MOLECULAR DOCKING SIMULATIONS ON WINDOWS AZURE BASED CLOUDS

T. Kiss[1], P. Borsody[1], G. Testyanszky[1], S. Winter[1], P. Greenwell[2], H. Heindl[2], S. McEldowney[2]

[1]Centre for Parallel Computing, School of Electronics and Computer Science, University of Westminster, London, United Kingdom
[2] School of Life Sciences, University of Westminster, London, United Kingdom
115 New Cavendish Street, London, UK, W1W 6UW
T.Kiss@westminster.ac.uk

Running molecular docking simulations is widely utilized by research and industry in a wide area of disciplines, including drug design, environmental studies, or even psychology. As these simulations require intensive computing power, several distributed computing platforms have already been investigated to support such scenarios [1], [2].

The European VENUS-C (Virtual multidisciplinary EnviroNments USing Cloud Infrastructures) project developed and deployed a production quality cloud computing service for research and industry communities. The project has extended the concept of the Windows Azure Generic Worker in order to support the execution of massive scientific experiments. As part of this project, molecular docking applications have been ported and prototyped on Windows Azure based cloud computing resources.

The paper describes the architecture requirements and implementation of molecular docking simulations on the Venus-C platform. The implemented solution enables bio-scientists to submit, monitor and retrieve the results of virtual screening or random blind docking experiments to Azure based virtual computing resources from a user friendly graphical user interface. Bio-scientists can easily deploy a bundled end-user interface to their own desktop or laptop machine, and remotely manage the experiments through Web Services based communication. The client components transfers input files from the researcher's local machine to the Azure container storage blob, submits jobs to the generic worker's queue running in the Azure cloud, polls the Azure database for the job status, and finally downloads the results. The managements of the virtualised infrastructure and the submission and queuing of the jobs are completely transparent from the users' point of view.

The developed solution has been successfully tested by the biologist end-users and its performance has been evaluated. A complete analysis and benchmarking of the Venus-C enabled solution based on over 20,000 docking experiments will be provided. The performance and reliability of the new cloud-based solution will be compared to previously implemented service and desktop grid based solutions [3], [4]. The usability of the implementation, including both technical characteristics and access to and purchasing of cloud resources will be evaluated. The core Windows Azure infrastructure and its Venus-C extended components will analysed from both the developers' and also from the end-users' point of view.

[1] N. Jacq, J. Salzemann, F. Jacq, Y. Legre, E. Medernach, J. Montagnat, A. Maaß, M. Reichstadt, H. Schwichtenberg, M. Sridhar, V. Kasam, M. Zimmermann, M. Hofmann, and V. Breton, "Grid-enabled virtual screening against malaria," *Journal of Grid Computing*, vol. 6, pp. 29–43, 2008.

[2] E. Tantoso, H. Wahab, and H. Chan, “Molecular docking: An example of grid enabled applications,” *New Generation Computing*, vol. 22, pp. 189–190, 2004, 10.1007/BF03040958.

[3] T. Kiss, P. Greenwell, H. Heindl, G. Terstyanszky and N. Weingarten: Parameter Sweep Workflows for Modelling Carbohydrate Recognition**,** Journal of Grid Computing, Vol 8, No 4, 587-601, 2010. DOI: 10.1007/s10723-010-9166-8. SpringerLink, 25 Oct 2010.

[4] T. Kiss, G. Szmetanko, G. Terstyanszky, P. Greenwell, H. Heindl, *Molecular docking simulations on a combined desktop and service grid infrastructure.* In: Proceedings of Third AlmereGrid Desktop Experience workshop: desktop grid applications for eScience and eBusiness, 30 March 2010, Almere, The Netherlands.