COST Training School on Molecular and Material Science Grid Applications

Trieste, Italy

March 29 – April 1, 2010

Molecular Dynamics on the Grid

A. Costantini

[1] Dept of Chemistry, University of Perugia, Perugia, Italy

[2] Dept. of Mathematics and Informatics, University of Perugia, Perugia, Italy

[3] COMPCHEM Virtual Organization

alex@dyn.unipg.it

Summary

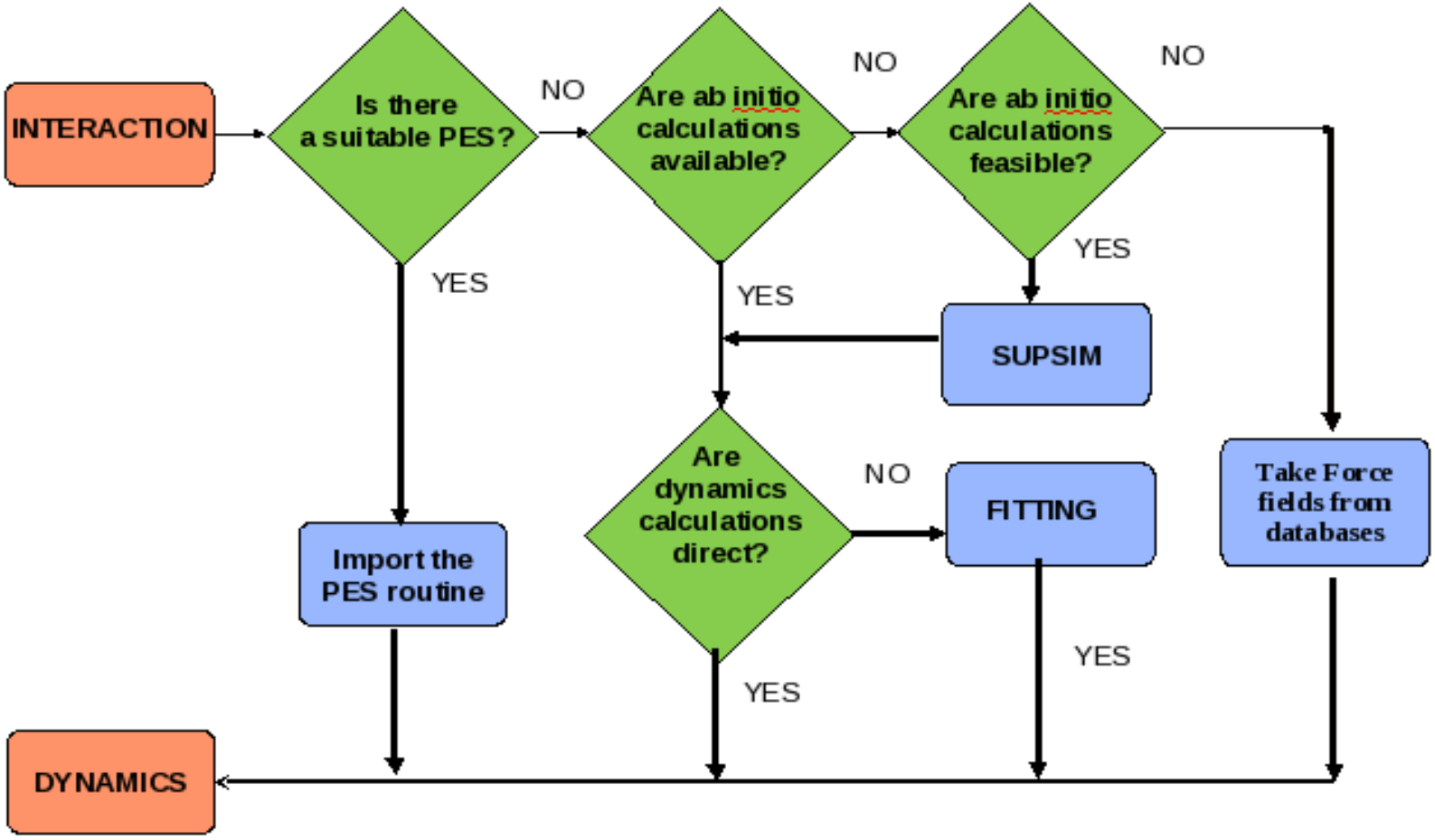
- GEMS and the extension to MD
- Porting of GROMACS MD package on the Grid environment
- Merging HPC and HTC resources
  - A new distribution scheme
- Structuring DL_POLY
  - Future prospects

Grid Enabled Molecular Simulator

- Specialized in dealing with atom-diatom quantum reactive scattering problems
- Works within the limits of Born-Oppenheimer approximation
- Structured as a workflow and divided in three blocks
  - INTERACTION
  - DYNAMICS
  - OBSERVABLES

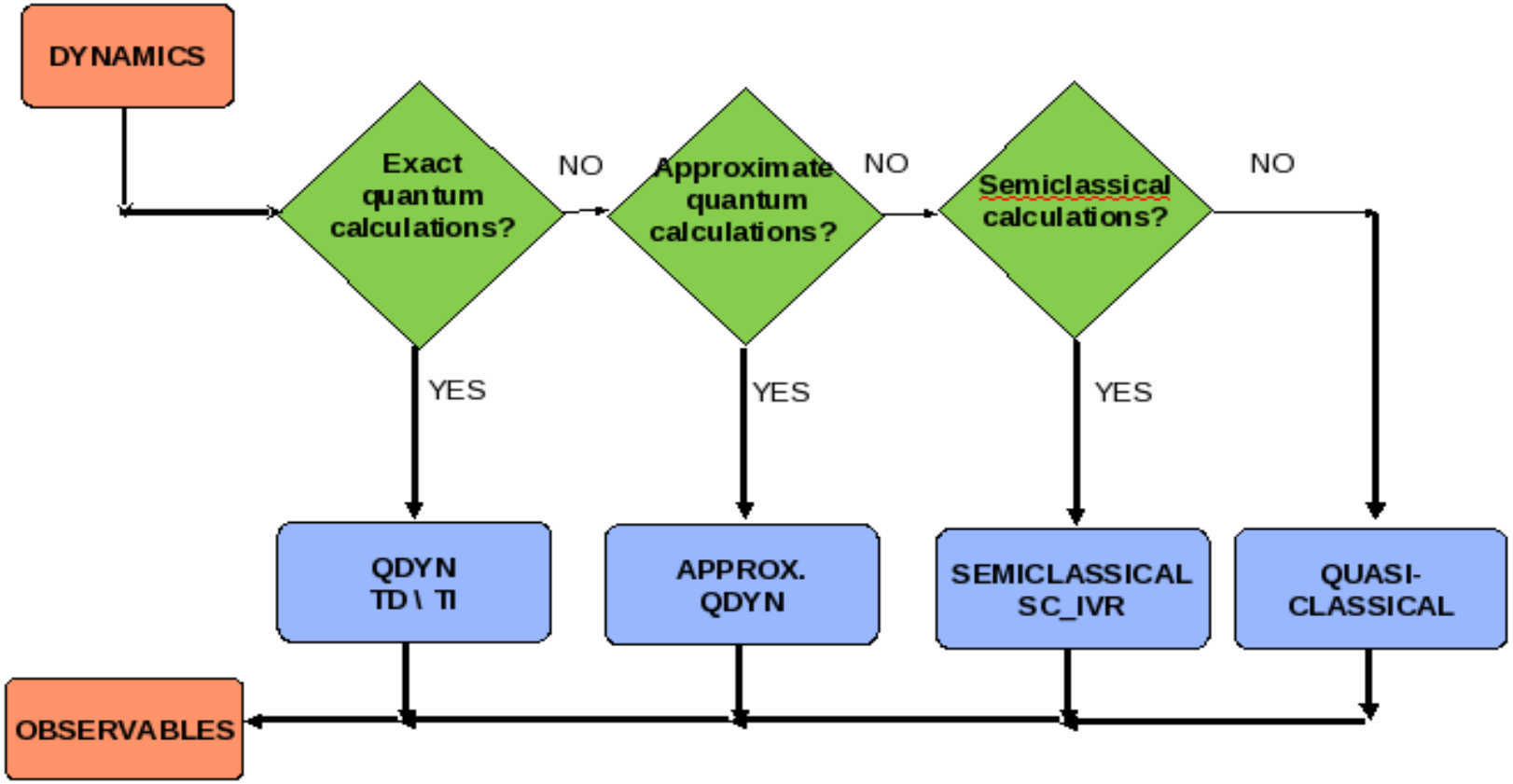
Structure of GEMS

- Devoted to the building of the eigensolutions of the electronic system



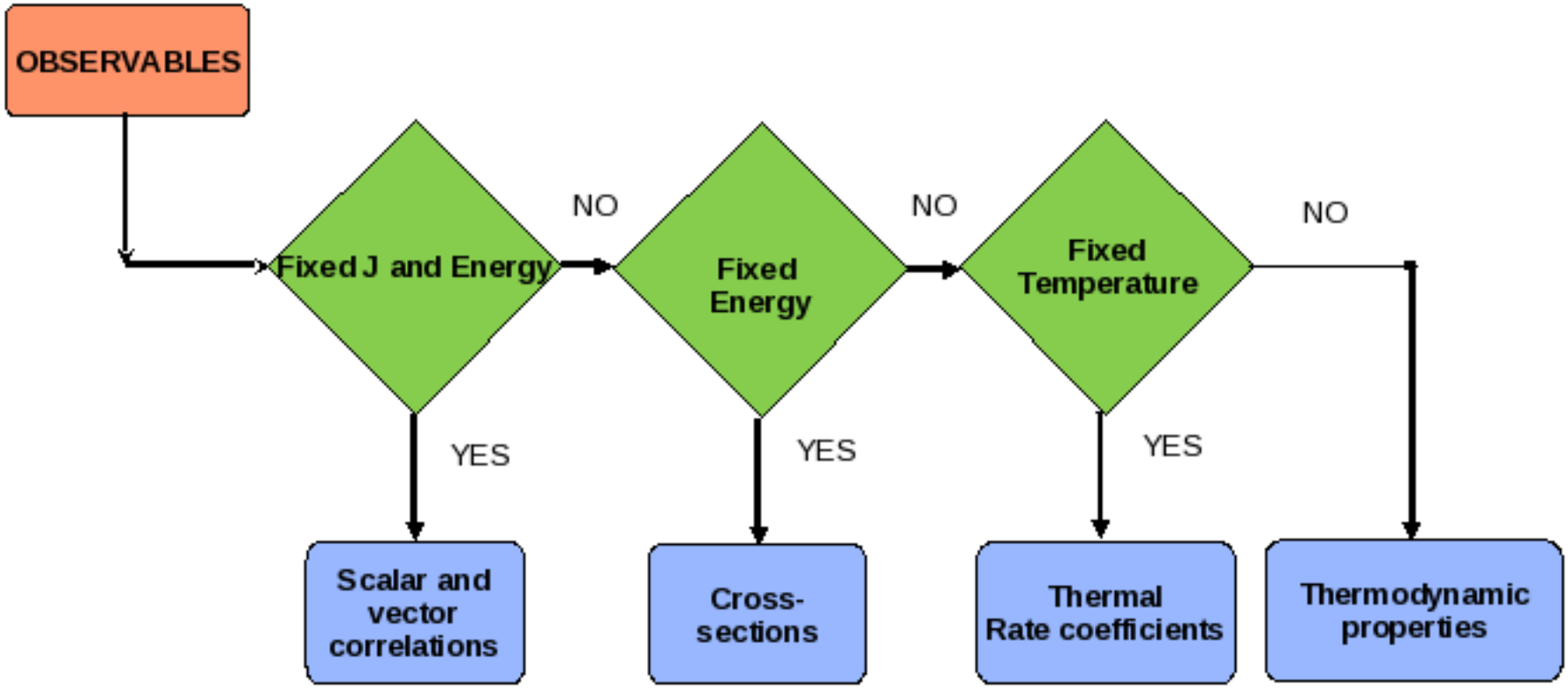
Structure of GEMS

- Integrates the equations of motions of the nuclei



Structure of GEMS

•Builds up the macroscopic properties and experimental observables



On the extension of GEMS

- GEMS deal recently with the effect of external fields
  - Extension of GEMS to MD calculations for large systems
- Porting of GROMACS package in to the Grid environment
  - Develop a set of workflows and Grid Enabled visualization tools

GROMACS code

- Molecular Dynamics package
  - Primarily designed for biochemical molecules
  - Perform well on scalar machines
  - Satisfactory scale up on parallel machines
  - Can be statically compiled assuring binary compatibility

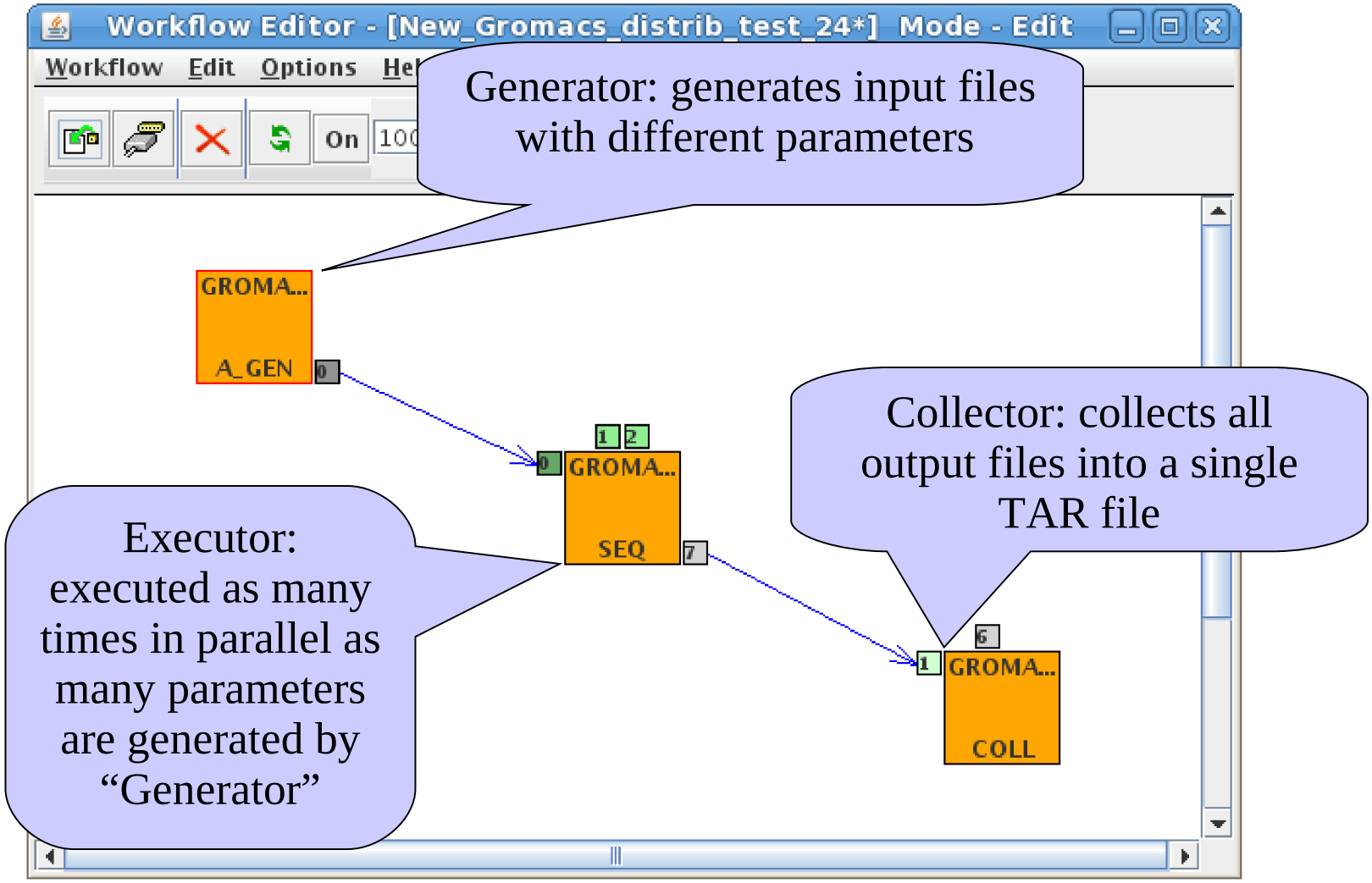
GROMACS code II

- Significant Memory demand
- Consuming a large amount of CPU time
- In a “Parameters Study” fashion each simulation is independent from the other
  - Can be executed several times for different sets of input parameters
- Can be structured as a workflow

The Gridification of GROMACS

- The porting has been carried out making use of the P-GRADE Grid Portal 2.7 implemented in COMPCHEM
  - Open source tool
  - Collect Grid resources
  - The generic application structure is a workflow
  - Specific graphical tools can be added
  - Does not require to modify the original code

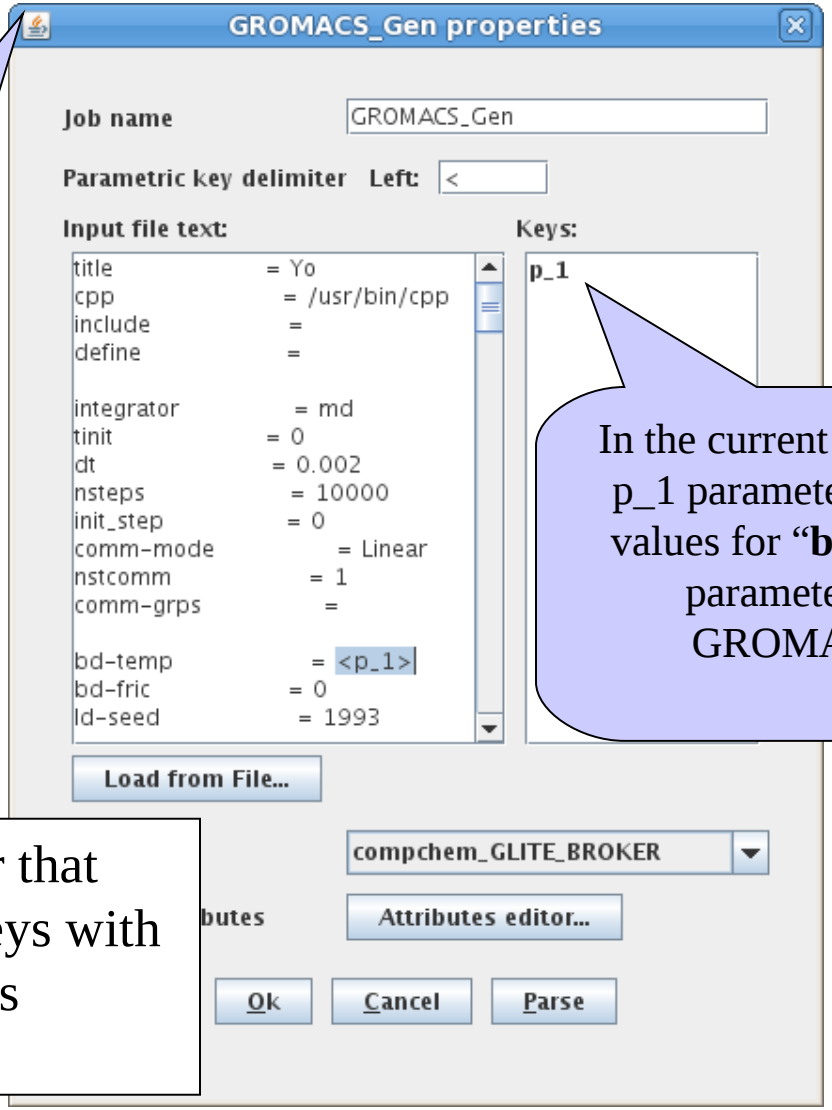
GROMACS workflow



GROMACS Generator

Template text with keys. Keys will be replaced with actual numbers by the Generator during the execution of the workflow.

All the possible combinations of the replaced template are written into separate files.



In the current workflow p_1 parameter defines values for "**bd-temp**" parameters of GROMACS.

Generator job is a macro processor that generates text files by replacing keys with actual values in a template which is defined by the user.

GROMACS Executor

- Manage the execution (in parallel) of the GROMACS jobs generated by the Generator
- Is made of a bash script
  - Download GROMACS executable from a central location
  - Configure the environment variables
  - Execute GROMACS on the EGEE Grid environment
  - Store the output files in the SE

GROMACS Collector

- Is made of a bash script
  - Collects the results of the parameter study workflow in a single zipped file
  - Store the file directly on the server
  - Able to crate a user friendly filtered results

Benchmark test & platform used

Benchmark

- "Water energy minimization"
  - GROMACS Tutorial
  - Different jobs works with different temperatures

Platform

- COMPCHEM VO computational resources (EGEE Grid)

COMPCHEM VO

COMPCHEM VO (http://compchem.unipg.it)

deploy Computational Chemistry applications running on a segment of the production EGEE Grid

- Runs on the EGEE production Grid from the end of 2004
- 80 total users; 40 active users
- 8000 CPUs (~10% of the EGEE resources)

COMPCHEM VO: Users & Resources

Organization	User	Contact email
University of Barcelona	Margarita Alberti	maw <at>qf.ub.es
University of the Basque Country (Spain)	Ernesto Garcia	qfpgapae <at>vc.ehu.es
University of Perugia (Italy)	Dimitris Skouteris. Leonardo Arteconi, Alessandro Costantini, Max Porrini	{dimitris, bodynet, alex, max} <at>dyn.unipg.it
Cyfronet (Poland)	Mariusz Sterzel	msterzel <at>cyfronet.pl
IMIP- Italian National Research Council (Italy)	Domenico Bruno	d.bruno <at>area.ba.cnr.it
CESGA (Spain)	Javier Lopez	jlopez <at>cesga.es
University of Vienna (Austria)	Hans Lishcka	hans.lischka <at>univie.ac.at
University of Cyprus (Cyprus)	Constantinos Zeinalipour-Yazdi	zeinalip <at>ucy.ac.cy
ENEA (Italy)	Carlo Scio	scio <at>frascati.enea.it
Vilius University (Lithuania)	Gintaras Urbelis	gintaras.urbelis <at>chf.vu.lt
Center fo Parallel Computers (Sweden)	Olav Vatras	vahtras <at>pdc.kth.se
University of Crete and FORTH (Greece)	Stavros C Farantos	farantos <at>iesl.forth.gr

COMPCHEM VO: Applications

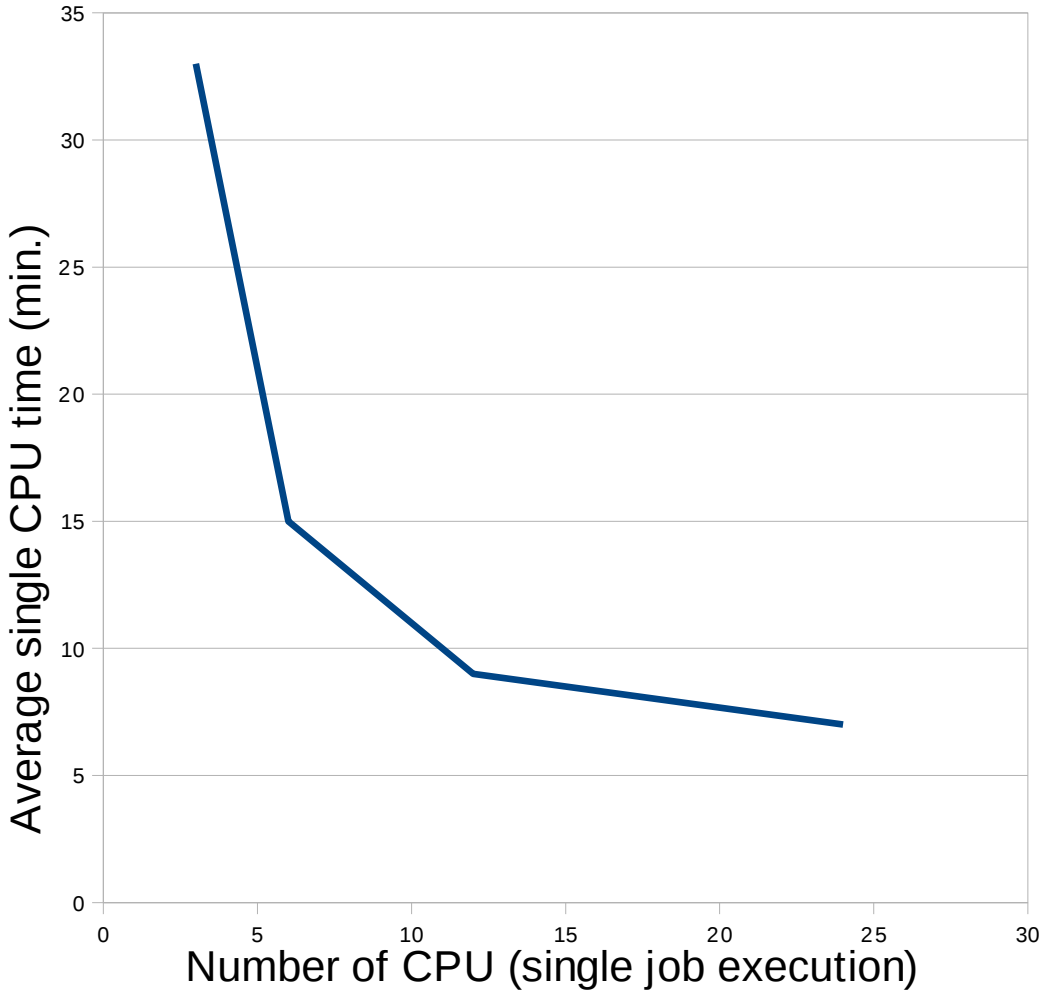
COLUMBUS **Vienna (Austria)** **high-level *ab initio* molecular electronic structure calculations.**
GAMESS **Catania (Italy)** **high-level *ab initio* molecular quantum chemistry**

ABC **Perugia (Italy), Budapest (Hungary)** **quantum time-independent reactive dynamics**
RWAVEPR **Perugia (Italy), Vitoria (Spain)** **quantum time-dependent reactive dynamics**
MCTDH **Barcelona (Spain)** **multi-configurational time-dependent Hartree method**
FLUSS **Barcelona (Spain)** **Lanczos iterative diagonalisation of the thermal flux operator**
DIFF REAL WAVE **Melbourne (Australia)** **quantum differential cross-section**

VENUS **Vitoria (Spain)** **classical mechanics cross sections and rate coefficients**
DL_POLY **Iraklion (Greece)** **molecular dynamics simulation of complex systems**
GROMACS **S. de Compostela (Spain)** **molecular dynamics simulation of complex systems**
CHIMERE **Perugia (Italy)** **chemistry and transport eulerian model for air quality simulations**

Performances

Result of GROMACS



Execution of 4 GROMACS jobs in a parameter study fashion on the EGEE Grid platform using 3, 6, 12 and 24 CPUs respectively

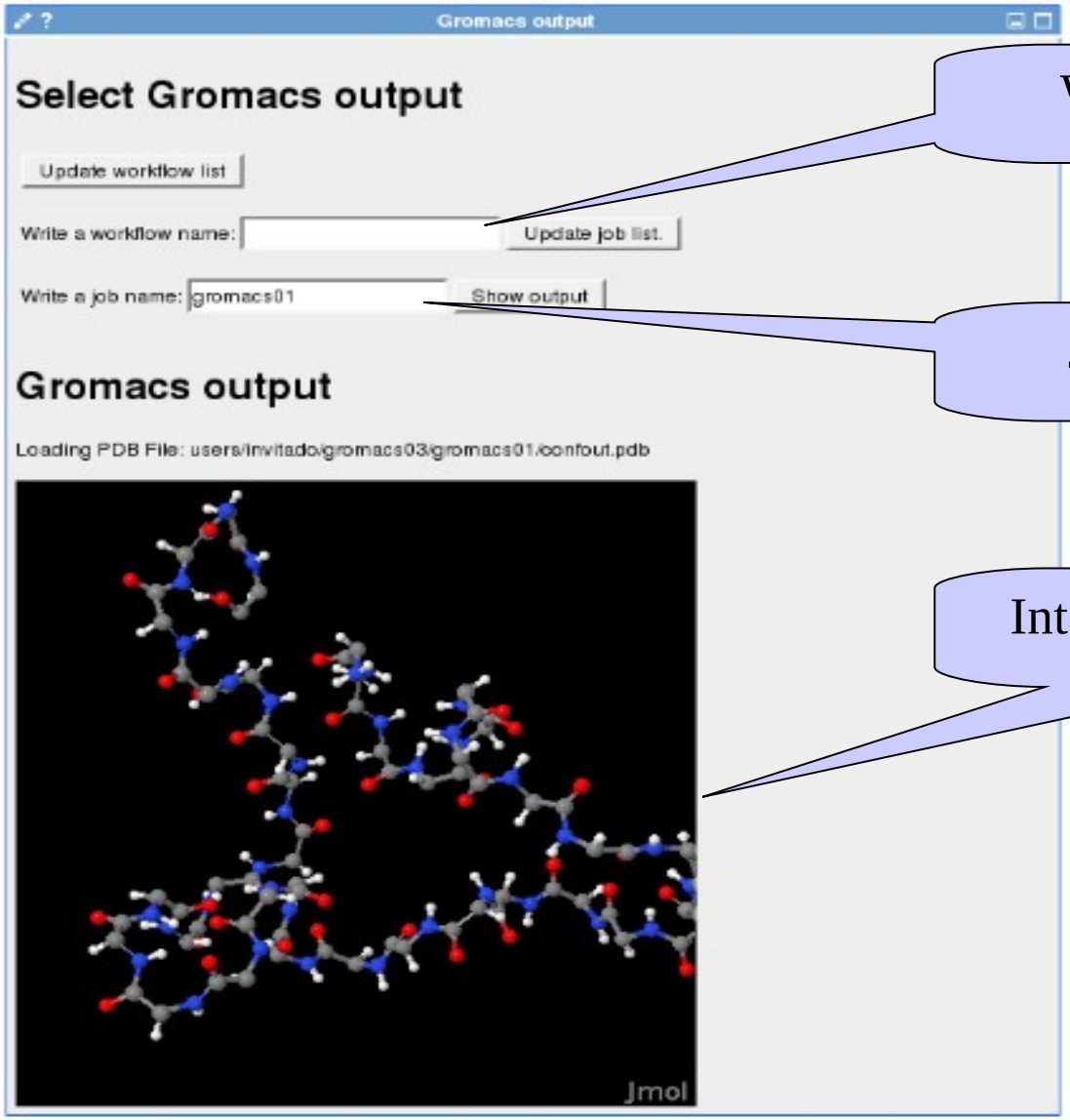
Better speed-up can be achieved increasing the number of CPU for a single execution

Portlet development: Visualization

Added value

- Prototypical visualization tool developed using Portlet
  - Components managed and displayed in a web portal
  - Based on Jmol open source java viewer
    - Contains a Web browser applet
- Interactive 3D rendering of the GROMACS output
  - No need to download the output files

Portlet development: Visualization



Workflow selection

Job name selection

Interactive 3D rendering

A new distribution scheme*

Motivations

- Computational Chemistry applications may have different requirements
  - Use of different computational resources
- Different resources usually accessed using multiple access points
  - User needs to use different architectures in a easy and transparent way

* In collaboration with the Supercomputing Center of Galitia (CESGA)

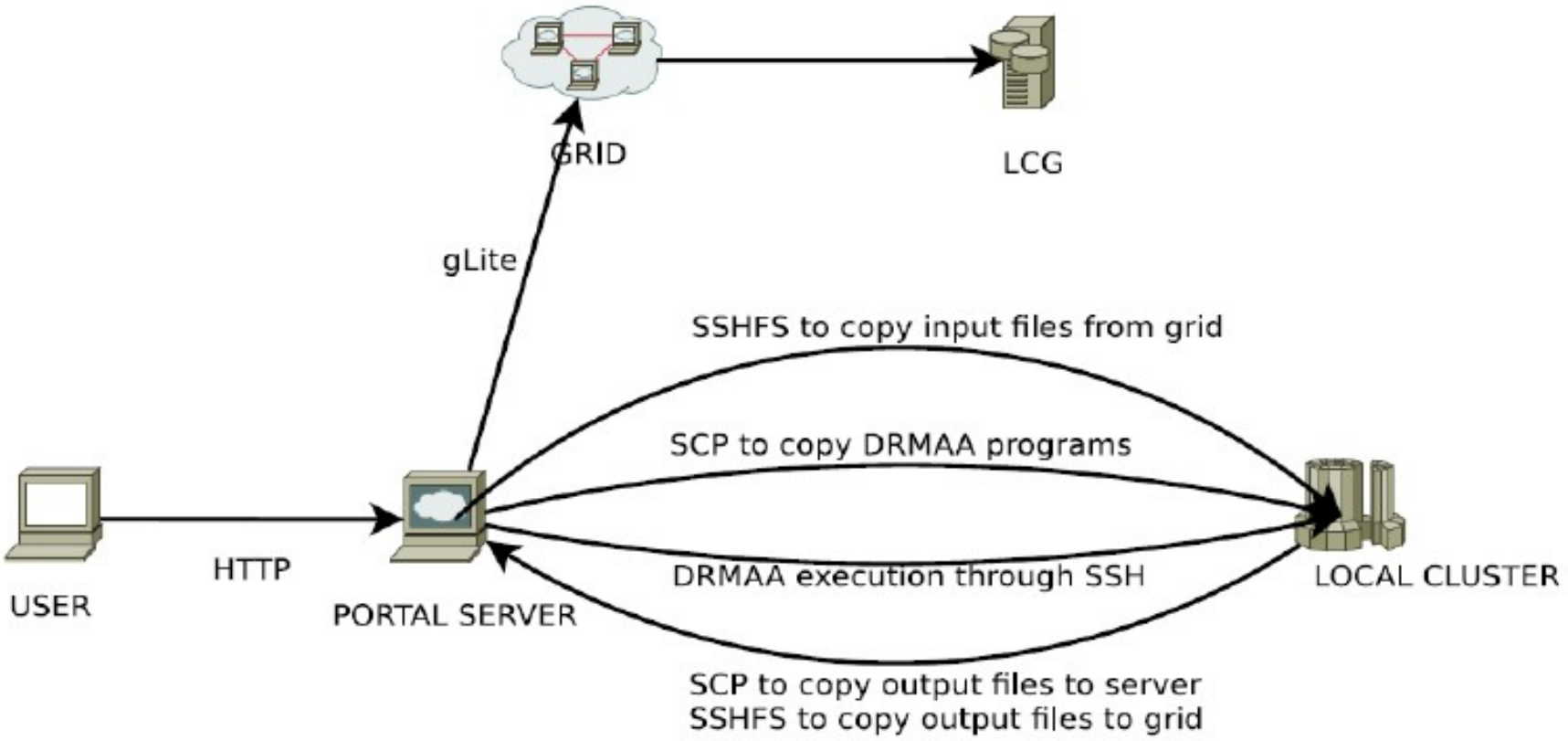
A new distribution scheme: Using multiple platforms

- 2 new workflow has been developed and tested
  - Made up of a chain of 3 jobs
    - Based on the available GROMACS tutorials
    - Output files of the previous used as input for the next
  - Different computational platforms involved
    - CESGA Server
    - Local cluster
    - Grid platform

A new distribution scheme: Portal extra functionalities

- Extra functionalities added to P-GRADE
  - File Management and Communication
    - SSH protocol to connect different platforms
  - Job Submission and Monitoring
    - DRMA (Distributed Resources Management Application) API to use local resources
  - External Services
    - Called via web

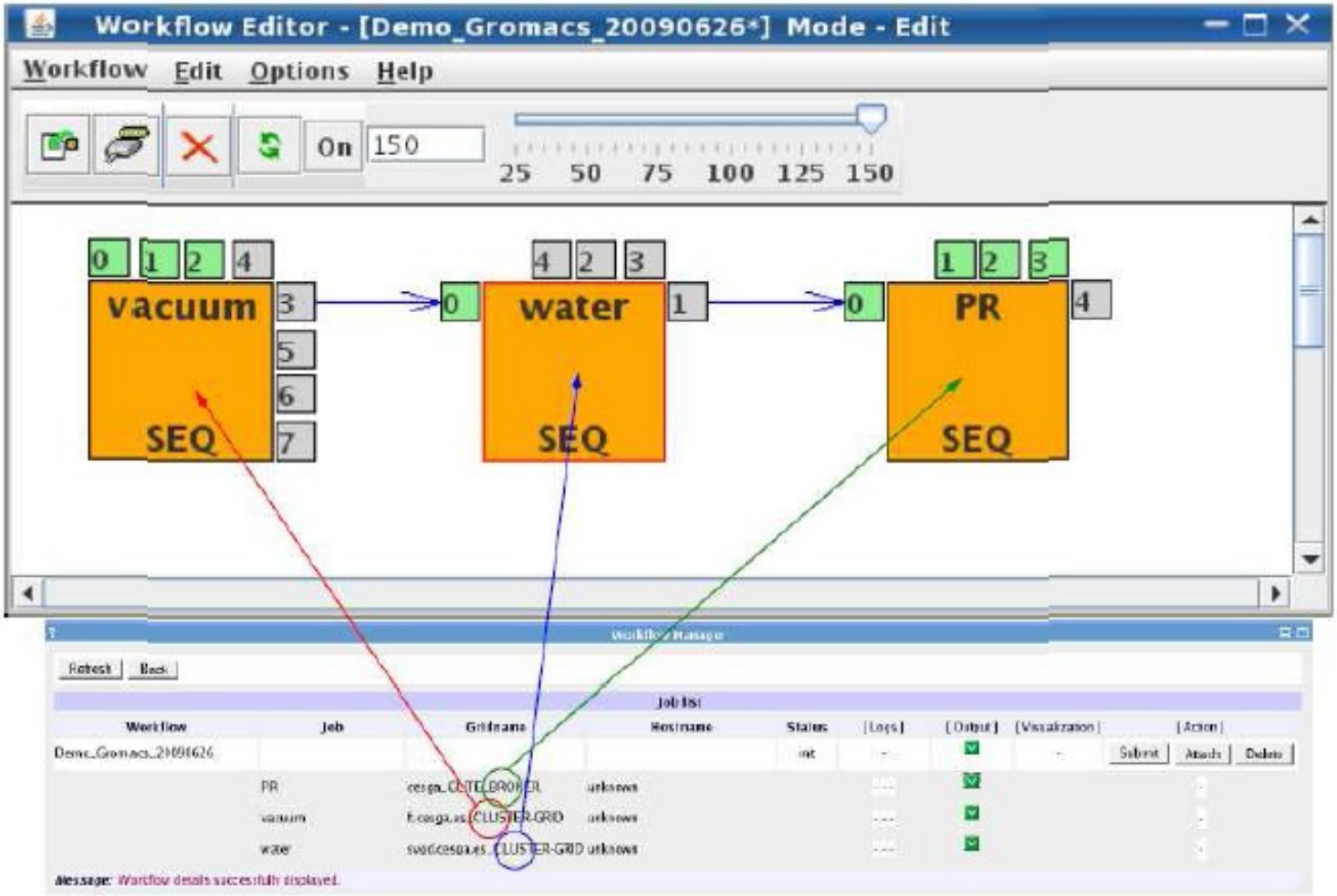
A new distribution scheme: Portal extra functionalities



A new distribution scheme: first workflow articulation

- The workflow representation running GROMACS
  - Vacuum
    - Running on a Local Cluster (LC1)
    - Perform energy min. of the proposed structure
  - Water
    - Running on a Local Cluster (LC2)
    - Perform energy min. of the solvated system
  - Relaxation
    - Running on the EGEE Grid
    - Perform solvent and Hydrogen atom positions relaxation

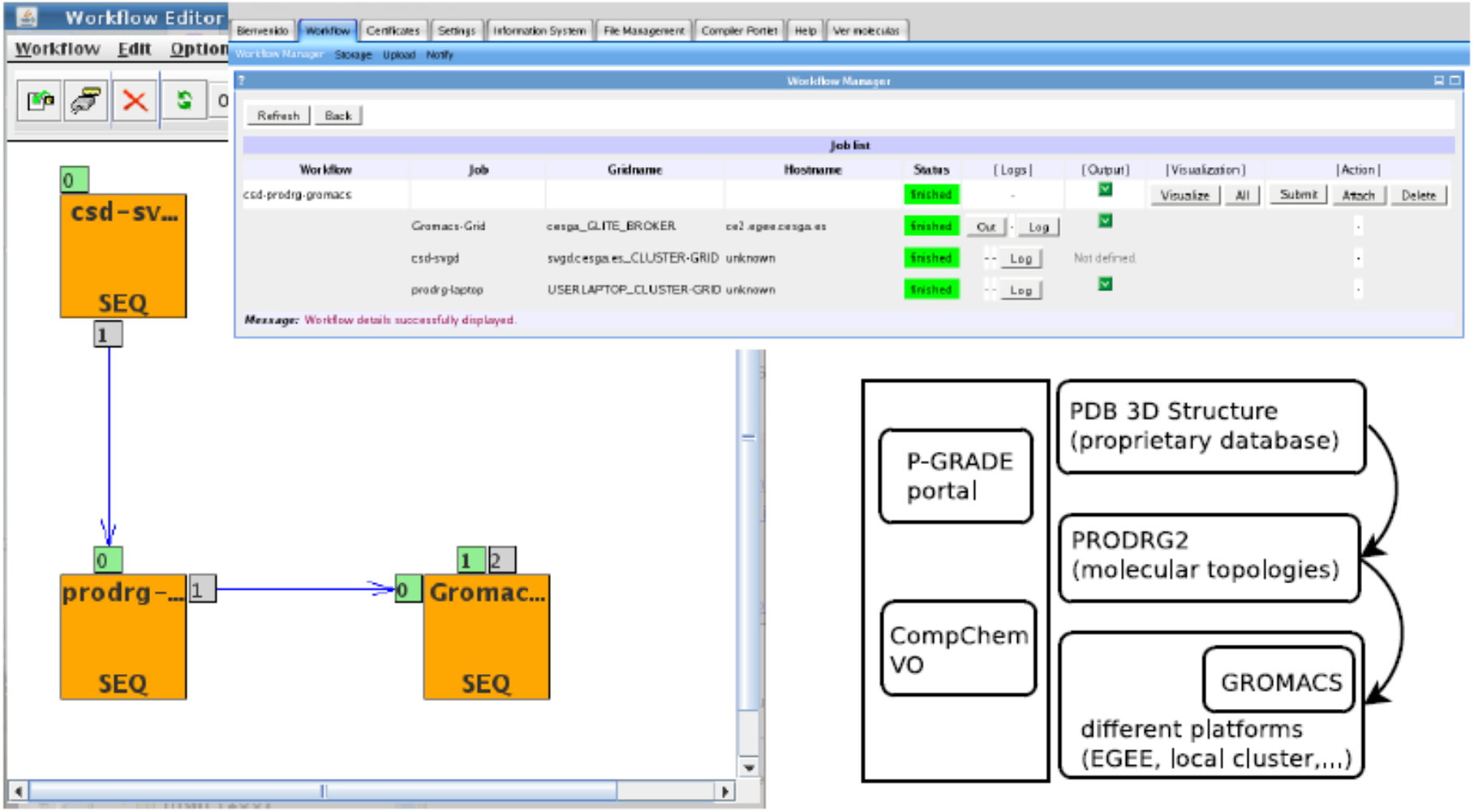
A new distribution scheme: first workflow articulation



A new distribution scheme: second workflow articulation

- The workflow representation: tools and applications used
  - CSD (Cambridge Structural Database)
    - Accessed through CESGA Server
    - Query performed to looking for PDB files
  - PRODRG
    - Running on a Local Cluster
    - Provide GROMACS ITP files (molecular topologies) from each query
  - GROMACS
    - Running on the EGEE Grid
    - Perform MD calculations from each ITP file

A new distribution scheme: second workflow articulation



Future work

- Support for data conversion
  - Open Babel
  - Q5 COST
- Scheduling based on “on-demand” resources
  - Scheduler able to submit job on the most adequate computational resource
- Runtime visualization
  - Develop a grid-tool able to retrieve outputs in real time from a running application

Summing up

- 3 workflows developed for GROMACS MD package have been discussed
  - "parameter study" using Grid resources
  - "chain" jobs using different computational platforms
- Web oriented visualization interfaces have been deployed and presented
  - Support of interactive 3D rendering
- The GROMACS workflow prototypes can be improved and exported to other applications

DL_POLY Code

- The other suite of computer programs considered for implementation on the EGEE Grid for a cooperative use is DL_POLY software package

  - Developed by Daresbury Laboratory

  - Developed for MD calculations

  -Designed to facilitate MD simulations of complex systems

  - Native parallel

- Prompted the investigation of the distribution strategies of DL_POLY

On the porting of DL_POLY: a preliminary study

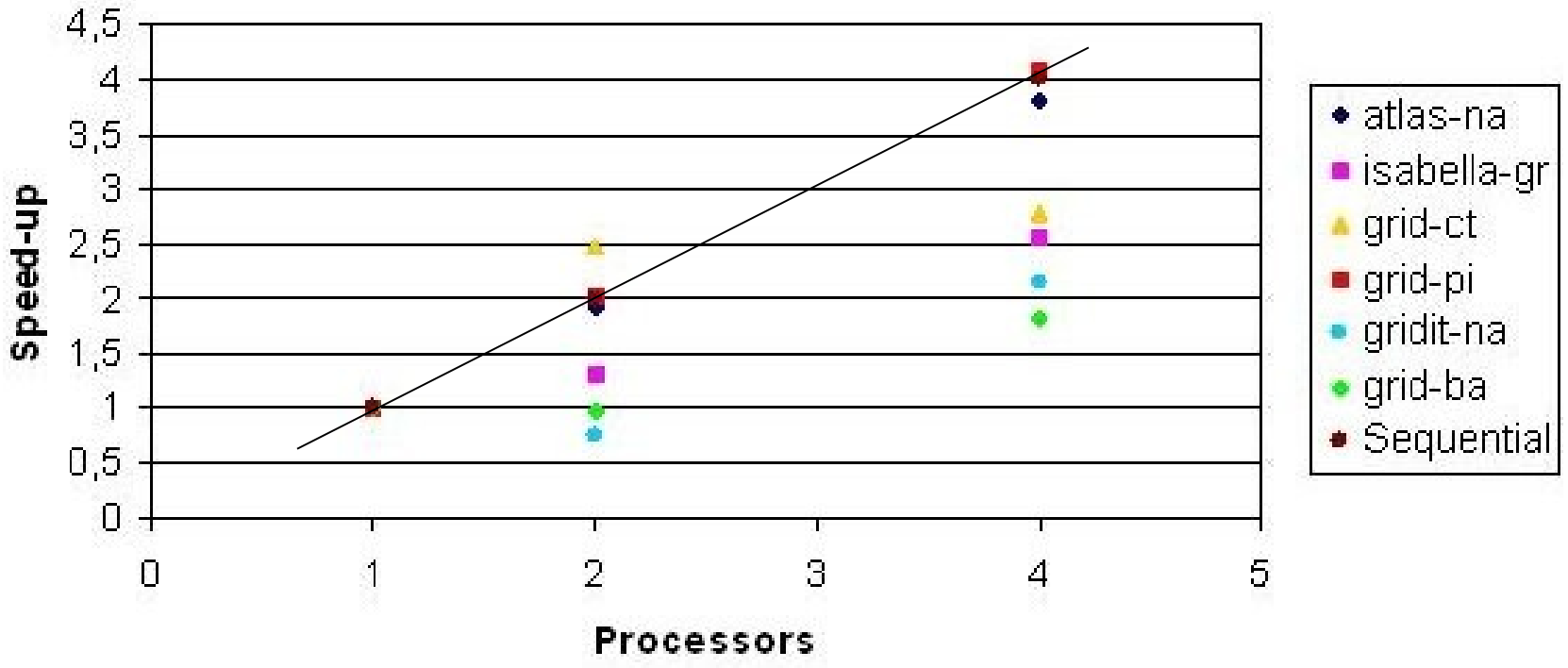
Performance of the used platform

- Preliminary indications were obtained running a test case on six different EGEE-Grid clusters

- The calculations ran sequentially on one node and in parallel on 2 and 4 nodes

- Related speed up was evaluated

Performances



-Parallel performance of each cluster not necessary depend only from the time sharing regime adopted by them

-Detailed evaluations of the parallel performance of the clusters were obtained restricting parallel calculations to two nodes

Statistical analysis

Job status	Number	%
Success	37	74
Aborted	13	26

Aborted reasons	Number	%
Communic. Error	8	62
DL POLY error	3	23
Scheduler error	2	15

-Statistical analysis were calculated on 50 parallel jobs

-More than 20% of the abortion is due to internal DL_POLY reasons

An alternative distribution scheme

-Communication or scheduling failures remedied re-running the related computation

-DL_POLY internal failures imply that starting conditions are inappropriate

-Different starting conditions can be obtained randomly altering the configurations having the same energy

  -randomly inverting the sign of a fixed number of momentum components of the particles of the system

-As a result a portion of the phase space sampled changes without altering the total energy of the particles

The distribution algorithm

To this end the core script of DL_POLY software package was modified in order to implement a coarse grained task farm model

- a Master process distribute the whole program to all the available nodes
- typical SPMD (Single Program Multiple Data) technique
- simple to implement and suited for the present structure of the EGEE Grid

The distribution algorithm

The algorithm was tested on a DL_POLY demo application on 8 nodes biproc. PentiumIV cluster

-the program scales very well; the trend is almost linear up to 8 processors

-the modifications of the momenta may lead to unstable configurations; the failure rate is however scarcely sensitive to the number of atoms randomly chosen

% **of changes**	300 steps		1000 steps	
	Term.	**Failed**	**Term.**	**Failed**
25 %	10	0	10	0
50 %	10	0	9	1
75 %	9	1	8	2
100 %	9	1	8	2

Future work

- Develop related workflow(s) for DL_POLY starting from those developed for GROMACS
  - Workflow for long time simulations
    - Short CPU time on the Grid (24-48 hours)
  - Workflow for merging different computational platforms
    - LC to perform starting conditions simulations
    - Grid to perform long time simulations
- Develop Grid (and Web) enabled visualization tools for DL_POLY