GPGPU Enhanced Protocols for Modeling of Biomolecules and Nanostructures

Shura Hayryan[1], Ming-Chya Wu[1,2], Chin-Kun Hu[1], Zuzana Gažová[3], and Tibor Kožár[1]

[1] *Insitute of Physics, Academia Sinica, Taipei, 11529, Taiwan*
[2] *RCADA and Department of Physics, National Central University, Chungli 32001, Taiwan*
[3] *Department of Biophysics, Institute of Experimental Physics, Slovak Academy of Sciences, Watsonova 47, 04001 Košice, Slovakia*

Abstract

Computational modeling of large biomolecular assemblies and nanoparticles requires powerful resources for both numerical number crunching and data storage. The recent development in general-purpose computing on graphics processing units (GPGPU) significantly accelerates the large-scale data processing. Several molecular modeling programs were updated for GPGPU environment with different execution efficiency. The speedup of certain molecular dynamics programs was remarkable. The running speed of these programs on powerful GPU clusters could compete with powerful supercomputers. Quantum chemical program codes transformed into GPGPU should benefit from accelerated data processing in a similar way. The presented study is aimed at performance testing of such GPGPU-related computational protocols devoted for simulation of biomolecular systems and nanoparticles. In addition, we were interested also in high-throughput virtual screening. We were using the AutoDock Vina program in order to find preferential binders to selected galectins, i.e. carbohydrate-binding proteins.

Introduction

Molecular modeling of macromolecules and their interaction profiles with other molecules as well as nanoparticles requires large computational resources. This is true for high-throughput calculations where millions of ligand molecules interacting with proteins have to be analyzed, as well as for high-performance runs where complex molecular dynamics simulations or quantum chemical calculations are required in order to get the inside view into such complex tasks as enzyme kinetics. Many of the calculations that were formerly done on supercomputers recently became doable on ordinary laboratory clusters equipped with GPGPU technologies.

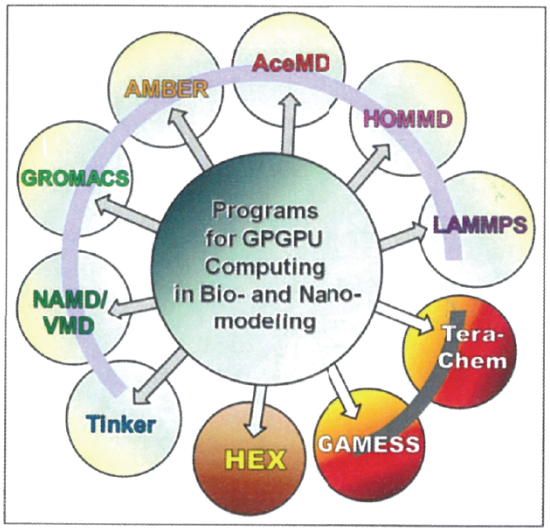

Figure 1: Examples of GPU-updated molecular modeling programs. The majority of programs support molecular dynamics simulations. Two other programs are quantum chemical ones, whereas HEX supports protein-protein docking.

Computational Methods and Protocols

Figure 1 represents an overview of selected programs used in molecular modeling that were updated to benefit from the GPGPU resources. The majority of these programs relates to molecular dynamics simulations. The Newtons's equation of motion is solved there in an approximate way of force field calculation of the molecular potential. The typical functional terms here account for bond deformations, bond angle distortions, torsion potentials, electrostatic interactions, nonbonded interactions (e.g. Lennard-Jones potential) and hydrogen bonding. The parametrization for these terms is usually empirical and atom/functional group related.

ACEMD [1] is a commercial molecular dynamics simulation package optimized for GPGPU environment. The AMBER 11 [2] and the GROMACS [3]–[6] programs were recently also upgraded for GPGPU.

The NAMD [7] and VMD [8] programs were among the first ones significantly benefiting from the GPU environment. NAMD is typically used for MD simulations of large molecules within an explicit solvent box [9].

The latest version of HOOMD (Highly Optimized Object-oriented Many-particle Dynamics) [10] is available at `http://codeblue.umich.edu/hoomd-blue/index.html`. HOOMD was originally proposed for polymer simulations and designed to run in the GPU environment. The recent version of LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator (`http://lammps.sandia.gov/index.html`) has enhanced GPU/CUDA support as well.

There were only two quantum chemical (QC) programs benefiting from GPGPU available in the time of writing this report. The first QC program written directly for GPU CUDA is the TeraChem software [11]–[14] commercially available

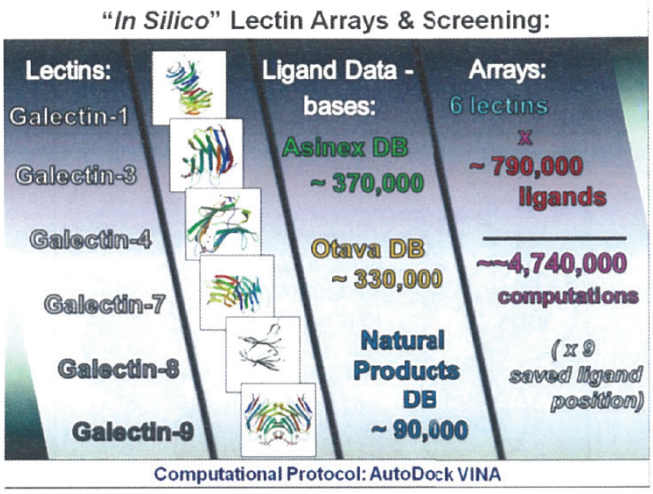

Figure 2: Complexity of the high-throughput molecular docking tasks. More than 42 millions of docking poses were calculated and evaluated in order to select the most potent set of ligand molecules.

at `http://www.petachem.com/`. From the other QC programs only GAMESS [15, 16] and its PC version GAMESS-UK supposed to be mentioned here as the second set of QC programs recently updated for GPU.

The protein-ligand or protein-protein interaction energies can be evaluated in a very fast way through molecular docking protocols with empirically parametrized functions. There are several programs available for this task and their list and usage overview was summarized in our former article [17]. In the recent report we will concentrate on the AutoDock VINA [18] program.

AutoDock VINA, although in contrast with the older AutoDock version, was not updated for GPGPU yet, but runs very efficiently for complex virtual screening tasks. We were using it as shown schematically on Figure 2 for finding potent binders to different carbohydrate-binding proteins. The ribbon-like representation of the structures of these lectins is illustrated on figure 2 as well. The experimental structures of the 6 galectins were extracted from the Protein Data Bank [19]. Small-molecule databases such as Asinex, Otava and the database of Natural products (with number of molecules as shown on Figure 2) were used for screening. The total number of docking runs was close to 4.75 million and there were 9 lectin-ligand complexes that resulted from each computational run. Our own software was used in order to process and analyze such large arrays and to select the preferential binders for further visualization, analysis and consequent computational tasks.

Results and Discussion

Although the GPGPU field is undergoing significant development, the implementation of the corresponding programs is still not a simple task. This is caused

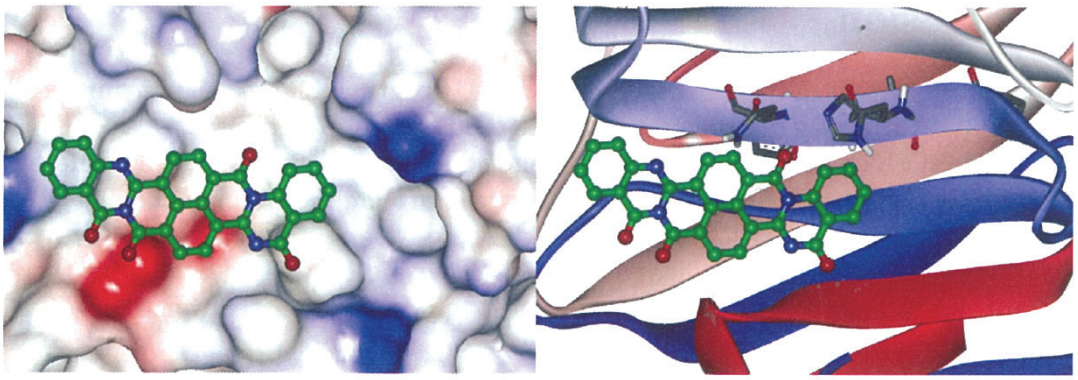

Figure 3: Example of the lowest-energy docking pose of ZINC05220992 molecule bound into Galectin 1 (PDB code 1GZW). This ligand is shown in ball and stick representation with the polar hydrogens. The water-accessible molecular surface with the ligand in the binding site is shown on the left figure, whereas the ribbon representation of the protein with the amino acids interacting with the ligand is shown on the right one.

by different program requests on CUDA/CULA and NVIDIA driver versions for different programs. We succeeded to get all programs shown on the scheme on Figure 1 to be compiled and executed successfully on the GPGPU platform. The performance of these programs was in according to our expectations: the GPGPU speedup is significant and is worth the problem solving required relating to all compilation difficulties and obstacles. For example, we monitored an eight-fold speedup in the case of NAMD for a solvated virus running on one processor of Intel Core i7 950 at 3.07 GHz, versus by adding one NVIDIA GPU (GTX 285 with 240 cores). Running the same task on a Supermicro machine (8 simultaneous threads on Intel Nehalem/CoreI7 Xeon E5620 at 2.4GHz with two NVIDIA TESLA C1060/240 cores) the speedup was even larger, approaching 26 times. This is a very promising result that initiated our interest towards simulations of amyloid aggregates of biological importance.

In the case of virtual screening, we were able to complete it in a few weeks. The complexity of the virtual screening relates not only to the number of calculations required but also on analysis and visualization task. The last is illustrated on Figure 3 where the best docking pose of ZINC05220992 molecule is shown.

The same molecule was shown to be the best binder for additional two lectins under study, i.e. Galectin 7 and Galectin 9. Further study is required in order to find more selective binders to these lectins. This could be accomplished either by screening additional databases or using different approaches (e.g. "core hopping"). A commercial version of the latest program from Schrodinger LLd. was also already updated for GPGPU and is available in the Schrodinger Suite April 2011 release. Based on the Schrodinger's GPGPU roadmap, this is their first program released for GPCPU (see `https://docs.google.com/document/d/1AMxX2LRRbtnkKOugxgGycUh1IAMElHXwR9V6BEYwmHY/edit?hl=en&authkey=CI_nxM4C&pli=1` for details).

Conclusion

The testing results of the implemented GPGPU programs showed significant speed-up for the majority of the implemented and tested programs. Although the AutoDock VINA program was not intended to run on GPGPU, we made significant performance throughput by calculating around half million docking poses per day. The protein structures were kept rigid during the docking runs. MD simulations of selected protein-ligand complexes will be completed as the next step in our study in order to gain more detailed information on the dynamics of the ligand binding to proteins of interest.

Acknowledgments. This work was supported by the Slovak Academy of Sciences (VEGA grants No. 2/0073/10 and 2/0079/10, Centre of Excellence Nanofluid), Slovak Research and Development Agency under the contract No. APVV-0509-07, Ministry of Education Agency for Structural Funds of EU in frame of project 26220120033, from the SAS Slovakia – NSC Taiwan joint research grant and by Taiwan National Science Council grants 100-2923-M-001-003-MY3, NSC 100-2811-M-001-057 & NCTS (Taipei).

References

[1] M.J. Harvey, G. Giupponi, G. De Fabritiis, *ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale*, J. Chem. Theory Comput., 5 (2009) 1632–1639.

[2] D.A. Case, T.A. Darden, I. T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R.Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. KolossvÃˇry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P.A. Kollman, *AMBER 11*, in, University of California, San Francisco, 2010.

[3] H.J.C. Berendsen, D. Vanderspoel, R. Vandrunen, *Gromacs – a Message-Passing Parallel Molecular-Dynamics Implementation*, Comput. Phys. Commun., 91 (1995) 43–56.

[4] E. Lindahl, B. Hess, D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, Journal of Molecular Modeling, 7 (2001) 306–317.

[5] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, *GROMACS: Fast, Flexible, and Free*, J. Comput. Chem., 26 (2005) 1701–1718.

[6] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, J. Chem. Theory Comput., 4 (2008) 435–447.

[7] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, *Scalable molecular dynamics with NAMD*, J. Comput. Chem., 26 (2005) 1781–1802.

[8] W. Humphrey, A. Dalke, K. Schulten, *VMD: visual molecular dynamics*, J. Mol. Graph., 14 (1996) 33–38, 27–38.

[9] J.E. Stone, D.J. Hardy, I.S. Ufimtsev, K. Schulten, *GPU-accelerated molecular modeling coming of age*, J. Mol. Graph. Model., 29 (2010) 116–125. [

[10] J.A. Anderson, C.D. Lorenz, A. Travesset, *General purpose molecular dynamics simulations fully implemented on graphics processing units*, J. Comput. Phys., 227 (2008) 5342–5359.

[11] T.J. Martinez, I.S. Ufimtsev, *Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation*, J. Chem. Theory Comput., 4 (2008) 222–231.

[12] I.S. Ufimtsev, T.J. Martinez, *Graphical Processing Units for Quantum Chemistry*, Comp. Sci. Eng., 10 (2008) 26–34.

[13] I.S. Ufimtsev, T.J. Martinez, *Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation*, J. Chem. Theory Comput., 5 (2009) 1004–1015.

[14] T.J. Martinez, I.S. Ufimtsev, *Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics*, J. Chem. Theory Comput., 5 (2009) 2619–2628.

[15] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, *General atomic and molecular electronic structure system*, J. Comput. Chem., 14 (1993) 1347–1363.

[16] M. Gordon, M. Schmidt, *Advances in electronic structure theory: GAMESS a decade later, Theory and Applications of Computational Chemistry: the first forty years*, CE Dykstra, G. Frenking, KS Kim, GE Scuseria (editors), 1167–1189.

[17] S. Hayryan, M.-C. Wu, C.-K. Hu, Z. Gazova, T. Kozar, *Modeling of structure, folding and interactions of biomolecules in the era of GPU computing*, in: L. Hluchá, P. Kurdel, J. Sebestyánová (Eds.) 7th International Workshop on Grid Computing for Complex Problems, Institute of Informatics, Slovak Academy of Sciences, Bratislava, 2011, pp. 36–44.

[18] O. Trott, A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*, J. Comput. Chem., 31 (2010) 455–461.

[19] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *The Protein Data Bank*, Nucleic Acids Res, 28 (2000) 235–242.