PC GRID use cases and requirements
Hiroyuki Takashima, Ph.D
OGF22, Enterprise Grids Requirements RG
Cambridge, 25 February 2008

Agenda
Enterprise Grids Requirements RG

- Overview of Novartis computing
- Scenarios
- Approaches
- Case studies and gap analysis
- Key technical requirements
Overview of Novartis GRID computing
*Enterprise Grids Requirements RG*

Novartis Institutes for Biomedical Research (NIBR) site map

Users and applications:
- Group A
- Group B
- Group C
- Group D
- Group E

Cluster existing HPC infrastructure:
- Large shared Linux Clusters
- Shared PC grid
- Shared Multi-processor servers
- External Collaborations

Job submission protocol

HPC technologies integration/sharing layer

4 | PC GRID use cases and requirements | H. Takashima | 25 Feb 2008 | OGF22
Overview of Novartis GRID computing

*Enterprise Grids Requirements RG*

The 22nd Open Grid Forum - OGF22
Cambridge, MA, USA

**Wednesday, February 27**
2:06 pm - 2:30 pm.

**Pharma, Biotech and Life Sciences Workshop** (Ballroom A/(B/C)) (6 of 6)

Steve Litster Ph.D, Novartis Institutes for BioMedical Research

Main topic: Storage virtualization

---

**Scenarios (Business cases)**

*Enterprise Grids Requirements RG*

- Medicinal chemistry
- Structural biology
- Bioinformatics and knowledge mining
**Scenarios (Business cases)**

*Enterprise Grids Requirements RG*

- **Structural biology**
  - Protein structure determination
  - Ligand interaction analysis

- **Target structure based drug design**

- **Bioinformatics and knowledge mining**
  - Target findings

- **Medicinal chemistry**
  - Docking simulations
  - Virtual screening
  - Lead optimization

**Approaches (global project: “ENGINE”)**

*Enterprise Grids Requirements RG*

- Enterprise next generation environment (ENGINE) from 2000 to 2003
  - Globally integrated hardware and platform
    - Windows XP Desktop PC, Laptop PC and Windows 2K/2K3 servers
  - Full Active Directory (AD) implementation
Approaches (global project: “ENGINE”)
Enterprise Grids Requirements RG

Life cycle management for…
- PC hardware
  - Distribution of hardware specific images (base of SLA)
- OS and service packs by OSD
  - Version management
- Applications by ESD with DIPR process (Development, Integration, Productive, Retirement)
  - Global Application Suite and Hot fix
  - Business specific Application portfolio
  - Business applications

Ownership (roles and responsibilities)
- Application owner, Application portfolio manager, GRID service manager etc…..
Approaches (Local build in environments)
Enterprise Grids Requirements RG

- Linux clusters
  - Site oriented (Basel, Cambridge etc.)
  - Test cluster in Tsukuba
    - 20 Linux PC controlled by Sun Grid Engine (SGE)
    - Logically separated private segments by VPN router

---

Approaches (Local build in environments of Tsukuba)
Enterprise Grids Requirements RG
Case studies and gap analysis

Enterprise Grids Requirements RG

- High-Throughput Docking for *in silico* screening
- Protein structure determination
- Vector/Parallel/Distributed jobs bench marks

---

**Case Study 1: High-Throughput Docking for *in silico* screening**

Enterprise Grids Requirements RG

- **Target structure determination**
- ** millones compounds**
- **Tens thousands of compounds**
- **Tens of compounds**
- **Bio assays**

**Protein Databank, X-ray, NMR, Molecular Modeling, etc.**

**Linux clusters, and/or 2700 PC GRID**

**Controlled by Sun Grid Engine and/or United Devices Grid Engine**
Case Study 1: High-Throughput Docking for in silico screening

Enterprise Grids Requirements RG

Discovery of a Potent and Selective Protein Kinase CK2 Inhibitor by High-Throughput Docking


Case Study 2: Protein structure determination

Enterprise Grids Requirements RG

- X-ray
- NMR
- Electron microscopy
- Computer modeling
Case Study 2: Protein structure determination  
*Enterprise Grids Requirements RG*

**Experimental data**
- Protein sample preparation, NMR experiments, Data analysis

**Molecular Dynamics calculations**

- Initial structures
  - Energy minimized structures
  - Refined structures
  - Linux clusters 20 CPUs
  - Controlled by Sun Grid Engine
  - Protein Databank, Medicinal chemistry

**Case Study 2: Protein structure determination  
*Enterprise Grids Requirements RG*  
Distributed Computing and NMR Constraint-Based High-Resolution Structure Determination: 
Applied for Bioactive Peptide Endothelin-1 to Determine C-terminal Folding.

**Structural precision**

**Figure 1.** (a) The precision (stranged mean values of all backbone atoms of ET-1) have been calculated for 29 minimum energy structures out of various numbers of calculated structures with 50-ps simulated annealing at 
initial temperature 5000 K. (b) Overlay for all residues 1 to 21. 
(c) Overlay for residues 3 to 15. 
(d) The same plot as (a), expanding the horizontal axis.

**Figure 2.** (a) Overlay of 20 minimum energy structures of ET-1, pairwise backbone rms 0.58 ± 0.26 Å. 
(b) The lowest energy structure of (a), with ribbon drawing, schematically representing α-helix and β-turns in red and green, respectively.

**PDB ID: 1v6r**
Case Study 2: Protein structure determination

Solution NMR Structure Investigation for Releasing Mechanism of Neocarzinostatin Chromophore from the Holoprotein.

High-Resolution Protein Structure Determination by NMR

Structure of Cytocrome c552 from a Moderate Thermophilic Bacterium, Hydrogenophilus thermoluteolus: Comparative study on the thermostability of cytochrome c.

Key technical requirements

**Limitation as is**
- Difficult license management
  - Number of CPUs is limited by software licenses
- Scientific application platform
  - Unix/Linux base to Windows base software
- Network performance
  - Especially at remote sites
- Overhead of parallel calculations
- Calculation precisions
  - With or without induced fitting, Free energy calculations, Protein dynamics

PDB ID: 1o5p
PDB ID: 2ai5
Key technical requirements

Enterprise Grids Requirements RG

- Requirements to be
  - Precision as high as possible (depending on computer power)
    - Expand available applications
    - Real time simulations
  - More computer power on demand (from user’s point of view)
  - Software license virtualization

Thank you for your collaboration!