GVSS: A new virtual screening service against the epidemic disease in Asia

Hsin-Yen Chen
ASGC
ISGC 2010
Virtual screening Activities

• Look for the favorable drug leads to combat Avian Flu and dengue fever viruses by the in silico simulation
• Fostering biomedical grid activity and e-Science collaboration in Asia
Influenza A pandemic

H1N1 1918 1957
H1N1 1933
H2N2 1968
H3N2 1977
H1N1 1997
H9N2 1999
H5N1 2003
H7N7 2004
H5N1 2005

113 deaths/204 cases

Apr 21, 2006

http://www.who.int/csr/disease/avian_influenza

FP7-INFRA-223791
Distribution of dengue fever in the world

Map produced by the Agricultural Research Service of the US Department of Agriculture
• **Screening** is the first measure to take for the biological activity of each compound in a large compound collection against a disease target.

• **High-Throughput Screening (HTS)** is the process of assaying a large number of potential effectors of biological activity against targets.

---

**Drug Discovery**

- Target selected
- Assay developed
- HTS (High Throughput Screening)
- HTS hits confirmed
- Chemistry begins
- Target structure obtained
- Development candidates is taken forward

2-4 years
Why we need Screening

$10^{100}$

Known Compounds: $10^7$
Virtual Screening

- **Virtual Screening** is defined as “automatically evaluating very large libraries of compounds” using computer programs.

- can be put in the early phase of the drug discovery procedure to shorten the obtainment of useful information of active compounds.

- can be used to funnel out the majority of less potential from a huge chemical space or in-house collections.

- can be used to study effects of mutation and provide information for the decision of assay development and medicinal chemistry in drug discovery procedure for the sake of performance.
Why Grid?

- Grid can provide
  - Large scale and on-demand resources
    - Computing resources (computational grids)
    - Storage resources (data grids)

- Grid can shorten the time of process of virtual screening.
  - Grid-enabled massive molecular docking
    - Massive molecular docking is one of the time consuming process in the virtual screening.
    - Massive molecular docking requires intensive computing resources.
    - Massive molecular docking is a type of high throughput application and can be divided into many independent tasks.
• **Features**
  
  – **Grid-enabled massive molecular docking**
  – Emulating wet-lab refinement simulation process.
    – Initial docking -> filtering -> refinement docking
  
  – **Friendly graphical user interface**
    – Graphical user login/logout and proxy delegation.
    – Graphical configuration for docking job submission.
    – One-click job submission
    – Easy and visualizable job monitoring.
  
  – **Docking result visualization**
    – Visualize the best complex of the docking.
    – Visualize the ligand conformation.
    – Generating histogram for the docking result.
300,000 compounds => 2000 focus library => 123 potential hits => more analysis in wet lab.

**WISDOM**

- Total number of completed dockings: $2 \times 10^6$
- Estimated duration on 1 CPU: 88.3 years
- Duration of the experience: 6 weeks
- Cumulative number of the Grid jobs: 54,000
- Max. number of concurrent CPUs: 2,000
- Crunching rate: 912
- Approximated distribution efficiency: 46%
- Approximated throughput: 2 sec/docking

**DIANE/GANGA**

- Total number of completed dockings: 308,585
- Estimated duration on 1 CPU: 16.7 years
- Duration of the experience: 4 weeks
- Cumulative number of the Grid jobs: 2580
- Max. number of concurrent CPUs: 240
- Crunching rate: 203
- Approximated distribution efficiency: 84%
- Approximated throughput: 10 sec./docking

> 100 CPU years in 6 weeks on the EGEE grid
• HTC via Grid Application Platform
A JAVA-based user interface was developed to facilitate job submission and data management of large-scale molecular docking on the GRID platform.

Thanks to DIANE framework and GAP applications, the GVSS allows submitted jobs to be split into multiple independent subtasks and run to complete, which has the potential of efficient utilization of the availability of the GRID resource for the massive molecular dockings empowered by Autodock3.

The GVSS hides the complexity of deploying large-scale molecular docking on the GRID while provides users more flexible control over their docking jobs on the GRID.

Provide on-line Avian Flu & Dengue targets and ZINC compounds library (> 300,000).

User prepare the himself target/compounds gridmap files with on-line tools.
GAP Virtual Screening Service

Worker Agent submission
Ask for a TASK and Return the Result
Access application metadata
Access application metadata
Download the output and visualize the result
Virtual Screening Service Client GUI

EUAsiaGRID

GridFTP Storage

DIANE Master
Agent Factory
Tasks Queue

diiane

GAP Grid Application Platform

FP7-INFRA-223791
Dengue virus

http://en.wikipedia.org/wiki/Aedes

J. Lescar et al. / Antiviral Research 80 (2008) 94–101
Dengue NS3 protease
Pilot Run Results to validate the parameters
Joint Computing Resources & Users

• Accumulating Computing Resources in EUAsia VO: 268 cpu-cores (100 – ASGC(TW), 2 – TH, 4 - VN, 18 – MIMOS(MY), 80 – UPM(MY), 64 - CESNET(CZ))
  – lcg-infosites --vo euasia ce

• Registered VQS account:
  – 6 users (TW)
  – 17 user (PH, 15 in AdMU, 2 in ASTI)
  – 2 user (TH, 1 in NECTEC, 1 in HAIL)
  – 1 user (MY, UPM)
  – 1 user (ID, ITB)
  – 2 user (VN, IAMI)
  – 1 user (FR, HealthGrid)
### Dengue Fever Data Challenge / resources & 1st result

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of completed docking jobs</td>
<td>300,000</td>
</tr>
<tr>
<td>Estimated needed computing power</td>
<td>4,167 CPU*days</td>
</tr>
<tr>
<td>Duration of the experiment</td>
<td>60 days</td>
</tr>
<tr>
<td>Cumulative computing results</td>
<td>42.5 GB</td>
</tr>
<tr>
<td>Total Computing Recourses in EUAsia VO</td>
<td>268 Cores</td>
</tr>
<tr>
<td>Number of used Computing Elements</td>
<td>6</td>
</tr>
</tbody>
</table>

**FP7-INFRA-223791**
Structural clustering

Examples of most frequent hits

FP7-INFRA-223791
Prepare the in vitro work

- Extract the top-most 10% compounds according to the binding energy profiles from this data challenge.
- The most potential 200 compounds from structural clustering are subjected to the first phase wet-lab assay.
- Dr. Ying-Ta Wu (GRC, AS Taiwan) and Dr. Kim (Chonnan Univ. Korea) will co-work on the assay.
Summary

• The stability and reliability of in-silico high-throughput drug screening was intensively executed and the results showed that the complete process is faster if compared with run on sequential computing resources.

• GVSS is presently supporting friendly user interface of in-silico drug discovery on the neglected and emerging diseases, such as Avian Influenza and Dengue Fever on production e-infrastructures (EGEE and EUAsiaGrid).

• The GVSS service is freely available on the web (http://gap.grid.sinica.edu.tw)
Acknowledgements

AS, Taiwan
Mason Hsung
Li-Yung Ho*
Hurng-Chun Lee*
Dr. Ying-Ta Wu
Dr. Simon Lin

LCG-ARDA, CERN
Jakub Moscicki
Dr. Massimo Lamanna

Genomics Research Center, Academia Sinica
National Science Council, Taiwan