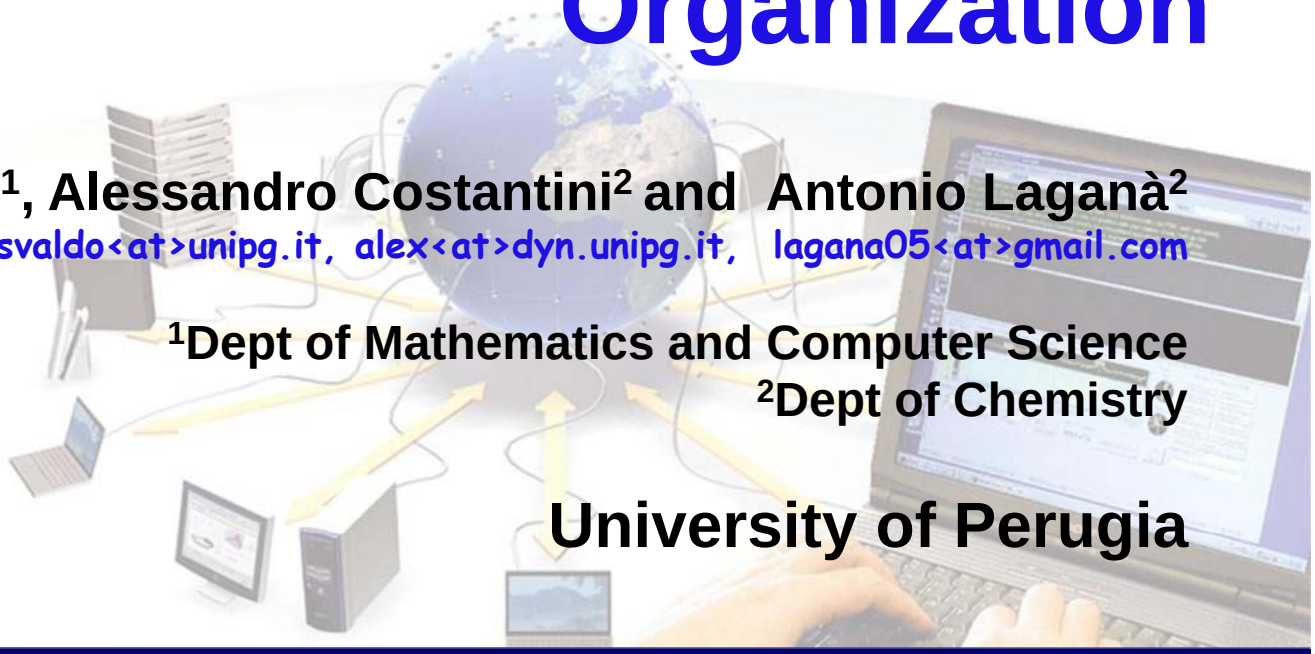
Enabling Grids for E-science in Europe

www.eu-egee.org

EGEE 07 Conference
Budapest (HU)
October 1st – 5th 2007

The CompChem Virtual Organization



Osvaldo Gervasi[1], Alessandro Costantini[2] and Antonio Laganà[2]
osvaldo<at>unipg.it, alex<at>dyn.unipg.it, lagana05<at>gmail.com

[1]Dept of Mathematics and Computer Science
[2]Dept of Chemistry

University of Perugia

EGEE is a project funded by the European Union under contract IST-2003-508833

CompChem VO

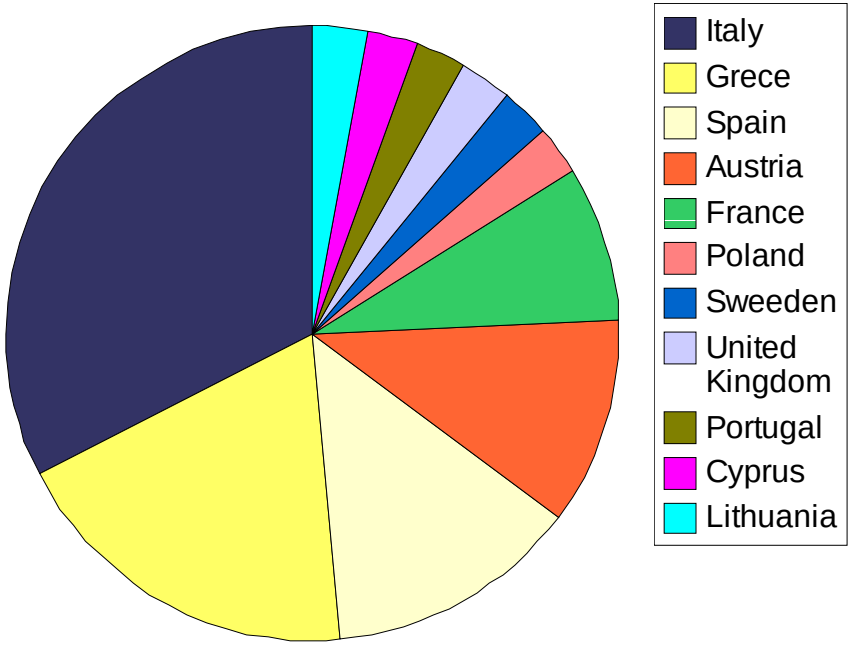
- CompChem VO is running in EGEE production Grid since 2004 to support Computational Chemistry applications
- We contribute to the EGEE production Grid with 2 clusters:
  - ce.grid.unipg.it (LCG 3.0) 12 nodes (biproc) PIII
  - cex.grid.unipg.it (Glite 3.0) 8 nodes (biproc) Xeon
- To simplify the new users startup we made available to all CompChem users a UI to access the Grid: ui.grid.unipg.it

The CompChem users

Enabling Grids for E-science in Europe

Country	# of users	Organization
Italy	12	Univ. Perugia, CNR-IMIP, ENEA
Grece	7	FORTH Crete, Univ. of Crete
Spain	5	Univ. Basque Country, Univ. Bar
Austria	4	Univ. Vienna, Univ. Innsbruk
France	3	CNRS
Poland	1	Cyfronet
Sweden	1	KTH Royal Institute of Technolog
United Kingdom	1	Imperial College London
Portugal	1	IRICUP
Cyprus	1	Univ. Cyprus
Lithuania	1	Univ Vilnius

Users per country



37 users

11 countries

COST in Chemistry D37 action: *GridChem*

Enabling Grids for E-science in Europe

- Most of the CompChem users started the collaboration in the COST in Chemistry D23 Action, called *Metachem*
- COST is one of the longest-running instruments supporting co-operation among scientists and researchers across Europe with 35 member countries.
- In the COST in Chemistry program has been recently activated the Action D37, called *GridChem*
  - devoted to deploy the most important computational chemistry codes in the Grid and the existing e-infrastructures
- EGEE and CompChem VO will support the activities of the Action and new users are expected to join the VO.

CompChem Applications

Enabling Grids for E-science in Europe

- **RWAVEP** University of Perugia (Italy): provides a quantum time-dependent treatment for the evaluation of the reactive properties of molecular systems.
- **DL_POLY** CCLRC (UK): performs the molecular dynamics simulation of complex systems. It has been designed to run in parallel and is a "de-facto" standard in the computational chemistry and computational biology communities.
- **Columbus** University of Vienna (Austria): is a collection of programs for high-level *ab initio* molecular electronic structure calculations (primarily for extended multi-reference (MR) calculations on electronic ground and excited states of atoms and molecules).
- **NewtonX** University of Vienna (Austria): is a general-purpose program package for excited-state molecular dynamics, including non-adiabatic methods (Tully's surface hopping).
- **Venus** CDSSIM - Texas Tech University (USA) Venus calculates the cross sections and rate coefficients for elementary chemical reactions by simulating the collisions between atoms and molecules whose initial conditions are sampled using a Monte Carlo scheme.
- **ABCtraj** University of Perugia (Italy) ABCtraj calculates the observables of the atom-diatom reactions in gas phase. The events are generated using Monte Carlo techniques. The program is linked to a molecular virtual reality environment that shows in virtual monitors the outcomes of the simulation.

…… and several user-developed codes ……

Columbus

Enabling Grids for E-science in Europe

- Quantumchemical Ab-Initio Software for high-level Calculations
- Very demanding on resources
- Set of (fortran-)executables
- Heavy usage of BLAS and LAPACK
- Parallel version available
- http://www.univie.ac.at/columbus

NewtonX

Enabling Grids for E-science in Europe

- Is an Ab-Initio Molecular dynamics package
- Calls external program for energy-/gradient-calculation (Columbus, Turbomole, ...) in each timestep
  - It may make hundreds of call
  - Must be installed on the same node of the external program
- Calculates **independent** trajectories (100-1000)
  - Distribution to independent nodes (no inter-node communications)

Main characteristics of CompChem applications

Enabling Grids for E-science in Europe

- Both CPU-bound and I/O intensive jobs are present
- Parallel jobs: some programs have been structured to run in parallel (MPICH APIs).
- Interactive jobs: GEMS needs in/outbound connectivity (portnumbers range: 20000-25000)
- We are implementing Web Services
  - to prepare the input data and execute the programs on the Grid
  - to enable the users to interpret the information obtained from the simulations running on the Grid using Web3D and visualization tools

How to join CompChem VO

- Each partner may be involved at different levels (ranging from the case where the user uses the available software for its own purposes to the case where a given laboratory promotes projects, make available new programs, etc)
- Reference URL: **http://compchem.unipg.it**
- The user must have a valid X.509 digital certificate and asks to join the VO through the URL: https://voms.cnaf.infn.it:8443/voms/compchem/webui/request/user/create
- **After acceptance she/he access the Grid through a UI (UipnP, ui.grid.unipg.it, etc)**

http://compchem.unipg.it

Enabling Grids for E-science in Europe

EGEE Home | Technical Pages Home | Calendar | Agenda | Glossary

Enabling Grids for E-sciencE

CompChem Virtual Organization

Information Society Technologies

CompChem Info Sheet | CompChem FAQ | Report a problem on CompChem to the EGEE Suppor Team (GGUS)

The CompChem Virtual Organization has been established in the EGEE Production Grid environment to support the Computational Chemistry applications. The collaboration has been defined in a **Memorandum of Understanding**. The text of the MoU is available here.

To be enabled to EGEE Grid you need a valid X.509 Certificate. If do not have it, you must request it to the Certification Authority that serves your Institution or your Country. If in trouble check the LCG User registration guide to find the most suitable Certification Authority for your needs.

To join CompChem VO, please connect with a web browser equipped with your personal X.509 certificate to the following URL: https://voms.cnaf.infn.it:8443/voms/compchem/webui/request/user/create.

A short description of the operations to be performed to start using CompChem VO is available here.

If you are an administrator of an EGEE Grid production site, please follow the following instructions to support the CompChem VO applications in your site.

The CompChem VO Manager

VO Software management

- To use the Grid efficiently we are pushing the researchers responsible for the named packages to use the available instruments to install the software in the EGEE sites.
  - Local installation on each site of interest supporting the VO (not all sites have the right characteristics for a given package)
  - The compilation of the program on the target architecture improves the performances
  - The dependencies of the program can be managed efficiently
  - The process is controlled by few VO members (experts in the named programs)
- The procedure has been implemented for the Columbus program and will be extended to the more relevant programs available in the VO

VO software management

Enabling Grids for E-science in Europe

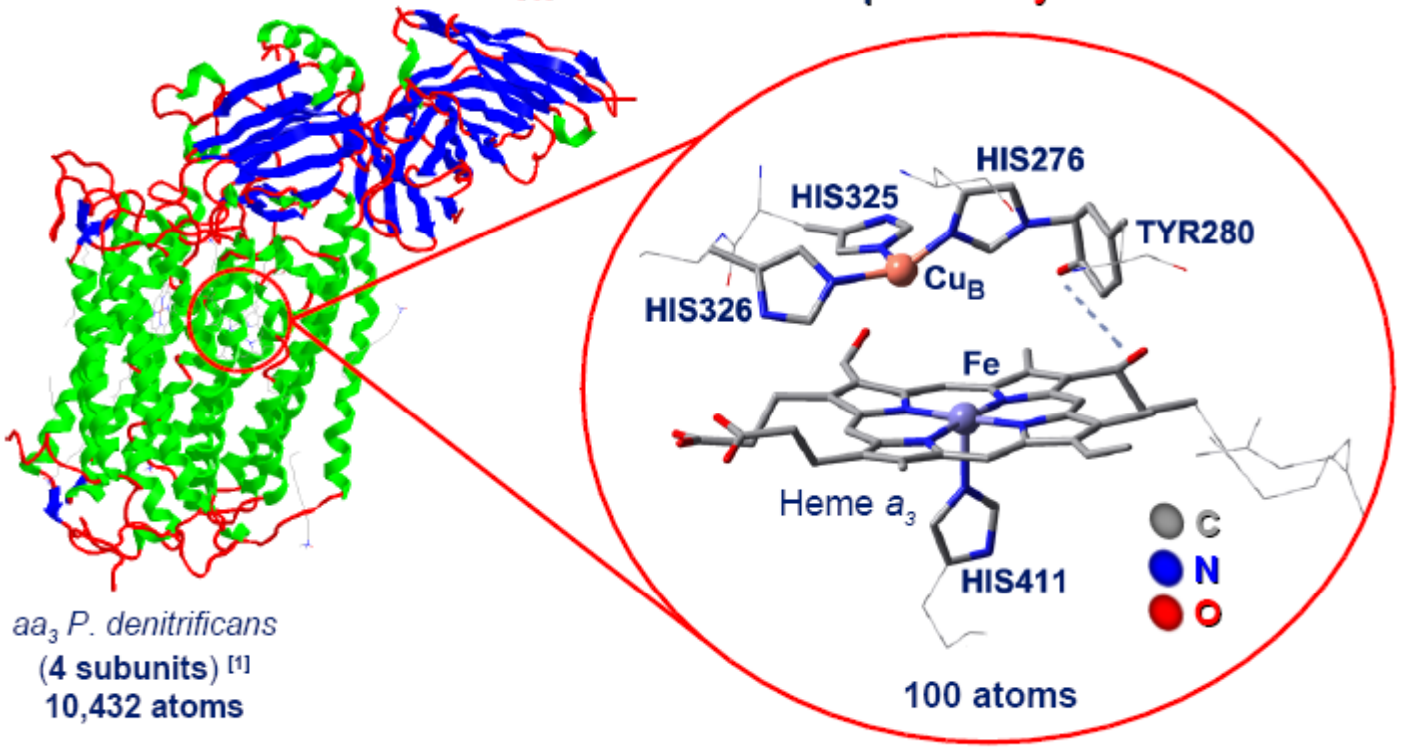
- The users responsible for the VO packages are part of the SoftwareManager group
- They can grant the permission to write on a system area reserved to the VO of each site. No other special permissions are required
- The manager stores the source on the Storage Element, aentually making replicas of the file on several SE.
- She/he launches jobs on the selected sites that will install and verify the package.
- She/he modifies the VO tags to announce the availability of a given program on the selected sites.

One of the next computational campaigns

Enabling Grids for E-science in Europe

Cytochrome c Oxidase – *a big system to study !*

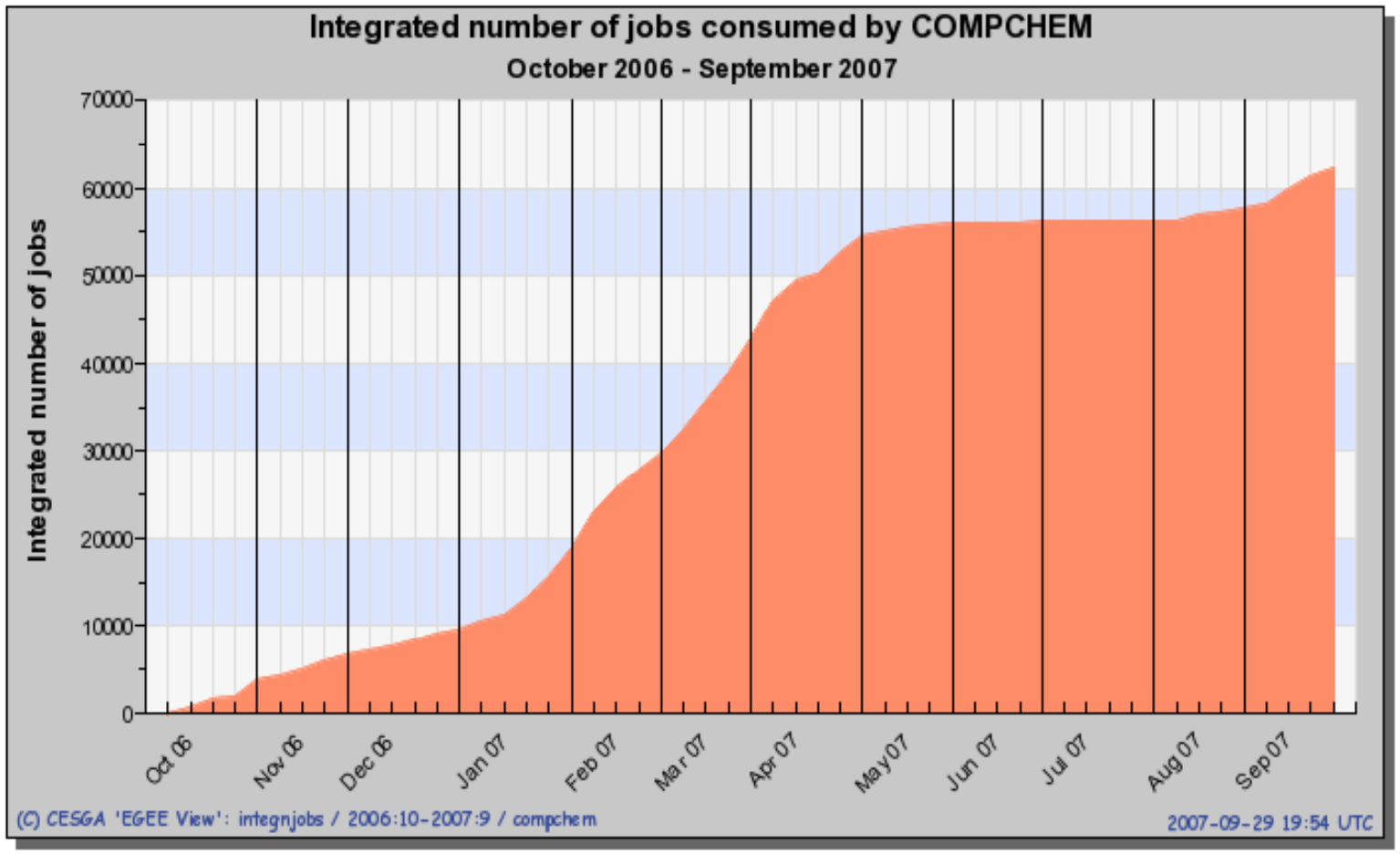
Where does molecular oxygen reduction take place in Cyt. *c* Oxidase ?



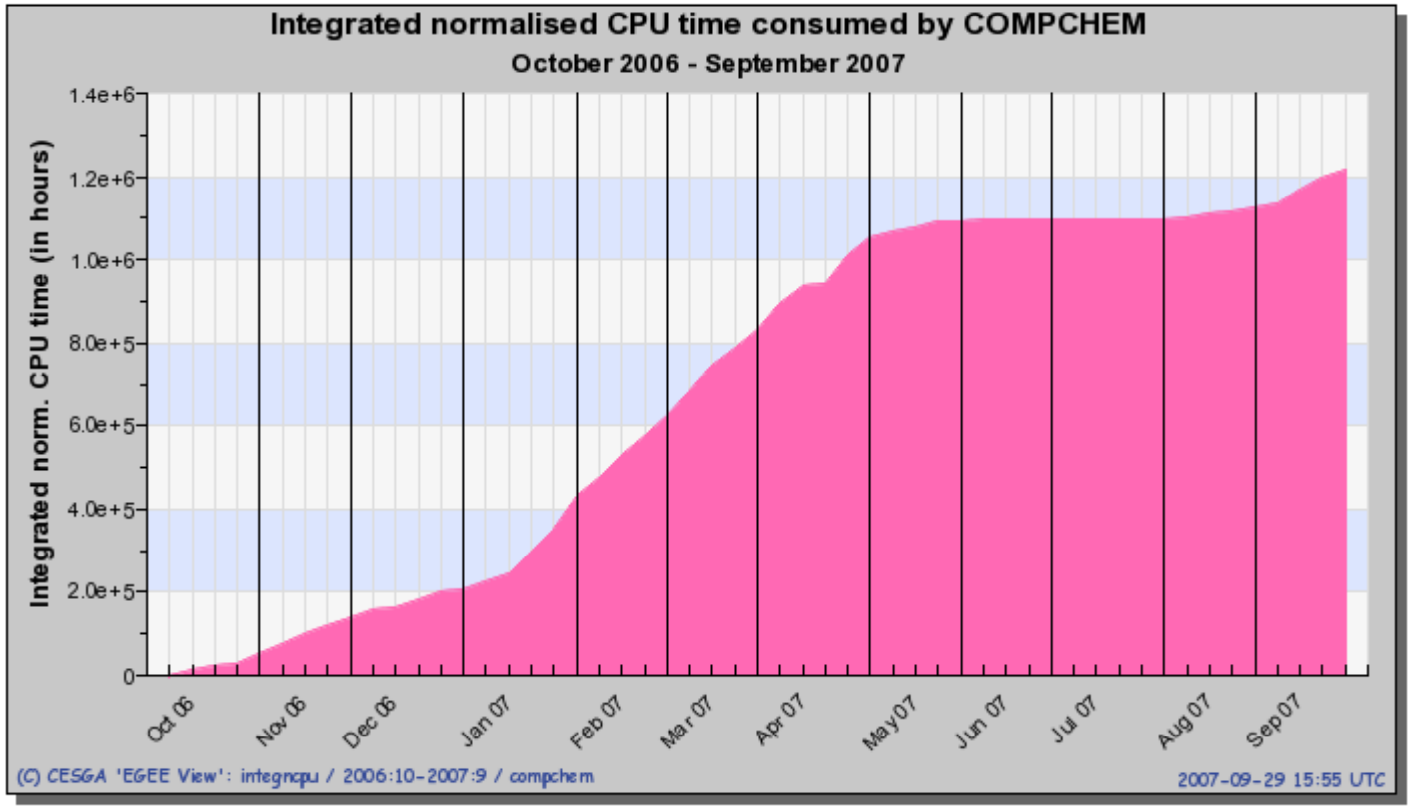
aa_3 P. denitrificans
(**4 subunits**) [1]
10,432 atoms

[1] Ostermeier, C.; Harrenga, A.; Ermler, U.; Michel, H. Structure at 2.7 A resolution of the Paracoccus denitrificans two-subunit cytochrome *c* oxidase complexed with an antibody FV fragment. *Proc.Natl.Acad.Sci.USA* 1997, *94*, 10547-10553

Some statistics: integrated N of jobs



Hours of CPU per week



Distribution of CPU time by disciplines and dates

Enabling Grids for E-science in Europe

Normalised CPU time [units 1K.SI2K.Hours] by DISCIPLINE and DATE														
DISCIPLINE	Oct 06	Nov 06	Dec 06	Jan 07	Feb 07	Mar 07	Apr 07	May 07	Jun 07	Jul 07	Aug 07	Sep 07	Total	%
Astrophysics	923	15,528	514	4,696	52,041	56,677	73,980	76,958	52,506	120,387	48,828	35,421	538,459	0.49%
Biomed. & Bioinfor.	1,193,476	1,820,575	1,992,461	1,232,751	99,696	54,132	102,085	598,448	672,440	583,880	766,283	776,015	9,892,242	9.01%
Comput. chemistry	41,665	96,866	72,375	197,765	227,291	295,718	324,318	216,982	160,201	425,381	186,781	323,629	2,568,972	2.34%
Earth sciences	1,378	10,349	4,403	4,829	5,916	50,165	2,599	993	3,888	3,374	15,975	28,206	132,075	0.12%
Finance	0	0	0	0	0	52	2,018	1,433	10,763	0	0	0	14,266	0.01%
Fusion	3,694	352	21	20	616	1,232	1,608	38,009	78,793	40,076	22,678	145,269	332,368	0.30%
Geophysics	5	6	0	1	6	9	0	2	0	1	0	0	30	0.00%
High-energy physics	4,900,059	4,950,452	5,792,277	7,891,885	9,330,004	7,013,390	8,605,886	6,385,734	8,538,528	10,579,619	9,055,485	7,869,544	90,912,863	82.83%
Infrastructure	2,354	7,053	10,089	10,733	20,463	22,575	23,563	10,795	19,153	23,553	17,944	20,480	188,755	0.17%
Others Disciplines	176,751	252,792	193,273	201,270	309,101	288,260	315,151	389,761	448,261	542,042	698,535	513,666	4,328,863	3.94%
Unknown Discipline	11,340	21,379	16,759	5,224	29,296	34,235	19,395	67,892	157,385	183,797	168,728	138,922	854,352	0.78%
Total	6,331,645	7,175,352	8,082,172	9,549,174	10,074,430	7,816,445	9,470,603	7,787,007	10,141,918	12,502,110	10,981,237	9,851,152	109,763,245	
Percentage	5.77%	6.54%	7.36%	8.70%	9.18%	7.12%	8.63%	7.09%	9.24%	11.39%	10.00%	8.97%		

Click here for a csv dump of this table

Chart of the cumulative CPU vs DISCIPLINE and DATE

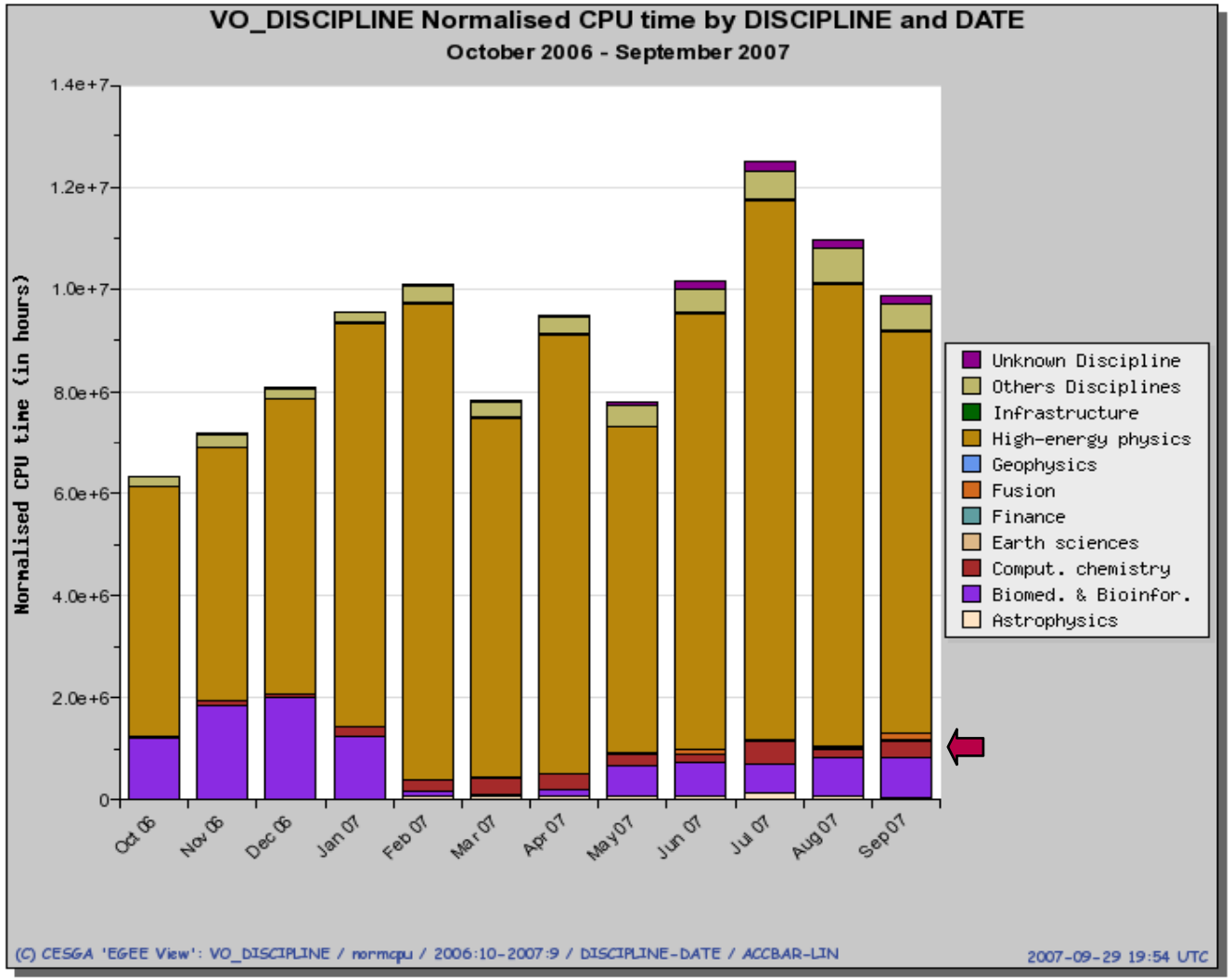
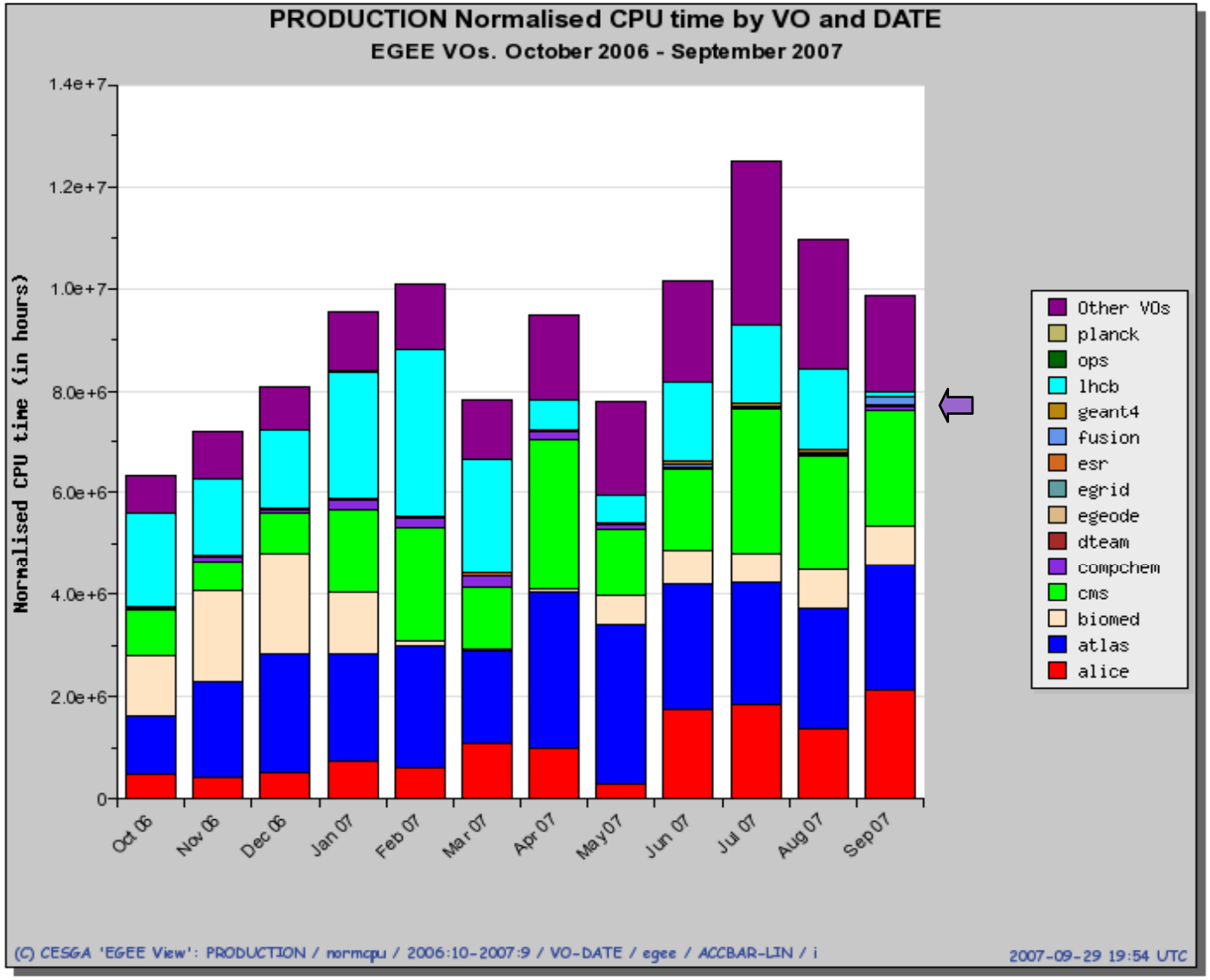


Chart showing the Cumulative Normalised CPU time grouped by DISCIPLINE and DATE.

Chart of the cumulative CPU vs VO and DATE



Conclusions

Enabling Grids for E-science in Europe

- The CompChem VO users have proved to be very active and to achieve important goals using the EGEE infrastructure.
- UNIPG will become member of the EGEE III
- CompChem and GAUSSIAN VOs will cooperate for developing new tools for the users
- New users and programs are expected in the next months.