Research Experiences for High School Students Project Description

Project Title: Develop Algorithms for the Analysis of 3D Protein Structures in the Protein Data Bank

1. Overall Research Project

The RCSB Protein Data Bank at UCSD (http://www.rcsb.org) provides worldwide access to three-dimensional structures of proteins, DNA, and RNA. Large-scale analyses and queries of this database is a “Big Data” problem and we use the Apache Spark (http://spark.apache.org/) platform to run these calculations in a distributed way on a computer cluster. This project involves mining of 3D structure information as well as text mining and machine learning.

This is an exciting opportunity to apply your computer science and math skills and gain hands-on experience in a professional team environment of software engineers and scientists.

2. Number of Students to be supported: 2

3. Name of Lead person: Dr. Peter Rose, RCSB Protein Data Bank, San Diego Supercomputer Center, UCSD

4. Plan to Integrate Student into Group Activity

The student will be a part of the research team working on the larger scale project that includes the project described here. He or she will attend the group meetings and communicate with the team members using other methods of communication. The student will work closely with the lead person and the other personnel involved.

5. Student Prerequisite

Demonstrated programming skills in Java (preferred) or other programming language and interest in algorithm development. Strong background in Math, Biology and/or Chemistry is a plus. Good communication and writing skills required.

6. Number of hours per week: 40 hours

7. Relevant link:

View some of our associated projects:
http://bioinformatics.sdsc.edu/projects

RCSB Protein Data Bank:
http://www.rcsb.org/pdb/home/home.do

San Diego Supercomputer Center:
http://www.sdsc.edu/