PC GRID use cases and requirements

Hiroyuki Takashima, Ph.D

OGF22, Enterprise Grids Requirements RG

Cambridge, 25 February 2008

NOVARTIS
INSTITUTES FOR
BIOMEDICAL RESEARCH

Agenda

Enterprise Grids Requirements RG

- Overview of Novartis computing
- Scenarios
- Approaches
- Case studies and gap analysis
- Key technical requirements

NOVARTIS

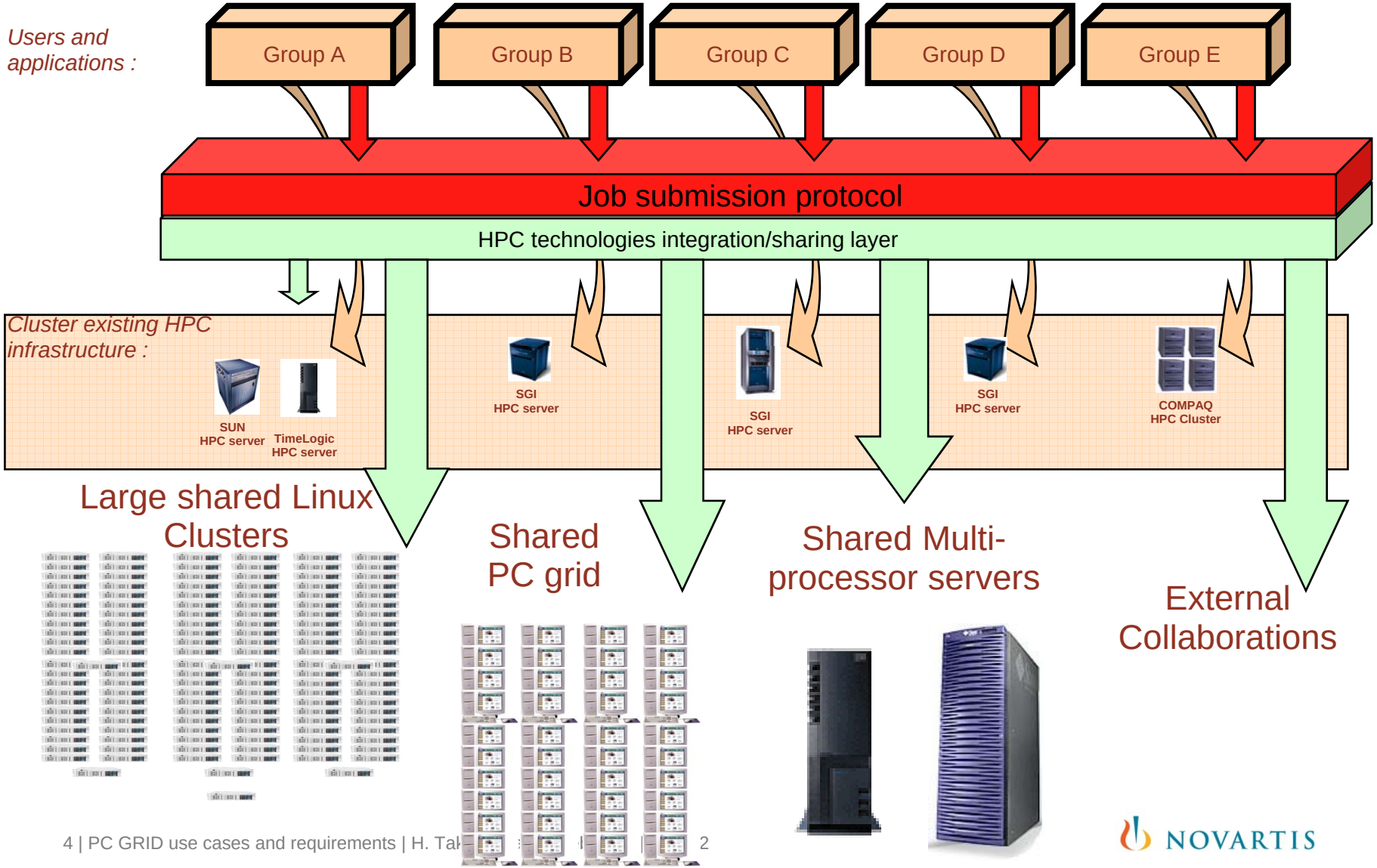
Overview of Novartis GRID computing

Enterprise Grids Requirements RG

Novartis Institutes for Biomedical Research (NIBR) site map

Overview of Novartis GRID computing

Enterprise Grids Requirements RG



NOVARTIS

Overview of Novartis GRID computing

Enterprise Grids Requirements RG

Steve Litster Ph.D, Novartis Institutes for BioMedical Research

Main topic: Storage virtualization

NOVARTIS

Scenarios (Business cases)

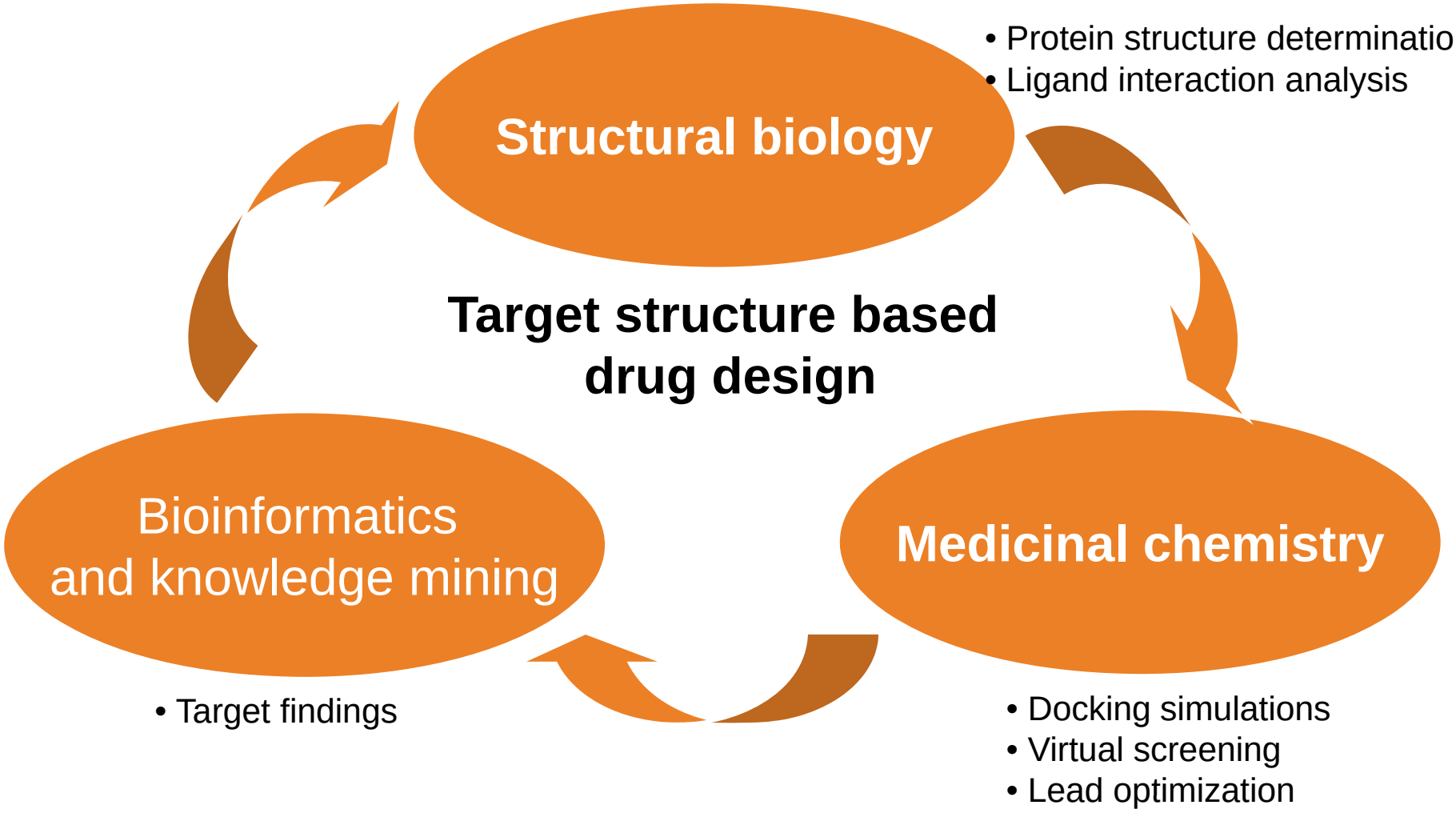
Enterprise Grids Requirements RG

- Medicinal chemistry
- Structural biology
- Bioinformatics and knowledge mining

NOVARTIS

Scenarios (Business cases)

Enterprise Grids Requirements RG



NOVARTIS

Approaches (global project: “ENGINE”)

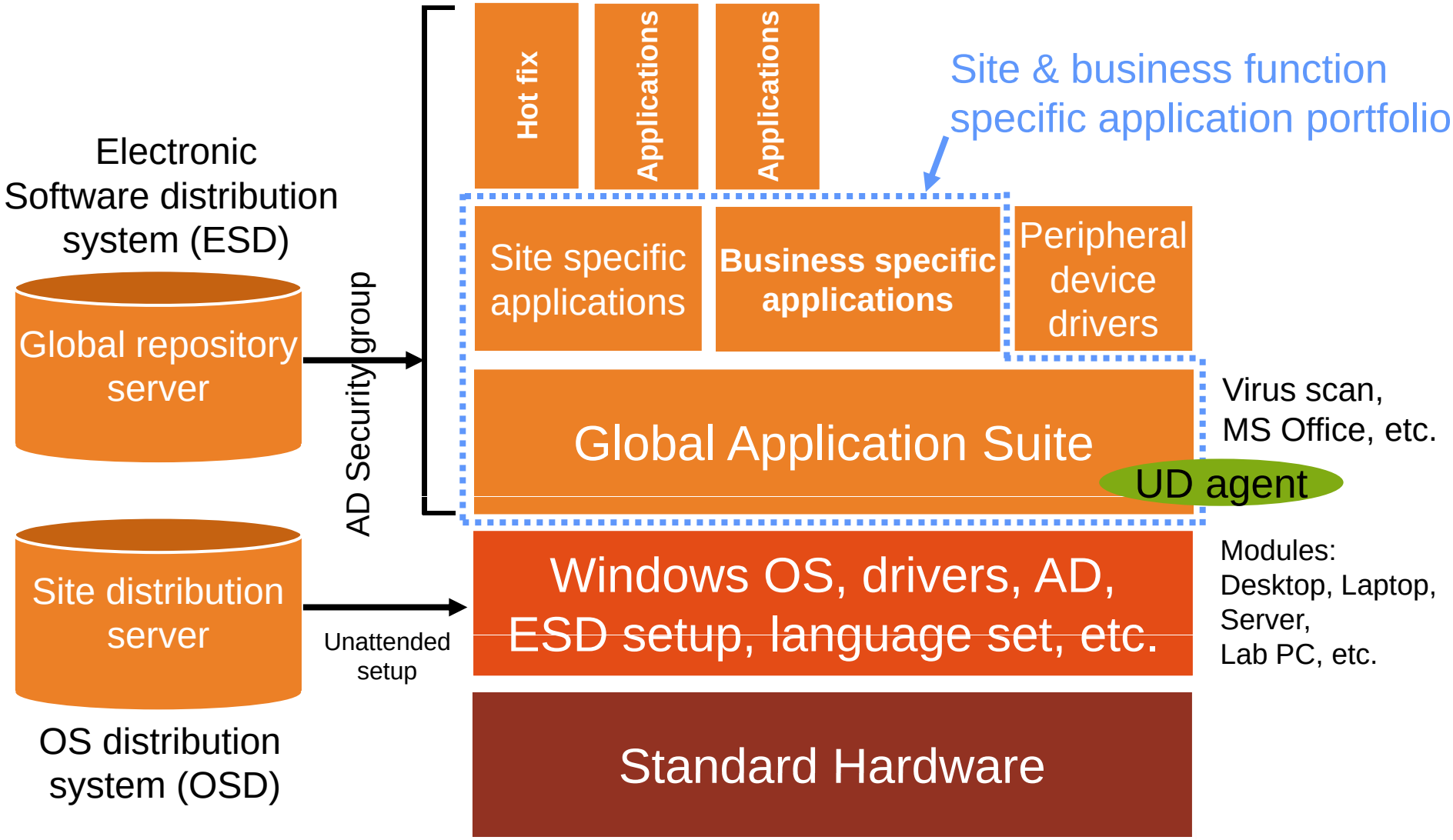
Enterprise Grids Requirements RG

- 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

NOVARTIS

Approaches (global project: “ENGINE”)

Enterprise Grids Requirements RG



NOVARTIS

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…..

NOVARTIS

Approaches (Local build in environments)

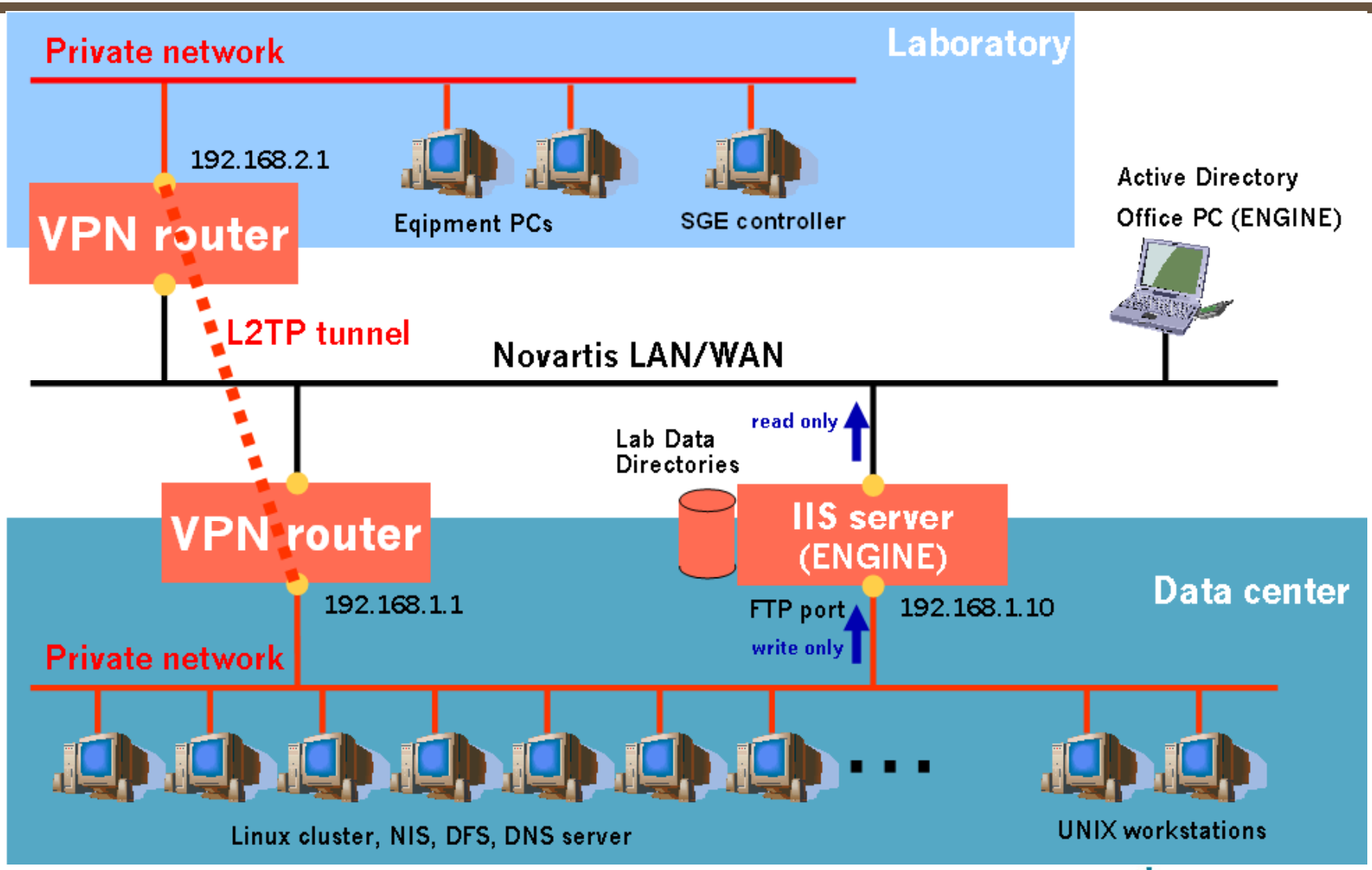
Enterprise Grids Requirements RG

- 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

NOVARTIS

Approaches (Local build in environments of Tsukuba)

Enterprise Grids Requirements RG



NOVARTIS

Case studies and gap analysis

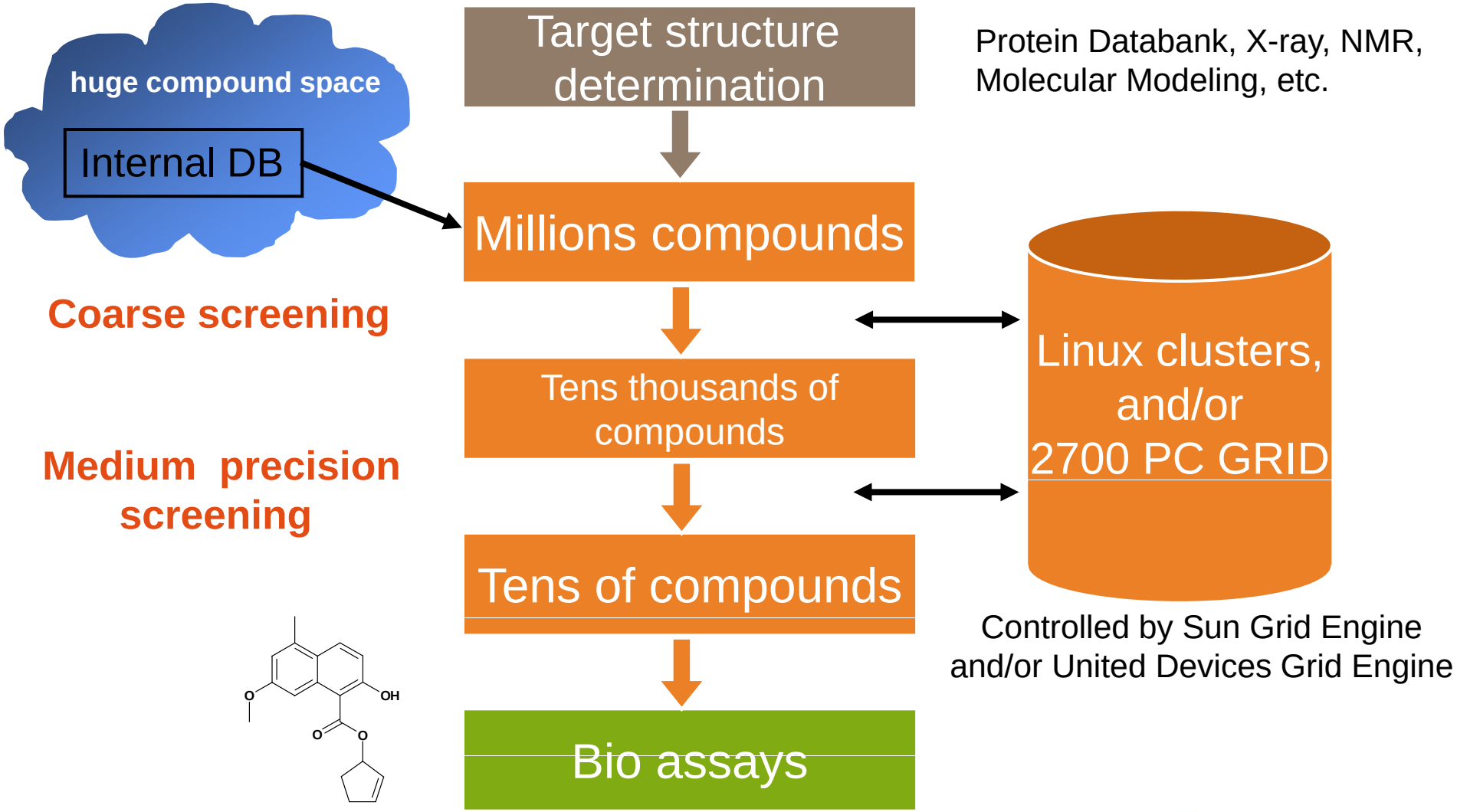
Enterprise Grids Requirements RG

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

NOVARTIS

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

Enterprise Grids Requirements RG



NOVARTIS

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
Vangrevelinghe, E., Zimmermann, K., Schoepfer, J., Portmann, R., Fabbro, D., and Furet, P. (2003) *J. Med. Chem.* **46**, 2656-2662.

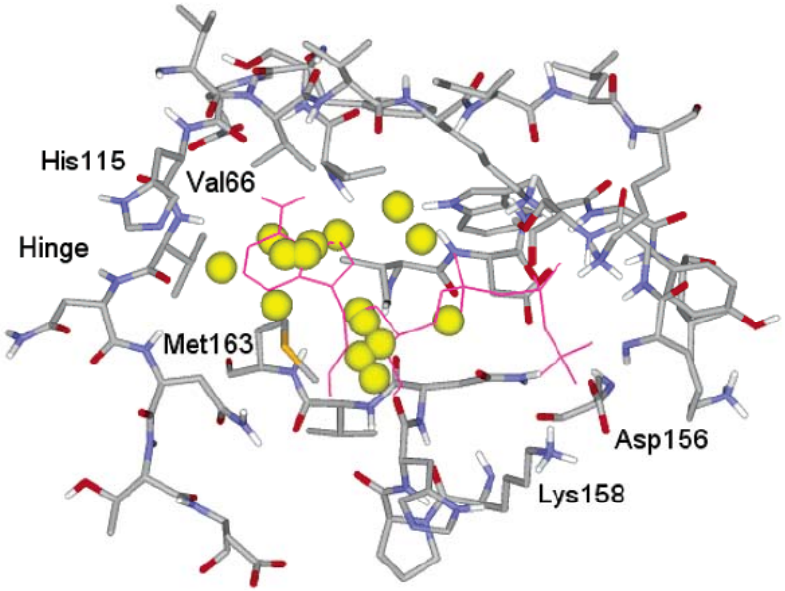


Figure 2. Overlay of the docking site points used (yellow spheres) and the ATP analogue extracted from the crystallographic template structure 1DAW (magenta) inside the active site of the homology model of human CK2α.

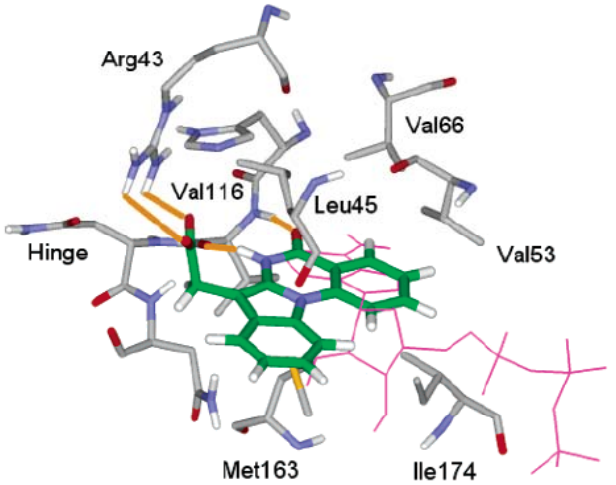


Figure 3. Relative binding modes of compound **4** (green) and the ATP analogue extracted from the crystallographic template structure 1DAW (magenta) in the human CK2α active site. The hydrogen bonds formed with Arg 43 and Val 116 are indicated as orange lines.

NOVARTIS

Case Study 2: Protein structure determination

Enterprise Grids Requirements RG

- X-ray
- NMR
- Electron microscopy
- Computer modeling

NOVARTIS

Case Study 2: Protein structure determination

Enterprise Grids Requirements RG

Experimental data

Protein sample preparation,
NMR experiments, Data analysis

Initial structures

huge
conformational space

Energy minimized
structures

**Molecular Dynamics
calculations**

Refined structures

Linux clusters
20 CPUs

Controlled by Sun Grid Engine

Protein Databank,
Medicinal chemistry

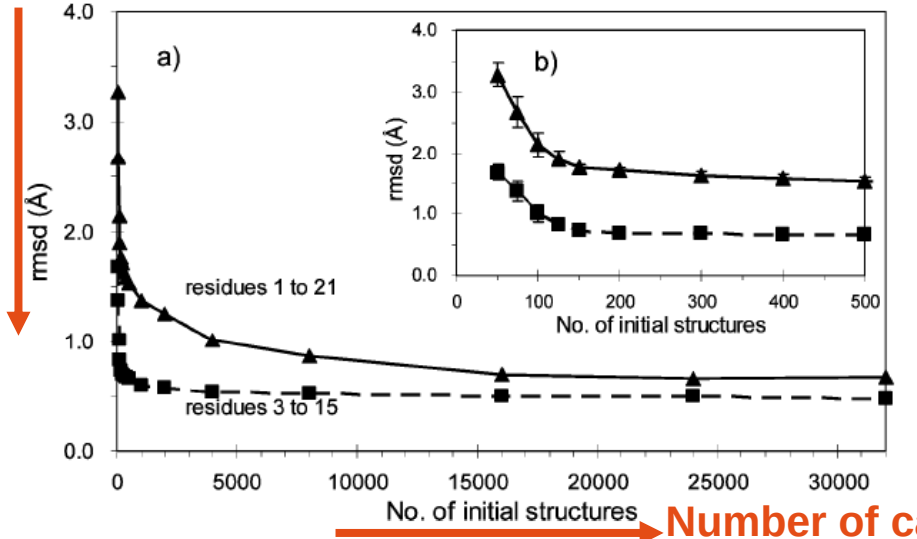
NOVARTIS

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.
Takashima, H., Mimura, N., Ohkubo, T., Yoshida, T., Tamaoki, H., and Kobayashi, Y. (2004) *J.Am.Chem.Soc.* **126**, 4504-4505.

Structural precision



Number of calculations

Figure 1. (a) The pairwise averaged rmsd values of all backbone atoms of ET-1 have been elucidated for 20 minimum energy structures out of various numbers of calculated structures with 80-ps simulated annealing at initial temperature 5000 K. (▲) Overlay for all residues from 1 to 21. (■) Overlay for residues 3 to 15. (b) The same plot as (a), expanding the horizontal axis.

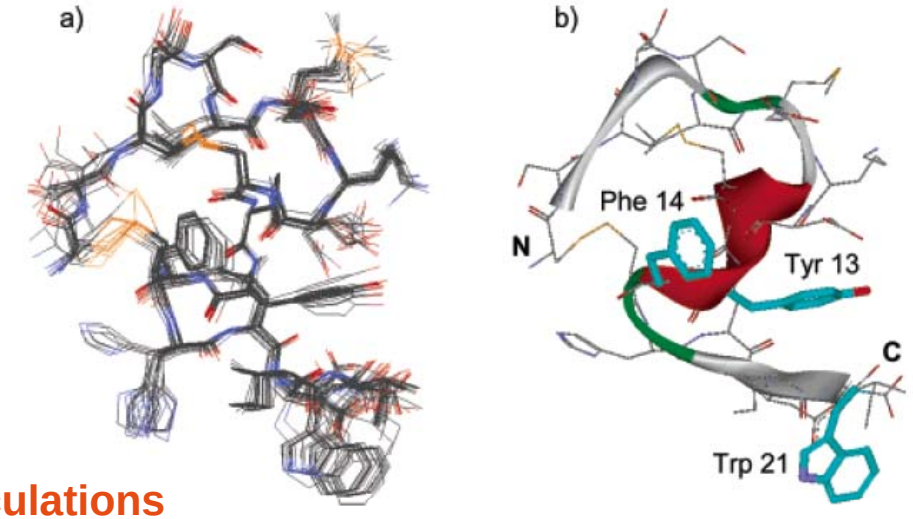


Figure 2. (a) Overlay of 20 minimum energy structures of ET-1, pairwise backbone rmsd 0.58 ± 0.28 Å. (b) The lowest energy structure of (a) with ribbon drawing, schematically representing α-helix and β-turns in red and green, respectively.

PDB ID: 1v6r

NOVARTIS

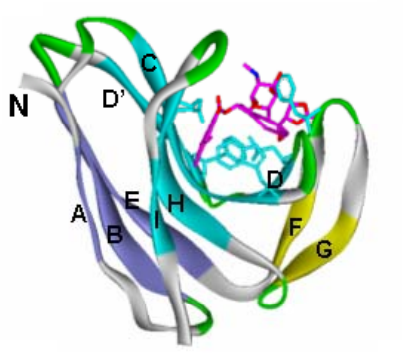
Case Study 2: Protein structure determination

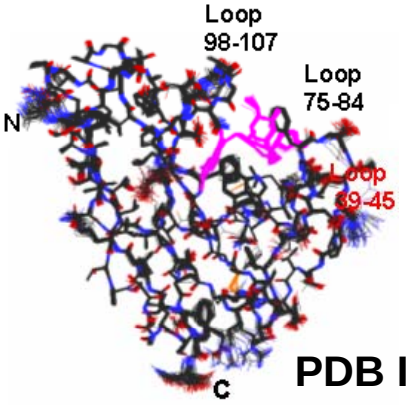
Enterprise Grids Requirements RG

Solution NMR Structure Investigation for Releasing Mechanism of Neocarzinostatin Chromophore from the Holoprotein.
Takashima, H., Yoshida, T., Ishino, T., Hasuda, K., Ohkubo, T., and Kobayashi, Y.
(2005) *J. Biol. Chem.* **280**, 11340-11346.

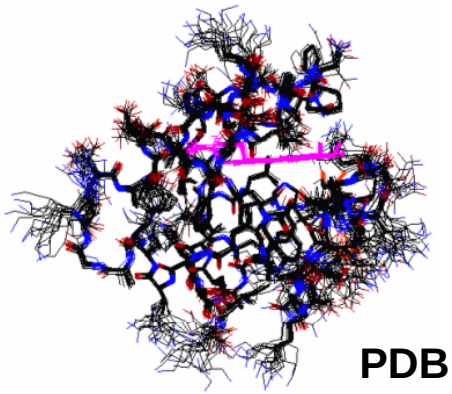
High-Resolution Protein Structure Determination by NMR
Takashima, H. (2006) in *Annual Reports on NMR Spectroscopy* **59**, 235-273. printed by Elsevier Ltd., London

Structure of Cytocrome c552 from a Moderate Thermophilic Bacterium, Hydrogenophilus thermoluteolus: Comparative study on the thermostability of cytochrome c.
Nakamura, S., Ichiki, S., Takashima, H., Uchiyama, S., Hasegawa, J., Kobayashi, Y., Sambongi, Y., and Ohkubo, T. (2006) *Biochemistry* **45**, 6115-6123.





PDB ID: 1o5p



PDB ID: 2ai5

NOVARTIS

Key technical requirements

Enterprise Grids Requirements RG

- 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 Force fields, Solvent and more……

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

NOVARTIS

Thank you for your collaboration!

NOVARTIS