Scalability of Molecular Dynamics Simulations of Lipopolysaccharide Membranes Using the GROMACS Software Package on Different Architectures

Frederico J. S. Pontes, Victor H. Rusu, Roberta P. Dias, Gabriel A. C. Hora, Denys E. Santos, Laércio Pol-Fachin, Roberto D. Lins, Thereza A. Soares

Departamento de Química Fundamental, Universidade Federal de Pernambuco, Recife, PE, 50740-560, Brazil.

{pontesfrederico,victorrusu,rdiasufpe,laerciofachin}@gmail.com, {roberto.lins,thereza.soares}@ufpe.br

Abstract. We present benchmarks and efficiency estimates for the GROMACS package performing molecular dynamics (MD) simulations of lipopolysaccharide membranes on different computer platforms (Intel Xeon, BlueGene/P, AMD Opteron Interlagos and Barcelona). Satisfactory parallel scaling was achieved in all architectures, even when it was compiled without vector instruction support on BG/P. GROMACS v 4.x scaled well up to 2048 cores for a system containing over one million atoms. Atom/core performance at different platforms was also assessed. It shows that MD simulation of large complex biological systems can greatly benefit from petascale computing and, if adequately ported, can take appreciable advantage of hybrid hardware.

Introduction

Lipopolysaccharide (LPS) membranes are the main component of the external wall of Gram-negative bacteria, where they play a pivotal role in the host cell defense against xenobiotic agents and antibiotics. LPS are large molecules consisting of a Lipid A portion and a polysaccharide chain (O-antigen) and oligosaccharide component (outer and inner core portions). The absence or presence of the O-antigen, a glycan polymer, portion determines if the LPS is classified as rough or smooth respectively. Since the O-antigen is a very exposed region on bacteria cell, this region becomes target for antibiotics recognition [Raetz and Whitfield 2002]. The inner and outer core regions are attached to Lipid A and are basically composed of unusual sugars as heptoses and keto-deoxyoctulosonate [Caroff and Karibian 2003] attached with chemical groups as phosphates, aminoacids and ethanolamine. The Lipid A component, a phosphorylated glucosamine disaccharide with multiple acyl chains of variable length, is the main responsible of the endotoxicity of the Gram-negative bacteria [Rietschel *et al* 1994]. A general structure of a rough penta-acylated LPS of *Pseudomonas aeruginosa* is indicated on Figure 1. Chemotypes derived from the mutational removal of sugar moieties from rough LPS are indicated in the lateral label. Supramolecular LPS and chemotypes structures change accordingly to environmental conditions (*e.g.* temperature, specific ion concentration), exhibiting transitions from lamellar to non-lamellar aggregation forms [Brandenburg *et al* 2003, Ernst *et al* 1999]. The biological relevance associated to the experimental difficulty to determine microscopic properties of LPS membranes calls for a complementary theoretical description of these systems. However, in order to describe biologically relevant processes, LPS membranes require simulation in the order of hundreds nanoseconds to a few microseconds.

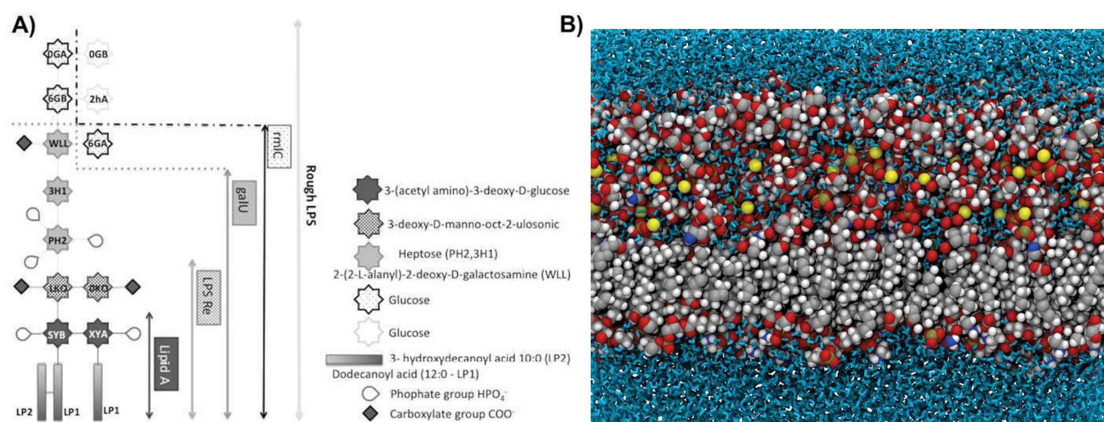

Figure 1. A) Scheme of the Rough LPS structure from *Pseudomonas aeruginosa*. Labels refer to monosaccharide units, phosphate and carboxylate groups. Vertical arrows describe the Rough LPS chemotypes simulated. B) Representative molecular model of a LPS membrane (atoms are color coded as: grey: carbon; red: oxygen; blue: nitrogen; brown: phosphorous; white: hydrogen; yellow: calcium ions; water molecules are shown in cyan).

We have developed and validated atomistic models for a number of LPS and Lipid A chemotypes. The models were able to describe the influence of temperature, ion valence and cross-linking ability on the supramolecular aggregate structure of LPS and Lipid A membranes as a function of their chemotypes [Kirschner *et al* 2012, Pontes *et al* 2012, Nascimento *et al* 2014]. MD simulations of these systems were carried with the GROMACS software package [van der Spoel *et al* 2005]. GROMACS is a MD package that explores parallel computer performance through efficient algorithms and code capabilities [Hess *et al* 2008, van der Spoel *et al* 2005]. Use of LINCS constraints in parallel [Hess 2008] and virtual sites [van der Spoel *et al* 2005] are examples of successful implementations, which enable the use of larger timesteps for the integration of Newton's motion equations. Parallel scheme using OpenMP multi-threading associated with use of domain decomposition has been shown to be more efficient than particle decomposition for parallel scaling and load balancing [Hess *et al* 2008]. In despite of these advances, MD simulations of charged membrane systems remain a challenge, not only due the electrostatic long range treatment of these systems, but also due to the simulation length required to estimate some of the typical lipid movements in bilayer membranes [Machán and Hoff 2010]. We have adopted a generalized reaction-field treatment with a cutoff of 1.4 nm for the long-range electrostatic interactions. In what follows, we compare the computing performance of GROMACS for simulations of different chemotypes of LPS membranes running on four architectures.

Results

We have carried out simulations in four different platforms: CENAPAD/PE (QuadCore Intel Xeon 5355, 2.66GHz, Infiniband Dual, and GROMACS 4.5.4), Intrepid (BlueGene/P (BG/P) 64-bit Power A2 16 cores/node, 1.6 GHz, 5D Torus 10GB/s and GROMACS 4.5.5), HPC2N (AMD Opteron 6238, 12 cores/node, 2.6 GHz Infiniband 40Gb/s, and GROMACS 4.6.2) and Chinook (2 Quad-core AMD Barcelona/node 2.2-gigahertz processors Infiniband DDR Interconnect 16 GB/s and

GROMACS 4.5.5). Simulation performance (ns/day) is presented as a function of the ratio between number of atoms per core on four different architectures (Figure 2). For the present simulations, BG/P and AMD Interlagos showed similar performance, followed by AMD Barcelona and Intel Xeon 5355. The lower performance of the latter was expected since this is a considerably older processor. Performances on HPC2N and Intrepid have largely benefited from increased inter-node communication, indicating that this is an important hardware parameter for efficient simulations. Scaling on BG/P architecture is nearly linear until 2048 cores. Roughly 50% efficiency is lost going from 2048 to 4096 cores. The inclusion of vector instruction support in recent compilation of GROMACS on BG/Q architecture is expected to offer improvement on this end. However, benchmarks are currently being obtained. Inter-node communication was not measured in this study, but it is well known to be inversely proportional to the number of cores per node. Moreover, the unexpected poor performance of 800 atoms/core in AMD Interlagos and AMD Barcelona are explained by the ratio between the number atoms of solute and solvent in the simulation box of two different systems, 0.71 vs ca. 0.18 for all the other systems.

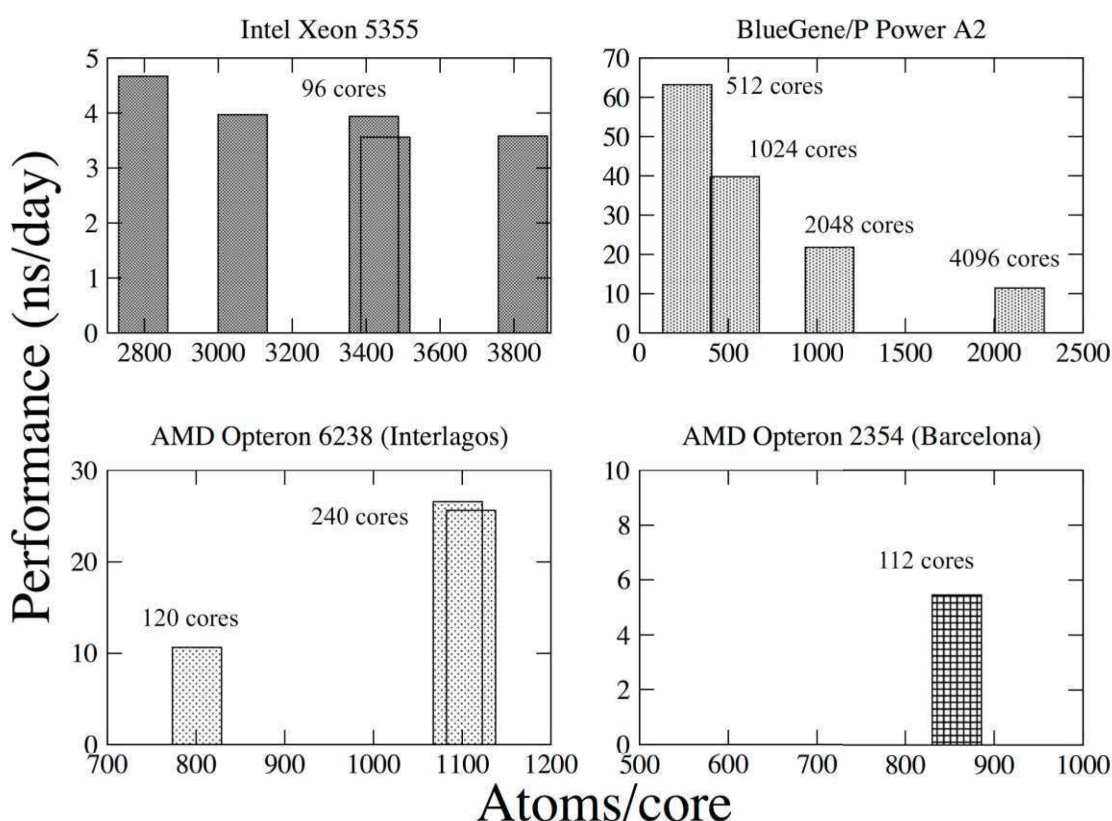

Figure 2. Atom per core performance (in ns/day) for MD simulations of different LPS membranes using four different architectures. (All benchmarks for the Intel Xeon 5355 were obtained using 96 cores).

Conclusions

MD packages for biomolecular simulations have reached maturity, and for this reason have greatly benefited from the use of many cores on different architectures.

Relative performance varies as a function of code, electrostatic treatment, cutoffs and force field granularity. The knowledge of the relative performance and bottlenecks of a given code on different hardware platforms are an essential step towards developing codes that run efficiently on hybrid architectures. For membrane simulations, the newer BG hardware has shown comparable performance to the latest AMD processors. In addition, fast inter-node connection is the most important hardware factor determining performance.

Acknowledgments

This work was developed with financial support from FACEPE, NanoBiotec-BR/CAPES, CNPq, INCT-INAMI and nBioNet. Computational resources were provided by Argonne Leadership Computing Facility, located at Argonne National Laboratory, the Swiss Supercomputing Centre (CSCS), the University of Umeå Supercomputing Centre (HPC2N) and CENAPAD-PE.

References

Brandenburg, K., Andra, J., Müller, M., Koch, M. H. and Garidel, P. (2003). *Carbohydrate Research*, pages 2477-2489.

Brandenburg, K., Mayer, H., Koch, M. H., Weckesser, J., Rietschel, E. T. and Seydel, U. (1993). *European Journal of Biochemistry*, pages 555-563.

Caroff, M. and Karibian, D. (2003). Structure of bacterial lipopolysaccharides. *Carbohydrate Research*, pages 2431-2447.

Ernst, R. K., Yi, E. C., Guo, L., Lim, K. B., Burns, J. L., Hackett, M. and Miller, S. I. (1999). *Science*, pages 1561-1565. Hess, B., Kutznet, C., van der Spoel, D. and Lindahl, E. (2008). *Journal of Chemical Theory and Computation*, pages 435-447.

Hess, B., Kutzner, C., van der Spoel, D. and Lindahl, E. (2008). *Journal of Chemical Theory and Computation*, pages 116-122.

Kirschner, K. N, Lins, R. D., Maass, A. and Soares, T. A. (2012). *Journal of Chemical Theory and Computation*, pages 4719-4731.

Machán, R. and Hof, M. (2010). *Biochimica et Biophysica Acta*, pages 1377-1391.

Nascimento, A., Pontes, F. J. S., Lins, R. D. and Soares, T. A. (2014). *Chemical Communication*, pages 231-233.

Pontes, F. J. S., Rusu, V. H., Soares, T. A. and Lins, R. D. (2012). *Journal of Chemical Theory and Computation*, pages 3830-3838.

Raetz, C. and Whitfield, C. (2002). *Annual Review of Biochemistry*, pages 635-700.

Rietschel, E., Kirikae, T., Schade, F., Mamat, U., Schmidt, G., Loppnow, H., Ulmer, A. J., Zähringer, U., Seydel, U., Di Padova, F., Schreier, M. and Brade, H. (1994). *FASEB Journal*, pages 217-225.

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. and Berendsen, H. J. C. (2005). *Journal of Computational Chemistry*, pages 1701-1718.