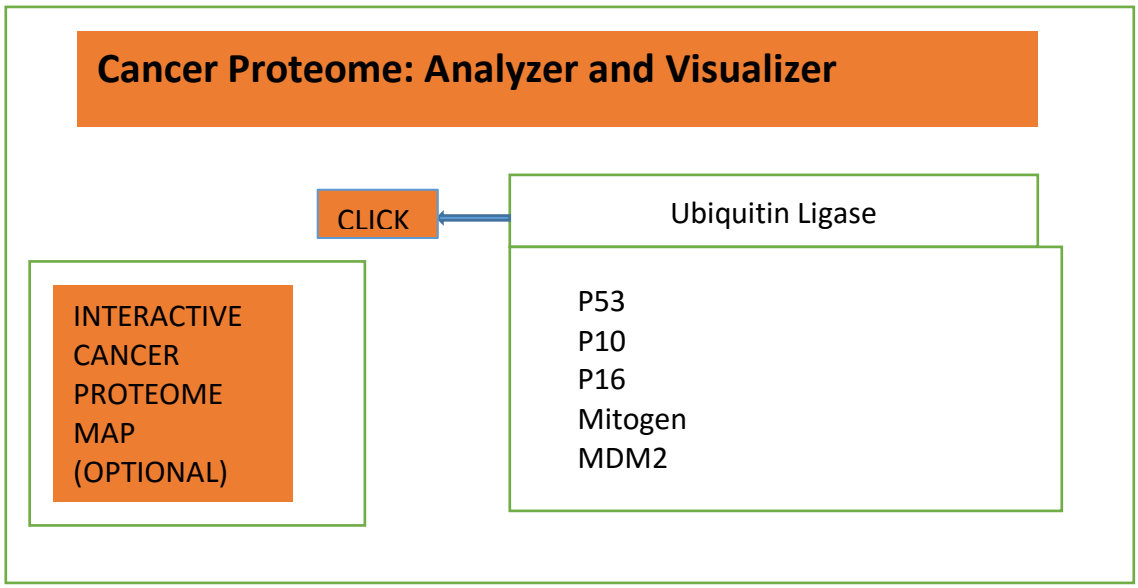
During development each of these options will be explored, and the most efficient one would be utilized for development

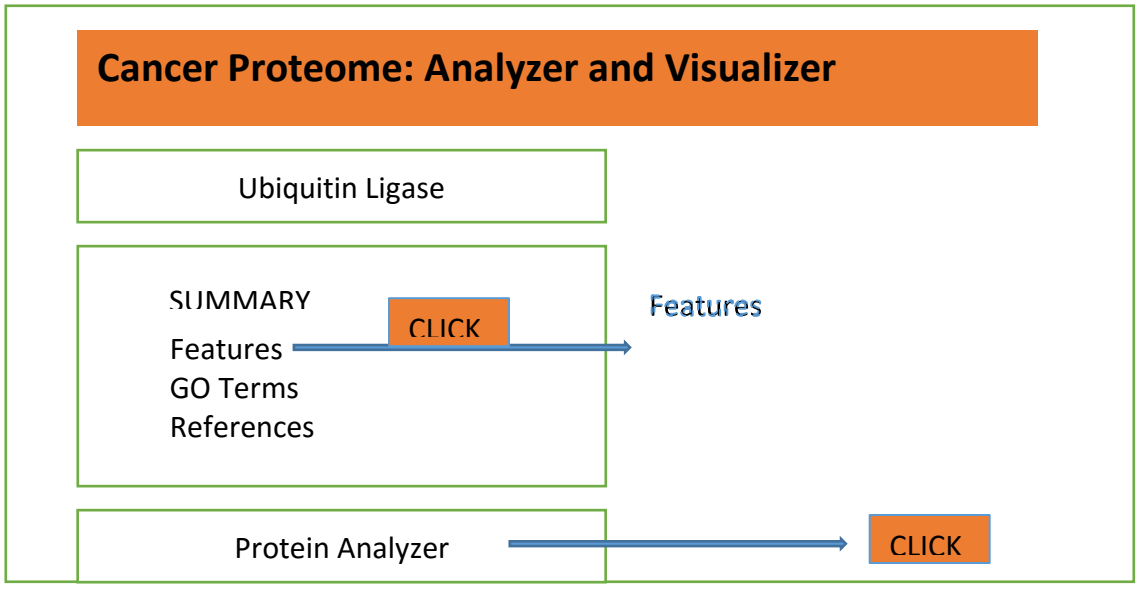
Using the parsed PDB data some quantitative analysis will be done, which includes atomic distance calculations, computations on the center of the molecule, and the whole molecule. The user will be able to see an interactive animation of the molecule, which would allow them to further explore the biomolecule.

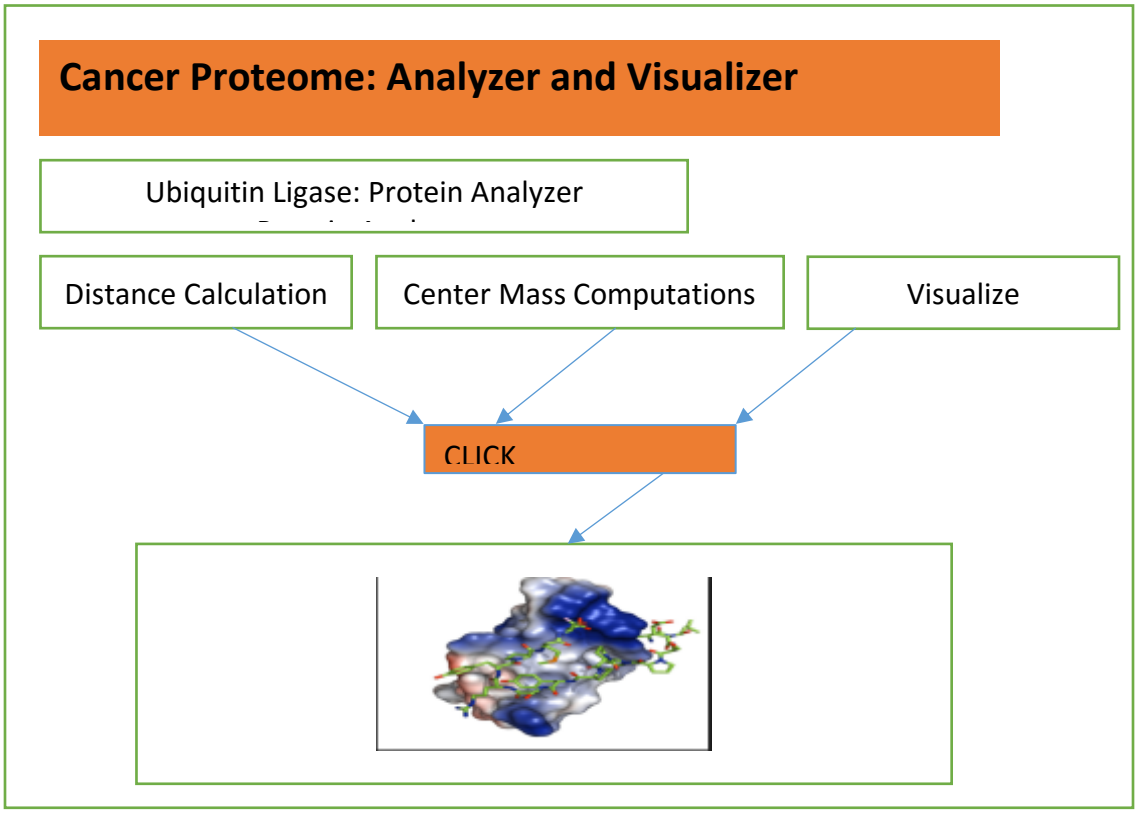
Additional Design Suggestions

Creating an interactive cancer map for all the selected cancer proteins, as they relate to each other using the JavaScript visualization library D3.js

Wireframes


Cancer Proteome: Analyzer and Visualizer
CLICK
Ubiquitin Ligase
INTERACTIVE CANCER PROTEOME MAP (OPTIONAL)
P53
P10
P16
Mitogen
MDM2


Cancer Proteome: Analyzer and Visualizer
Ubiquitin Ligase
SUMMARY
Features
GO Terms
References
CLICK
Features
Protein Analyzer
CLICK


Cancer Proteome: Analyzer and Visualizer
Ubiquitin Ligase: Protein Analyzer
Distance Calculation
Center Mass Computations
Visualize
CLICK