Mobile Interaction and Query Optimization in a Protein-Ligand Data Analysis System

Marvin Lapeine[*], Katherine G. Herbert[*], Emily Hill[*], Nina M. Goodey[+]

[*]Dept. of Computer Science
Montclair State University
Montclair, NJ, 07043
{lapeinem1, herbertk, hillem}@mail.montclair.edu

[+]Dept. of Chemistry and Biochemistry
Montclair State University
Montclair, NJ, 07043
goodeym@mail.montclair.edu

ABSTRACT

With current trends in integrating phylogenetic analysis into pharma-research, computing systems that integrate the two areas can help the drug discovery field. DrugTree is a tool that overlays ligand data on a protein-motivated phylogenetic tree. While initial tests of DrugTree are successful, it has been noticed that there are a number of lags concerning querying the tree. Due to the interleaving nature of the data, query optimization can become problematic since the data is being obtained from multiple sources, integrated and then presented to the user with the phylogenetic imposed upon the phylogenetic analysis layer. This poster presents our initial methodologies for addressing the query optimization issues. Our approach applies standards as well as uses novel mechanisms to help improve performance time.

Categories and Subject Descriptors

H.2.8 [**Database Applications**] – **Scientific Databases**

General Terms

Management, Measurement, Performance, Design, Reliability.

Keywords

Bioinformatics, pharmaceutical informatics, phylogenetics, data integration, query processing.

1. INTRODUCTION

The discovery of drugs with the desired pharmacological profiles is critical for human health and yet time consuming and expensive. The DrugTree Project creates a toolkit for scientists interested in understanding the broader implications of the relationship between phylogenetics and the binding between a homologous set of enzymes and their corresponding ligands and inhibitors. Phylogeny is a useful context in which to view these relationships: As a protein evolves, one feature that changes is the binding pocket and hence binding specificity. Consequently, evolutionary relationships can provide predictive power to establish the binding between a given ligand and a homolog based on known binding relationships within a protein family. The DrugTree Project has completed a prototype World Wide Web-based computing system that integrates both phylogenetic data and analyses about enzymes with known information about their

ligands and inhibitors. Currently, no one data repository integrates the drug-target, protein-ligand curated datasets with a large, popular protein database like UniProt and then gives tools to allow users to view these datasets in a phylogenetic-meaningful context. The DrugTree tool integrates data from UniProt (http://www.uniprot.org/) and the BindingDB (http://www.bindingdb.org/) databases to allow the user to create trees with data from both UniProt, with its massive non-redundant database, and the data from the known inhibitor repositories.

2. MOBILE PLATFORM

In studying usage of such a tool, the benefits of a mobile platform have become evident. While many practitioners have integrated technology readily into their experimental environment, many others are working in environments that are not conducive to a computational platform. Most researchers separate their "wet lab" activities from searching computational databases, especially in smaller labs. In polling our colleagues and collaborators who might be interested in this tool (about 10 scientists), none of them use laptops let alone desktops in their experimentation phase, preferring to complete experimentation and then add a computational phase after completion of the experimental phase. Also, data like the Ki, IC50 and Kd values are not always easily accessible, often requiring the researcher to do extensive literature searches to obtain these key values regarding the effectiveness of the ligand in inhibiting the protein. A mobile platform would allow for better integration of the experimental and data analysis activities in protein ligand binding in the lab while not needing the lab "real estate" that a laptop or desktop require. Our team has observed the usefulness of the model and is looking at possible extensions of this work into understanding invasive species studies and species survival studies in particular identified regions. When considering the environments this platform has the potential for working within, mobility becomes a key concern.

Considering the nature of this platform, where the user needs to see potentially massive data sets visualized with a histogram, both querying and visualization issues become concerns. When adding the invasive and survivability of species applications, these problems become compounded due to the data collection environment. For example, we can waterproof an iPad to go up to 9 feet in water with a solar powered charger as opposed to the challenges presented with water-proofing a laptop. Even the most computationally sophisticated platforms offer challenges in addressing this problem. Our contribution is to look at these problems, present our observations and the solutions we have attempted, including mechanisms for improved querying.