Journal of Optoelectronics and Advanced Materials Vol. 7. No. 6, December 2005, p. 3101 - 3106

PERFORMANCE ANALYSIS OF PARALLEL MOLECULAR DYNAMICS SIMULATION FOR BIOMOLECULAR SYSTEMS

A. Butu*

National Institute for Research and Development of Biological Sciences, Splaiul Independenței 296, Bucharest, Romania

Although the functional dynamics of the biomolecules goes from pico to miliseconds, the time needed to compute the molecular dynamics trajectory is much longer. One method to shorten this time is the parallel simulation. This method follows from the Amdahl's Law which is governing the parallel processing. This way, there were realized a series of molecular dynamics parallel simulation experiments on two biomolecular systems. The first system has 16993 atoms and is composed by one protein with 130 amino acid residues (2021 atoms) solvated in a water box. The length of one side of the water box is 57 Å. The second system has 8933 atoms and is made up by one peptide with 9 amino acid residues (140 atoms) solvated in a water box having the side length 46 Å. The aim of these experiments was to set the number of PCs in a cluster for which the ratio time simulation /number of PCs in cluster is optimal.

(Received May 27, 2005; accepted November 24, 2005)

Keywords: Parallel simulation, Molecular dynamics, PC-cluster performances, Biomolecule

1. Introduction

The molecular dynamics follows the temporal evolution of a microscopic model system through numerical integration of the equations of motions for all freedom degrees. The major advantage of molecular dynamics is that it provides detailed information of short-time molecular motions. Simulations of molecular dynamics is an important challenge for scientific people and its limitation resides in computer time consideration. Molecular dynamics simulations can be performed in a microcanonical ensemble (constant number of molecules, volume, and total energy), a canonical ensemble (constant number of molecules, volume, and temperature), as well as in an isothermal-isobaric ensemble (constant number of molecules, pressure, and temperature). We have shown in [1] the importance of Monte Carlo methods in the field of structural molecular biology. The computational complexity of the behavior of molecular systems increases considerably with the molecule size. At this moment the problem is to study the dynamic of large systems consisting of several thousands of atoms. In order to reduce a simulation duration it is useful to use parallel computers. Parallel processing on a problem is governed by Amdahl's Law [2], which states that the time is divided into two components: one which describes the calculations made on each processor and the other which describes the necessary time to communicate the data between processors. Whenever the necessary time for communication is longer than the time for the calculations, the processor spends more time waiting than working.

2. Materials and methods

The program used for the molecular dynamics simulation was CHARMM version 30 [3, 4]. The water was modeled using the TIP3P potential [5].

* Corresponding author: alina_butu@yahoo.com

We worked on two biomolecular systems. The first system was composed by one protein, the human lysozyme, 1LZ1 [6]. The structure file for 1LZ1 was taken from PDB [7]. The Protein is made up by 130 amino acid residues (2021 atoms) and it was added 8 Cl^- ions to get zero electrical charge. The protein was solvated in a water box with 57×57×57 Å dimensions, and the final model, is noted S1 (Fig. 1), has 16993 atoms.

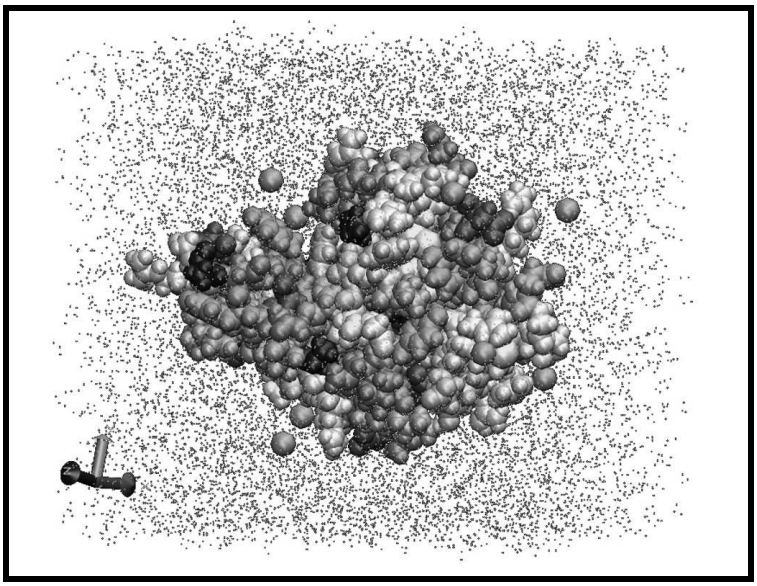

Fig. 1. The solvated protein - biomolecular system with 16993 atoms (S1).

The second system is composed of one peptide with 9 amino acid residues (140 atoms) solvated in a water box with 46×46×46 Å dimensions. This model, is noted S2 (Fig. 2) and has 8933 atoms. The water boxes were equilibrated with CHARMM.

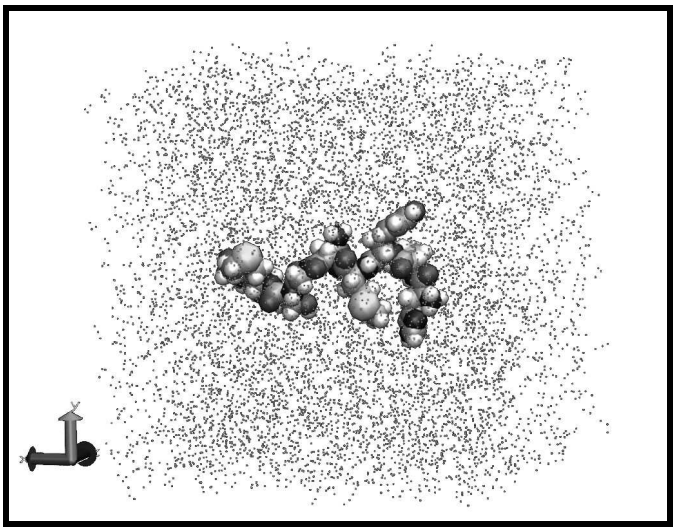

Fig. 2. The solvated peptide - biomolecular system with 8933 atoms (S1).

The simulation was performed in the NPT ensemble for system S1 and in NVT ensemble for system S2. The non-bonded cutoff distance was 14 Å. Bonds containing hydrogens were constrained with SHAKE [8], allowing a 1-fs timestep. The systems was subjected to minimization, heating end equilibration. After these stages, the production dynamics input files were tested for 24 hours in real time on each cluster.

In the present work we are using a large-scale atomic/molecular massively parallel simulator (LAMMPS) [9] LAMMPS is a classical molecular dynamics code suitable for modelling large molecular systems. LAMMPS is capable of modelling a variety of molecular systems such as bio-membranes, polymers, liquid-crystals, and zeolites. The code computes two kinds of forces: (1) short-range forces such as those due to van der Waals interactions and molecular bond stretching, bending, and torsions, and (2) long-range forces due to Coulombic effects. In the latter case, LAMMPS uses either Ewald or particle particle/particle-mesh (PPPM) techniques to speed the calculation [10] LAMMPS achieves parallelism by a spatial-decomposition of the workload which enables it to run large problems in a scalable way where both memory cost and per-timestep execution speed scale linearly with the number of atoms that are being simulated [11].

Because we used Ewald summation algorithms [10], the number of PCs in the clusters tested with system S1 was a power of 2 (1, 2, 4, 8, 16). The visualization of the biomolecular system in dynamics was done using VMD [12].

We used 18 PCs having the following configuration: Pentium IV at 2,8GHz, with 256M RAM, HDD 40G, network board 100 Mb. We also used a switch of 100 Mb with 24 ports. In order to avoid the fluctuations in the electricity network we used 8 UPS APC of 650 VA which could sustain 2 computers for about 5 minutes.

3. Results and discussion

The number of PCs in a cluster is noted with n. The same parallel molecular dynamics simulation input file ran 24 hours on each cluster. 5 readings of the number of steps were done, for the steps run at one hour distance and another reading for the number of steps run after 24 hours. The results were averaged in order to obtain the average number of steps made in one hour (Nn). The results obtained for an average number of steps per hour by each one of the 5 clusters tested are shown in Fig. 3.

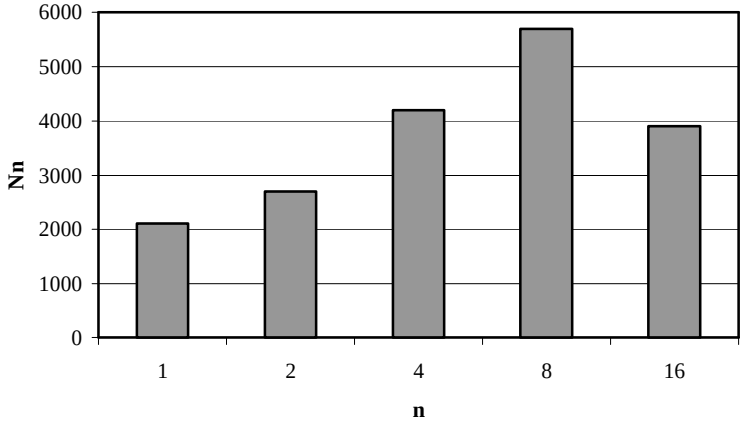

Fig. 3. Number of steps run per hour in parallel molecular dynamics simulation of solvated protein system (S1).

As we can see, the cluster made up by 8 PCs realizes the biggest average number of steps per hour. The cluster of 16 PCs realizes even less steps than the one with 4 PCs. In this case it is

obvious that we deal with a case when the data transmission takes more than processing them and therefore we meet a limitation of the cluster performances.

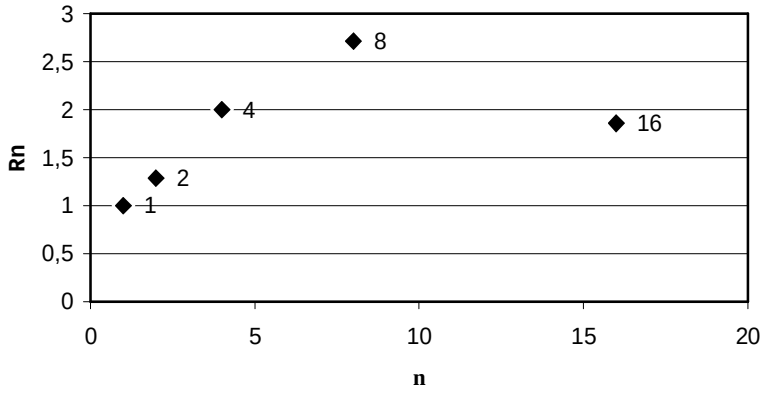

Fig. 4. Number of steps run on each cluster related with number of steps run on one single PC in parallel molecular dynamics simulation of solvated protein system (S1).

Calculating the ratio between the average number of steps per hour of a cluster with n PCs and the average number of steps of one single PC we obtain the computing capacity of the clusters against one single PC (Rn) (Fig. 4). This way it comes out that 4 parallel PCs work as much as 2 individuals, 8 PCs outrun 2,5 individual PCs, and the result for 16 PCs in smaller than the one for 2 individual PCs.

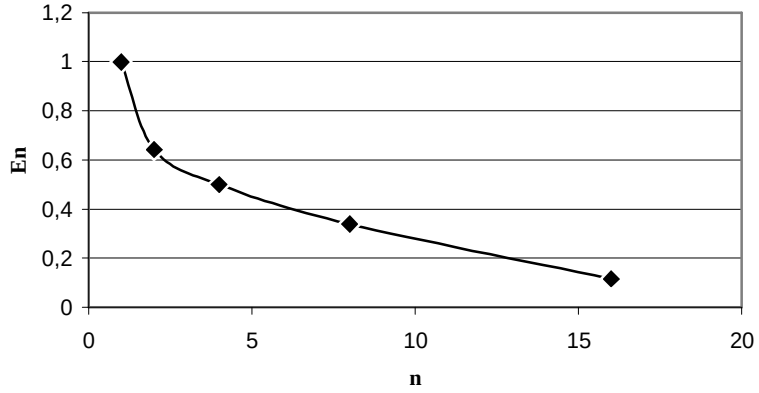

Fig. 5. Efficiency of clusters in parallel molecular dynamics simulation of solvated protein system (S1).

The efficiency of the use of computing capacity of the clusters was obtained both by calculating the ratio Rn/n and by reading the value CPU (s) used reported by the clusters. The curve presented by these parameter is illustrated in (Fig. 5) being obvious the decreasing of the efficiency in the same time with the increasing of the number of PCs in the cluster.

For S1 the best performance was obtained for the cluster of 8 PCs. This uses about 34% of the computing resources available. The cluster of 16 PCs is the less efficient, using only 12% of the computing resources.

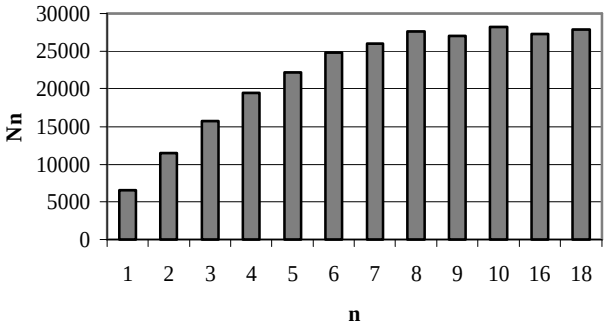

Fig. 6. Number of steps run per hour in parallel molecular dynamics simulation of solvated peptide system (S2).

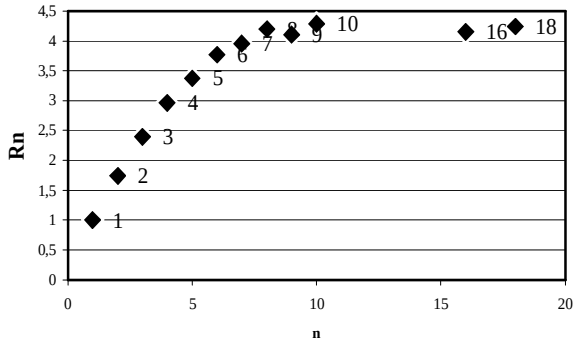

Fig. 7. Number of steps run on each cluster related with number of steps run on one single PC in parallel molecular dynamics simulation of solvated peptide system (S2).

The second biomolecular system, S2, gave the possibility to test clusters of 1, 2,…18 PCs. It started with one PC, and then followed clusters of 2,3,…PCs. When the average number of steps for a cluster began to decrease, the cluster of 18 PCs was tested, and then the one of 16 PCs.

It was read the number of steps run at one hour distance, 5 times, and the number of steps in 24 hours. By averaging it was computed the average number of steps run by a cluster per hour. In this case the average number of steps per hour was maximum for the cluster of 10 PCs (Fig. 6). The values for average number of steps per hour for the clusters of more than 10 PCs are on the same floor.

The calculation of the computing capacity of the clusters related to the capacity of one single PC (Rn) (Fig. 7) shows, in this case, that the cluster of 8 PCs overruns the capacity of 4 PCs that work individually. This value is a maximum and the clusters with more PCs present values close to this one and are on the same floor. In this case also, the data transmitting time grows larger than the computing time and it comes out the clusters performance limitation.

The curve of the efficiency of the clusters is computed just like the one for the system S1, and has the same form. This time, the cluster of 8 PCs uses 53% of the resources, and the cluster of 18 PCs uses 24%.

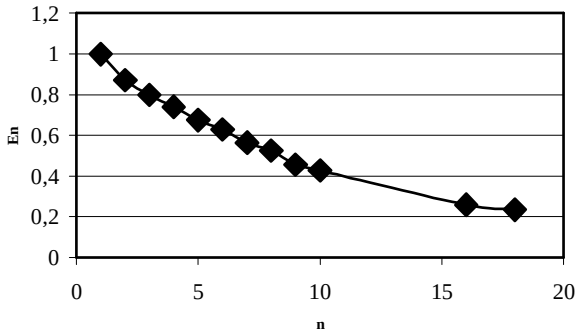

Fig. 8. Efficiency of clusters in parallel molecular dynamics simulation of solvated peptide system.

4. Conclusions

Both for parallel molecular dynamics simulation of solvated protein system (S1), and for parallel molecular dynamics simulation of solvated peptide system (S2) the best performances were obtained for the cluster of 8 PCs. This cluster ran the biggest average number of steps per hour in both cases. For S1 the cluster used 34% of the resources, while for S2 there were used 53 % of the resources. This difference is due, in part, to the different number of atoms of the systems (16993 against 8933), and to the different simulation parameters (NPT and NVT ensemble). For the clusters with more than 8 PCs the data communication time becomes larger than the time necessary for computing for both tested systems. Therefore the simulation duration on these clusters grows dramatically for S1 and enters on a floor for S2.

When the research supposes to simulate in the same time more structures that are alike a better choice is the use of a cluster of 4 PCs instead of clusters of 8 PCs.

A very good example is our case, when we need to simulate 4 S1 systems and we have 16 computers available for this experiment. Supposing that it is realized a trajectory of 3ns for each system, simulation duration for a system S1 will be 22 days on a cluster of 8 PCs and 30 days on a cluster of 4 PCs. In these conditions there are necessary 44 days to realize a 4×3ns of simulation on the cluster of 8 PCs and 30 days on the cluster of 4 PCs.

References

[1] A. Butu, J. Optoelectron. Adv. Mater. **7**(3), 1563 (2005).
[2] G. M. Amdahl, in AFIPS Conference Proceedings vol. 30 (Atlantic City, N.J., Apr. 18-20). AFIPS Press, Reston, Va., 1967, p. 483.
[3] B. R. Brooks, R. E. Bruccoleri, B. D. Olafsen, D. J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. **4**(2), 187 (1983).
[4] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrak, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, J. Phys. Chem. B. **102**, 3586 (1998)
[5] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[6] P. J. Artymiuk, C. C. Blake, J. Mol. Biol. **152**, 737 (1981).
[7] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, Nucleic Acids Research, **28,** 235 (2000).
[8] J. P. Rykaert, G. Ciccotti, H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[9] S. J. Plimpton, B. A. Hendrickson, J. Comput. Chem. **17(3),** 326 (1996).
[10] S. J. Plimpton, R. Pollock, M. Stevens, in Proc. of Eighth SIAM Conf. on Parallel Processing for Scientific Computing, Minneapolis, MN, March 1997.
[11] S. J. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[12] W. Humphrey, A. Dalke, K. Schulten, J. Molec. Graphics. **14**, 33 (1996).