Request for Supplemental Allocation

Principal Investigator: Shantenu Jha[1,2]
Co-Principal Investigator: Joohyun Kim[1]
Co-Principal Investigator: Yaakoub El Khamra[3]
[1]*Center for Computation & Technology, Louisiana State University, Baton Rouge, USA*
[2]*Rutgers, State Univeristy of New Jersey, USA*
[3]*Texas Advanced Computing Center TACC, University of Texas, Austin, USA*

01 April 2011

1 Summary

We would like to request a supplemental allocation on TeraGrid resources. In the first year of our multi–year allocation (TG-MCB090174: *Scale-Up and Scale-Out of Ensemble-based Simulations*) we requested 5 million SUs on Ranger and 3 million SUs on Kraken. Our allocation was awarded half the requested SUs on Ranger (2.5 million) and 2 out of 3 million SUs on Kraken.

As we have pursued our science unhindered, we have run-out of SUs half-way through the year. We have over a dozen publications appearing (or soon to appear) in conferences and journals based on research conducted with our allocation, and would want to continue with our research at the same pace.

We document progress made along multiple fronts in the short period of time in the attached progress report. As it stands, our research will stall in the third week of April and remain that way for several months. We request 1.5 million SUs on Kraken and 1 million SUs on Ranger to tide us over to the next allocation renewal cycle, when we will be eligible to request an advance on our second–year allocation. This will allow us to continue our work and productivity for the next few months.

2 The Case for Continuity

As outlined in the progress report, we have delivered impressive scientific and technological advances in the short period since this grant was awarded. This includes over a dozen publications appearing in journals/conferences or in preparation. Importantly, we are on the trajectory that we were aiming for and are set to deliver on the goals that we hoped the allocation would facilitate. Specifically, the uninterrupted continuation is important as, (i) Project 2 forms the basis for *specialized* runs on the DE Shaw Anton machine, to which we will have access starting in Q2 of 2011, (ii) Projects 3 is an important component of the International Interoperability Project (between TeraGrid and DEISA), (iii) Project 5 will lead to the timely delivery of an infrastructure that in turn will be used by multiple biomolecular simulation groups on the TeraGrid for efficient and effective execution of ensemble-based simulations. We would like to maintain the pace of the progress we are making; additionally, it is important to mention that several graduate dissertations and papers critically depend upon non-disruption. In the next 3-6 months, we anticipate 3 Graduate student led publications and 2 theses (1 PhD (Wei Huang) and 1 Masters (Abhinav Thota)) based upon a continued allocation.

Table 1 shows the current status of the main stages of the projects in the allocation and the estimated computational requirements. The resource justification is as follows:

Project 2 Resource Justification: According to our recent benchmarks on Ranger, when using 32 cores, the time taken per MD step is approximately 0.06s for a SAM-I aptamer RNA; thus the wall clock time required to complete 1ns is 0.34 day; in other words for a 56K system, 1 ns simulations require $\approx$ 300 CPU hours. Thus each 100 ns simulation requires approximately 30,000 CPU hrs. Analysis [1] results indicate that more than 300 ns trajectory

Table 1: Status of subprojects and estimated requirements

Type of Calculation	Method or Package	HPC Resources To Be Used	SUs required
Atomistic MD Simulation MM-PBSA	NAMD AMBER	Ranger	1000K
AEE/EGFRs (50K atoms)	Amber	Kraken	1500K
Total SUs		Ranger/Kraken	2500K

is desirable for observing meaningful conformational dynamics. Therefore, without additional post-analysis including MM-PBSA calculations, a rough estimation suggest that we can obtain about 100 simulations of a similar system with 900,000 SUs (See `http://staging.teragrid.org/userinfo/aus/namd_benchmark.php` Boltzmann Ensemble sampling. And, analysis require 10-100 times of sampling time. Therefore, we expect to consume 100 K SU for 5 - 50 calculations that combine the sampling and the analysis. The total requested to finish this project is 1 million SUs on Ranger.

Project 3 Resource Justification:

The epidermal growth factor receptor (EGFR) is an especially important enzyme target in lung cancer therapy because it mutates and/or is overexpressed in most non-small cell lung carcinoma (NSCLC) tumours. Inhibition of kinase activation of EGFR is a frequently used method to suppress its functions [**?**]. The majority of tyrosine kinase inhibitors (TKIs) are ATP-competitive inhibitors which bind in the ATP-binding site. Molecular dynamics (MD) simulations will be used to study the structural and energetic properties of inhibitor-EGFR complexes. The binding affinity of inhibitors to EGFRs will be calculated by molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) methods [**?**]. This molecular level study is one component of the EU FP7 ContraCancrum (Clinically Oriented Translational Cancer Multilevel Modeling) project which aims at developing a composite multilevel platform for simulating malignant tumor development and pharmacologic responses to a therapeutic intervention (http://www.contracancrum.eu). We have employed large scale MD techniques using both TeraGrid and DEISA resources in order to study the interactions of inhibitors with wild-type and mutant EGFRs [**?**]. A better understanding of the reasons for the success or failure of a therapeutic intervention will help us in the selection of subgroups of patients who are most likely to respond to specific drugs, and paves the way for personalized treatment [**?**]. We have already performed a preliminary study of different inhibitors (AEE788, AFN941 and gefitinib) with EGFR which we now intend to extend to look at a wider variety of inhibitors and EGFR mutations and to probe longer time scale motions of the protein. Planned simulations include ensembles of 50, 50,000 atoms with 25 runs each. Each simulation lasts for 4ns and runs for 9 hours on 128 cores. We therefore request 1.5 million SUs on Kraken to finish this project. The ensemble size and number of runs are typical of established studies [**?**, **?**].

In conjunction with the scientific questions we are addressing, we are also involved in the development of a runtime execution system (DARE) that uses the Simple API for Grid Applications (SAGA: http://saga.cct.lsu.edu/) and SAGA-based Pilot-Job (BigJob) to allow the running and coordination of hundred if not thousands of large-scale ensembles across resources, both on the TeraGrid and the EU DEISA network, as part of the NSF-HPCOPS funded Interoperability Project. Significant progress has recently been achieved in allowing the use of SAGA to interoperate between these two different grids. A key goal of an extended allocation would be to further develop this infrastructure and assess performance using real scientific workloads, and make it available for the broader & larger community of biomolecular simulators.

References

[1] W. Huang, J. Kim, S. Jha, and F. Aboul-ela, "A mechanism for s-adenosyl methionine assisted formation of a riboswitch conformation: A small molecule with a strong arm," *Nucleic Acid Res.*, vol. 37, no. 19, pp. 6528–6539, 2009.