



**CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY**

Name(s) Kevin Wu	Project Number S1434
Project Title A Novel Framework and Algorithm for Ultra-fast Protein Shape Comparison	
Abstract Objectives/Goals Scientists study the structural similarity of proteins for many purposes, such as developing new medicines, finding protein evolutionary relationships in the protein families, etc. With more than 100,000 protein structures already discovered, the comparison of protein structures becomes more complex and time consuming. Current protein comparison takes about 12 hours to compare one protein structure against all protein structures. A fast comparison algorithm to filter out any proteins that are definitely similar or not similar to a given protein is an effective approach to speed up the process. My engineering goals/objectives are: 1. Design and develop a framework to allows scientists to develop and test new protein comparison algorithms more conveniently and efficiently 2. Design and develop new algorithms to achieve fast and accurate protein shape comparisons.	
Methods/Materials Step sequence of Framework: The Framework procedure is as below: 1. Pull the Query and Rep protein data stored locally of over the network. 2. Extract the 3D points of the Query and Rep proteins. 3. Apply the Fingerprinting method to all of the proteins. 4. Applies the Scoring method, comparing Query proteins with Rep proteins. 5. Store the pairwise scores. 6. Use the pre-saved TM scores to plot the statistics results based on the scores. Algorithm Method Input: 1. Query Protein List; 2. Selected cluster of Representative Proteins. Output: 1. Pairwise Scores; 2. Statistics Plot. Materials: 1. Super Computer Center of UCSD. 2. Access to the RCSB PDB. 3. 27,000 Representative Proteins Data. 4. Facility for conducting the research: Skaggs School of Pharmaceuticals. 5. A laptop computer. 6. Java Programming Language. 7. Eclipse IDE for Java.	
Results 1. The framework is able to run multiple algorithms at the same time. Each algorithm yields one plotted result set. This provides the developers a more flexible and efficient way so that one time execution can output multiple results for multiple algorithms.	
Summary Statement The purpose of fast protein shape comparison is to speed up the development of new medicines, predict new protein functions, and discover evolutionary relationships of proteins.	
Help Received I would like to thank the Skaggs School of Pharmaceuticals at University California of San Diego for letting me to use Protein Data Bank to conduct my research under the supervision of Dr. Peter Rose.	