Drug Discovery using Grid Technologies

Yuichiro Inagaki
Biotechnology division
Fuji Research Institute co.
Outline

- Needs for grid technologies in drug discovery
- g-Drug Discovery system
- Test calculation results
Needs for grid technology in drug discovery

- Increase in number of both drug candidate compounds and target
  - $10^7$ molecules $\times 10^3$ conformations
  - screening throughout a family: Kinases, GPCRs,…

- Various type of calculation
  - Druglikeness screening
  - ADME/Tox screening
  - Conformational search
  - Pharmacophore screening
  - Docking
  - Molecular Orbital methods
“g-Drug Discovery”

- Funded by Japan Science and Technology Agency (JST)

Components
- DB system
- Conflex-G
- Xsi-G
- REMD
- FMO
DrugML
a XML Schema for drug design

- Use tags of CML as much as possible
- Conformers
- Complex
- Descriptors
Omni-RPC
a Grid RPC system for Parallel Programming

- Supports typical master-worker grid applications such as docking simulation.
- Users can use the same program for both clusters and grids.
- Supports a local environment with "rsh", a grid environment with Globus, and remote hosts with "ssh".
- OmniRPC inherits its API from Ninf, the programmer can use OpenMP for easy-to-use parallel programming because the API is designed to be thread-safe.
- For a cluster over a private network, an agent process running the server host functions as a proxy to relay communications between the client and the remote executables.
Xsi 2.0

- Combines Ligand Based Drug Design and Structure Based Drug Design
- Montecarlo, minimization, docking by MMFF94s force field
- 2D & 3D descriptors
- Statistics, Clustering, Similarity
- Machine Learning by support vector machine
LigandAlignment

- Optimizes similarity between pharmacophore map and ligand
- Pharmacophore map can be defined by physico-chemical properties and voids
  - VDW, hydrophobicity, HD, HA, aromaticity, electrostatic...
- 0.6 sec/1 alignment (viracept)
Map of binding site

HIV protease and inhibitor (DMP323)
Alignment onto binding site

Binding Site Map

Alignment of JG-365
Pharmacophore screening by LigandAlignment

97 random compounds + 5 known HIV protease inhibitors

- Hit rate (10% ranked DB) ~ 50%
Docking flow

Viracept

- Docking flow
- MonteCarlo
- Finding binding site
- Calculate WHIMs
- Calculate WHIMs
- Calculate similarities
  Between ligand and pocket
- Sort ligands by similarities
- Alignment by using WHIMs

Master

Workers

- Docking
- Docking
- Docking
- Docking
Calculation Environment

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total: 1 master + 71 workers
Speedup of calculations
Docking results

X-ray (yellow)
Comp.(white) RMSD: 1.77
Summary

- Hit rate more than 50% can be achieved
- Protein family screening
- LigandAlignment on grid necessary
Acknowledgements

- Hiroshi Chuman (Tokushima Univ.)
- Umpei Nagashima (AIST)
- Hitoshi Goto (Toyohashi Univ. of Technology)
- Mitsuhisa Sato (Tsukuba Univ.)
Back up Slides
リガンドアライメントによるファーマコフォアマップの最適化の様子。
(r)参照分子(ベンゼン)のMS(原子質量)のマップ
(c1)候補分子(トルエン)の最適化前のMS(原子質量)のマップ
(c2)候補分子(トルエン)の最適化後のMS(原子質量)のマップ

トルエン分子のベンゼン環の配置がベンゼン分子のベンゼン環の配置に近くなるように最適化されている。図はマップの等数値面を描いたもの。格子点数は32*32*32。
1. Introduction
A number of computer resources, such as CPUs and storages, can be connected over networks to construct a huge virtual computing environment using grid technologies. Our project “og-Drug Discovery” aims to develop a platform for drug design using grid technologies, on which various analysis and calculations are conducted, such as molecular mechanics method, replica exchange method, docking with proteins, molecular orbital method, and 3-dimensional quantitative structure activity relationship. In this poster we will present the following things:

- DrugML: The markup language for drug discovery
- Database system for drug discovery
- Docking calculations using grid technology
- Ligand-receptor docking simulation using “Xsi” and “OmniRPC”.

2. DrugML (Drug Markup Language)
DrugML (Drug Markup Language) is the markup language for drug discovery whose specification has been decided upon newly by our project. It is defined by XML Schema [10], so we can validate it’s file strictly by using the existing XML parser (such as Xerces [9]). DrugML imports tags from CML (Chemical Markup Language) [8] as much as possible.

- Tag “universe” can represent the snapshot of two or more molecules.
- Tag “conformation” can represent the 3-dimensional structure of molecule and universe.

3. DB system for drug discovery
We have developed database system in Fig.1, which adopted DrugML as the data structure to store. Fig.3 is the abstract structure of the system and Fig. 4 is the framework which enables us to exchange the kind of database easily. We assume the native XML database, but it has not had standard query language such as SQL and the way of connection such as ODBC (Open DataBase Connectivity) of relational database (RDB). This framework is able to absorb those differences of each XML database. We have implemented this framework by C++ and native XML database Xindice [9] (Fig.5).

4. Docking calculations using grid technologies
We have performed docking calculations between viracept (Fig.7) and HIV protease, using Xsi and OmniRPC. It is only the conformation of HIV protease that we assumed, and neither the 3-dimensional structures of viracet nor the place of binding site with HIV protease assumed. Fig. 4 shows the flow chart of calculations. We generated 10000 initial conformations of viracept by Monte-Carlo methods and found binding site of receptor by using ProShape [4]. Each conformation are aligned by using WHIM descriptors [5][6]. Our experiment’s environment is described in Table 1, and The result of calculations is showed in Fig. 9 and Fig. 10.

- Grid RPC system
  OmniRPC is a grid RPC system which enables seamless parallel programming in cluster and grid environment ([1][2][3],Fig.6).

- Suite for virtual screening
  Xsi (ku-su-shi, [7]) is a suite for virtual screening based on Molecular Mechanics (MM) which has been developed by us. This is the following features
  - Exhaustive Search, Monte-Carlo simulation, and Docking simulation based on Molecular Mechanics
  - Various descriptor, similarity, clustering, superimpose
  - Composite of two conformations is 1.77 Å. The actual time was based on the time measured by one node of Xsi.

References

Fig. 1. Platform for drug discovery
Fig. 2. Data structure of DrugML “Any tag of cml:molecule” means that any element under “molecule” of CML may exist in that place.
Fig. 3. Abstract structure of DB system
Fig. 4. Framework of DB System.
Fig. 5. Implementation of DB system
Fig. 6. OmniRPC
Fig. 7. Viracept, which is a kind of HIV protease inhibitor
Fig. 8. Flow chart of calculations
Fig. 9. (Left figure) Speedup of calculations. This graph is obtained by 10 initial conformations (that is, each execute 240 docking simulations). The contents of the parenthesis of the X-axis express the number of CPUs. The Speedup was based on the time measured by one node of Xsi.
Fig. 10. (Right figure) Comparison of viracept of computation (white line) and X-ray (yellow line). RMSD of two conformations is 1.77 Å. (Left figure) Complex of HIV protease of X-ray (stick line) and the calculated viracept (line and ball).
Flow of Screening

- 3D structure generation by Conflex-G
- Screening by pharmacophore (Xsi)
- Docking (Xsi-G)