Opening the data bank -- scientists try to match new protein structures

November 7 2011

Imagine playing Go Fish with 3,000 cards. Scientists at Rochester Institute of Technology and Dowling College are engaged in a similar game with higher stakes. Instead of cards, they are matching the protein to the job it performs in the human body. Their research could lead to drugs that target proteins and switch on or off specific functions associated with various diseases.

The three-year study is funded by a $417,000 grant from the National Institutes of Health's National Institute of General Medical Sciences and furthers research started in 2006 by Paul Craig, professor in the Department of Chemistry at RIT, and Herbert Bernstein, professor in the Department of Mathematics and Computer Science at Dowling College, in Long Island, N.Y.

Proteins are like nanomachines in the human body that carry out their missions—such as digesting fats and sugars in muscles, or carrying oxygen—and then rapidly degrade. Likewise, rogue proteins can do terrible things. A misshaped protein is like a hastily folded roadmap shoved into a glove compartment, leading to detours down roads of disease. Creutzfeldt-Jakob Disease, or mad cow disease, for instance, results from proteins misfolded into infectious prions.

Students from both universities work with Craig and Bernstein to identify the biological functions of each protein based on the shape or fold of its enzymes. Approximately 75,000 compounds in the Research Collaboratory for Structural Biology Protein Data Bank—a repository of

biological structures of proteins, nucleic acids and complex assemblies—have already been assigned functions. The RIT-Dowling team will investigate the structural coordinates of the remaining 3,000 proteins in the bank.

They will compare proteins from the bank with a library of 400 protein motifs associated with known functions from the Catalytic Site Atlas database, focusing on amino acids and their spatial relationships that dictate a protein's specific mission in the human body. The team will verify and statistically rate the matches using three-dimensional modeling and recommend two or three candidates as possible functions for each protein. The project uses existing active-site templates and new templates created by RIT students.

"We are concentrating on similarities between the active sites—the places where proteins interact with other molecules," Craig says. "Using active sites allows for much faster processing than finding structural similarity between the entire backbone of two proteins."

Craig is a biochemist who uses computers to address biological questions; Bernstein is a computer scientist who applies algorithms and database design to biological problems. They hope to produce two papers annually throughout the duration of the project.

"We want to be part of the conversation," Craig says.

"We collaborate on problems in molecular graphics and bioinformatics," Bernstein adds. "Our students actively contribute to our research and present their work at scientific meetings. The Internet allows us to function as a coherent group despite the distance between our campuses."