

# Searching For A New Drug on Computational Grid

---

**Habibah A. Wahab, PhD**

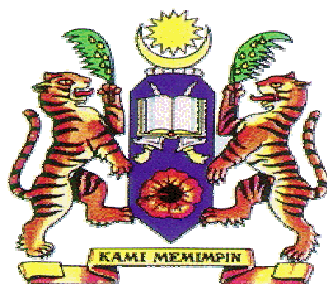
**&**

**Erwin Tantoso**

**School of Pharmaceutical Sciences**

**Universiti Sains Malaysia**

**Penang, Malaysia**



*GGF7, Tokyo, Japan. 6<sup>th</sup> March, 2003*

# Introduction

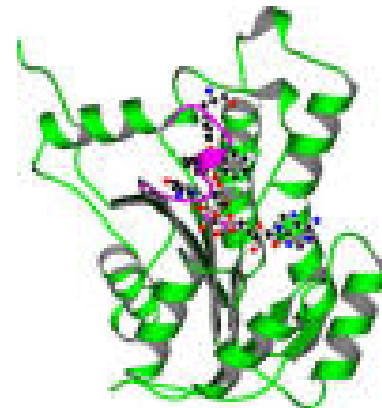
---

## □ Drug discovery

- Complex and expensive undertaking.
- ~\$800mil (Tuft Center)
- 12-15 years.

## □ Effort to cut down the research timeline and cost

- by reducing wet-lab experiments
- use computer modelling.





# Outline:

---

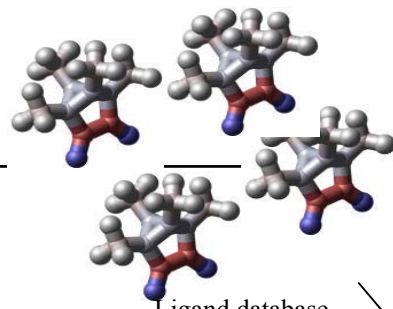
- **Brief introduction to the application (area)**
- Reasons for using Grids?
- Problems of:
  - building a testbed/production grid
  - writing/running your applications



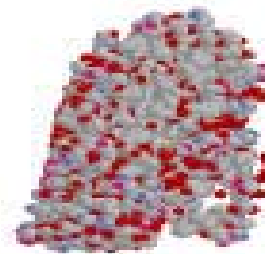
# Area of application:

---

- Computational Biology
  - Specifically structure-based drug design involving protein-drug binding
  - One of the methods in rational drug design
  - Can be applied in molecular biology
  - Enzyme-substrate/inhibitor complex
  - Protein function, mutation and resistance.



Ligand database



Target Protein

**Protein-Ligand docking**



Ligand docked into protein's active site



# Outline:

---

- Brief introduction to the application (area)
- **Reasons for using Grids?**
- Problems of:
  - building a testbed/production grid
  - writing/running your applications



# Reasons for using Grid

---

- ❑ In usual virtual drug discovery, thousands of compounds screened onto its target.
- ❑ Each compound-target interaction requires extensive computation, typically about a day calculation on a workstation.
- ❑ Protein-ligand studies – very popular in Malaysia.
- ❑ Promote computational modelling in these studies.
- ❑ Make the Computation looks effortless for biologist – i.e. they just send the input files to the program and the program will do the rest.

# Reasons for using Grid

---

- Workstation is very expensive in Malaysia
- Computation on the grid environment provide opportunities for:
  - **Harnessing CPU cycle**
  - **Accessing remote databases**
  - **User friendly – instructions are clear**
  - **Effortless computation**
  - **Easy obtaining the output**

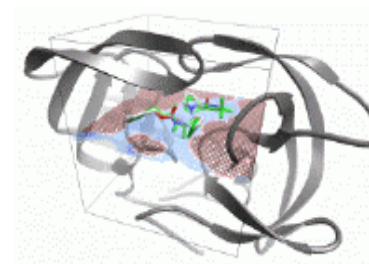
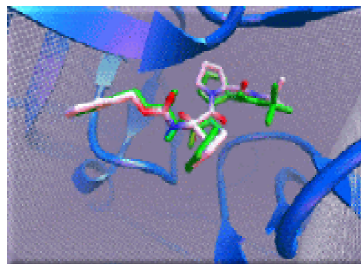


# Searching for a drug on the Grid

---

## □ Molecular Docking

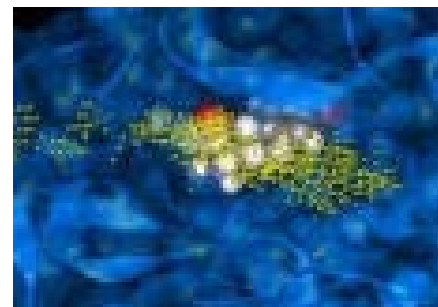
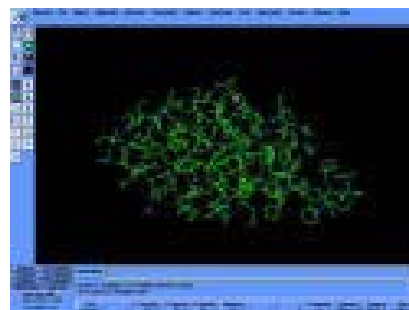
- Autodock 3.05



## □ Visualization

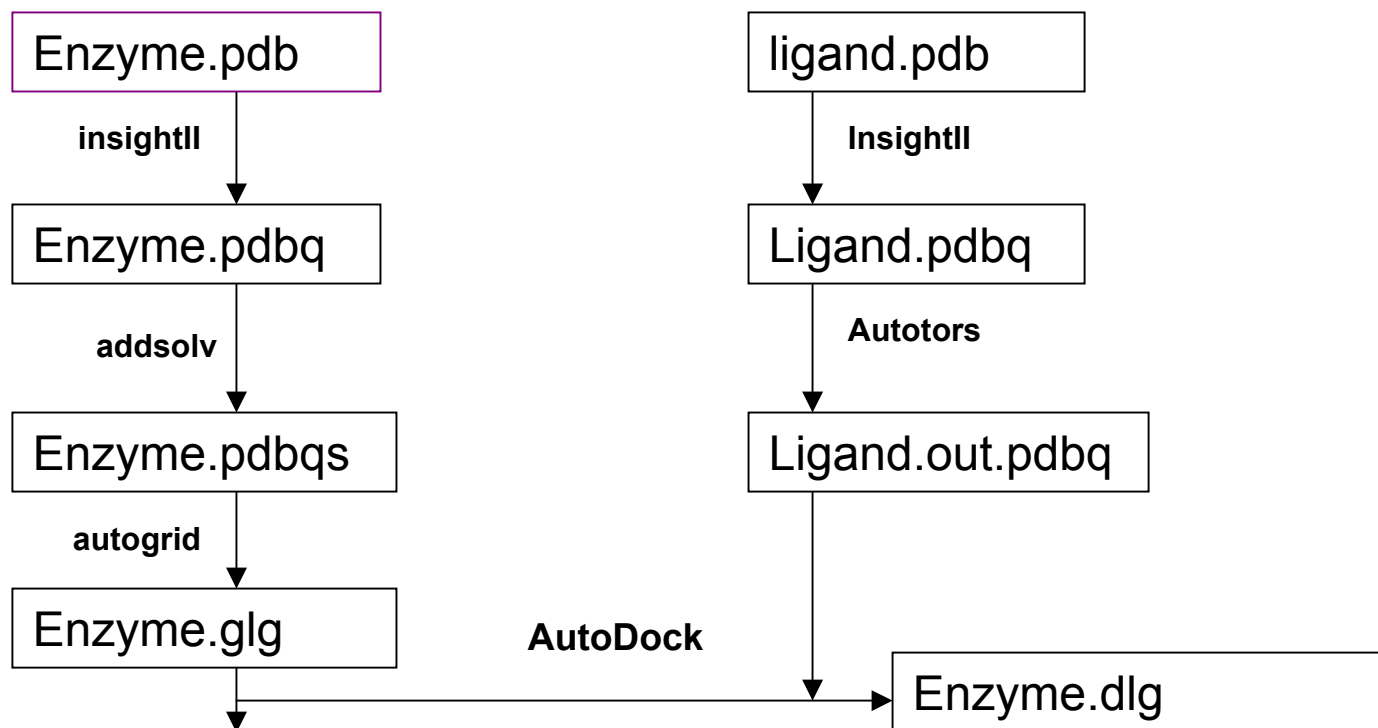
- Rasmol

- Chimes



# Setting-Up AutoDock

---



# Molecular Docking Introduction Page

e-Science Grid Portal - version alfa.099 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites History Print

Address <http://grid.cs.usm.my:8080/eportal/jsp/integrated/index.jsp> Go

**e-Science Grid PORTAL**

Home Logout User Profile File Transfer Queue Info Submit Job 27 November, 2002 Wednesday

## Molecular Docking Introduction

### General Information

- About
- Resources
- FAQ
- Help

### Archives

- News
- Documents
- Status Report
- Download

### Applications

- Iterative Solver
- Molecular Docking

### Links

- GCORP
- Globus
- NCSC
- CPM

### Account Information

- Check Credit
- Transaction History

### Overview

This page is mainly for submitting molecular docking job. In this application, we have one ligand file (from the user) and a number of macromolecule files (provided by PDB web site).

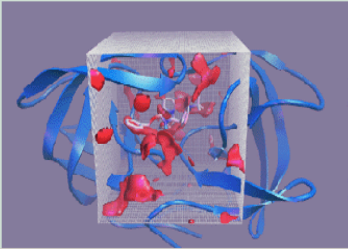
The application will perform docking ligand on each of the macromolecule files. The type of macromolecules is provided by the user.

The docking software that we are using here is [AutoDock3.05](#).

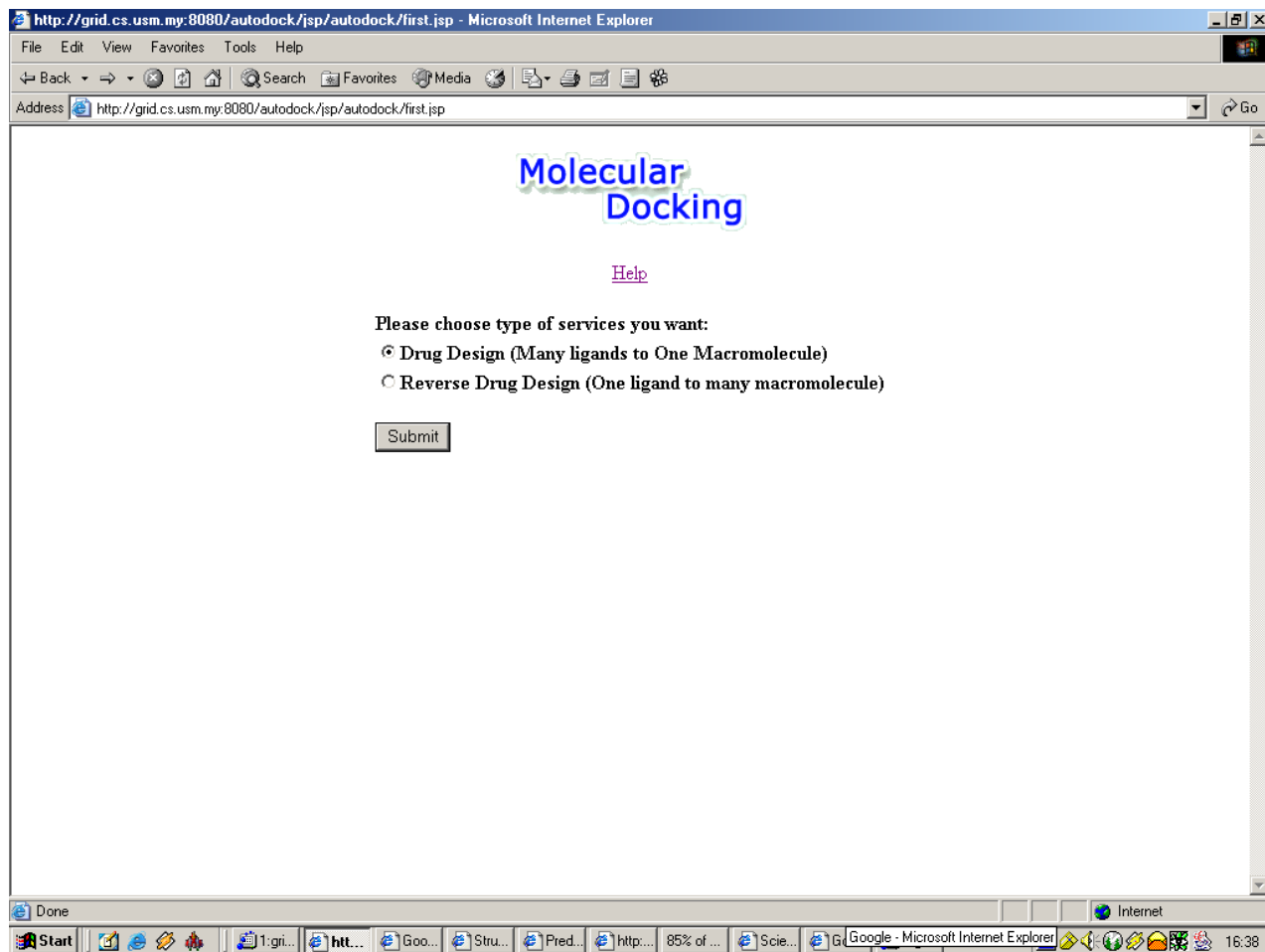
### Molecular Docking

Docking is the study of how drug (small molecule) bound to an enzyme or protein receptor. It is just a prediction of whether one ligand (drug) can be bound to the receptor or not. To know it, we can just look at the free energy of the docking result and the binding energy. Usually the negative free energy indicates that this process can be done and the ligand can be docked into the receptor.

When we said about molecular docking, it is highly related to drug design.



## □ Input Page



## □ Input Page

http://grid.cs.usm.my:8080/autodock/jsp/autodock/first.jsp?numUpload=5&submitNum=Submit&submitT - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media Print Mail News RSS Feeds

Address http://grid.cs.usm.my:8080/autodock/jsp/autodock/first.jsp?numUpload=5&submitNum=Submit&submitType=Submit&type=Drug+Design Go

# Molecular Docking

[Help](#)

*Please Upload Your ligand File to Us !*

Ligand FileName : C:\Documents and Sett Browse...

Ligand FileName : C:\Documents and Sett Browse...

Ligand FileName : C:\Documents and Sett Browse...

Ligand FileName : C:\Documents and Sett Browse...

Ligand FileName : C:\Documents and Sett Browse...

Submit

Done Internet

Start 1.gri... htt... Goo... Stru... Pred... http... 85% of ... Scie... Goo... 2.gri... Doc... 16:41

## □ Input Page (cont..)

http://grid.cs.usm.my:8080/autodock/jsp/autodock/input.jsp - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print View Source Settings

Address http://grid.cs.usm.my:8080/autodock/jsp/autodock/input.jsp Go

### Molecular Docking

[Help](#)

**Trading Specifications:**

Budget:  tokens

Deadline:

Date

Time  :

**Optimization Mode:**

☒ Speed Optimization

☐ Cost Optimization

Ligand FileName :

Please select your macromolecule :

Please input the PDB ID or name :

Please specify your search method to perform docking:

☒ Genetic Algorithm

☐ Simulated Annealing

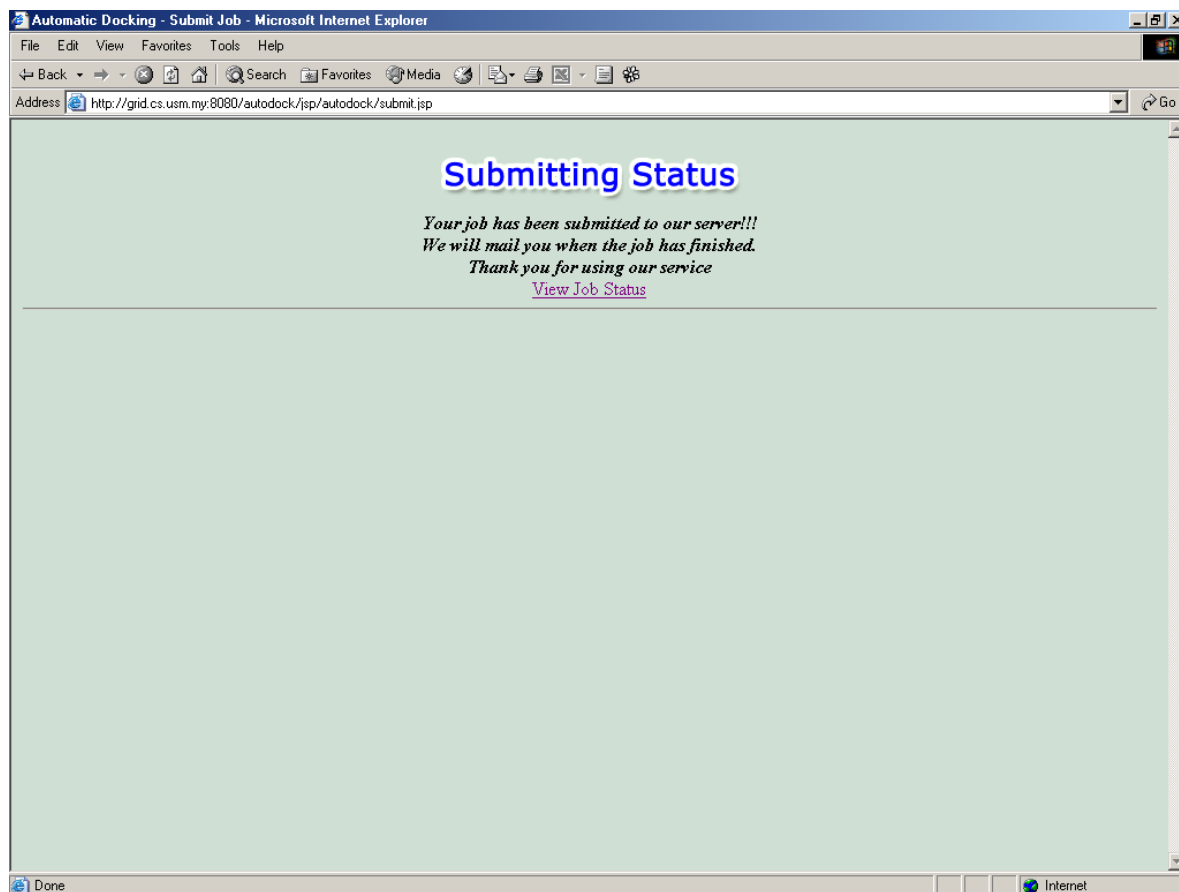
Number of Run you would like to perform:

Your email-address:

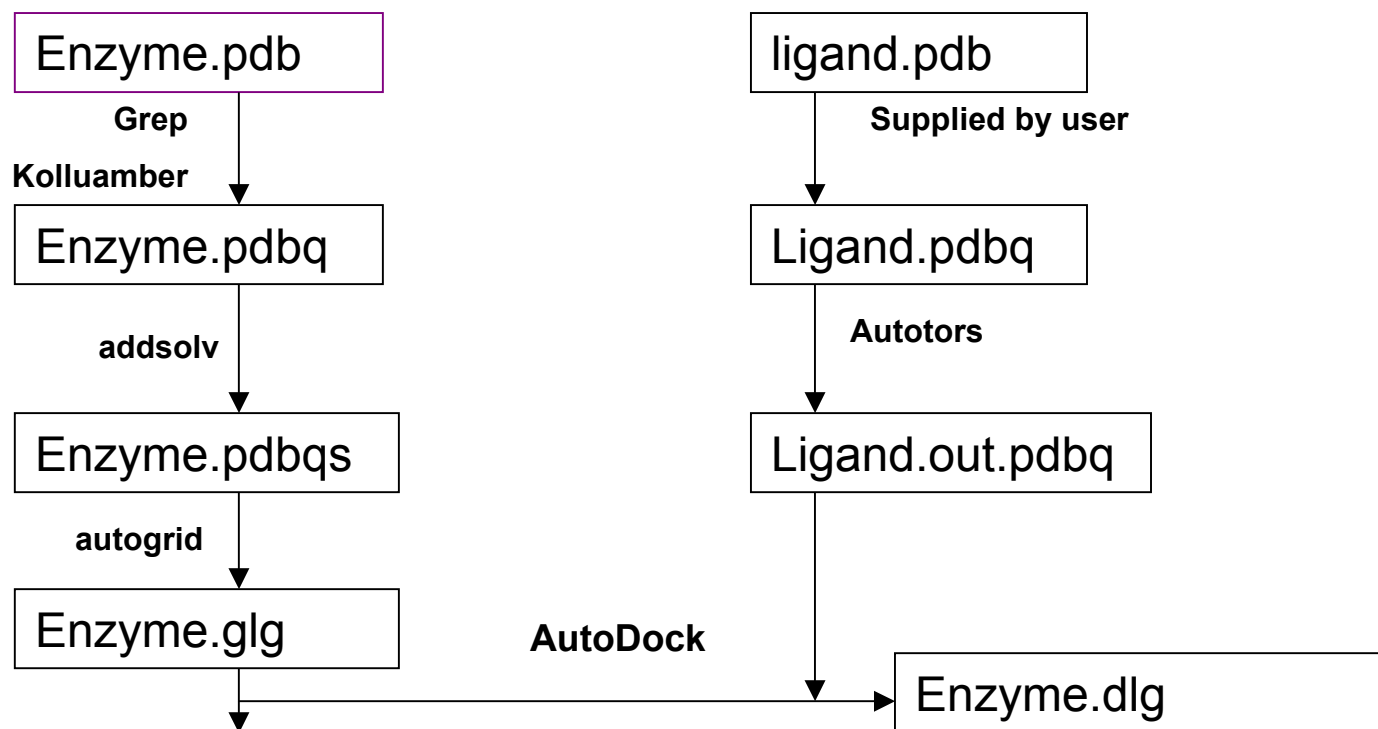
Done Internet

## □ Submitting Job

---



# AutoDock on the grid





## □ Results

http://grid.cs.usm.my:8080/autodock/jsp/autodock/result.jsp - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS

Address http://grid.cs.usm.my:8080/autodock/jsp/autodock/result.jsp Go

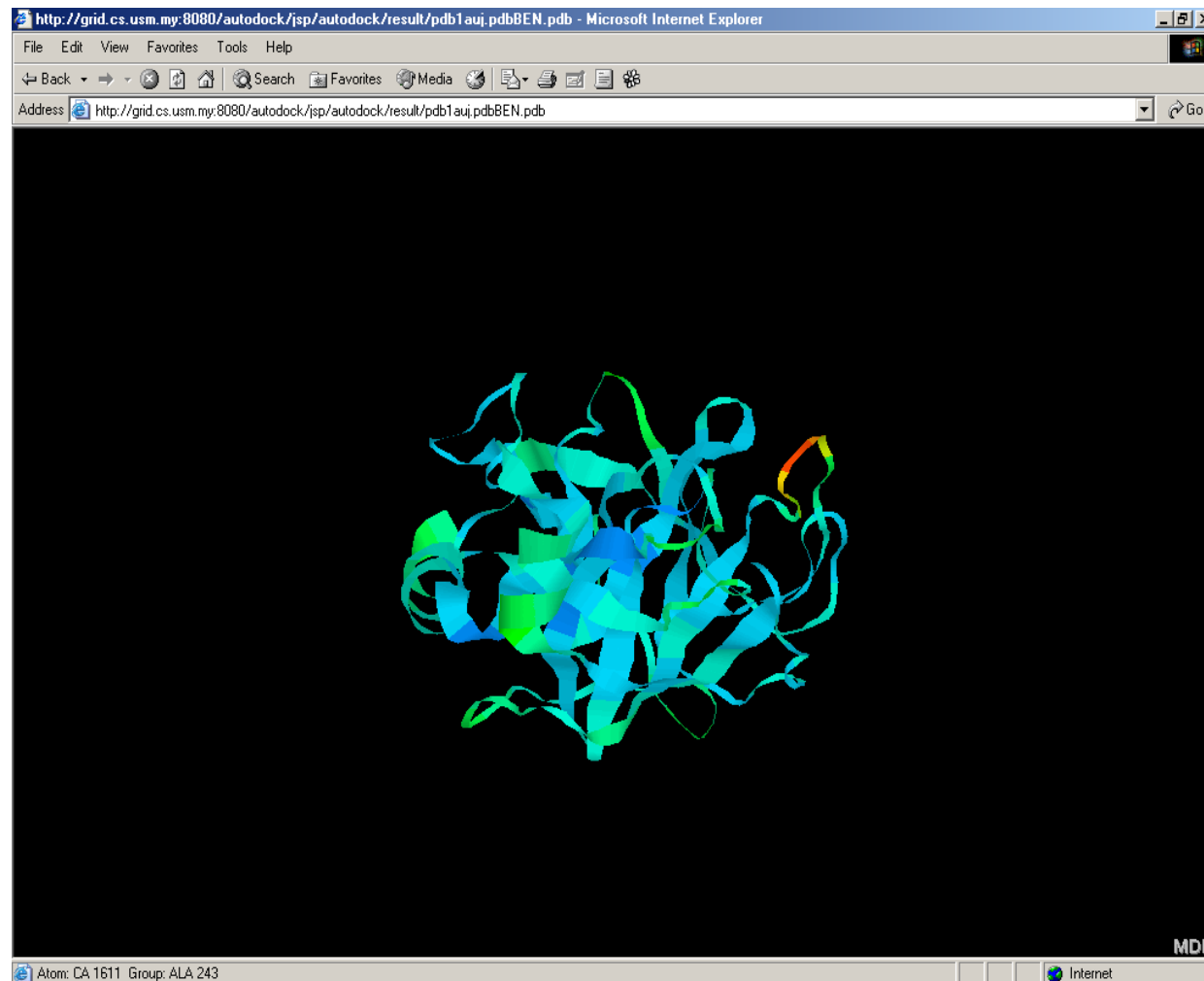
# Results

[Return](#)  
[Download All](#)

Dock Log File	Docked-Ligand Coordinate	Macromolecule-Ligand Structure File	Visualization
<a href="#">pdb1auj.pdb.dlg</a>	<a href="#">pdb1auj.pdb.dlg.pdb</a>	<a href="#">pdb1auj.pdbBEN.pdb</a>	<input type="button" value="View"/>
<a href="#">pdb1az8.pdb.dlg</a>	<a href="#">pdb1az8.pdb.dlg.pdb</a>	<a href="#">pdb1az8.pdbBEN.pdb</a>	<input type="button" value="View"/>
<a href="#">pdb1eb2.pdb.dlg</a>	<a href="#">pdb1eb2.pdb.dlg.pdb</a>	<a href="#">pdb1eb2.pdbBEN.pdb</a>	<input type="button" value="View"/>
<a href="#">pdb2tbs.pdb.dlg</a>	<a href="#">pdb2tbs.pdb.dlg.pdb</a>	<a href="#">pdb2tbs.pdbBEN.pdb</a>	<input type="button" value="View"/>
<a href="#">pdb3ptb.pdb.dlg</a>	<a href="#">pdb3ptb.pdb.dlg.pdb</a>	<a href="#">pdb3ptb.pdbBEN.pdb</a>	<input type="button" value="View"/>

Internet

## □ Results (in visualization form)



# Benchmarking....

---

- Testing on 3 SGI machines
- Testing on 3 Linux Machines
  - Pentium IV 1.70 GHz , 256 MB Memory
  - Pentium III 1000 MHz , 512 MB Memory
- Data / Molecules:
  - Macromolecule of wild and mutant type
  - 14 types of ligands
  - Total number of docking =  $14 \times 2 = 28$



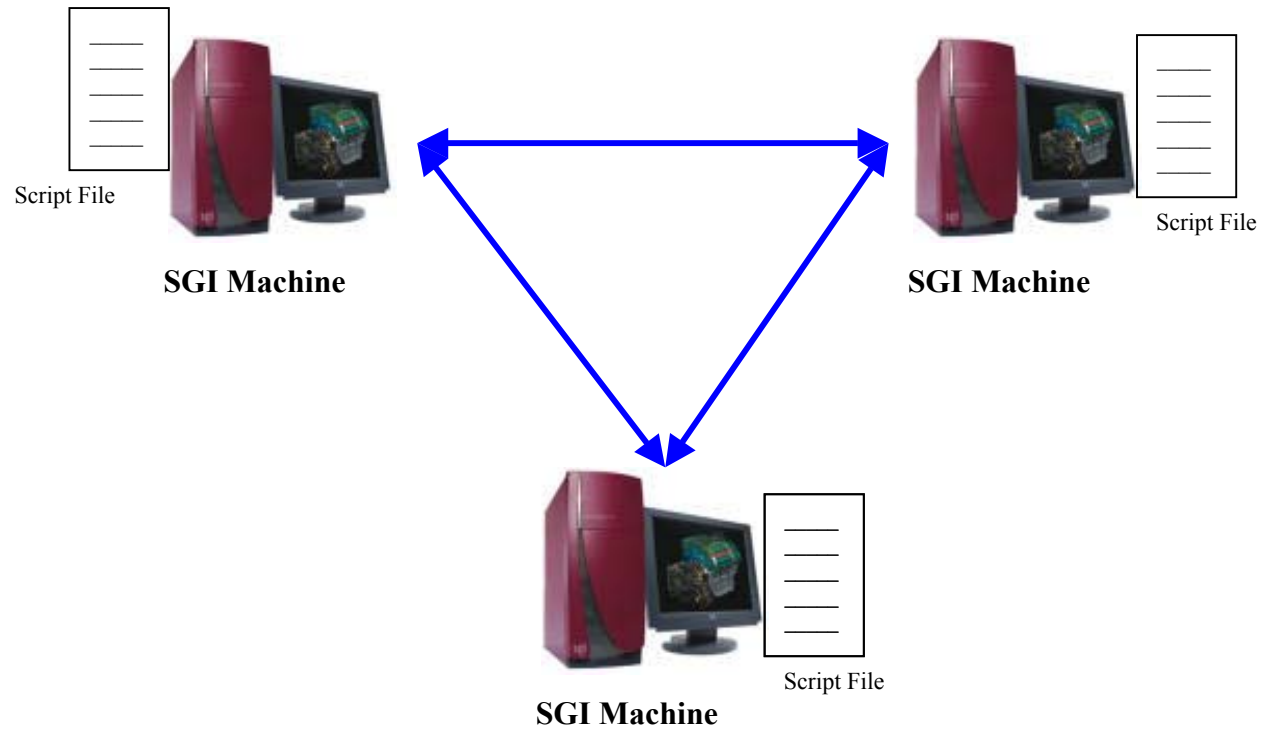
## Searching For A New Anti-tuberculosis Drug on the GRID

---

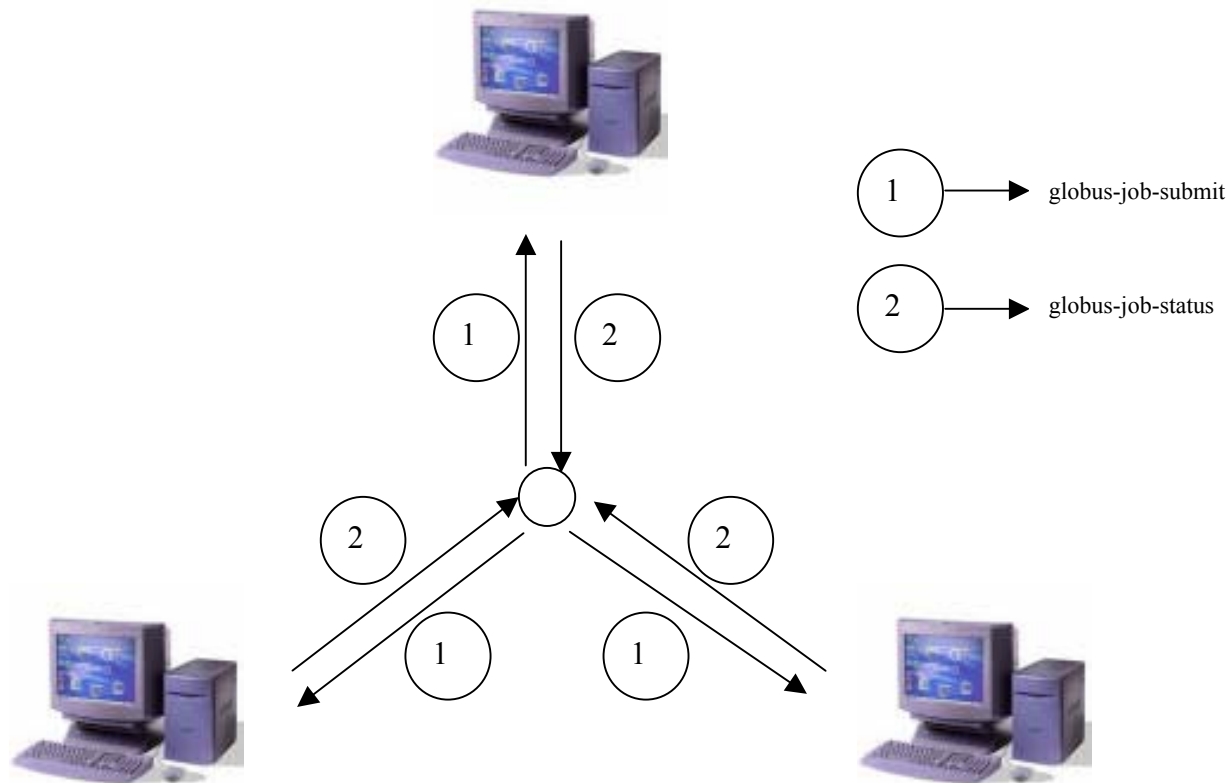
- One third of the world population is infected.
- Multi-drug resistant especially in HIV/AIDS patients
- As much as 30% on INH (front-line drug) resistant strain
- Use AutoDock3.1 to understand drug(s) interaction with its MTB protein target.
- 16 INH derivatives synthesised in USM subjected Molecular Docking experiment on the grid to choose which of the derivatives would be a good alternative to INH

# Testing Method (on SGI)

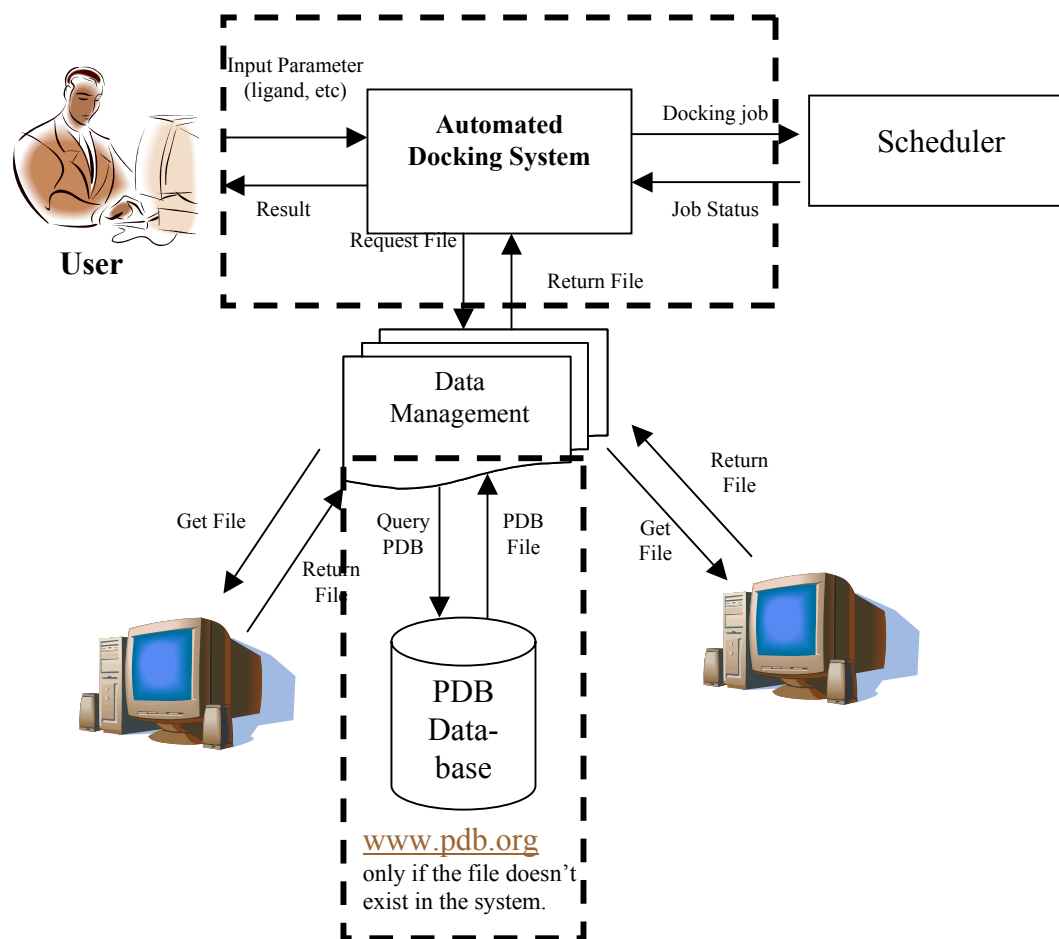
---



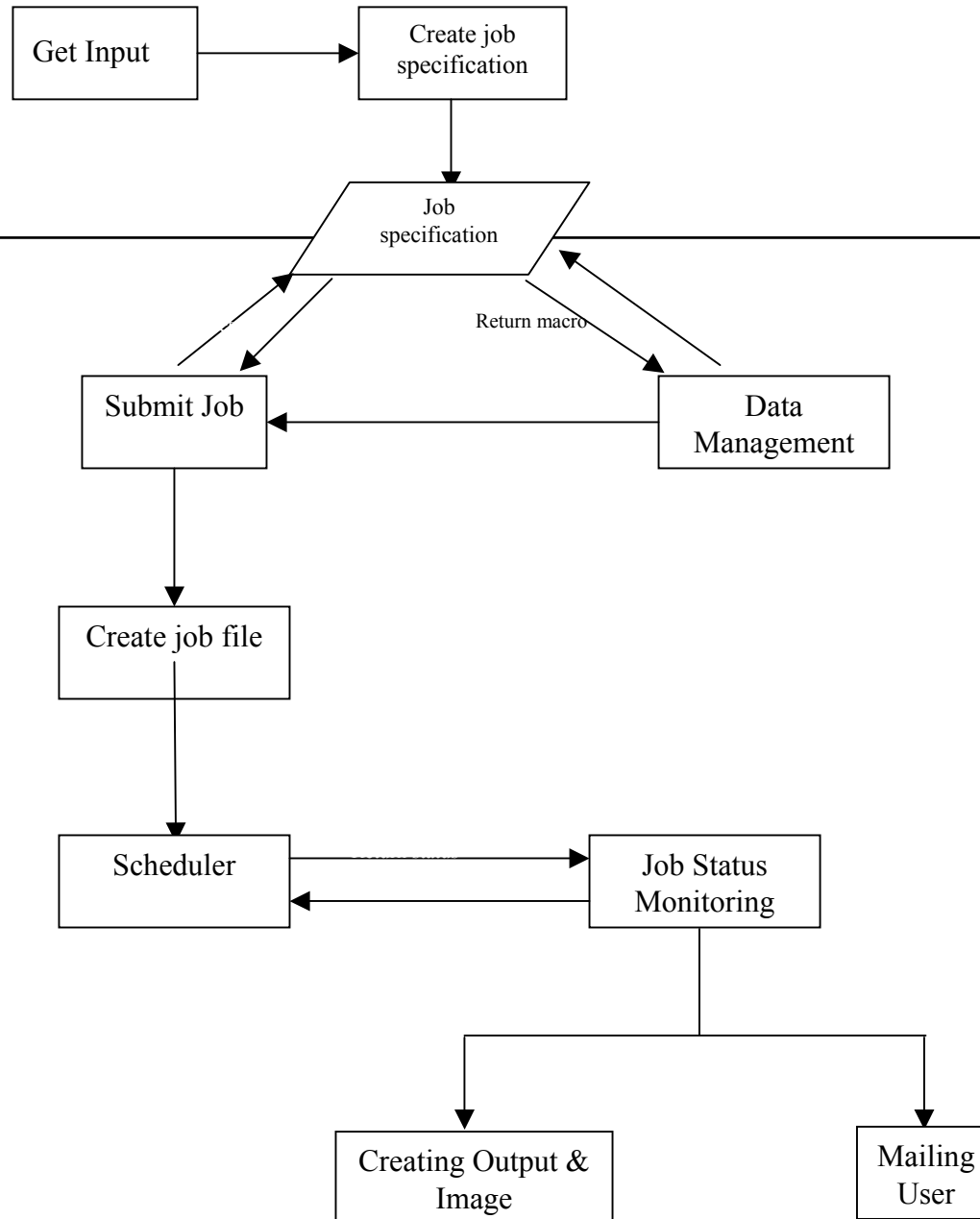
# Testing Method (on Linux)



## Overall Flow of Automated Docking System

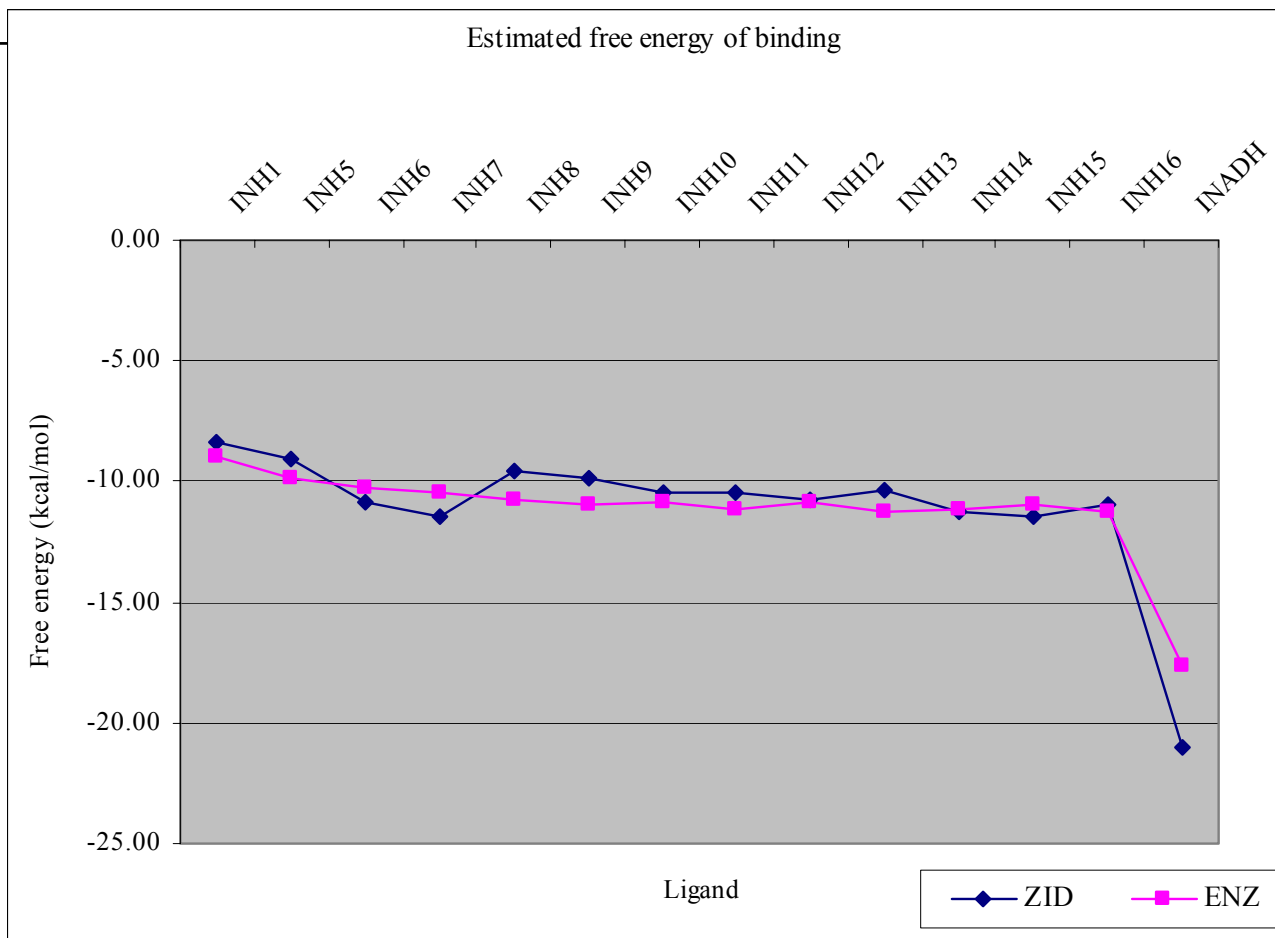


# Automated Docking System: The steps in Automated Docking

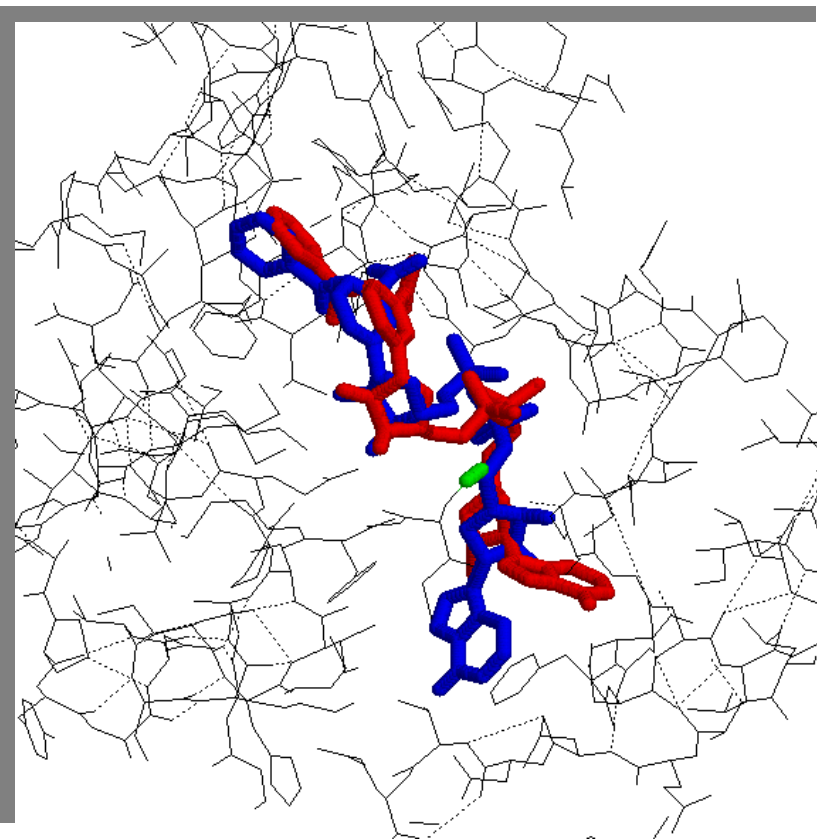
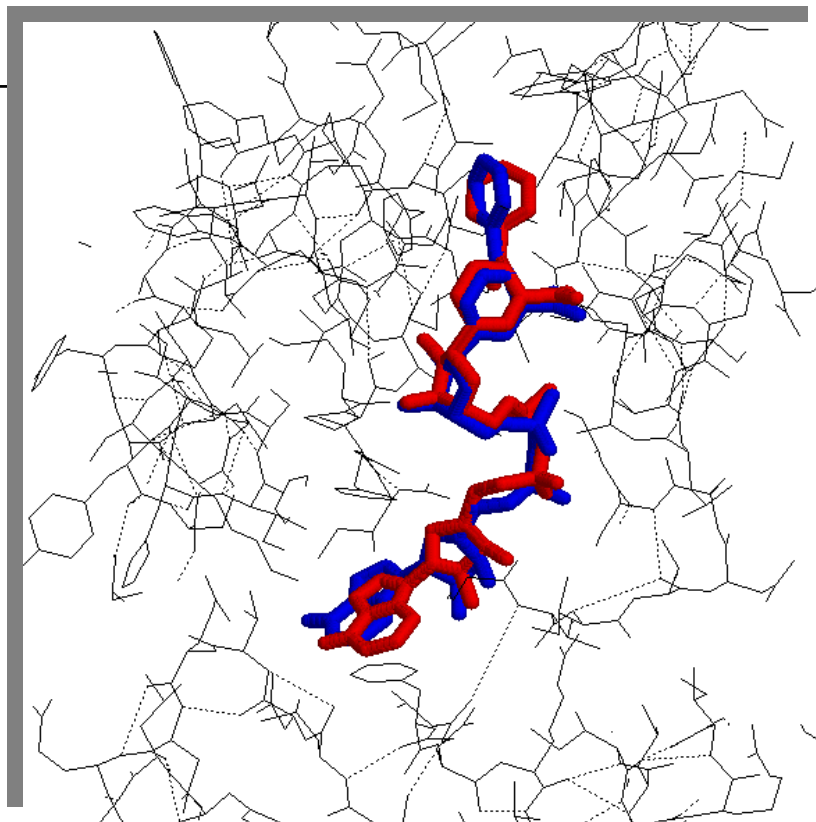




# Result



# Result



# Result (total time)

On SGI	On Linux PC
220h 24min 59.34s  (1 machine= 73h 8min 21.11s)	333 hrs 1 min 22 s  (1 machine= 111h 0min 27s)



# Result (\$/time)

---

On SGI	On Linux PC
\$1.85/min  (1 machine= ~USD25K)	\$0.075/min  (1 machine= ~USD 1.5K)



# Outline:

---

- Brief introduction to the application (area)
- Reasons for using Grids?
- **Problems of:**
  - **building a testbed/production grid**
  - **writing/running applications**

# Problems:

---

## □ Testbeds:

- This is a our pleminary work in integrating an application on the grid environment.
- Our test-beds are very minute, only local PC – homogenous platform (only 3 to be precise!).
- Globus was installed in each PC and job submission is through local server.
- Still working on building testbed – 2\*16 nodes clusters are still in tender process.
- Looking to tap ApGrid and PRAGMA testbeds.

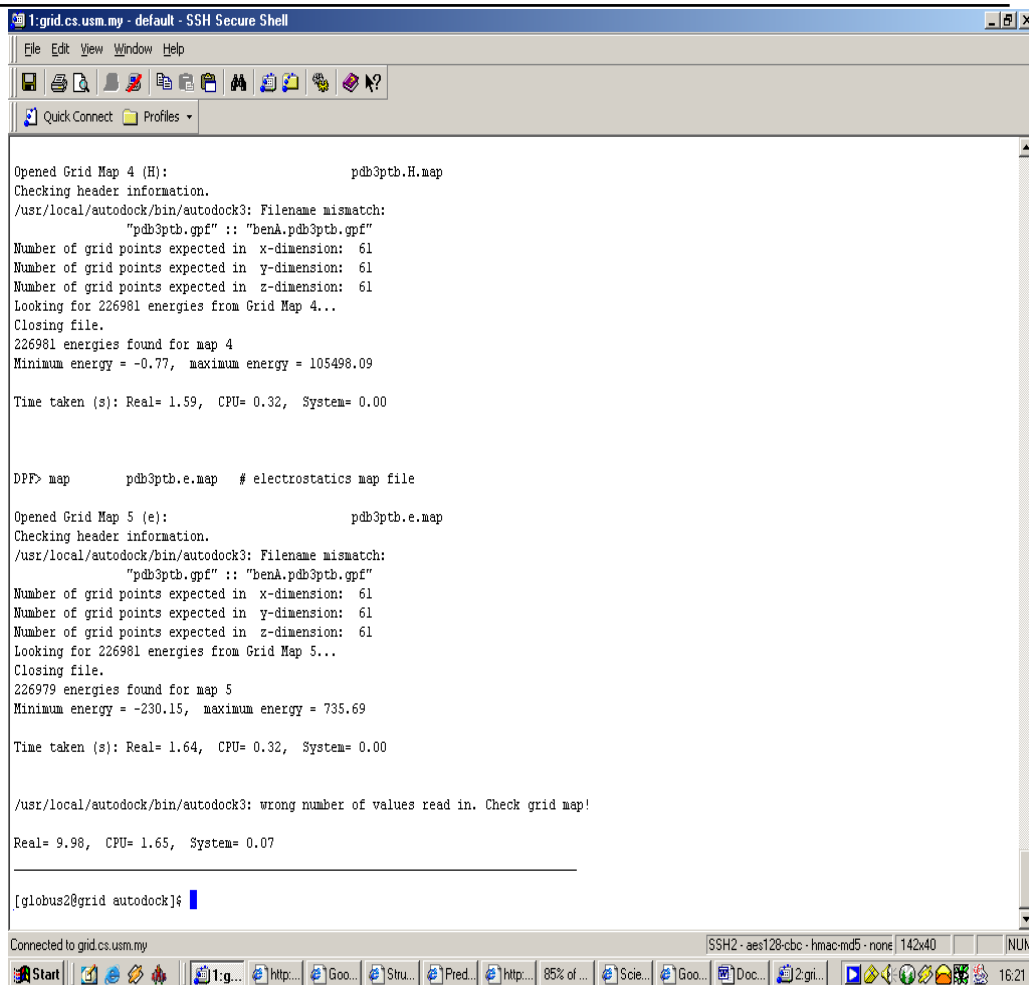
# Running the Application

- ❑ Job scheduler not working properly
- ❑ Problem in submission:
  - Problem with Globus:
    - ❑ Unability to submit job due to authentication problem. (Authentication operation failed: Error Code 7)
    - ❑ Even after checking the grid-map file, job still failed to be submitted.
    - ❑ Proxy must be started otherwise job fails.

```
grid.cs.usm.my - default - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid104-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid105-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid106-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid107-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=globus toolkits 2" globus
bash-2.05$ exit
logout
Connection to 10.207.132.8 closed.
[globus2@grid autodock]$ ls
~          CreateOutput.java  getOutput.sh  input.html  Mail.class  pdb1eb2.pdb  result.jsp
3PTB_.fasta.txt  distFile.sh  Get.sh       input.jsp   mail.jar    pdb2tbs.pdb  statusFile.sh
activation.jar    execute.sh   help.html    intro.html  Mail.java   pdb3ptb.pdb  submitDrug.sh
autodock         FileUpload.class  hs_err_pid15307.log  jobstatus.jsp  Makefile    prepare      submit.jsp
benA.pdbq        FileUpload.java  hs_err_pid24450.log  JobThread.class  moving1.sh  ReadSpecification.class  submit.sh
coba.sh          FileUpload.java.bak  humour.txt  JobThread.java  moving.sh   ReadSpecification.java  tomat
compress.sh      ftp.jar         initjob.bak  ligandfile   pdblauy.pdb  ReadSpecification.java.bak  transfer.sh
CreateOutput.class  getFile.sh    initjob.sh   macrofile    pdblaz8.pdb  remove
[globus2@grid autodock]$ grid-cert-info -subject
/O=Grid/0=Globus/OU=cs.usm.my/CN=globus toolkits 2
[globus2@grid autodock]$ globus-job-submit 10.207.132.8 /bin/date
GRAM Job submission failed because an authentication operation failed (error code 7)
[globus2@grid autodock]$ ssh globus@10.207.132.8
The authenticity of host '10.207.132.8 (10.207.132.8)' can't be established.
RSA key fingerprint is 32:1c:72:3f:9e:36:63:31:d2:de:05:0e:e0:22:28:24.
Are you sure you want to continue connecting (yes/no)? yes
Failed to add the host to the list of known hosts (/home/globus2/.ssh/known_hosts).
globus@10.207.132.8's password:
Last login: Mon Mar  3 16:06:44 2003 from grid
Sun Microsystems Inc. SunOS 5.9  Generic May 2002
bash-2.05$ cat /etc/grid-security/grid-mapfile
"/O=Grid/0=Globus/OU=cs.usm.my/CN=gridsun108-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid101-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid102-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid103-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid104-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid105-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid106-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=grid107-globus" globus
"/O=Grid/0=Globus/OU=cs.usm.my/CN=globus toolkits 2" globus
bash-2.05$
```

# Running the Application

- Other problems:
  - Job did not finish properly. The output file is not as what is supposed to be.
  - Job was not submitted at all.
  - output overwritten if the same user submit the job.



```
1:grid.cs.usm.my - default - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles

Opened Grid Map 4 (H):                pdb3ptb.H.map
Checking header information.
/usr/local/autodock/bin/autodock3: Filename mismatch:
"pdb3ptb.gpf" :: "benA.pdb3ptb.gpf"
Number of grid points expected in x-dimension: 61
Number of grid points expected in y-dimension: 61
Number of grid points expected in z-dimension: 61
Looking for 226981 energies from Grid Map 4...
Closing file.
226981 energies found for map 4
Minimum energy = -0.77, maximum energy = 105498.09

Time taken (s): Real= 1.59, CPU= 0.32, System= 0.00

DPF> map                pdb3ptb.e.map # electrostatics map file

Opened Grid Map 5 (e):                pdb3ptb.e.map
Checking header information.
/usr/local/autodock/bin/autodock3: Filename mismatch:
"pdb3ptb.gpf" :: "benA.pdb3ptb.gpf"
Number of grid points expected in x-dimension: 61
Number of grid points expected in y-dimension: 61
Number of grid points expected in z-dimension: 61
Looking for 226981 energies from Grid Map 5...
Closing file.
226979 energies found for map 5
Minimum energy = -230.15, maximum energy = 735.69

Time taken (s): Real= 1.64, CPU= 0.32, System= 0.00

/usr/local/autodock/bin/autodock3: wrong number of values read in. Check grid map!

Real= 9.98, CPU= 1.65, System= 0.07

[glabus2@grid autodock]$
```

Connected to grid.cs.usm.my

SSH2 - aes128-cbc - hmac-md5 - none 142x40 NUM

Start | 1:g... | http... | Goo... | Stru... | Pred... | http... | 85% of ... | Scie... | Goo... | Doc... | 2.gri... | 16.21



# Future Works

---

- ❑ Improvement of the submitting the same job. Meaning that user can still submit the same job anytime and the system will be able to execute it.
- ❑ Improve the visualization of the result given once the docking job has finished.
- ❑ Performance Evaluation, by doing comparison between sequential and distributed version in the grid environment.



# Acknowledgement

---

- Dr. Yoshio Tanaka & AAPS WG
- Prof. Shinji Shimojo
- Tomomi Takao
- Osaka University
- PRAGMA
- Choong Yee Siew
- TB Research Group
- Grid Computing Research Group



**THANK YOU**