CALIFORNIA STATE SCIENCE FAIR
2016 PROJECT SUMMARY

Name(s)

Jennifer R. Pan

Project Number

S0527

Project Title

Computationally Revealing Protein Targets for Metal-based Drugs

Abstract

Objectives/Goals

Knowing the structural properties and cellular distributions of metal-protein interactions is useful for metal-based drug design. Cysteine (Cys) residues are reactive towards metals and conserved in protein evolution. Thus, Cys-rich domains (protein structures) are ideal targets of metal-based drugs. With the advance of technology, computational tools like molecular-dynamics (MD) simulations and application protocol interfaces (APIs) have provided an affordable way to gain insight into the nature of metal-protein interactions. My objective is to computationally discover the cellular distributions and structural characteristics of Cys-rich protein domains to reveal effective targets for metal-based drugs.

Methods/Materials

I applied MD-simulations to metal-Cys protein models from PDB database. Through simulations, I created a scientific standard of what defines a Cys-cluster by finding the average distance between Cys-residues. I applied the standard to an algorithm I developed using the concept of "k-means clustering" so that I can identify Cys-clusters in any protein model. I applied this algorithm to protein models in the CATH database to find which have Cys-clusters. I used statistical analysis to determine the Cys-models# structures and correlated the models' ID to the SubCellLoc database to find their locations. I created a database that effectively combines information about Cys-domains and their structure and location.

Results

I identified 11,406 Cys-rich structures from 173,207 domains and investigated their location and structure. By visual verification of the clusters, I found that my algorithm can accurately identify Cys-domains. Cys-domains are closely related to the structures of Arc repressor-like domains, four helix bundle, and zinc finger, and have functions like metal homeostasis and proteolysis, implying a critical role in metal-protein interactions. Metallodrug development can be enhanced as scientists know where to target drugs and what structure/ligands the drug should have to best bind to Cys-clusters.

Conclusions/Discussion

Through my algorithm to identify clusters, examination of dynamic protein structures, and data-correlation to discover their structure and localizations, I have computationally explored these domains, and their structure and localizations may uncover the great potential of metallodrug targets that are particularly sensitive to metals and expedite drug development.

Summary Statement

I developed an algorithm to identify Cys-rich clusters in any given protein model, and coded programs to correlate these clusters to their structural characteristics and subcellular localizations to enhance metal-based drug development.

Help Received

Dr. Shujian S. Tsuen from the University of Hong Kong provided technical support in helping me learn the various programming languages and guided me through my thought process.