A Generic Grid Environment for Central Europe
A Study of Pseudorotaxanes Behavior

Jan Kmuníček, Daniel Kouřil, Petr Kulhánek, and Martin Petřek

CESNET z. s. p. o., Zikova 4, 160 00 Praha 6, Czech Republic
{kmunicek,kouril}@ics.muni.cz, {kulhanek,petrek}@chemi.muni.cz

Abstract. Grid computing is currently seen as one of the key ingredients for fulfilling up to date science requirements. Requests for extreme computing power together with large storage capacities led to forming a research infrastructure available for everyday scientific work provided by computing/data Grid. This article describes VOCE - Virtual Organization for Central Europe - a production grid infrastructure available for researchers from the Central European region and its successful utilization for accomplishment of large scale computing challenge - detailed investigation of behavior of supramolecular complexes - pseudorotaxanes.

1 Introduction

Grids can be seen as huge distributed systems composed of organizationally independent elements that can be individual computing and storage facilities, middleware services and information services mutually interconnected by a computer network. The potential of Grid systems is immense and many large research projects benefit from their use. However, current Grids are very complex systems with long and slow learning curve causing they are only used by large user communities with enough resources for training their researchers. Also deployment and mainly everyday operation of core services that provide access to the Grid resources is a difficult task that requires skilled administrators who have to be well trained first. Therefore, the potential of current Grids is not fully utilized yet and especially small research groups, which do not have resources to build and administrate such a sophisticated infrastructure and train users are not able to even test the Grid environment and evaluate if it fits their needs.

In this paper we describe a grid infrastructure provided to all users from the Central Europe region, which is suitable for getting first experience and also for a production use. In the second part of the paper we describe a large chemical computation demonstrating the possibilities that the environment offers to their users.

2 VO for Central Europe – VOCE

Users in current Grids are organized within *virtual organizations* (VO) that reflect real hierarchy of users coupled to solve joint problems. Each VO provides

its users with resources necessary to solve their research problems. There are many mechanisms for accessing such resources varying according to the Grid environment used to configure the VO. One of the largest activities aiming at building and operation of a production grid infrastructure is EU project Enabling Grids for E-SciencE (EGEE)[1]. The operation activities of the project are organized according to the geographical principle segmenting the participants into several *federations* that operate their locale resources and provide support for users from their region.

In order to bring the Grid facilities closely to the end users the Central Europe (CE) federation consisting of institutions from Austria, Croatia, Czech Republic, Hungary, Poland, Slovakia, and Slovenia established and operates a VO available to all users from the Central European region – the Virtual Organization for Central Europe environment[2] (VOCE). This infrastructure is open for every user from the region and is meant as a general *catch-all* virtual organization serving users not covered by other application specific environment.

VOCE currently consists of computational resources and storage capacities provided by the Central European resource owners. Unlike majority of other virtual organizations, VOCE tends to be a generic virtual organization providing an application neutral environment. Primary purpose of VOCE is to provide an easy access for users who already use computing techniques to solve their problems but are not familiar with utilization of Grids. According to our experience, individual researchers and/or small research groups cannot often penetrate through the complex environment provided by the Grids to start their using. Such groups also usually do not have access to any established Grid infrastructure and are not willing or able to build a separate environment just to test it. Targeting at these user communities, VOCE tries to motivate new application communities to use the immense power provided by current Grids.

VOCE currently provides a complete grid infrastructure running all necessary grid services to serve scientific end users solving their research projects and problems. The core VOCE infrastructure is entirely independent of any services operated outside CE and thus can be configured very flexibly on demand if any special needs arise. End resources (i.e. computing and storage elements) are provided by the CE institutions that are distributed across all the region and are usually shared with other VOs supported by the sites. Currently VOCE have approximately one thousand of CPUs and thirteen TB of disk capacity available for their users.

2.1 Users' Utilization of VOCE Resources

VOCE is open for all users from the region who are interested in the Grid technologies. The only requirement for the users is a possession of a digital certificate issued by a certification authority approved by the International Grid Trust Federation (IGTF). This requirement makes the VOCE different from other similar

[1] `http://www.eu-egee.org/`
[2] `http://egee.cesnet.cz/en/voce/`

projects, which allow users to use arbitrarily certificates, often issued using very weak identity vetting procedures. Such users cannot be identified properly and it is hard (or even impossible) to trace them if they cause an incident. The VOCE users, on the other hand, can be easily contacted by the resource owners if they harmed the resource. Last but not least such a precaution helps educate users in proper credential management.

Applicants for VOCE membership are not required to have any prior experience with Grids. For such inexperienced users VOCE provides an ideal platform for getting experiences with Grid as the environment provides real services and not limited facilities. Moreover the VOCE partners also organize training events where users can learn how to use the Grid. After getting familiarized with the environment the users are ready to use VOCE for their own research. They gradually move from examples to real work and prepare their applications for the Grid environment. When the applications are tested and ready, the users use the VOCE infrastructure for everyday routine production. The VOCE resources provide efficient background for complex problems solving and VOCE experts are available to assist and help users. When the users gain enough knowledge with the Grid their needs may grow. If the features provided by VOCE do not fulfill their requirements, such users can establish a new VO providing dedicated resources and/or services for their purpose. In this case the VOCE experts can provide necessary expertise in building and operating the VO.

Main feature that distinguishes VOCE from other existing VOs is its application neutrality. VOCE is not bound to any particular application or application domain and therefore it is designed to allow users from arbitrary application area to test and use the environment. The main advantage of the application neutrality feature is the fact that VOCE becomes a place where different application requirements meet and expectations are to be fulfilled. This is notably important especially for complete grid novices who has no experience with work within a grid infrastructure. For such end users VOCE provides a place where a predefined set of ported applications from distinct application domains is accessible immediately after getting a VOCE account. Presently, VOCE provides a broad portfolio of available application programs from different application domains as are molecular mechanics and dynamics (Gromacs, *etc.*), quantum mechanics and dynamics (Abinit, USPP, *etc.*), computational chemistry visualizations (Gnuplot, Grace, PovRay, *etc.*) and physics (MPB, Octave). The VOCE application generality promises a smooth porting of programs from further application areas as astrophysics, bioinformatics, technical and material simulations and similar ones based on users real requests and wishes in an efficient way.

2.2 Application Tools

VOCE as a standard EGEE VO provides users with native gLite command-line tools to access the Grid environment, which may be a bit encouraging for complete grid newcomers. Therefore VOCE offers alternative ways how to perform standard operations related to day-to-day research work. The VOCE users can utilize a high-level command line tools as well as graphical user interfaces.

The GUIs are provided by portals that can be accessed using common WWW browsers. Portal gateways supported currently are P-GRADE [2] and GENIUS [3]. Portal P-GRADE is a workflow-oriented Grid portal that enables the creation, execution and monitoring of workflows in grid environments through high-level, graphical Web interfaces. Workflows developed in the P-GRADE portal are portable between different Grids without any re-learning of the new Grid system or re-engineering the application. Portal GENIUS was originally devolved to support training and demonstration Grid activities and is often used as a standard tool during training courses that aim at Grids utilization and allow smooth transition in production Grid usage.

In addition to the GUI methods, the VOCE environment offers unique Charon Extension Layer (CEL) system [1] that is a universal framework creating a layer upon the basic Grid middleware environment and making an access to the complex Grid infrastructure much easier compared to native middleware. The CEL system unifies the variability in Grid middleware (PBS, LCG/gLite, etc.) allowing transparent access to distinct Grids. Moreover, the CEL system offers easy access and utilization of heterogeneous Grids in a unique, easy and smoothly integrated way. The Charon Extension Layer toolkit provides a command-line oriented interface and is supposed for users that require a full control over their computational jobs. The CEL system provides a uniform and modular approach for (complex) computational jobs submission and management and forms a generic system for the use of application programs in the Grid environment independently of Grid middleware present at specific fabric infrastructure. CEL can be easily used for powerful application management enabling single/parallel execution of computational jobs without the job script modification.

3 VOCE Data Challenge

One of crucial factors showing the real usability of the VOCE environment is its ability to enhance solving of research tasks that require extreme computational resources. An example of such a large-scale application utilization can be demonstrated by accomplishments achieved by the computational chemistry community utilizing the VOCE grid environment.

3.1 Aim of the Study

One of the currently intensively studied areas of modern material design and drug discovery is chemistry of interlocked supramolecules and their mutual interactions. A pseudorotaxane – a supramolecular complex – and its behavior was thoroughly investigated during a VOCE data challenge. Pseudorotaxanes [4] are interlocked molecules in which macrocycle (the 'wheel') is threaded by a long 'axle' component. Our system consists of a molecule cucurbit[7]uril (CB7) and a 4,4'-bipyridinium derivate (see Fig. 1.). Experiments [5] show a switch-like movements along the axle in this supramolecular complex. To give a detailed insight to the switch-mechanism we calculated the free energy profile along reaction coordinate related to this movement.

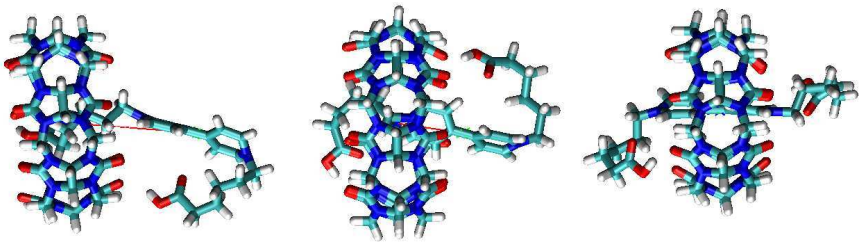


Fig. 1. Supramolecular complex of cucurbit[7]uril (CB7) and a 4,4'-bipyridinium derivate

3.2 Methodology and Requirements

We focused on two methods for the free energy estimations that are used currently — umbrella and bluemoon.

The main idea of the umbrella method is to add an artificial restraint to the system. This restraint keeps a value of a reaction coordinate in a small interval (so-called window) about a prescribed value. The whole interval is then covered by these windows. Windows should overlap partially. A molecular dynamic simulation is then performed in each window holding the reaction coordinate in the window. A histogram of reaction coordinate values for the window range is obtained from the each simulation. These histograms are consequently combined into a single one by the weighted histogram analysis method (WHAM) [6]. Then, relative frequencies are estimated from it and used to calculate free energy in whole range.

In the bluemoon method, the whole interval is covered by a satisfactory amount of points, not necessarily equidistantly distributed. A constraint molecular dynamic simulation is calculated for each point with the reaction coordinate being fixed at the point value. So-called mean force is followed during the simulation. Mean force denotes the force that has to be applied on the system in each time-step to hold the constraint fixed. The mean-value of this force quadrate with a derivation of the free energy with respect to the value of the reaction coordinate. The mean-values of forces are merged together from all windows to get the dependence of the derivative in the whole range. A numerical integration is then used and a resulting free energy profile is obtained. The key point of the method is constraint keeping and mean force estimating. This is solved by an iterative algorithm that adjusts system coordinates to reach the state where the difference between the real and desired values of reaction coordinate vanishes. The Lagrange multiplier that is used to get mean force estimate is retrieved from that algorithm as well as new atoms coordinates with fulfilled constraint.

The evaluation of the free energy is a computationally demanding task requiring extensive computational resources due to necessity to properly sample large phase-space. Both the umbrella and bluemoon methods as well as other that are

used nowadays (such as the adaptive biassing force method) require a calculation of a huge amount of middle-length molecular dynamic simulations which can run independently in parallel. Therefore, these types of computational tasks are very well suited to exploit large grid environments like VOCE. The complete solving of the described research problem comprises approximately hundreds thousands of CPU hours at 1.6 GHz CPU with 1 GB RAM.

3.3 Computational Details

There have been performed several simulations of pseudorotaxane complex in various environment. As experiments indicate a dependence of the molecular switch on the environment acidity simulations for both protonated and unprotonated system have been launched. The 'axle' molecule within the protonated system was enriched at the both ends by two hydrogen atoms (total charge 2+). These hydrogens are missing in the unprotonated system and the 'axle' molecule is neutral (total charge 0). Two types of explicit solvent - pure water (tip3p) and natural conditions environment (water and Na+, Cl- ions) were selected as an external environment. Moreover, a separate simulation within vacuum has been carried out.

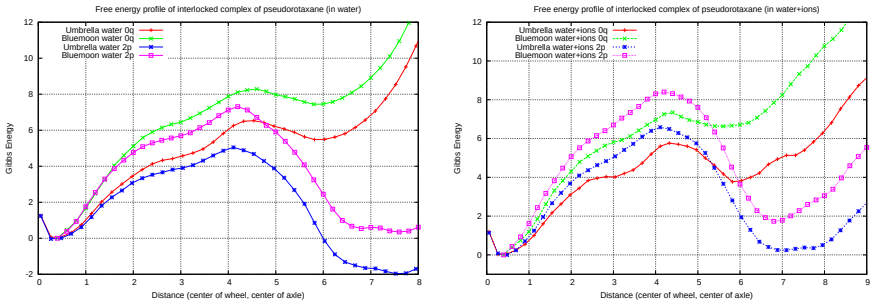
Charges and the input structure was obtained for both systems (protonated and unprotonated) from *ab initio* calculations using program Gaussian 03 [7] (basis 6-31G*). Subsequent parametrization for the molecular dynamics was carried out by program 'antechamber' from the AMBER 8 program package using the GAFF (General AMBER Force Field) parametric model. Program Solvate 1.0m2[3] was used for solvation of the model. Program AMBER 8 [8] was used for simulations themselves together with an additional library for calculations of free energy using methods Umbrella and Bluemoon. The reaction coordinate ξ was set as the distance of centers of individual parts of the complex ('wheel' and 'axle'). The same calculation of so-called main ridge of reaction coordinate was accomplished for both the explored methods. Such a calculation gradually increases the reaction coordinate from zero (interlocked state) up to 8 Å.

Individual subjobs suitable to be solved in the grid environment have been generated in consecutive steps of the main computational job. Each subjob simulated 250 ps of constrained or restrained molecular dynamics, respectively. The integration step was chosen as 2 fs with the SHAKE algorithm option switched on. Concerning the size of the system (17 000 atoms) the length of the particular subjobs was around 4 up to 5 days.

3.4 Results

The simulations confirm the fundamental impact of the environment acidity on the free energy profile of studied system and therefore the possible molecular motion during pH change, too. There is an evident energy minimum for interlocked state ($\xi = 0$) at the unprotonated system (0q) in the pure water. On the

[3] http://www.mpibpc.mpg.de/groups/grubmueller/start/projects/solvate/solvate.html



(a) water environment (b) natural environment

Fig. 2. Free energy profiles of unprotonated/protonated pseudorotaxane complex in water and natural (water and ions) environments.

other hand, the protonated system (2p) in the same environment allows also energetically identical (or even better) favorable state in which the 'axle' molecule is undocked ($\xi = 7.5$).

The situation within the natural environment differs as can be seen from Fig. 2(a) and Fig. 2(b). The global minimum for unprotonated state is confirmed again by both the used methods and the system tends to prefer a interlocked state ($\xi = 0$). While the situation is different for the protonated state (2p) where a minimum appears that is not energetically favorable compared with a interlocked state. When comparing both the methods it can be seen that for pure water i.e. homogeneous environment the lengths of the simulations are sufficient enough to properly sample the phase space and the graphs of both methods converge to a smooth curves that differ in a constant (cca 2 kcal). This is in agreement with continuity of the defined problem. There is no such convergence within the natural environment where freely moving ions play a crucial role and this results in the different shapes of curves especially for unprotonated state. This behavior can be related to insufficient lengths of simulations needed for sampling the phase space.

4 Conclusion

The VOCE infrastructure provides a complete Grid infrastructure available for all users from Central Europe that is based on the latest Grid technologies. The VOCE environment can be easily utilized by Grid novices to easy join and test the Grid services as well as for production use. We also demonstrated on an example of a large computational analysis that VOCE is a place where challenging applications requiring advanced computational chemistry techniques can be carried out and corresponding research problems successfully solved.

References

1. J. Kmuníček, P. Kulhánek and M. Petřek, "CHARON System - Framework for Applications and Jobs Management in Grid Environment", In *Krakow Grid Workshop 2005*, Krakow: Academic Computer Centre. 2006.
2. P. Kacsuk, G. Dózsa, J. Kovács, R. Lovas, N. Podhorszki, Z. Balaton and G. Gombás, "P-GRADE: a Grid Programming Environment", *Journal of Grid Computing*, **1**, 171-197, 2003.
3. G. Andronico, R. Barbera, A. Falzone, G. Lo Re, A. Pulvirenti and A. Rodolico, "GENIUS: a Web Portal for the Grid", *Nuclear Instr. and Methods in Physics Research, A*, **502/2-3**, 433-436, 2003.
4. Y. Ling, V. Šindelář and A. E. Kaifer, Poly(pseudo)rotaxane based on cucurbit[7]uril. *Polymer Preprints*, 46, 2, 1144-1145. 2005.
5. V. Šindelář, S. Silvi and A. E. Kaifer, Switching a molecular shuttle on and off: simple, pH-controlled pseudorotaxanes based on cucurbit[7]uril, *Chemical Communications*, *DOI:* 10.1039/b601959e, 2185–2187, 2006.
6. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman: Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *Journal of Computational Chemistry*, 16, 11, 1339-1350. 1995.
7. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, Jr. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.
8. D. A. Case, T.A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross and P. A. Kollman, AMBER 8, University of California, San Francisco. 2004.