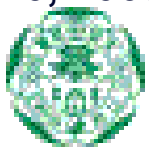
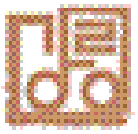
Status of Activities in KISTI within the WISDOM collaboration

Soonwook Hwang, Jincheol Kim, Sehoon Lee
KISTI

On behalf of the WISDOM collaboration

February 26, 2009



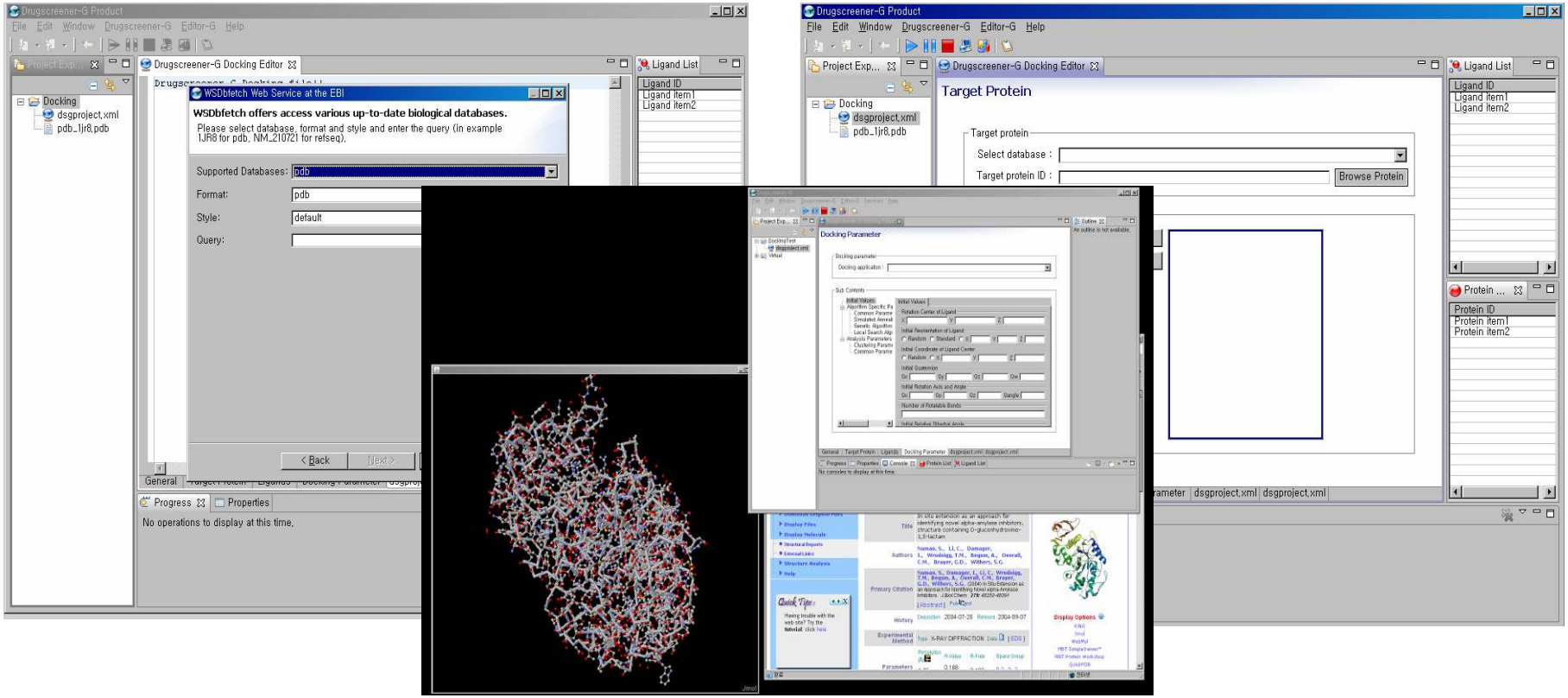


KISTI's Achievements within the WISDOM project in 2008

- Completed development of a prototype version of DrugScreener-G software suite
- Performed a part of the data challenge for human diabetes type II in collaboration with LPC/IN2P3 (target protein: alpha-amylase)
- Improved the performance of WISDOM Production Environment by adopting AMGA to store production data
  - Also added more enhanced features, such as enhanced monitoring and bookkeeping, more fault-tolerant operation by CE, RB blacklisting, etc.
- Development of new enhanced features of AMGA in collaboration with the AMGA team of CERN IT Division and INFN-Catania team
  - Native SQL support
  - WS-DAIR support

Introduction to DrugScreener-G

- DrugScreener-G: Integrated Environment for Grid-enabled Large-scale Virtual Screening



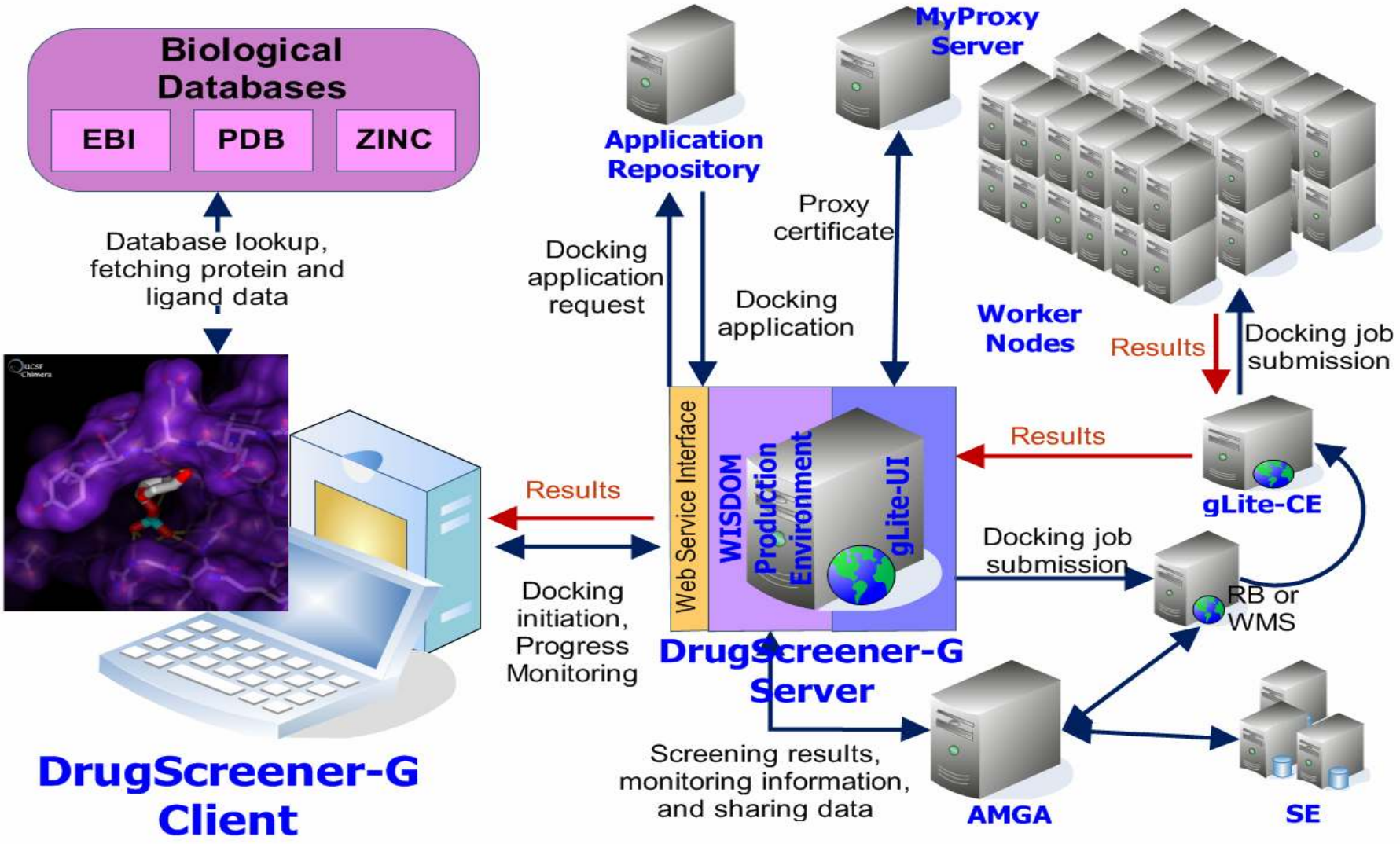
Motivation and Vision

- Motivation
  - **Scientists and researchers in drug discovery are usually not familiar with new concepts and techniques of computing and information technology such as Grid computing.**
  - New concepts and technical details to utilize Grid computing and infrastructure for large-scale virtual screening **should be hidden** in an integrated environment for Grid-based virtual screening.
  - To provide an integrated environment **easy to integrate various virtual screening methods and extensible to adopt new virtual screening methods easily.**

- Vision
  - Software platform to provide a **user-friendly integrated environment for Grid-enabled large-scale virtual screening for users without much knowledge of Grid computing** to exploit full power of Grid computing infrastructure for drug discovery

- Target users: Bioinformaticians, Biologists, Drug Chemists

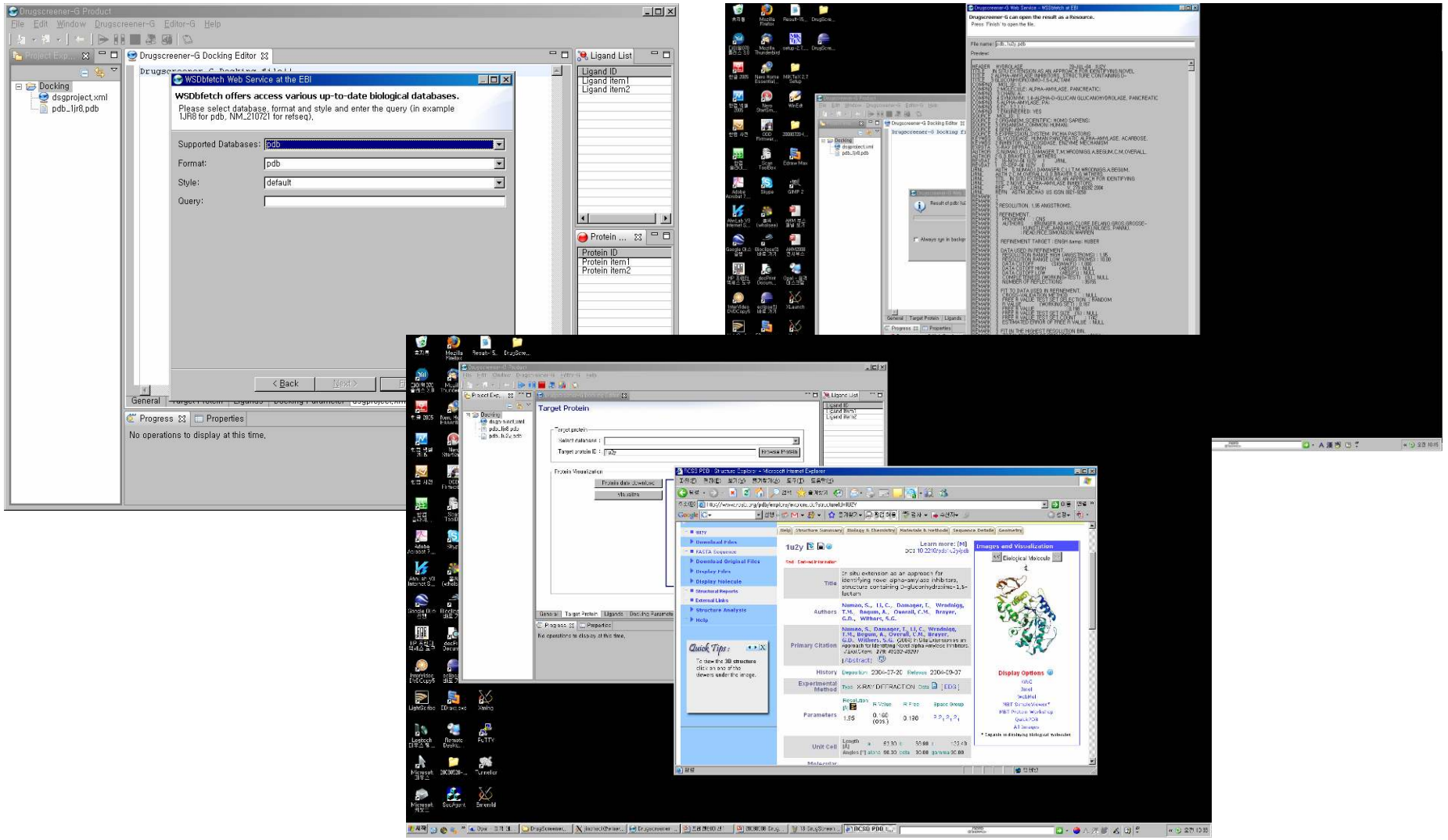
How does DrugScreener-G work with the EGEE infrastructure?



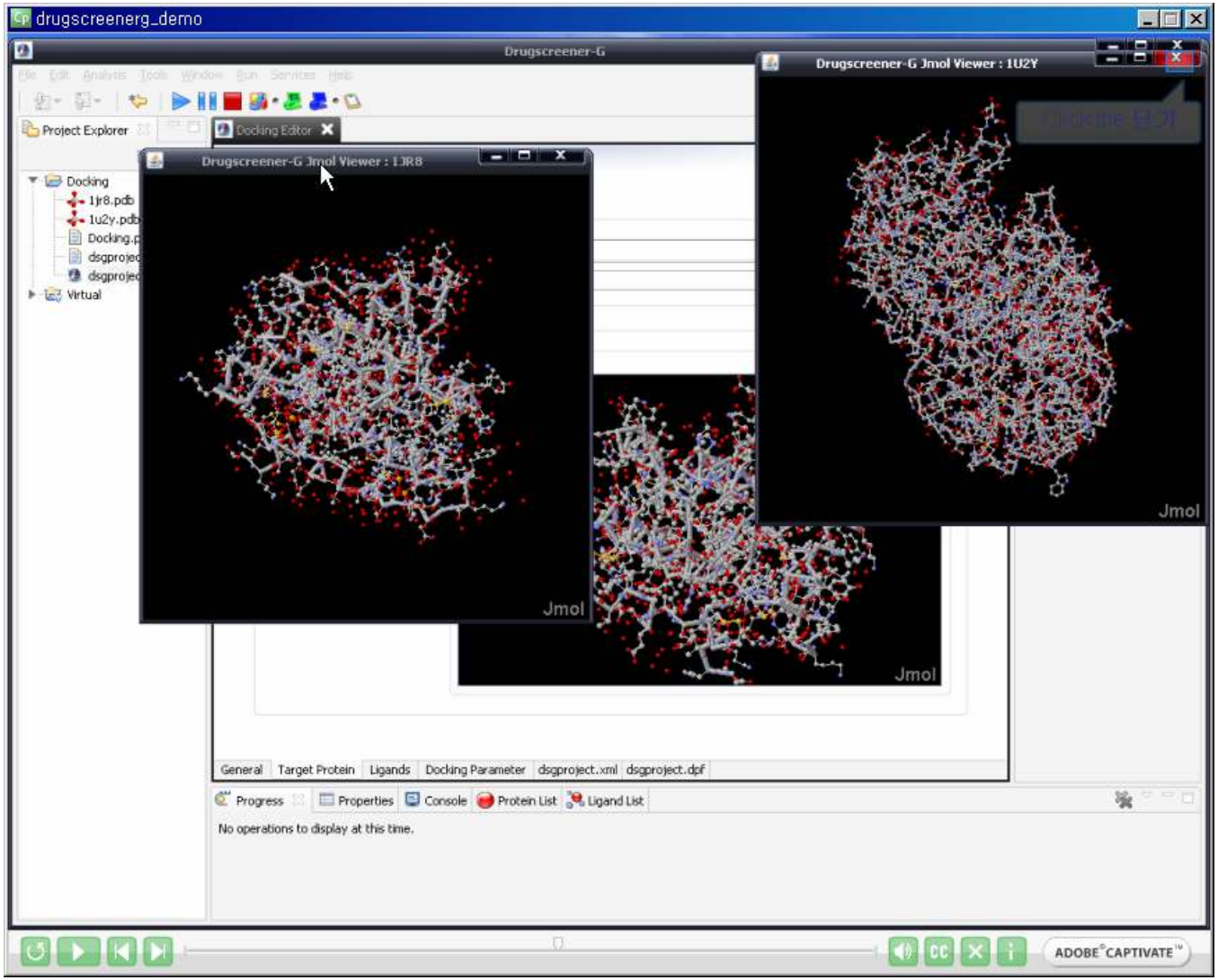
Architecture

- DrugScreener-G Client
  - Docking project editing and management, monitoring, postprocessing and analysis
- DrugScreener-G Server
  - Communicates with a client by web service
  - Management of jobs and data on Grid, or local computing farm and storage devices
- Plug-in is the basic unit of functional components on both of client and server

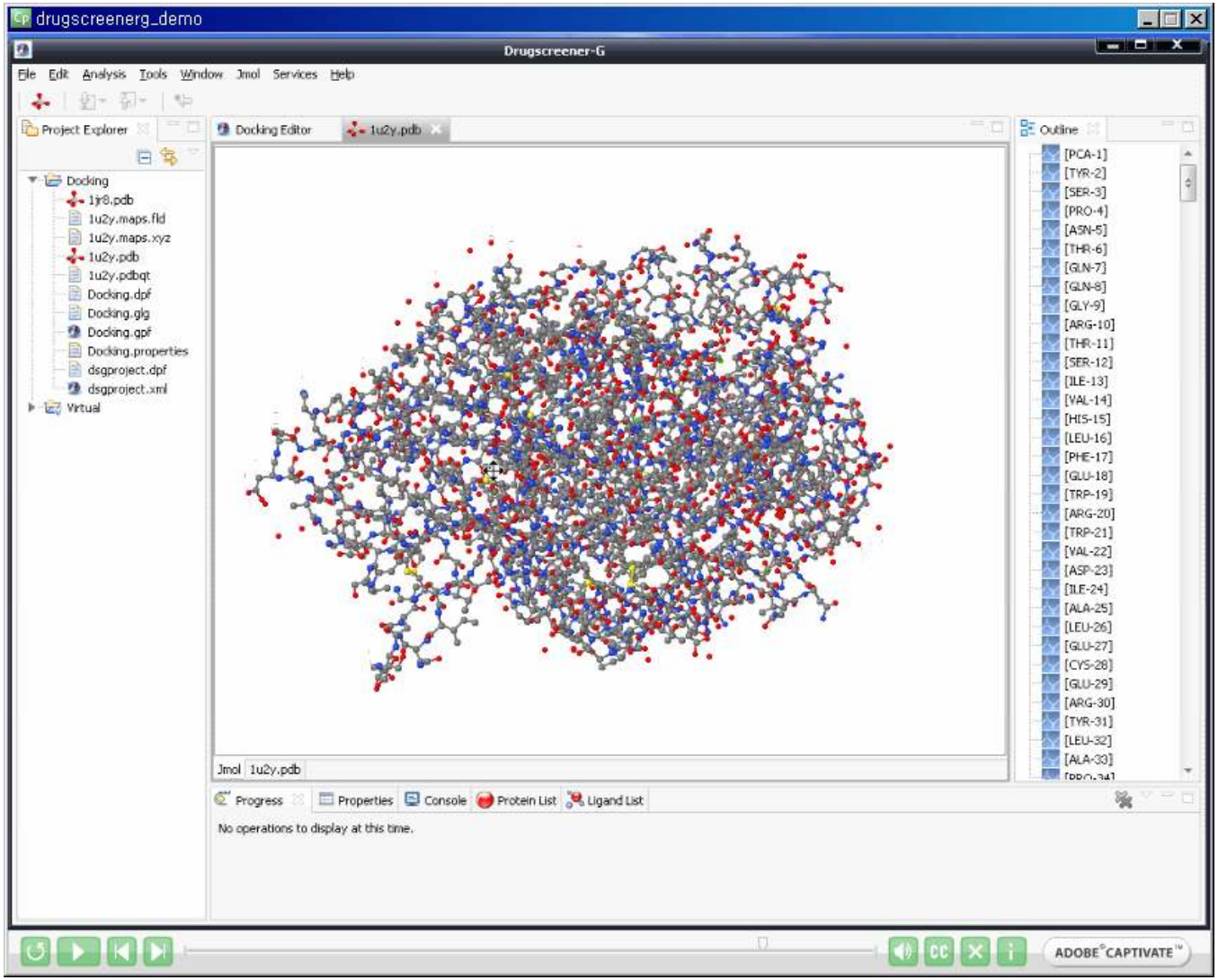
Notable features of DrugScreener-G – Accessing biological data



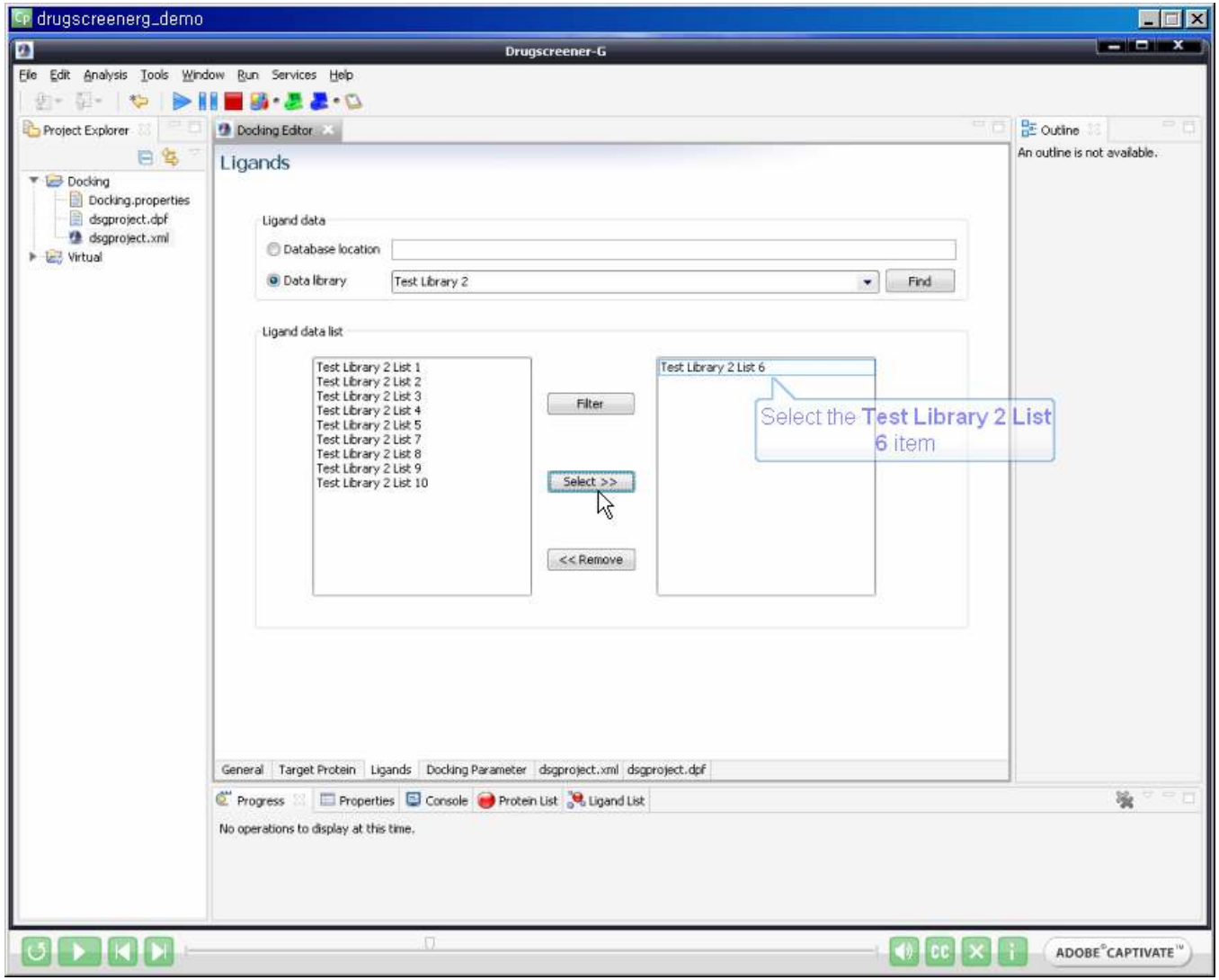
Notable features of DrugScreener-G – Protein visualization and modeling



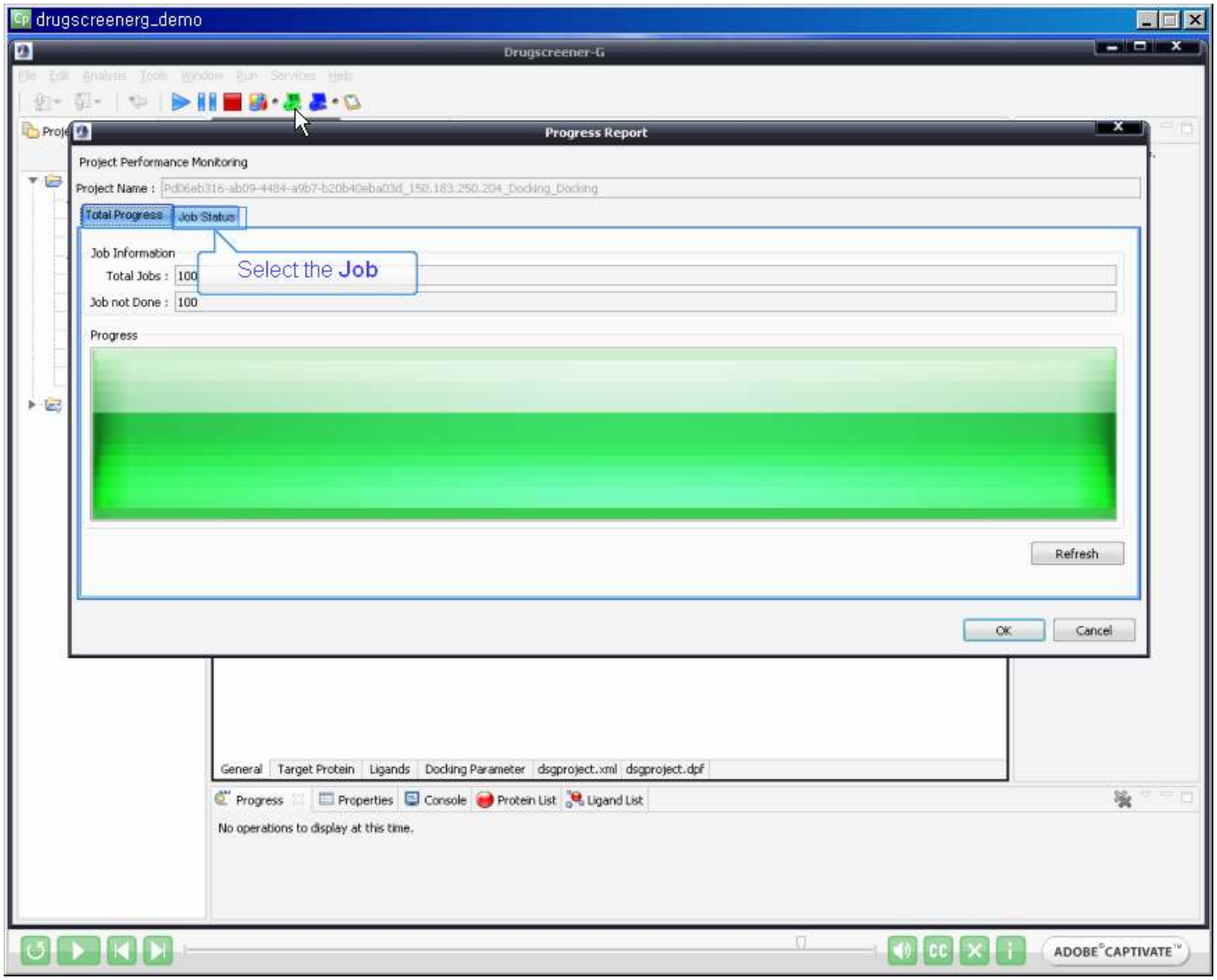
Notable features of DrugScreener-G – Protein visualization and modeling



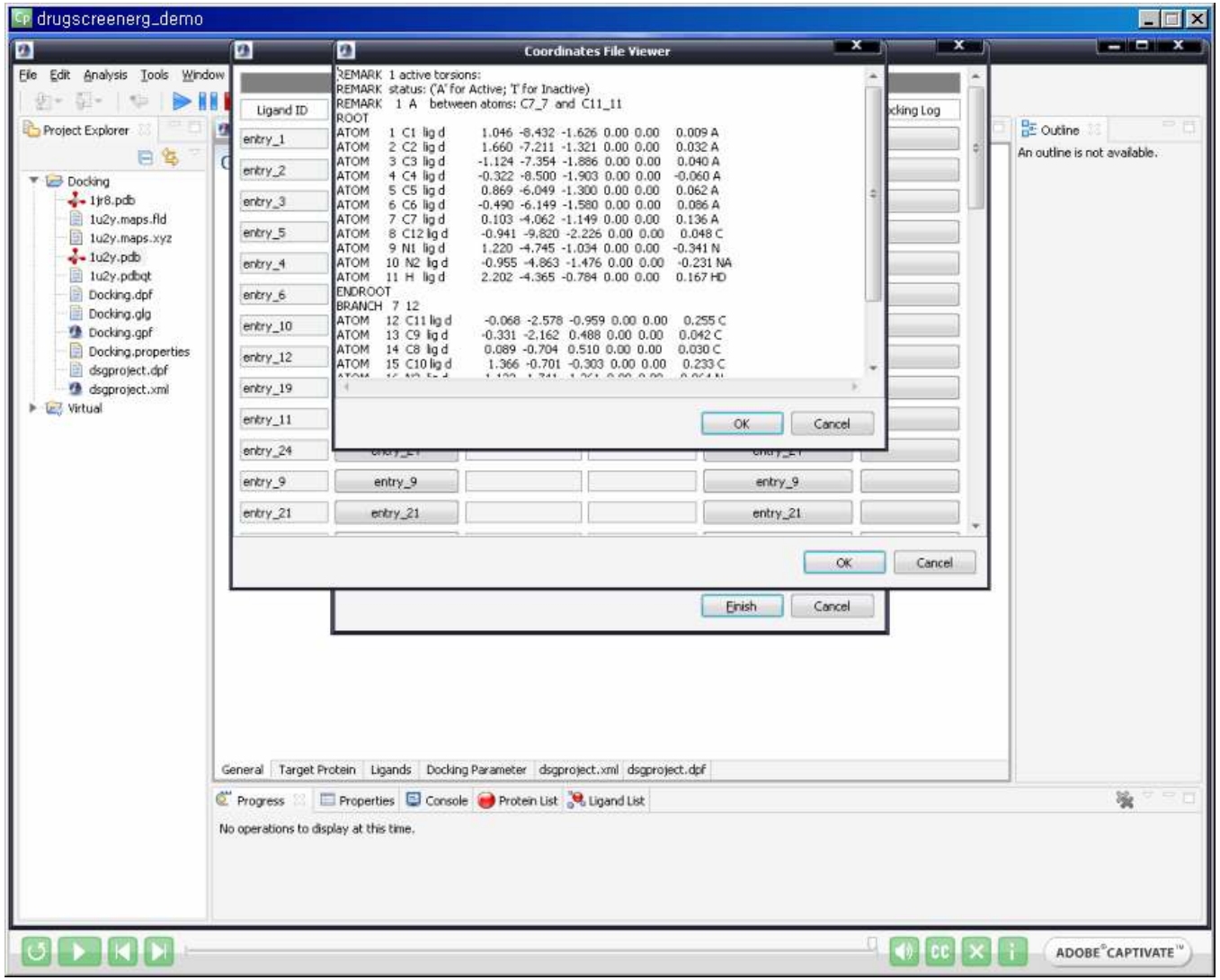
Notable features of DrugScreener-G – Ligand database management



Notable features of DrugScreener-G – Progress monitoring



Notable features of DrugScreener-G – Screening result view



Future work: more convenient features and tools for drug discovery users

- Support for R language for post-processing and analysis
  - Summary table of screening results as a data frame in R
- Workflow support
  - Motuer workflow engine for EGEE
- Consensus docking support
- Multi-step docking support with workflows
- Multiplatform support for Grid computing (Long Term Goal)
  - Based on the EGEE-DEISA interoperability model
  - Plans to develop resource selector of DrugScreener-G server for:
    - ✓ Globus Toolkit 4.0
    - ✓ UNICORE

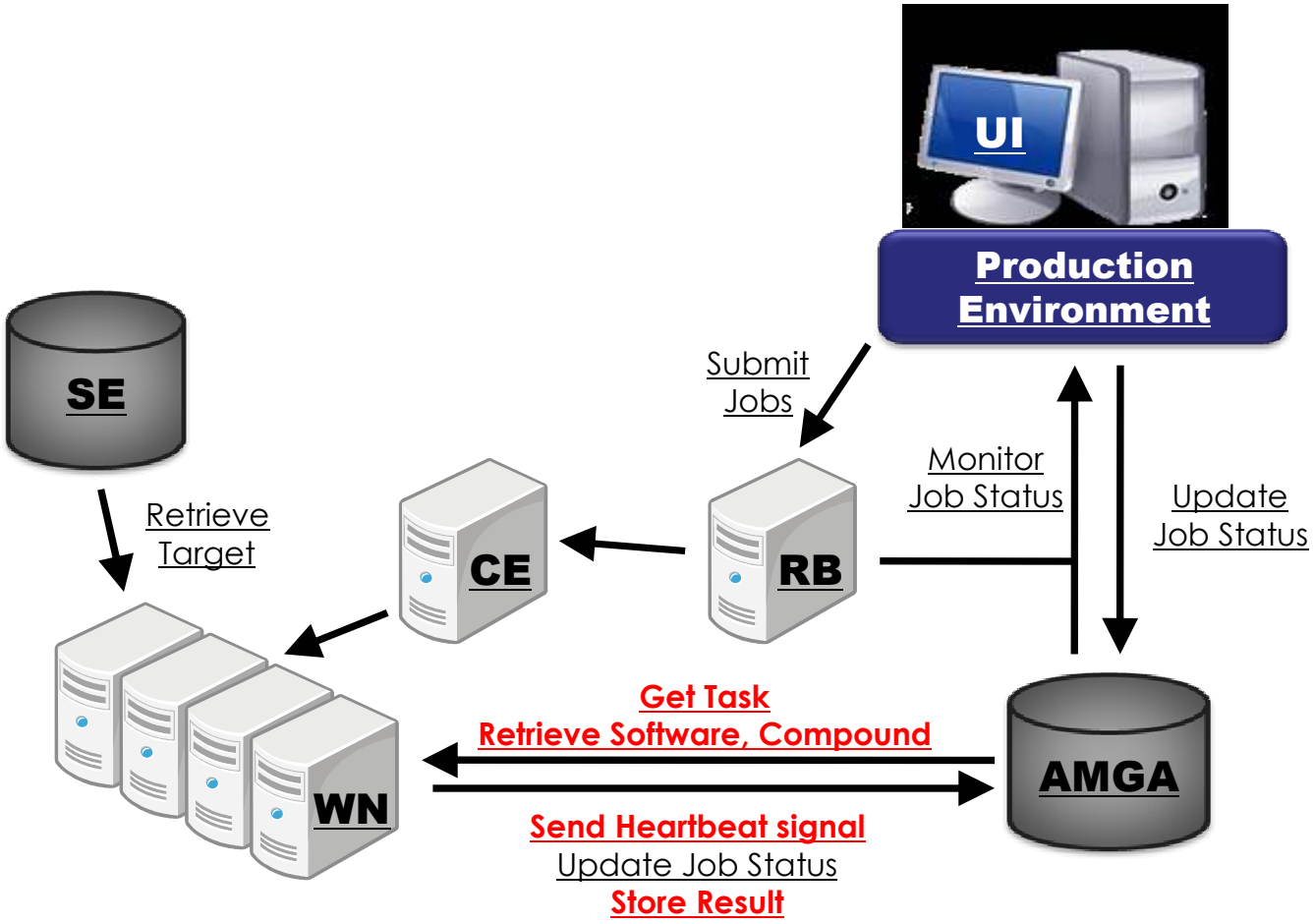
Future Work: Use of Molecular Dynamics

- Molecular Dynamics as a virtual screening tool
  - Screening drug candidates considering flexible conformation of proteins and compounds in solution with MD
  - Generating protein conformations with MD for flexible docking (Ensemble-based virtual screening with relaxed complex scheme)

- New developments for integration of Molecular Dynamics (future work)
  - User-friendly GUI for input file generation
  - Job management engine for Molecular Dynamics (as a plug-in for both of client and server)
  - Well-known molecular dynamics application support (CHARMM, AMBER, NAMD, etc.)

Performance enhancment of the WISDOM Production Environment

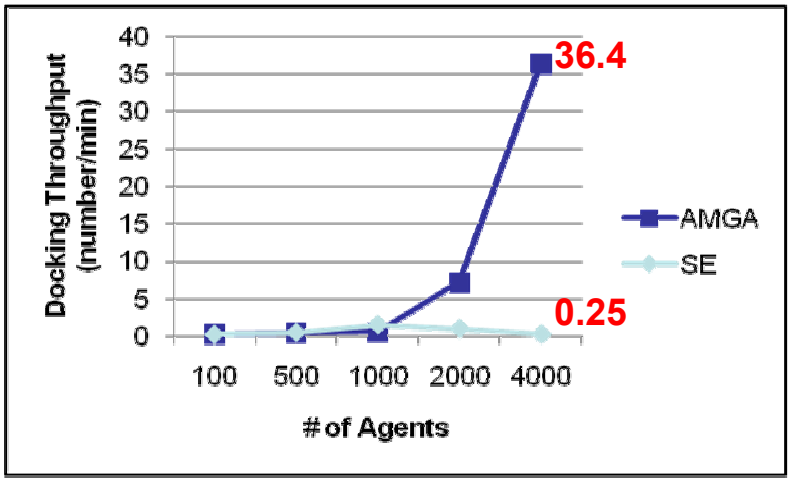
- Use of AMGA for task distribution and file store



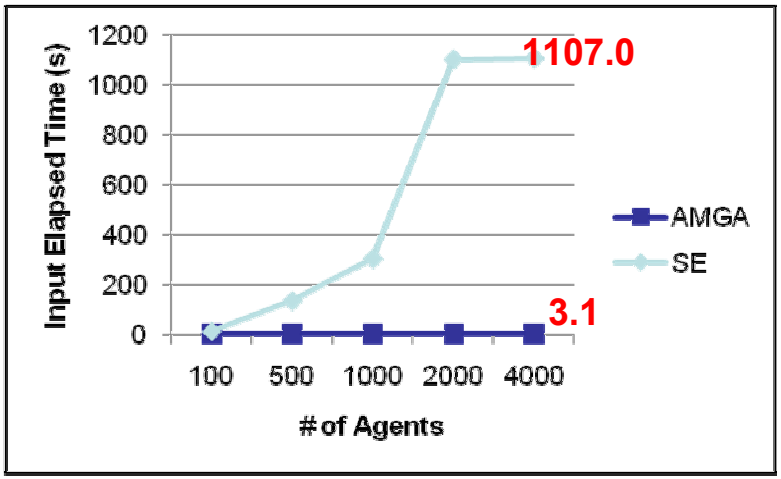
The exploitation of AMGA for storing ligands and docking results

- Why is AMGA used for the input/output data storing purpose?
  - AMGA performs way much better than SE in terms of throughput and latency

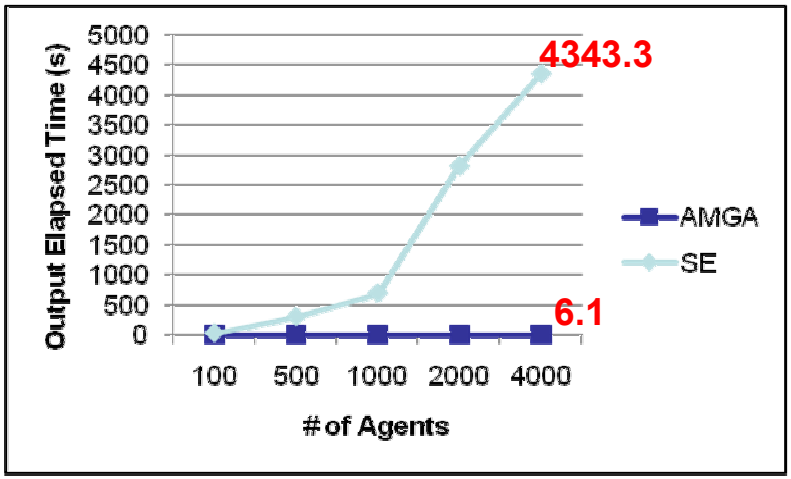
Docking Throughput



Time taken to retrieve a ligand



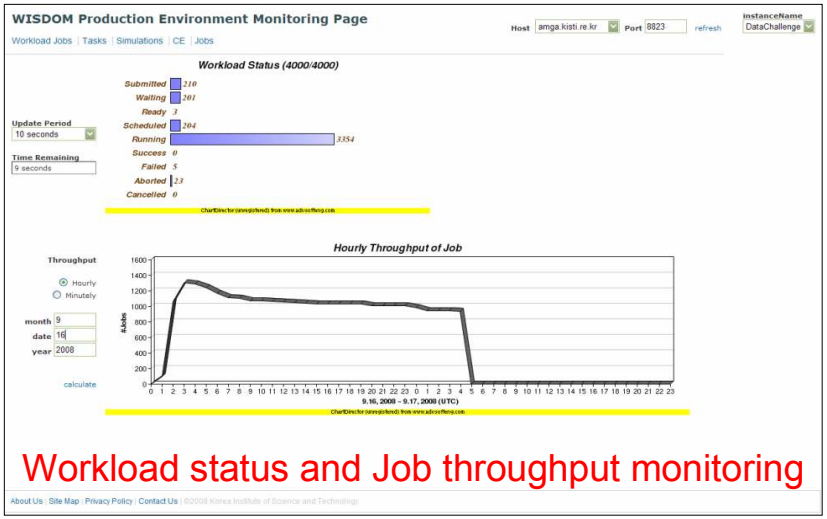
Time taken to store a docking result



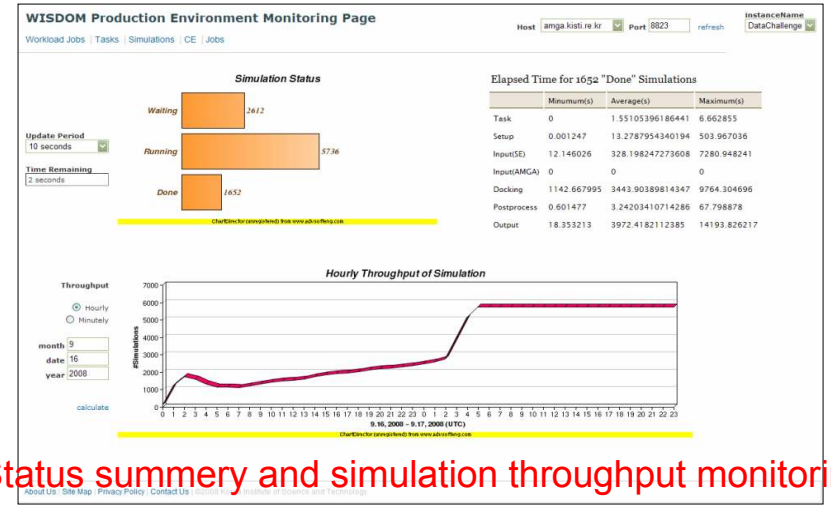
Additional Features to the Improvement of WISDOM PE by KISTI

- Flexible fault tolerance
  - Can recover a previous instance using data on AMGA
  - Can increase/decrease the number of agents of a running previous instance
- Improved Job distribution efficiency
  - Dynamic management of the list of available RB and CE
  - While the initial job submission keeps going, we can check the status of submitted jobs and resubmit failed jobs
  - Direct collecting the statistical data of sites from WN
  - Its own ranking based on previous statistical data
  - Timeout feature (submission timeout, heartbeat timeout)
- Enhanced monitoring and bookkeeping

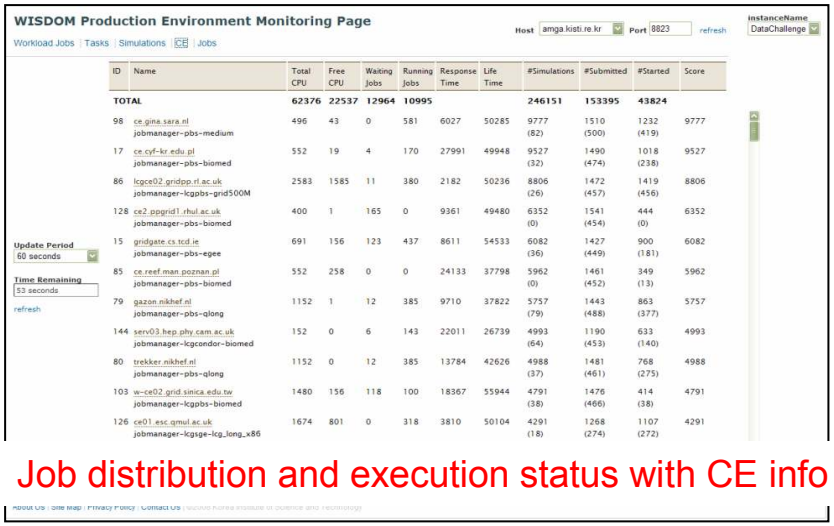
Enhanced monitoring and bookkeeping



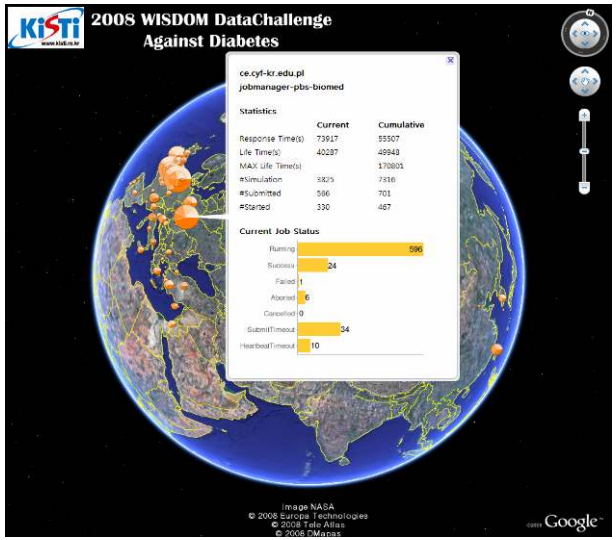
Workload status and Job throughput monitoring



Status summery and simulation throughput monitoring



Job distribution and execution status with CE info



Data challenge against Diabetes Type II with the improved WISDOM PE

- Data challenge to Human pancreatic amylase inhibitor (1u2y), a target protein of diabetes type 2, with 308310 chemical compounds
- Achieved significant improvement in efficiency and throughput

	WISDOM-II	DIANE	KISTI
Total number of dockings	156,407,400	308,585	308,310
Estimated duration on 1 CPU	413 years	16.7 years	39.0 years
Duration of the experiment	76 days	30 days	**2.4 days**
Cumulative number of Grid jobs	77,504	2,580	103,583
Maximum number of concurrent CPUs	5000	240	7,370
Number of used Computing Elements	98	36	127
Crunching Factor	1983	203	**5937**
Distribution Efficiency	39%	84%	**81%**

KISTI's contribution to AMGA included in a new version of AMGA

- AMGA1.9 released on Oct. 6, 2008
  - Support for the import of existing relational tables
  - Native SQL support
  - Multi-threaded Server
- Currently preparing AMGA 2.0 release
  - To be released by the end of this year
  - WS-DAIR support
  - Stability of WS-DAIR server needs to be improved before release

Summary

- KISTI's achievements in FKPPL in 2008
  - Completed development of a prototype version of DrugScreener-G software suite
  - Performed a part of the data challenge for human diabetes type II in collaboration with LPC/IN2P3 (target protein: alpha-amylase)
  - Improved the performance of WISDOM Production Environment by adopting AMGA to store production data
    - ✓ Also added more enhanced features, such as enhanced monitoring and bookkeeping, more fault-tolerant operation by CE, RB blacklisting, etc.
  - Development of new enhanced features of AMGA in collaboration with the AMGA team of CERN IT Division
    - ✓ Native SQL support
    - ✓ WS-DAIR support

Thank you for your attention!

Soonwook Hwang
Senior Researcher, Korea Institute of Science and Technology Information
hwang@kisti.re.kr (Soonwook Hwang)

Questions?

Backup slides