CompChem (RS) aplikacija u okviru projekta HP-SEE.
Dostupni računarski resursi, programi, prednosti i nedostaci

Branko J. Drakulić, Ivan O. Juranić

Centar za hemiju – IHTM, Univerzitet u Beogradu. Njegoševa 12, 11000 Beograd, Srbija
(*bdrakuli@chem.bg.ac.rs*)

Introduction

High-Performance Computing (HPC) Infrastructure for South East Europe's Research Communities (HP-SEE), supported by the European Commission through seventh framework project (FP7), link existing and upcoming HPC facilities in South East Europe in a common infrastructure, and offer operational solution for it. The initiative open the South East European HPC infrastructure to a wide range of new user communities, including those of less-resourced countries, fostering collaboration and providing advanced capabilities to researchers.[1] The three strategic groups, named as the virtual communities (VC), in computational physics, computational chemistry, and computational life sciences are included in the project.

Project offers computational resources and ported software applications on existing HPC infrastructure from Hungary to Azerbaijan. Along with this regular training and dissemination events were performed, aimed to introduce users in the basic of programming, usage of HPC recourses and software ported. The strong management of the project aim to obtain unified prerequisites on all HPC centers to facilitate porting of the software and their usage. In the same frame younger researcher should be trained to use ported software, perform basic programming and, consequently, be able to use similar HPC facilities in their further work.

Our group is the active participant of the project within Computational Chemistry virtual community. The aim of our application, named 'Quantum Mechanical, Molecular Mechanics, and Molecular Dynamics computation in chemistry' (CompChem(RS)),[2] is to use HPC resources, available in our country (mainly on the Institute of Physics, Belgrade), and resources of other partners, to improve and facilitate ongoing projects that use molecular-dynamics simulations, DFT and *ab initio* calculations. The other aim is the education in this field of postgraduate students.

In this communication we want to address a few points. The first is brief presentation of the software ported, including some examples of usage by our group, as well as the aim for recruitment of other researchers in Republic of Serbia that can become the users. The next is the presentation of HPC resources available by the other partners (Hungary, Romania, and Bulgaria offer BluGene/P super-computers, as an example). The last is the problem, lasting from the very beginning of the project, in the consent and recognition of the usefulness of available resources among different research groups that we met (means, among our colleagues that have a need for the similar resources, but did not use it, even if available). Such problem is, unfortunately, inherent to the area of the South-Eastern Europe, as well as to our scientific community in general.

Software ported

On the so far unique HPC facility in Serbia, PARADOX cluster on the Institute of Physics, Belgrade,[3] we ported source codes (programs) that are instantly ready-to-use, as is shown in Table 1. Scripts for the execution of the programs are available to users that have account on the cluster. The requisites to get access to HP-SEE infrastructure are given on the project Wiki page.[4] Any other software recognized as the useful for the users within VC Computational Chemistry, or the new ones, after registration, can be ported to the cluster. The prerequisite (generally within the project) is that software that should be ported is distributed for free for the academic community. For commercial software, user must provide the proof of license of the software version made for execution on more than one node (parallel execution).

So far we have a ported tools for massive classical molecular dynamics studies (NAMD), *ab initio* molecular dynamics (CPMD), classical DFT and *ab initio* calculations (Gaussian, Firefly). The last programs listed can be used for calculations on the semiempirical level of theory, too. For the molecular docking, AutoDock Vina is installed. For the modelling of the large set of compounds, including ligand-

based and structure-based virtual screening, we ported OpenEye applications. OMEGA is fast conformer generation tool, able to process tens of thousands of compounds per day. ROCS is shape comparison tool that can be used for virtual screening. EON compare electrostatic potentials of pre-aligned molecules, and often is used for the refinement of the ROCS results. FRED is docking program that performs fast exhaustive docking, and also can be used for virtual screening. To make comparison with other HPC centers in the project, we offer the list of programs related to chemistry, available on Bulgarian Supercomputer Center (BGSC), hosting IBM BlueGene/P.

Table 1. Software ported to PARADOX cluster.

Software/version[a]	Mode of execution[b]	Home Page
NAMD 2.8	P	www.ks.uiuc.edu/Research/namd/
Firefly 7.1G	P	classic.chem.msu.su/gran/firefly/index.html
CPMD 3.1	P	cpmd.org
Gaussian03	P	www.gaussian.com/
AutoDock Vina 1.0	P/S	vina.scripps.edu/
OpenEye Software OMEGA, EON, ROCS, FRED	P/S	www.eyesopen.com/products

a) The last stable version ported on the day of the submission of this communication; b) P – parallel, S – sequential.

BGSC offer NAMD, GROMACS, LAMMPS, and DL POLY for the molecular dynamics; CPMD, NW Chem, Quantum Espresso, and CP2K for the *ab initio* molecular dynamics simulation; GAMESS for the DFT and *ab initio* calculations.

Resources

Our home cluster PARADOX offer 89 worker nodes (2 x quad core Intel Xeon E5345 @ 2.33 GHz with 8GB of RAM) and 15 service nodes (Xeon-based nodes). About 80 nodes are always available for computation. Jobs are submitted using PBS queuing system. Number of nodes and number of processors per node can be chosen in the input script, as well as needed time of computation. Rare advantage is that there is no limitation on time of the computation. Such limitation is frequent on the big clusters. Paradox also offers large space for the permanent storage of the input/output data. Other resources within HP-SEE project is listed in Table 2.

Table 2. Illustration of the computing power offered within the HP-SEE project.

Name	Configuration	Processors*/ Threads**
Bulgarian Supercomputing Centre (BGSC) / BG	IBM Blue Gene/P	8192
HPCG at IICT-BAS / BG	Express 7000	576
HPCG at IICT-BAS / BG	NVIDIA GTX 295 (GPU)	1920
IFIN-HH / RO	Intel Xeon / AMD opteron	704
NCIT – UV / RO	IBM Blue Gene/P	4096
NCIT – UV / RO	Intel Xeon	400
NIIFI - SC / HU	AMD Opteron	798
Pecs - SC / HU	Intel Xeon	1152
Szeged - SC / HU	AMD Opteron	2304
Debrecen - SC / HU	Silicon Graphics (SGI)	1536

* Processors 850 MHz – 3.2 GHz; ** Threads reported for the Graphics processing units (GPU).

Usage examples

The one of the main aims of the project is to allow to participants to perform modelling of interest to their research in a full scale. So far we have been limited by the computational power, and able to perform very short simulations, or just to try the software of interest to us. Now we are able to use programs in their full power. As an example, molecular dynamics simulation, using biasing forces to map free-energy landscapes of compounds examined, were performed on the set of ~ 20 congeners in different explicit or implicit isotropic, or explicit anisotropic solvents, during 20-30 ns each. Small part of the results is published so far.[5,6] For such calculations about 10000 cpu hours were used on PARADOX, and the lot of the storage space. A bit (1/20), of the each run of such simulations, on the author lab computer (~ 5 GHz in two processors), takes about three days. For the different project we have need to perform docking of about 120 small organic molecules to acetylcholinesterase active site to obtain conformations suitable to be used in three-dimensional QSAR.[7] By using AutoDock Vina, which is very fast, on PARADOX this took half of the day. On the user lab computer few days will be spent for such calculation, using full processors power, and the lot of memory. Consequently, no other tasks could be done on that computer. As the last example, we examined dynamic behaviour of the protein isolated in the department of biochemistry of the Faculty of Chemistry.[8] Even protein is not big (140 amino-acid residues) large solvent cluster have been added, along with counter ions, to simulate ionic strength comparable with experimental conditions. Due to memory issues such calculation cannot be even started at the lab computer.

Number of users

So far, our part of the project counts eleven users, including two technical staff that did not use CPUs (also did not perform planed tasks at all). For the technical side of view we have high quality support from the researchers of Scientific Computing Laboratory – Institute of Physics, Belgrade. In Serbia, research community count much more users that devote their research time to computational chemistry (in the broadest sense). Many attempts to involve number of our colleagues, from the very beginning of the project, did not trigger their interest. Arguments for non-interest span range from asking 'Is it possible to use my terminal, and graphical user interface to submit and monitor my job, as that I perform computations in my own computer', to the inability to port software of interest in a way to be executed in parallel mode. There are many other subjective points of view, which are not suitable to be described in communication for the scientific conference. Number of training events so far, that instruct the users in using of HPC, from console commands, to non-trivial programming; as well as many information on usage of resources, in a cook-book way, on the project Wiki web site, freely available to all interested should be mentioned. There are a lot of researchers that did not give any comment, just cover their knowledge and skills (Figure 1).

Figure 1. The corresponding author view on the possible subjective attitude,
that triggers non-interest to project in our scientific community.

CompChem (RS) aplikacija u okviru projekta HP-SEE.
Dostupni računarski resursi, programi, prednosti i nedostaci

Projekt 'High-Performance Computing for the south eastern European Communities' (HP-SEE), podržan od strane Evropske komisije, povezuje postojeće resurse u oblasti računara visokih performanci od Mađarske do Azerbejdžana u zajedničku infrastrukturu. U okviru projekta postoje tri strateške grupe (virtuelne zajednice), Računarska fizika, Računarska hemija i Računarske prirodne nauke. Kao deo projekta, u okvru virtuelne zajednice Računarska Hemija, u Republici Srbiji postoji aplikacija 'Kvantno mehanička, molekulsko mehanička izračunavanja i simulacije molekulske dimnamike u hemiji' (skraćenica CompChem). U sopštenju su kratko prikazani dostupni računarski resursi u okviru celog projekta, programi povezamni sa hemijom instalirani na našem nacionalnom centru za izračunavanja visokih performansi, kao i nekim drugim centrima u okviru projekta. Na pristupačan način su opisane prednosti korišćenja dostupnih resursa i dat je kitički osvrt na mali interes naše naučne zajednice za korišćenje dostupnih računarskih kapaciteta/instaliranih programa.

References:
1. http://www.hp-see.eu/ - the project overview.
2. http://wiki.hp-see.eu/index.php/CompChem
3. http://www.scl.rs/
4. http://wiki.hp-see.eu/index.php/Access_Procedure
5. B.J. Drakulić, A.D. Marinković, I.O. Juranić *Tetrahedron Lett.* **53** (2012) 553
6. B.J. Drakulić, A. Pedrretti, M. Zloh, V. Slavnić, I.O. Juranić, M.M. Dabović "Range And Sensitivities of 2-[(Carboxymethyl)sulfanyl]-4-oxo-4-arylbutanoic Acids Property Spaces. Part 2. Multidimensional Free Energy Landscapes" 18th European Symposium on Quantitative Structure - Activity Relationships, "Discovery Informatics & Drug Design" (2010) Rhodes, Greece.
7. M.D. Vitorović-Todorović, I. Cvijetić, I.O. Juranić, B.J. Drakulić. "The 3D-QSAR study of 110 diverse, dual binding acetylcholinesterase Inhibitors, based on alignment independent descriptors (GRIND-2). The effects of conformation on predictive power and Interpretability of the models." (2012), submitted.
8. Article in preparation.