Journal of the Korean Magnetic Resonance Society **2014**, *18*, 69-73
DOI 10.6564/JKMRS.2014.18.2.069

Letter to Editor: Accelerating atomistic refinement of NMR structures using Graphics Processing Unit

Jun-Goo Jee*

Research Institute of Pharmaceutical Sciences, College of Pharmacy, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 702-701, Republic of Korea

Received Nov 01, 2014; Revised Dec 01, 2014; Accepted Dec 07, 2014

Conventional structure determination of proteins using NMR data is an iterative process where NOE assignments and structure calculation are tightly coupled.[1] Steady advances in algorithm development now permit automatic calculation of 3D structures, provided chemical shifts are assigned for most atoms and the NOESY data quality is sufficient to obtain structural restraints.[2] For performing a search of conformational spaces using a suitable algorithmic or software method, these methods must meet the experimental restraints as well as force field that defines the physical energies between atoms. Traditional software for NMR structure calculation use simplified force fields compared to atomistic molecular dynamics (MD) simulation to enhance the search efficiency of conformational space.[3-5] For instance, Lennard-Jones potential and solvation energies are not included. The simplified force fields permit fast calculations by higher temperature annealing. However, the geometries of the regions lacking structural restraints often diverge and are inaccurate. On the other hand, atomistic MD calculations are suitable for characterizing the regions where experimental restraints are insufficient. Most atomistic MD-driven calculations for NMR structure refinement approximate the solvation effects by using generalized Born implicit solvent (GBIS) model[6] because of the very long computational times required for calculating interaction energies with explicit solvents. Thus, GBIS attains the right balance between the computational times and structural accuracy. We have reported extensively on the advantages of using GBIS in refining protein, protein–protein complex, and membrane protein structures.[7-12] However, GBIS still requires more than two orders of magnitude longer computational times compared to conventional calculation for NMR structures.

The performance of scientific computing hardware has undergone tremendous advances in the last few years. Particularly notable among the recent major breakthroughs is the adaptation of Graphic Processing Units (GPUs), originally designed to accelerate the manipulation of images, to scientific computation. GPUs are more effective for treating large blocks of data in parallel than Central Processing Units (CPUs). Hence, atomistic MD simulations are greatly accelerated when run on a GPU; in fact, the latest single GPU cards often outperform CPU-based clusters. AMBER software package has been developed for atomistic MD simulations, as well as for calculating NMR structures, where both NOE and dihedral angle restraints can be easily incorporated. In addition, version 12 of the AMBER package of Compute Unified Device Architecture (CUDA) supports the running in GPU environments.[13,14] Inspired by the high performance of the latest GPU cards, we have endeavored to harness the power of GPU for fast and accurate NMR-structure refinement. We selected GTX 780 (NVIDIA® GeForce) from among the latest GPU cards because of its speed and ready

* Address correspondence to: **Jun-Goo Jee,** Research Institute of Pharmaceutical Sciences, College of Pharmacy, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 702-701, Korea; Tel: 82-53-950-8568; Fax: 82-53-950-8557; E-mail: jjee@knu.ac.kr

Table 1. Statistics for GBIS-refined GB1 structures

Parameters	Durations of restrained simulated annealing (ps)							
	50*	50	100	200	300	400	500	1000
AMBER energy (kcal/mol)	-2,061	-2,060	-2,065	-2,069	-2,070	-2,071	-2,074	-2,073
Backbone RMSD (Å) (mRMSD, residues 1-56)	0.25	0.33	0.23	0.30	0.33	0.29	0.31	0.27
Most favored region in Ramachandran plot (%)	93.2	93.3	93.8	93.1	94.1	93.3	93.0	93.6
MolProbity Clash score	0.70	0.64	0.53	0.94	0.53	0.70	0.47	0.47
rRMSD (Å) (2QMT, residues 1-56)	0.74	0.77	0.78	0.71	0.75	0.74	0.73	0.74

* indicates the data from 50 ps CPU-based calculation

Table 2. Statistics for GBIS-refined UBQ structures

Parameters	Durations of restrained simulated annealing (ps)							
	50*	50	100	200	300	400	500	1000
AMBER energy (kcal/mol)	-3,161	-3,158	-3,165	-3,169	-3,171	-3,173	-3,175	-3,177
Backbone RMSD (Å) (mRMSD, residues 1-70)	0.34	0.26	0.41	0.30	0.32	0.29	0.34	0.27
Most favored region in Ramachandran plot (%)	85.8	85.6	85.4	86.1	84.8	85.6	86.1	85.5
MolProbity Clash score	2.68	2.44	3.05	2.92	2.84	2.93	2.84	2.64
rRMSD (Å) (1UBQ, residues 1-70)	0.76	0.74	0.78	0.76	0.77	0.76	0.76	0.76

commercial availability. One can purchase a GTX 780 equipped PC with the costs less than $2,000. Out of the four available CUDA-coded AMBER versions (SPSP, SPFP, SPDP, and DPDP), we selected SPFP (Single Precision Fixed Point),[15] which uses a combination of single precision for calculation and fixed precision for accumulation. We applied the GPU-accelerated GBIS to two model proteins: GB1 and ubiquitin (UBQ). Both high-resolution X-ray structures (PDB IDs: 2QMT for GB1 and 1UBQ for UBQ) are available, which enables us to juxtapose the results by GBIS in a straightforward manner.

The mean running times for the retrained simulated annealing (rSA) period during the GPU-accelerated calculations were 519.1 and 384.0 ns/day for GB1 and UBQ, respectively. Comparing these times with the corresponding times of 2.4 and 1.2 ns/day for the CPU-based runs shows that using GPU leads to dramatic increases in the speed. The speed acceleration values were approximately 216- and 320-fold for GB1 and UBQ, respectively. In the latest benchmarks for AMBER atomistic MD simulations (http://ambermd.org/gpus/benchmarks.htm), approximately 55-, 370-, and 464-fold acceleration for TRPCage, myoglobin, and nucleosome, respectively, were observed on a GTX 780 GPU against a Xeon CPU E5-2650 (2.60 GHz). It has been concerned that the high performance enhancement by GPU CUDA codes resulting from using a single-point precision can be negated by the diminishing accuracy of the results. It is, therefore, critical to know the difference between the structures obtained by CPU- and GPU-based GBIS. The structures optimized on CPU and GPU at 50 ps were almost identical for both GB1 and UBQ within the error range (Tables 1 and 2). It can be reasoned by two. First, SPFP likely causes less deviation compared with pure single-point precision barely impairing speed. Second, the use of experimental restraints in NMR structure calculations will minimize the potential errors caused by single-point precision to the point where they are smaller than those caused by simplified force field.

Table 3. Statistics for GBIS-refined GB1 structures with sparse distance restraints

Parameters	Durations of restrained simulated annealing (ps)							
	50	100	250	500	750	1000	1250	1500
AMBER energy (kcal/mol)	-2,061	-2,027	-2,097	-2,112	-2,107	-2,125	-2,130	-2,133
Backbone RMSD (Å) (mRMSD, residues 1-56)	2.90	2.09	1.82	1.25	1.53	0.97	1.16	0.88
Most favored region in Ramachandran plot (%)	93.4	94.1	93.7	94.7	94.8	95.3	94.6	94.2
MolProbity Clash score	0.06	0.18	0.12	0.18	0.0	0.23	0.23	0.06
rRMSD (Å) (2QMT, residues 1-56)	3.16	2.12	1.65	1.23	1.25	0.98	1.17	0.91

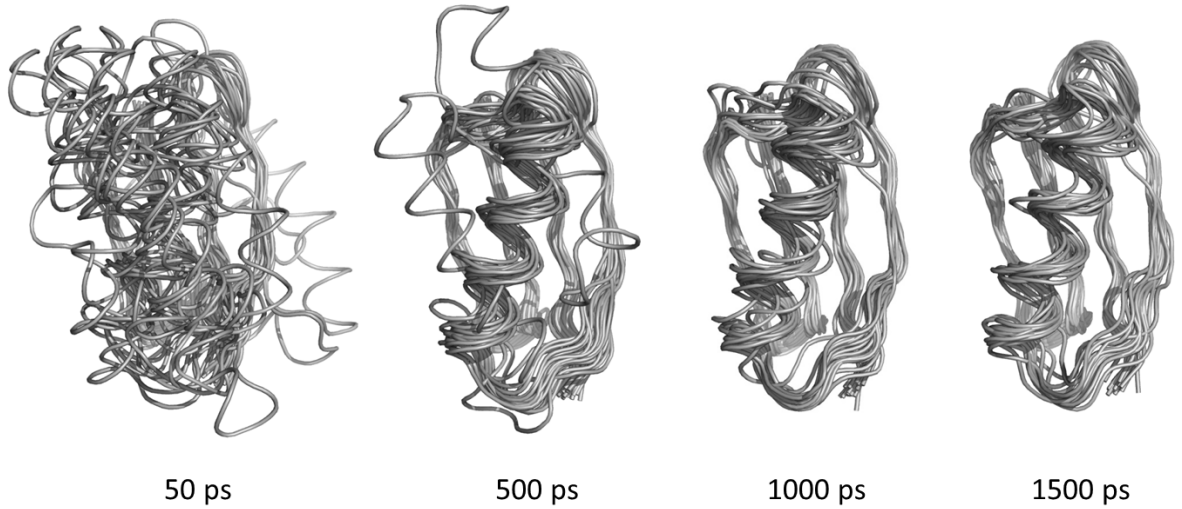

Figure 1. Ensembles of GB1 calculated with experimental sparse distance restraints and GPU-accelerated GBIS. Top 20 structures in each duration of restrained simulated annealing were overlaid with backbone atoms of residues 1-56. Corresponding lengths were labeled at the bottom. Pymol (http://www.pymol.org) prepared the figures.

We then investigated the potential improvement of the structure quality on increasing the length of the rSA duration. AMBER energies decreased slightly with time-step length (Tables 1 and 2). However, there was no apparent correlation between backbone root-mean–square deviation (RMSD) to the mean structure (mRMSD), backbone RMSD to the reference X-ray structure (rRMSD), the portions of the most favored regions in the Ramachandran plot, MolProbity clash scores, and the length. The lack of correlation of the backbone parameters (RMSD and the portion of the most favored regions) means that the geometries, at least in the backbone, were in close proximity to the optimal locations under the restraints. The large number of experimental distance and torsion angle restraints for both GB1 and UBQ might leave only a small margin for global improvement compared to other cases.

The effects of increments in time-steps were further studied with sparse experimental restraints. The distance restraints that one can quickly identify and prepare in the early stage of NMR data interpretation are those from amide protons. Knowing the accurate global fold of a protein only with the restraints is beneficial for subsequent analyses. Since GB1 includes 62 experimental HN-HN restraints, whereas UBQ does not, we refined the structures of GB1 under a series of rSA durations only with the HN-HN distance and backbone torsion angle restraints. Here the starting 100 structures for GBIS were also calculated by CYANA with the identical sparse restraints. The results clearly showed that a more extended time-step improves precision as well as accuracy of the resulting ensemble (Table 3). The values of mRMSD and rRMSD at 1500 ps were 0.88 and 0.91 Å,

respectively, which are comparable to those of conventional calculations. It is noteworthy that GPU-accelerated GBIS finished the calculation in a day with a single machine. Graphical representations reveal the improvements clearly (Fig.1). The tendency of improvements is consistent with our previous calculations with random omissions of distance restraints.[16]

In conclusion, we showed a dramatic acceleration of GBIS calculations by using GPU, which almost compensates the longer computational times required for GBIS. To the best of our knowledge, this is the first report of using GPU for refining the NMR structures. The 1500 ps calculation for 100 structures would be one of the longest atomistic refinements. This approach would permit a wider use of GBIS, even on moderately expensive hardware. GPU-accelerated GBIS could be particularly useful for biomolecules whose structures are difficult to determine by conventional NMR methods, such as membrane proteins and protein–protein complexes. It is difficult to obtain sufficient experimental restraints in the cases. The calculations with HN-HN restraints may approximate the situations. In this study, rSA was used for searching conformational space. One can combine sophisticated algorithm such as replica exchange[17] with GPU instead. Our results will be a meaningful guideline in applying GPU-based acceleration into calculating NMR structure.

Experimental Methods

AMBER structural refinement with GBIS consists of three stages: 1500-step minimization, restrained simulated annealing (rSA), and a second 1500-step minimization. In the rSA stage, the temperature rises to 1000 K during the first quarter; then, it is maintained at 1000 K during the second quarter, followed by cooling to 0 K during the second half. We used the ff99SB–ILDN all-atom force field[18] and the generalized Born model (*igb*=1 option), which is a pairwise descreening approach proposed by Hawkins, Cramer, and Truhlar.[19] The maximum distances for summing pairwise nonbonding interactions and effective generalized Born radii were set at infinity. The covalent bonds containing hydrogen were fixed by the SHAKE algorithm, resulting in an integration time-step of 2 fs. The force constants for distance and torsion angle restraints were 50 $kcal \cdot mol^{-1} \cdot Å^{-2}$ and 200 $kcal \cdot mol^{-1} \cdot rad^{-2}$, respectively. The initial structures for GBIS were generated by CYANA.[4] With the experimentally determined distance and torsion angle restraints, CYANA generated 100 structures that did not violate the input restraints greatly. There were a total of 584/101 and 1188/62 distance/backbone torsion angle restraints for GB1 and UBQ, respectively. For sparse distance restraints of GB1, 62 HN-HN restraints consisting of 37 sequential, 15 medium and 10 long-range data were extracted. Out of the 100 structures calculated by GBIS, we chose the 20 that showed the lowest AMBER energies as a final ensemble. As references, we performed 50 ps GBIS using the CPU version of AMBER with identical input files on an Intel® Xeon workstation with an E5-2650 (2.00 GHz) CPU. The ensembles resulting from both runs were compared in terms of AMBER energy, backbone root-mean–square deviation (RMSD) to the mean structure in ensemble (mRMSD), the portion of the most favored region in the Ramachandran plot, the MolProbity clash score[20,21], and RMSD to the reference X-ray structure (rRMSD).

Acknowledgements

This Research was supported by Kyungpook National University Research Fund, 2011.

References

1. Wüthrich, K. NMR of Proteins and Nucleic Acids; Wiley: New York, 1986.
2. Güntert, P. Eur Biophys J 2009, 38, 129.
3. Nilges, M.; Clore, G. M.; Gronenborn, A. M. FEBS Lett 1988, 239, 129.
4. Güntert, P.; Mumenthaler, C.; Wüthrich, K. J Mol Biol 1997, 273, 283.
5. Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Acta Crystallogr D Biol Crystallogr 1998, 54, 905.
6. Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. J Biomol NMR 2002, 22, 317.
7. Jee, J. G.; Ahn, H. C. Bull Korean Chem Soc 2009, 30, 1139.
8. Jee, J. G. Bull Korean Chem Soc 2010, 31, 2717.
9. Jee, J. G.; Mizuno, T.; Kamada, K.; Tochio, H.; Chiba, Y.; Yanagi, K.; Yasuda, G.; Hiroaki, H.; Hanaoka, F.; Shirakawa, M. J Biol Chem 2010, 285, 15931.
10. Sekiyama, N.; Jee, J.; Isogai, S.; Akagi, K.; Huang, T. H.; Ariyoshi, M.; Tochio, H.; Shirakawa, M. J Biomol NMR 2012, 52, 339.
11. Jee, J. G. J Kor Mag Res Soc 2014, 18, 24.
12. Jee, J. G. J Kor Mag Res Soc 2013, 17, 11.
13. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. J Comput Chem 2005, 26, 1668.
14. Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. J Chem Theory Comput 2012, 8, 1542.
15. Le Grand, S.; Götz, A. W.; Walker, R. C. Computer Physics Communications 2013, 184, 374.
16. Jee, J. G. Bull Korean Chem Soc 2014, 35, 1944.
17. Sugita, Y.; Okamoto, Y. Chem Phys Lett 1999, 141.
18. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Proteins 2010, 78, 1950.
19. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 19824.
20. Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. J Biomol NMR 1996, 8, 477.
21. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. Nucleic Acids Res 2007, 35, W375.