Chapter 6.7
A Graphical Workflow Modeler for Docking Process in Drug Discovery

Qiang Wang
Harbin Institute of Technology, China

Yunming Ye
Harbin Institute of Technology, China

Kunqian Yu
Chinese Academy of Science, China

Joshua Zhexue Huang
University of Hong Kong, China

ABSTRACT

A drug discovery process is aimed to find from a large set of molecules the candidate leads that have strong interaction with the target proteins. The process of drug discovery is characterized by its complexity in data and computation. A useful tool to simplify the handling of intensive data and complex algorithms is necessary for domain scientists to build proper drug discovery procedures, carry through the data intensive computation tasks and produce fruitful results. This chapter presents a graphical workflow modeler for domain scientists to perform drug discovery tasks on high performance grid computing grid platforms. A client/server system is described as the platform for implementation of the graphical workflow modeler. A case study on drug discovery for avian influenza virus is given to demonstrate the use of this tool in drug discovery research.

INTRODUCTION

Scientific research today has changed its way. Computational science, which focuses on simulation of complex phenomena in the last few decades, is being replaced today by e-science or data-centric science where scientific discoveries often result from data intensive computation and analysis. In this area, Grid computing is playing a key role, owing to its characteristic that it allows geographically distributed scientists to work col-

DOI: 10.4018/978-1-4666-0879-5.ch6.7

laboratively in a networked environment and share resources and expertise to solve, in a large scale, difficult scientific and engineering problems. The typical scientific research areas cover a wide range including gravitational-wave science (Deelman, Kesselman, Mehta, Pearlman, Blackburn, Ehrens, Lazzarini, Williams, & Koranda, 2002), neuroscience (Lathers, Su, Kulungowski, Lin, Mehta, Peltier, Deelman, & Ellisman, 2006), astronomy (Katz, Anagnostou, Berriman, Deelman, Good, Jacob, Kesselman, Laity, Prince, Singh, Su, & Williams, 2006), high energy physics (Deelman, Blythe, Gil, & Kesselman, 2004), etc. All are distinctly characterized with complex research problems that involve huge amount of data and requires a lot of computing resources to analyze.

Drug discovery is another such area that grid computing is used to carry out various computational processes for identification of active compounds against a given target from a large number of chemical compounds in molecule databases (Richards, 2002; Ren, Zhang, Wan, Huang, Xie, & Yang, 2006). In modern drug discovery, molecule docking is an important process which aims to determine the candidate lead components that have the strongest interaction with the target component by a series of complex computations. Docking process involves multiple steps, including 3D molecule modeling and representation, search of molecule databases for candidate molecules that best match the receptor structure, evaluation of candidate molecules with scoring functions, conformation determination, hit identification, and lead optimization. In each step, multiple techniques and algorithms are developed, and different techniques are implemented in different systems for drug discovery research and applications. Nevertheless, these techniques and algorithms are used in setting parameters, integrating data sources and performing data analysis tasks. It is time-consuming and costly for domain scientists to learn and master multiple techniques and systems to solve drug discovery problems. Therefore, development of an easy-to-use collaborative workflow modeler is needed by drug discovery practitioners as part of their sophisticated problem solving environment.

In this chapter, we present a software platform and a graphical workflow modeler for drug discovery. The goal of our work is to develop an effective and efficient system that supports drug discovery practitioners to build docking processes for virtual high throughput screening (VHTS) (Yoon, 2005) of potential inhibitors through drag-and-drop and connecting operations in a graphical interface with a mouse. In this graphical working environment, a docking process is presented as a workflow consisting of a set of functional nodes connected in a directed acyclic graph (DAG). Each function node is an implementation of a particular algorithm to perform one docking function such as search or scoring. When a workflow is formed and all parameters of each function node are set, the workflow is executed through a workflow engine that schedules the execution of each function in the underlying computing infrastructure, e.g., a grid computing infrastructure (Foster, & Kesselman, 1998; Berman, Hey, & Fox, 2003). Such a system could provide an efficient and easy way for modeling and managing scientific processes of experimental investigation, evidence accumulation and result validation. Processes themselves can then be modified, reused and shared through collaboration between interdisciplinary scientists.

The rest of this chapter is organized as follows. Section 2 gives the background information on China's Drug Discovery Grid (DDGrid[1]) project on which this work is rooted. Section 3 presents the client/server platform on which the graphical workflow modeler is implemented. The system architecture and functionality are described. The workflow patterns for supporting drug discovery processes are defined. Section 4 discusses implementation of the graphical workflow modeler from the viewpoint of users who would use this tool to carry out drug discovery research. A case study is presented in Section 5 to show how to use this tool to carry out a drug discovery process for