Machine Learning on the Free Energy Landscape of p53 Captured with Energy from MD Simulations

Nabeel Kemal[1], In Sub M. Han[2], Kelly M. Thayer[1,2*]

1. Department of Mathematics and Computer Science 2. Department of Chemistry

Introduction

In allosteric regulation, the energy landscape of a protein, a statistical representation of a proteins potential energy, can be altered to achieve various results. Understanding allostery provides exciting insight into the possibilities of more targeted and effective drugs made using allosteric design. To better understand the shifts in the free energy landscape that occur as a result of allosteric effectors, we can use vectors, programmed with magnitude and angle data from simulation outputs, and couple them with residues. This allows for an instantaneous visual representation of how a protein reacts to various allosteric effectors in a simulation. This visualization allows for continuous refinements in order to create allosteric effectors that most favorably shift the free energy landscape, with the eventual goal of engineering allosteric effectors to reactivate native functionality in proteins. Each vector will be centered on alpha carbon of a residue, using a polar coordinate system the vector will represent the net force felt by its respective residue. This net force on a residue is a product of its neighboring residues inflicting external forces upon it.

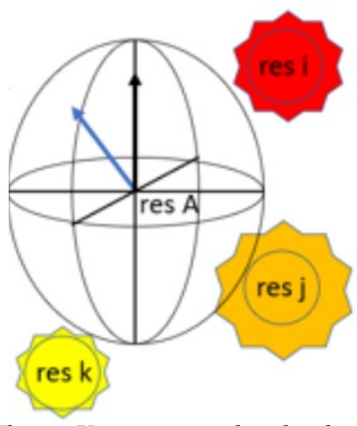


Figure 1: Example of a Thayer Vector, uses a local polar coordinate system centered on each residue's alpha carbon. Pictured with three neighboring residues i,j,k

The feasibility of this approach, in which we capture the free energy landscape using vectors will be demonstrated using the p53 tumor suppression protein. Mutations in p53 are present in nearly 50% of ovarian, esophageal, colorectal, head and neck, larynx, and lung cancers in humans[1]. Most of which are both lethal and undruggable. However, if this method proves to succeed in providing the missing link between identifying allosteric control points and which protein substate will be selected, it will be an integral step toward a new class of allosteric drugs with targeted control of the biological processes in any protein.

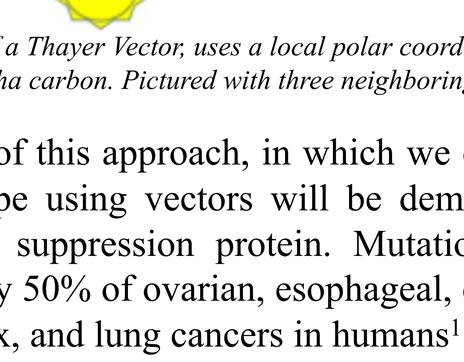


Figure 2: Difference in the Molecular Surface near Y220 position in p53 WT and p53 Y220C[2]

Acknowledgments

Thank you to my colleagues Prof. David Beveridge, Prof. Michael Weir and the rest of the M2M Lab for their helpful insight and thoughtful discussion. Many thanks as well to Henk Meij for maintaining Wesleyan's HPCC and finally thanks to the Wesleyan Summer Research Program. This research was funded by Grants NIH R15 GM128102-01, CNS-0619508, and CNS-095085.

References

1. Olivier, Magali et al. "TP53 mutations in human cancers: origins, consequences, and clinical use." *Cold Spring Harbor perspectives in biology* vol. 2,1 (2010)
2. Sundar, D., Yu, Y., Katiyar, S.P. *et al.* Wild type p53 function in $p53^{Y220C}$ mutant harboring cells by treatment with Ashwagandha derived anticancer withanolides: bioinformatics and experimental evidence. *J Exp Clin Cancer Res* **38,** 103 (2019).

Methods

To obtain the data necessary to create the vectors, cpptraj dihedral analysis and peptide interaction analysis was run on three 500ns trajectories of p53; a wild-type trajectory, a y220c mutated trajectory, as well as a drug-bound trajectory. From this data, sets of vectors for each trajectory is created.

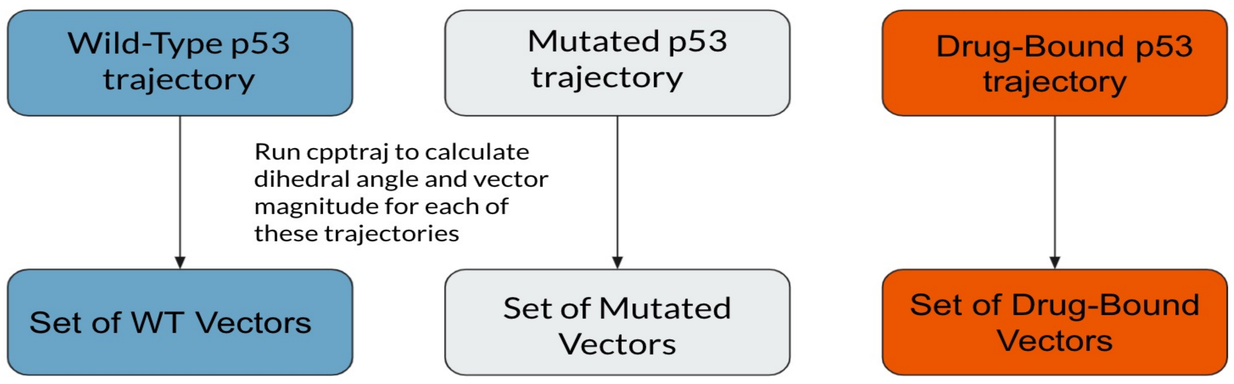


Figure 3: Flowchart Depicting Data Collection

From these vector sets, which contain phi-psi angles as well as a magnitude, the data is converted from the spherical coordinate system to the cartesian system. Thus, the data is now in the form of x,y,z vector components, which are combined to create our energy vectors. These vectors are then overlayed onto an alpha-carbon backbone derived using data from the 1TUP PDB file and plotted using a scatterplot. This is all done through a Python script that uses package MatPlotLib to take the magnitude and phi-psi angles and output vector representations of the free energy landscape.

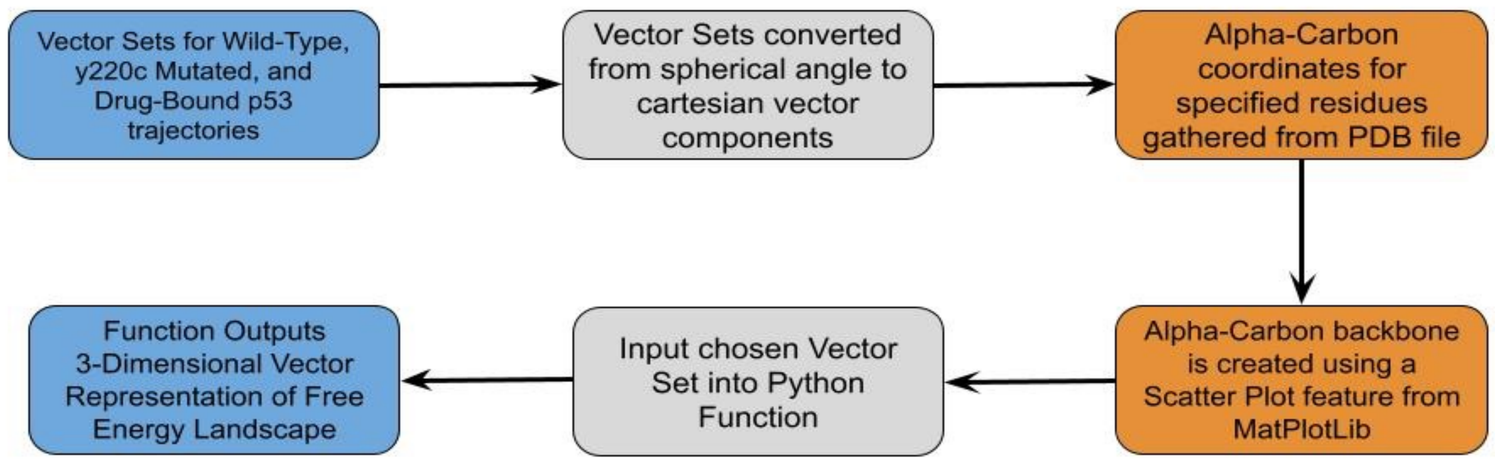
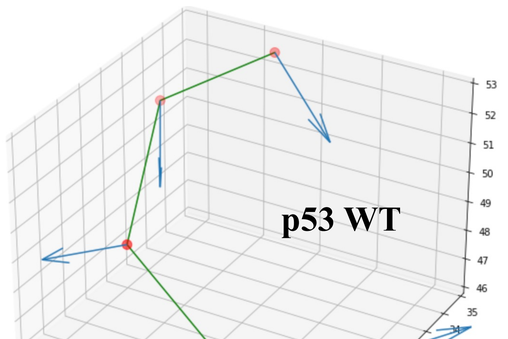


Figure 4: Flowchart Depicting Creation of 3-Dimensional Vector Representations using MatPlotLib

Results



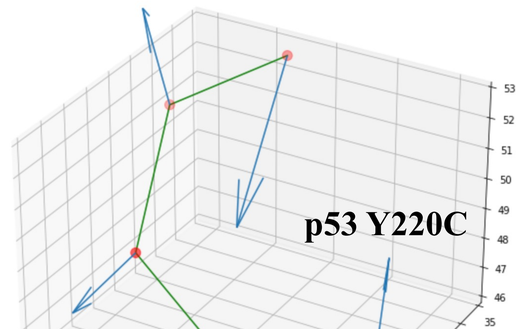


Figure 5: Vector Representations of the Free Energy Landscapes of Wild-Type p53, Drug-Bound p53, and Y220C Mutated p53. Displayed are Residues 218-223 encompassing the Y220C Mutation

Future Direction

Using the vector sets, which are each labeled either mutated or wild-type, the goal in the future will be to use the sets as inputs for an assisted machine learning model to predict the conformational substate of p53 from its free energy landscape. The conformational substate can be predicted using a linear regression model crafted with the labeled vector data sets.