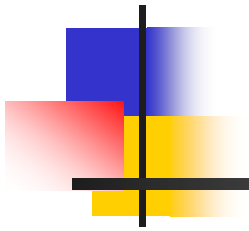
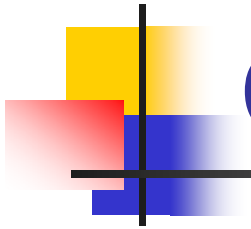


Drug Discovery using Grid Technologies

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

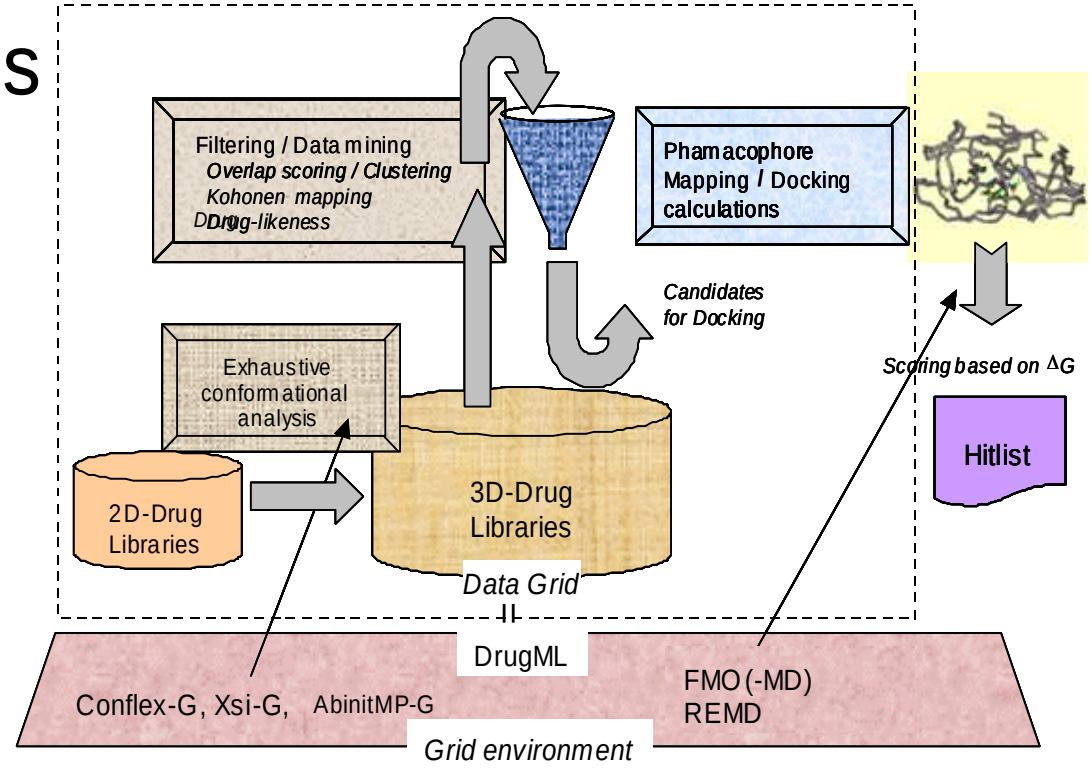
- Increse in number of both drug candidate compounds and target
  - 10^7 molecules ✕ 10^3 conformations
  - screening throughout a family: Kinases, GPCRs,…
- Various type of calculation
  - Druglikeness screening
  - ADME/Tox screening
  - Conformational search
  - Pharmacophore screening
  - Docking
  - Molecular Orbital methods

More CPU Power

Seemless connection

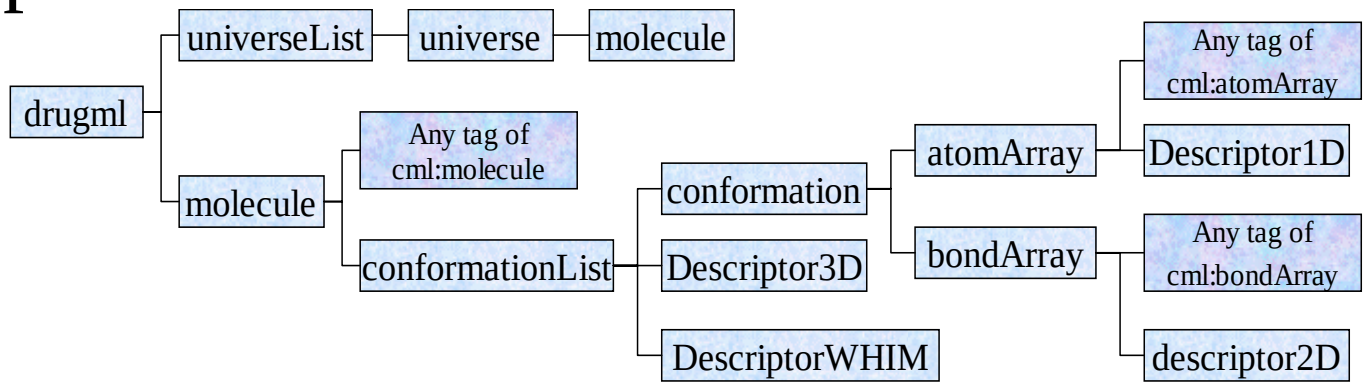
“g-Drug Discovery”

- Funded by Japan Science and Technology Agency (JST)
- Components
  - DB system
  - Conflex-G
  - Xsi-G
  - REMD
  - FMO

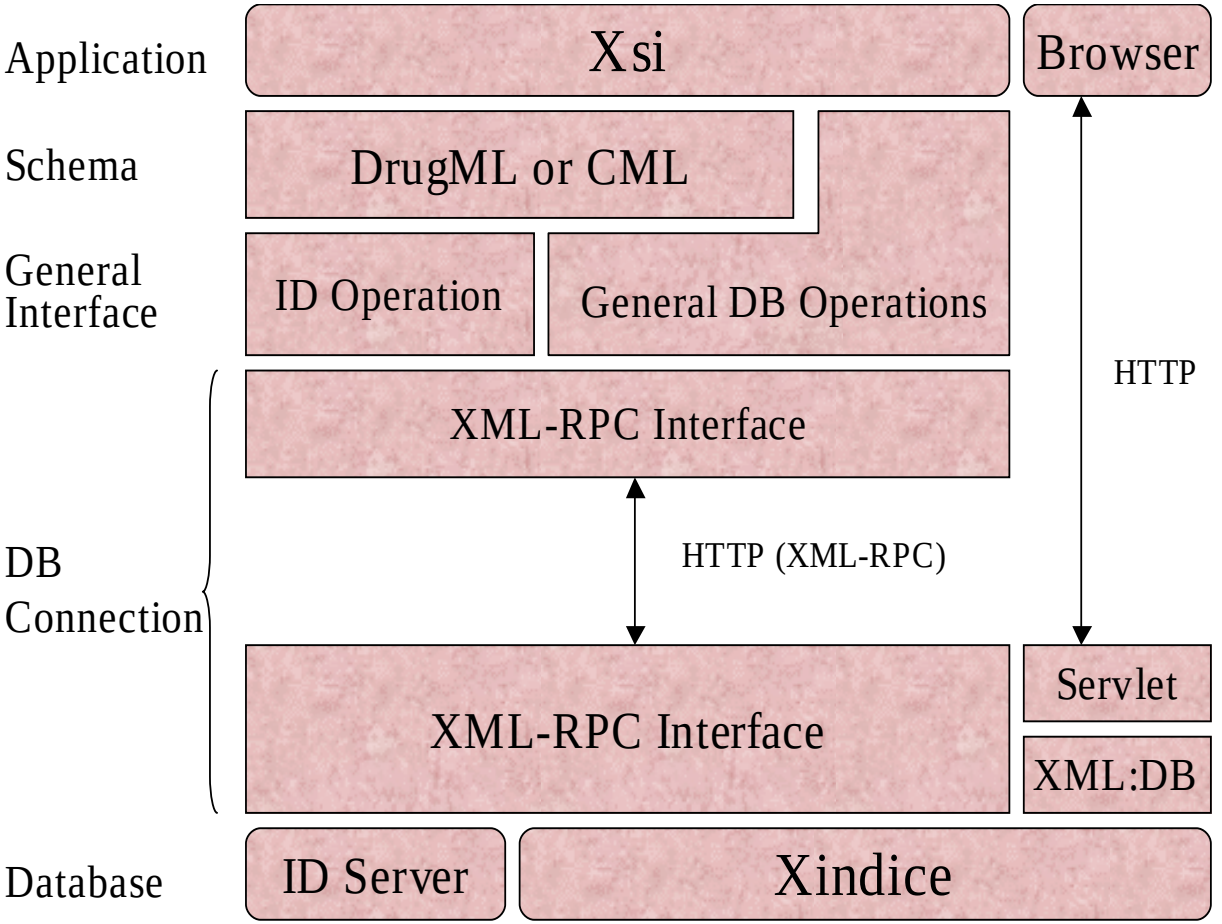


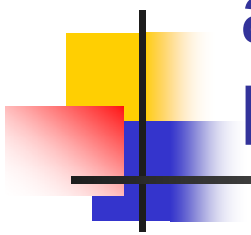
DrugML
a XML Schema for drug design

- Use tags of CML as much as possible
- Conformers
- Complex
- Descriptors



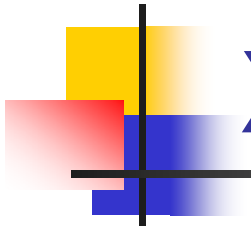
Structure of DB system





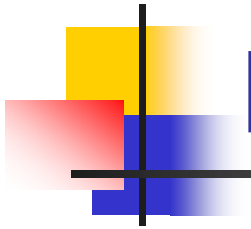
Omni-RPC
a Grid RPC system for Parallel Programming

- 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



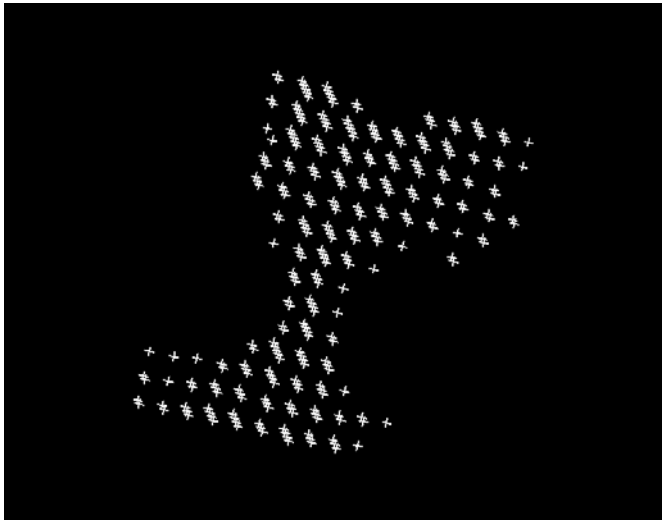
LigandAlignment

- Optimizes similarity between pharmacophore map and ligand
- Pharamacophore 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

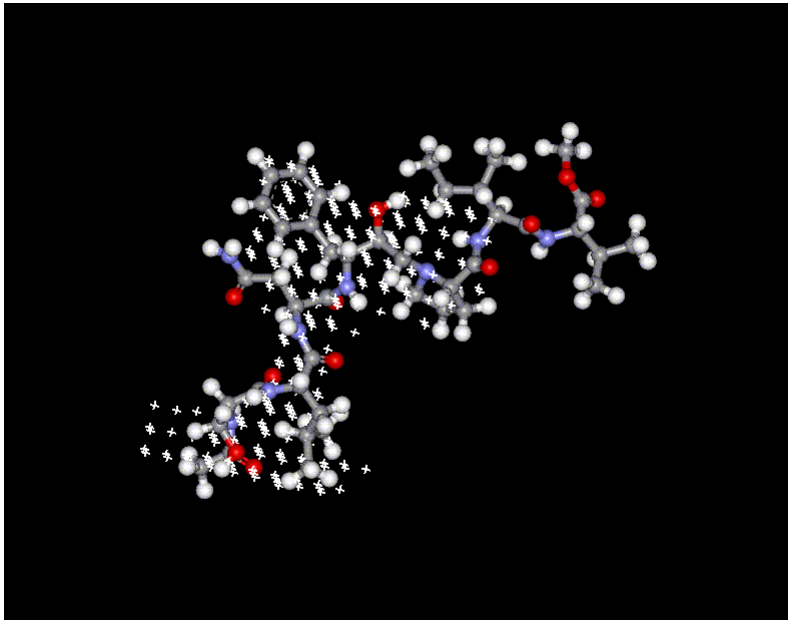
HIV protease and inhibitor (DMP323)

Alignment onto binding site



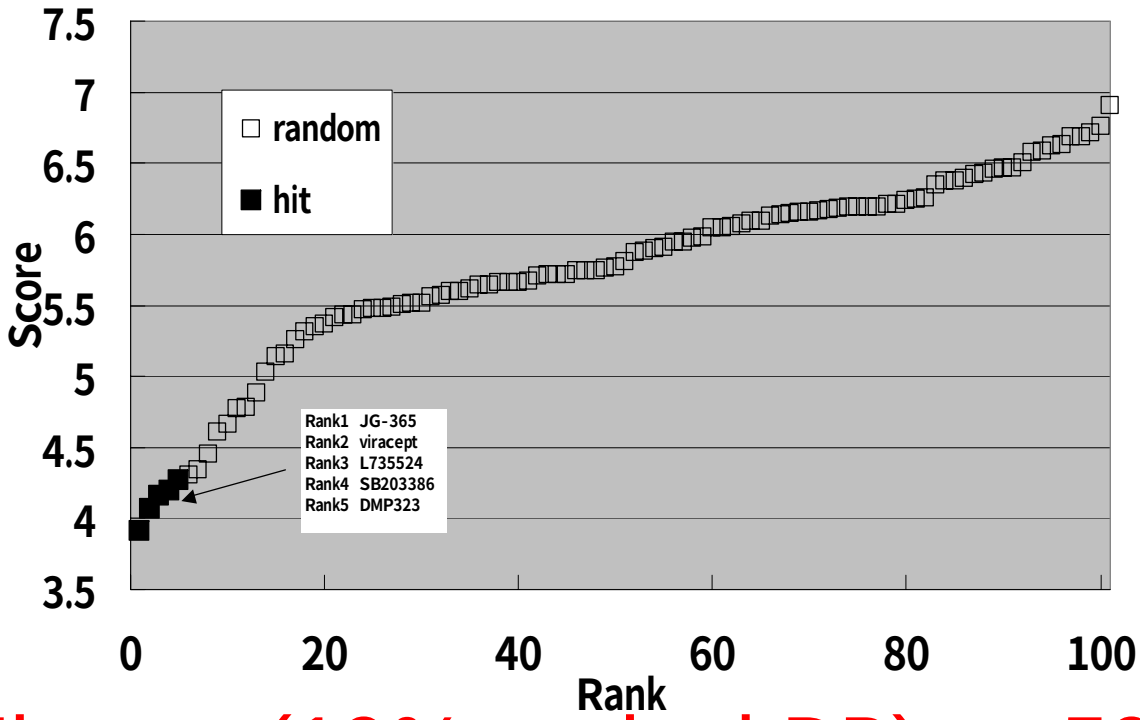
Binding Site Map

Alignment of JG-365



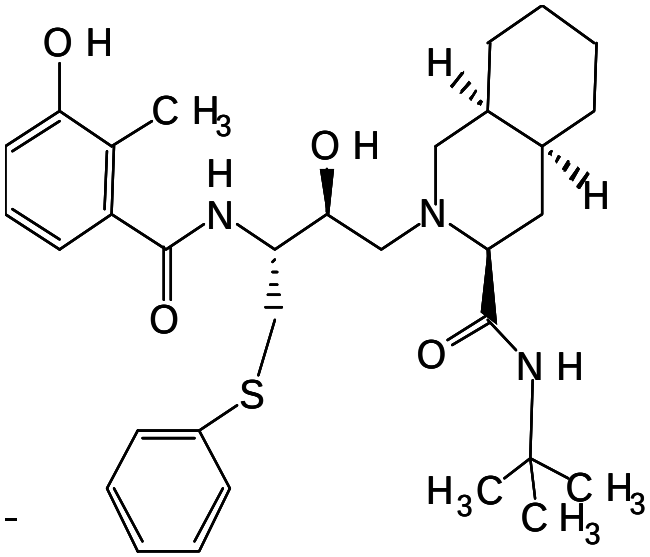
Pharmacophore screening by LigandAlignment

97 random compounds + 5 known HIV protease inhibitors

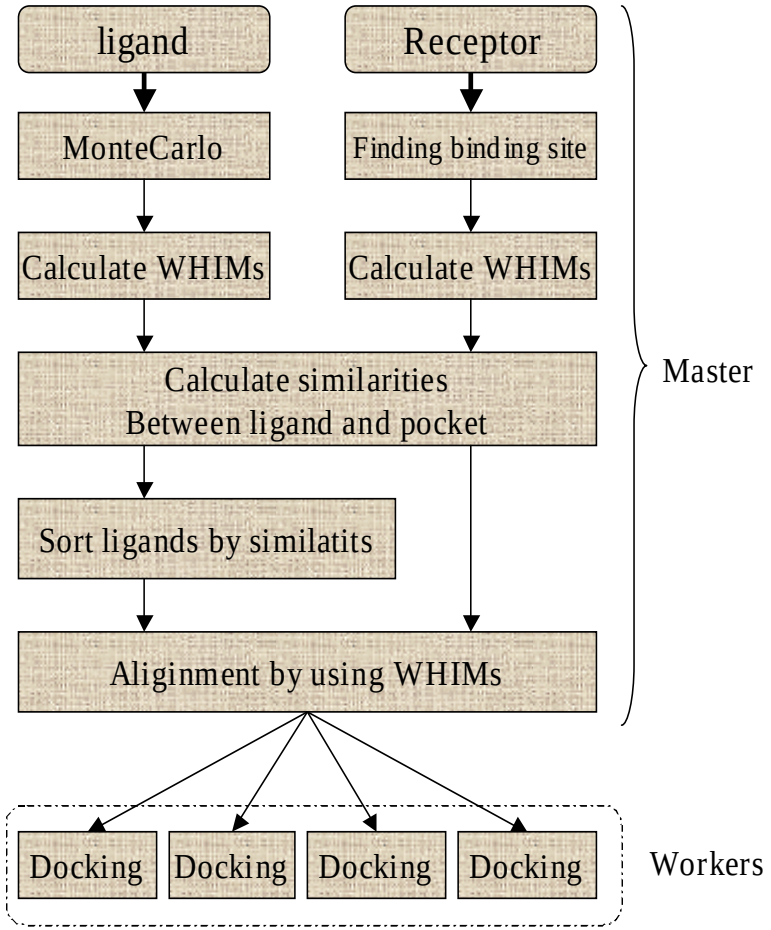


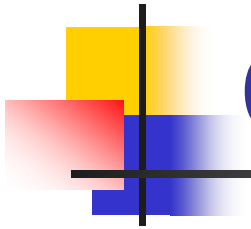
- Hit rate (10% ranked DB) ~ 50 %

Docking flow



Viracept



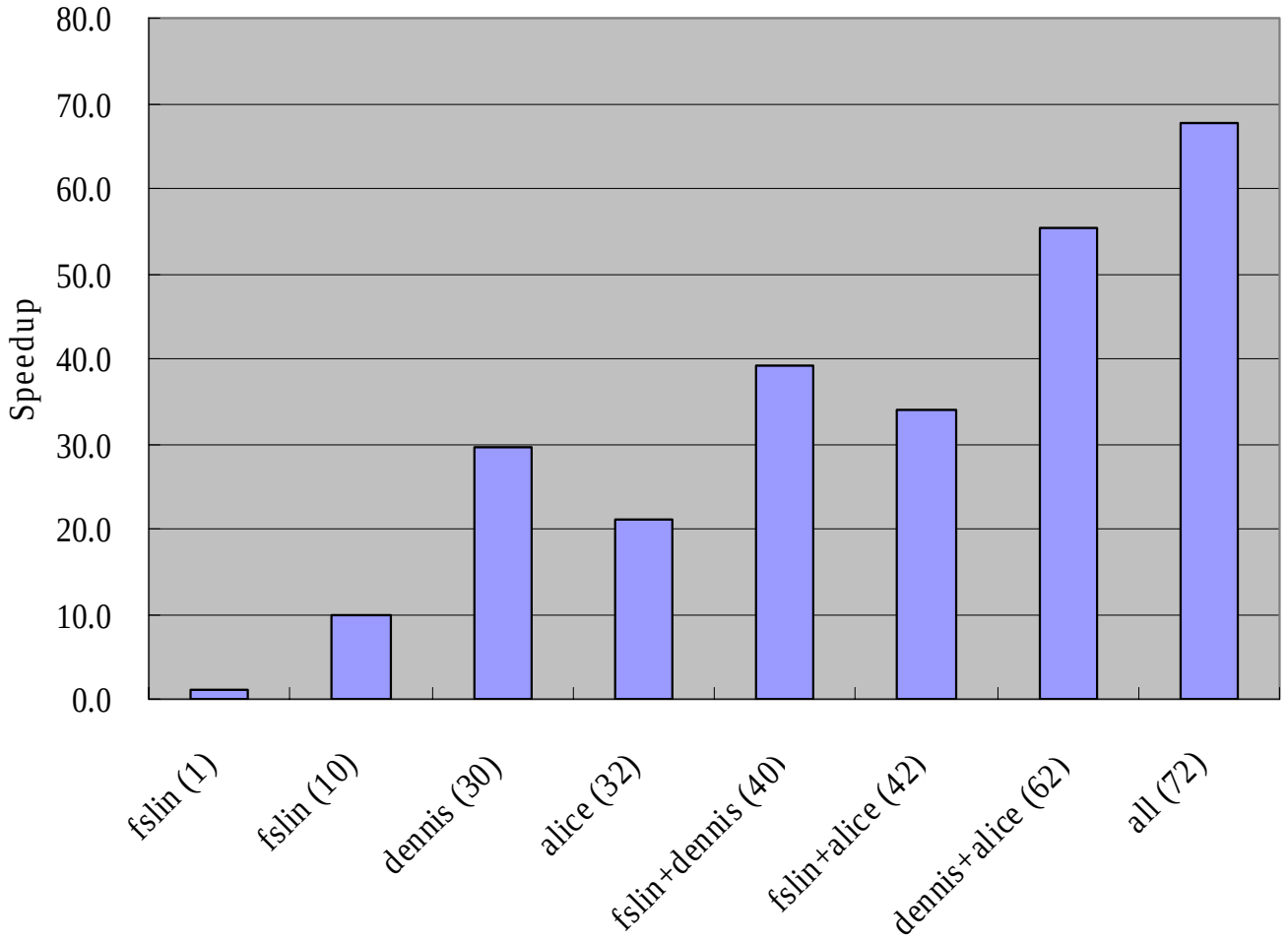


Calculation Environment

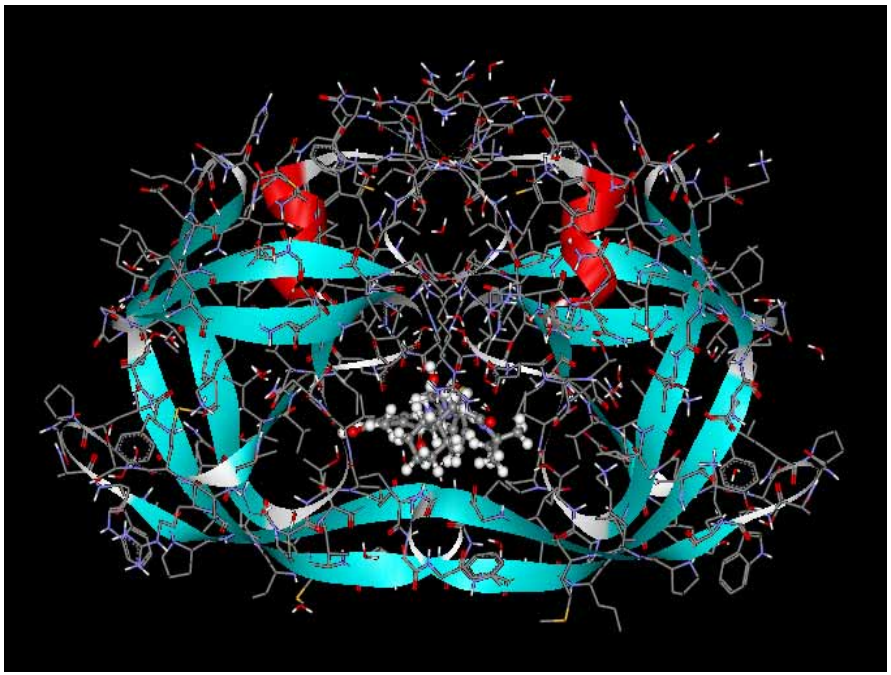
	location	CPU	number of nodes	RTT(ms)
fslin	Fuji-RIC (Tokyo)	Dual Xeon 2.4GHz	5	0.019
dennis	Tsukuba university (Tsukuba)	Dual Xeon 2.4GHz	10	27.2
		Dual Xeon 3.0GHz	5	
alice	Tsukuba university (Tsukuba)	Dual Athlon 1800+	16	27.2

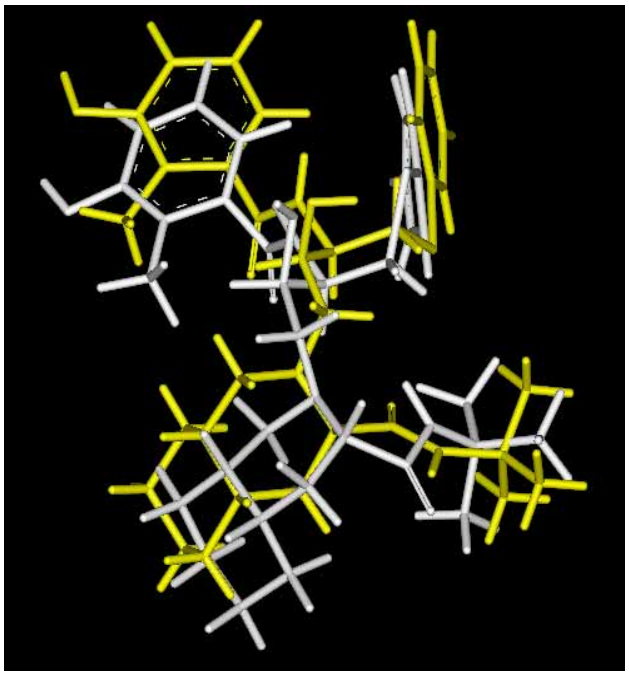
Total : 1 master + 71 workers

Speedup of calculations



Docking results





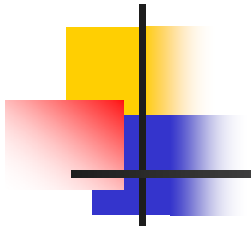
X-ray (yellow)
Comp.(white) RMSD:1.77 Å

Summary

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

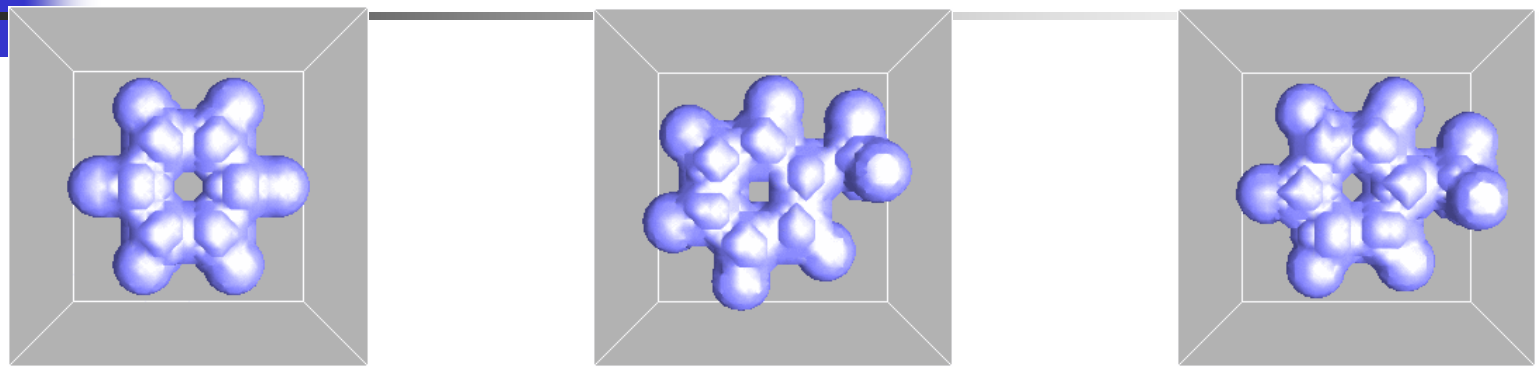
Acknowledgements

- Hiroshi Chuman (Tokushima Univ.)
- Umpei Nagashima (AIST)
- Hitoshi Goto (Toyohashi Univ. of Technology)
- Mitsuhisa Sato (Tsukuba Univ.)



Back up Slides

LigandAlignment



(r)　　(c1)　　(c2)

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。

Drug discovery using grid technologies and DrugML

CINF 47

Michiaki Hamada[1], Yuichiro Inagaki[1], Hitoshi Goto[2], Umpei Nagashima[3] , Shigenori Tanaka[4], and Hiroshi Chuman[5]

[1]Fuji Research Institute Corporation, [2]Toyohashi University of Technology, [3]National Institute of Advanced Industrial Science and Technology, [4]Toshiba Research and Development Center, and [5]Tokushima University

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 **"g-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** … Our database system which stores 3-dimensional structure of molecule
- **Docking calculations using grid technology** … Ligand-receptor docking simulation using "Xsi" and "OmniRPC".

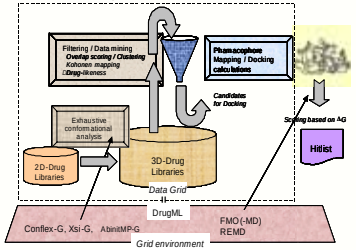


Fig. 1. Platform for drug discovery

References

[1] Mitsuhisa Sato, Taisuke Boku, Daisuke Takahashi, OmniRPC:a Grid RPC system for Parallel Programming in Cluster and Grid Environment 3rd International Symposium on Cluster Computing and the Grid (CCGrid2003), May 12 - 15, 2003, Tokyo, Japan.

[2] Mitsuhisa Sato, Motonari Hirano, Yoshio Tanaka, Satoshi Sekiguchi, OmniRPC: A Grid RPC Facility for Cluster and Global Computing in OpenMP WOMPAT 2001, 130-136.

[3] http://www.omni.hpcc.jp/OmniRPC/index.html.en

[4] http://csb.stanford.edu/koehl/ProShape/

[5] Todeschin, R. and Gramatica, P., New 3D Molecular Descriptors: The WHIM theory and QSAR Applications, In 3D QSAR in Drug Design Volume 2, Eds., Kubinyi, H., Folkers, G and Martin, Y.C., 355-380, KLUWER/ESCOM, Dordrecht, 1998.

[6] R. Todeschini, M. Lasagni, and E. Marengo, "New Molecular Descriptors for 2D and 3D Structures. Theory", J. Chemometrics, 8, pp.263-272, (1994).

[7] http://www.fuji-ric.co.jp/st/xsi/index.html

[8] http://www.xml-cml.org/

[9] http://xml.apache.org/

[10] http://www.w3.org/XML/Schema

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.
- **Tag "descriptor*"** can represent various descriptors of molecule, such as WHIM (Weighted Holistic Invariant Molecul descriptors.

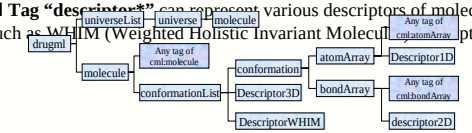


Fig. 2. Data structure of DrugML. "Any tag of cml:molecule" means that any element under "molecule" of CML may exist in that place.

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).

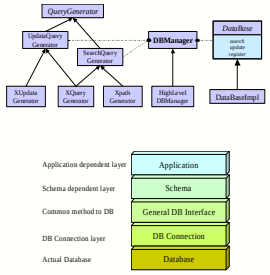


Fig. 3. Abstract structure of DB system

Fig. 4. Framework of DB System.

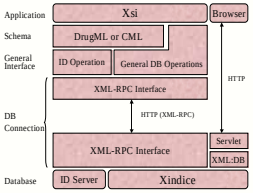


Fig. 5. Implementation of DB system

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 viracept 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 MonteCarlo methods and found binding site of receptor by using ProShape [4]. Each conformations 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) .

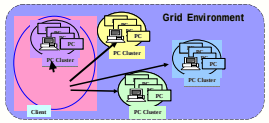


Fig. 6. OmniRPC

	location	CPU	number of nodes	RTT(ms)
fslin	Fuji -RIC (Tokyo)	Dual Xeon 2.4GHz	5	0.519
dennis	Tsukuba university (Tsukuba)	Dual Xeon 2.4GHz	16	27.2
		Dual Xeon 3.0GHz	5	
alice	Tsukuba university (Tsukuba)	Dual Athlon 1800+	16	27.2

Table 1. Experiment's environment. The master node is the top node of fslin.

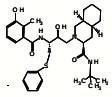


Fig. 7. Viracept, which is a kind of HIV protease inhibitor

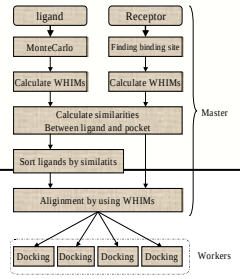


Fig. 8. Flow chart of calculations

- **Suite for virtual screenin**

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, MonteCarlo simulation, and Docking simulation based on Molecular Mechanics

✧ Various descriptor, similarity, clustering, superimpose

✧ Configurable by script

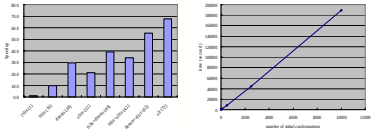


Fig. 9. *(Left figure)* Speedup of calculations. This graph is obtained by 10 initial conformations (that is, each execute 240 docking simulations). The contents of the parenthesis of the X-axis express the number of CPUs. The Speedup was based on the time measured by one node of fslin. *(Right figure)* The time which calculations took. The actual number of docking calculations is three times of the number of initial conformations.

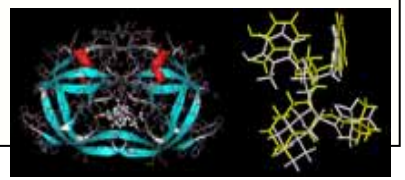
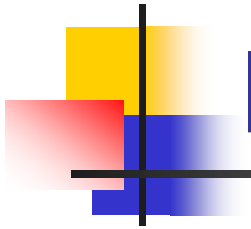


Fig. 10. *(Right figure)* Comparison of viracept of computation (white line) and X-ray (yellow line). RMSD of two conformations is 1.77 Å. *(Left figure)* Complex of HIV protease of X-ray (stick line) and the calculated viracept (line and ball).



Flow of Screening

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