

WebDock: A Structure-Based Drug Discovery Web Service

Joseph Gill[‡], Legand Burge[‡], Zengjian Hu[†], William Southerland[‡]

^{‡‡} Department of Systems & Computer Science, Howard University Washington, DC 20059

[†] Department of Pharmacology, Howard University Washington, DC 20059

[‡] Department of Biochemistry & Molecular Biology, Howard University Washington, DC 20059

± jgill@howard.edu, ¥ blegand@scs.howard.edu, † zhu@scs.howard.edu, ‡ wsoutherland@howard.edu

ABSTRACT

High-Throughput Docking has emerged as a well established method for identifying lead components in the process of searching for new drugs. The High-Throughput Docking Service provides researchers with the ability to do drug screening using tools usually reserved for large research centers. The service provides web-based access to large clusters of computers and extensive searchable molecular libraries. Along with this, it provides the tools necessary for Homology modeling, visualization and animation. In this paper we discuss the design and implementation of the High-Throughput Docking Service. In particular, we discuss the architectural design and its integration with Dock using XML and J2EE.

Keywords

Java, Database, Visualization, Web Services, Docking.

1. INTRODUCTION

A crucial step for any new drug discovery project is the fast and cost-effective identification of novel lead compounds for specific biological targets. Currently two alternative, yet complementary strategies are commonly used to discover novel lead compounds for molecular targets [1]. One is the experimental high-throughput screening (HTS) of large chemical databases to identify promising potential ligands, and the other is computational structure-based screening (virtual screening) using high-throughput docking (HTD) [2]. HTD consists of the virtual docking of ligands from large compound databases to a macromolecular target site and selecting a limited number of promising candidate molecules for further testing in order to identify novel lead compounds that have the desired biological activity. [3-5]

Both approaches have proven to be very effective in identifying novel lead compounds, but HTD has some advantages relative to HTS. First, screening of a large chemical library of structurally diverse compounds through a high-throughput experimental approach is often very expensive and requires large physical storage space for the vast arrays of equipment, supplies, and products. Computational HTD is a very effective means of finding lead compounds at a relatively low cost since it can screen large numbers of chemically diverse drug-like molecules without the need to synthesize and experimentally test

every screened compound. Second, HTS often does not result in viable leads even with its considerable cost and technical complexity which provided further motivation for the use of the computational alternative. The primary goal of HTD is to rapidly screen large chemical databases in order to eliminate the large numbers of negatives and to discover more than one class of potential lead compounds for subsequent optimization. The total number of compounds for experimental testing via HTD is small compared to HTS. Consequently, HTD has emerged as a very powerful tool and a well-established method in the drug discovery process.

In an effort to make HTD more accessible to a broader research community, we have developed WebDock, a web-based user friendly graphical user interface which provides integrated access to web applications and ligand databases that may be used in the HTD process as illustrated in Figure 1 below. The purpose of this communication is to introduce WebDock, a new tool to aid in the performance of structure-based drug discovery tasks on via the internet.

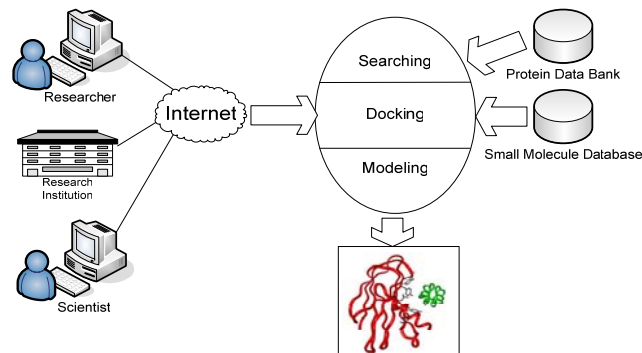


Figure 1. WebDock High-Level Overview

2. DOCKING PROCESS

The first requirement in HTD virtual screening is a searchable database of 3D small molecule structures. We have collected about 2 million chemical structures of both commercially and freely available compounds from various resources, and have subsequently converted them into 3D searchable databases. As shown in Figure 2, HTD investigations require a 3D structure of the biological target of interest. Structures determined experimentally by X-ray are generally preferred. The X-ray structures of some target proteins may be obtained from the Protein Data Bank (PDB)[7]. When the 3D structure of the desired target

protein is not available from either of these sources, it may be necessary to model the desired 3D structure with the aid of known structures of homologous proteins (templates) via homology modeling. A critical step in the development of a homology model is the alignment of the target and template sequences.

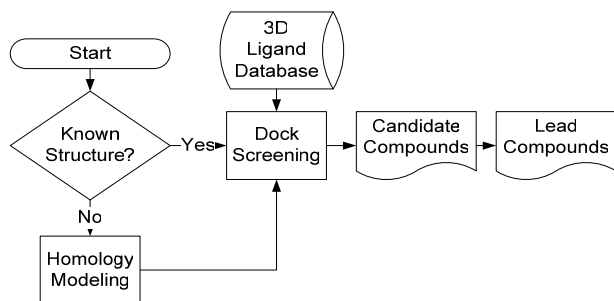


Figure 2. HTD Process

WebDock uses the widely distributed DOCK program [6], as the screening engine to perform HTD. The DOCK searching algorithm generates a set of spheres as the 'negative image' of the protein binding site. DOCK, uses an incremental ligand-search method to take into account the conformational flexibility of ligands, and a force-field-based scoring function to evaluate the binding affinity of each ligand.

3. DESIGN

WebDock is coded in a combination of C and Java and was designed to provide a structure-based drug discovery environment for the web. The typical structure-based drug discovery process using HTD requires 3D searchable small molecule libraries, homology modeling (if the 3D structure of the target protein binding site is not available), and ligand-protein docking. WebDock is capable of managing all these steps as a single process through its web interface, rather than the individual execution of each application. WebDock's interface was organized to follow the logical order of operations involved in the basic structure-based drug discovery procedure. Its modular software architecture allows smooth integration of small molecule searchable 3D databases, homology modeling tools, ligand-protein docking programs, and molecular visualization. We designed WebDock for scalability and high throughput utilizing a 16-node commodity cluster for executing the docking process.

4. PERFORMANCE

We have tested the screening efficiency of WebDock and it can be seen that it takes less than 2 seconds to process a compound. Screening a million compounds on a single cluster node is no longer a considerable effort. If 30 molecules per minute are processed on a single node, then ~43,000 molecules can be processed per day. If 5 Nodes are assigned to the problem, then approximately 215,000 compounds may be screened per day or approximately ~1.5 million compounds can be screened per week as shown

below in Figure 3. Therefore, HTD on clusters holds significant potential to contribute to HTD efforts.

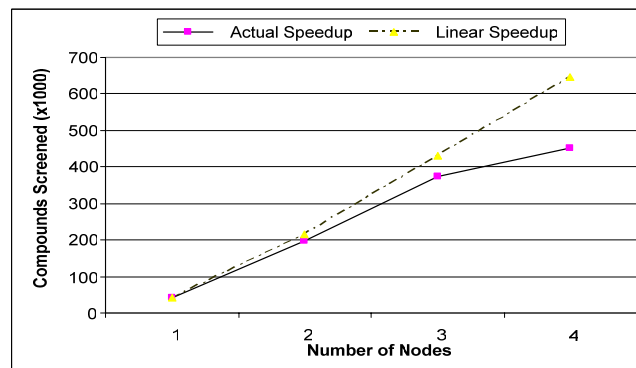


Figure 3. Performance Speedup

5. CONCLUSION

HTD has been applied in pharmaceutical research for almost two decades, providing an opportunity for the identification of active compounds [2]. We have presented here a new research tool, WebDock, which has been developed as a web-based application for structure-based drug design. It is not intended as a replacement for existing molecular modeling and computational chemistry programs, but instead is a general tool for both novices and experts and is applicable to a wide range of computer-aided drug discovery applications. We anticipate that WebDock will provide a convenient tool for structure-based drug discovery and aid in the performance of HTD investigations.

6. REFERENCES

- [1] Bajorath, J. Integration of virtual and high-throughput screening. *Nature Rev. Drug Discov.* 2002, 1, 882–894.
- [2] Bredel, M.; Jacoby, E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.* 2004, 5, 262-275.
- [3] Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening — an overview. *Drug Discov. Today* 1998, 3, 160–178.
- [4] Langer, T.; Hoffmann, R. D. Virtual screening: an effective tool for lead structure discovery. *Curr. Pharm. Design* 2001, 7, 509–527.
- [5] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking & scoring in virtual screening for drug discovery: methods & applications. *Nat. Rev. Drug Discov.* 2004, 3, 935-949.
- [6] Sadowski, J.; Schwab, C.H.; Gasteiger, J. 3D Structure Generation and Conformational Searching. In *Computational Medicinal Chemistry and Drug Discovery*; Bultinck, P.; DeWinter, H.; Langenaeker, W.; Tollenaere, J. P. Eds.; Dekker Inc.: New York, 2003.
- [7] Berman, H. M; Bhat, T. N.; Bourne, P. E.; Feng, Z.; Gilliland, G.; Weissig, H.; Westbrook, J. The protein data bank and the challenge of structural genomics. *Nature Struct. Biol.* 2000, 7, 957–959.