

Accelerating Drug Discovery: Methods for Effective Virtual Screening and Scaffold-Hopping

George Karypis

Department of Computer Science & Engineering

University of Minnesota

karypis@cs.umn.edu

<http://www.cs.umn.edu/~karypis>

Abstract

Discovering new drugs is an expensive and challenging process. Any new drug should not only produce the desired response to the disease but should do so with minimal side effects and be superior to the existing drugs in the market. One of the early steps in the drug development process is the identification of the chemical compounds (hits) that display the desired and reproducible behavior against the biomolecular target of interest. In recent, the development of high-throughput screening (HTS) technologies has made it possible to experimentally assay tens of thousands of compounds and has significantly enhanced our ability to quickly identify hit compounds. Though in principle, HTS techniques can be used to test each compound against every target, it is never practically feasible for the following reasons. First, the number of chemical compounds that have been synthesized or can be synthesized using combinatorial chemistry techniques is extremely large. Evaluating this large set of compounds using HTS can be prohibitively expensive. Second, not all biological assays can be converted to high throughput format. Third, in most cases it is hard to find all the desirable properties in a single compound and chemists are interested in not just identifying the hits but studying what part of the chemical compound leads to desirable behavior, so that new compounds can be rationally synthesized. As a result, we have witnessed the rapid development and wide-spread adoption of computational techniques that build models to correctly assign chemical compounds to various classes of interest or to retrieve potential drug-like compounds. These techniques are used extensively at various phases during the drug development process and are designed to supplement and enhance results obtained from HTS efforts.

In this talk we present our ongoing research in developing computational methods for chemical compound classification and retrieval. Specifically, the talk will focus on the problems of developing machine learning and retrieval techniques to quickly screen large compound libraries to identify those compounds that are most likely to exhibit the desired properties. A key challenge in developing effective computational methods stems from the fact that the properties of a chemical compound are strongly related to their chemical structure. We will present our work on developing sub-structure-based approaches that map the compound's molecular graph into a multi-dimensional descriptor space that captures the key properties of the underlying compounds. In addition, we will present our work on developing novel methods to determine the similarity between compounds that take into account additional information from the library's similarity network. We will show that these indirect similarity methods can effectively identify *target-hopping* compounds that are structurally diverse but still possess the same biological properties—a desired characteristic when trying to intelligently explore the vast chemical space towards the goal of identifying a diverse set of promising drug candidates. Finally, we will conclude our talk by describing some of the research opportunities and challenges as they relate to the emerging and exciting field of chemical genetics.