Journal of the Korean Magnetic Resonance Society **2014**, *18*, 24-29
DOI 10.6564/JKMRS.2014.18.1.024

Effects of force fields for refining protein NMR structures with atomistic force fields and generalized-Born implicit solvent model

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 April 15, 2014; Revised May 2, 2014; Accepted June 10, 2014

Abstract Atomistic molecular dynamics (MD) simulation has become mature enabling close approximation of the real behaviors of biomolecules. In biomolecular NMR field, atomistic MD simulation coupled with generalized implicit solvent model (GBIS) has contributed to improving the qualities of NMR structures in the refinement stage with experimental restraints. Here all-atom force fields play important roles in defining the optimal positions between atoms and angles, resulting in more precise and accurate structures. Despite successful applications in refining NMR structure, however, the research that has studied the influence of force fields in GBIS is limited. In this study, we compared the qualities of NMR structures of two model proteins, ubiquitin and GB1, under a series of AMBER force fields—ff99SB, ff99SB-ILDN, ff99SB-NMR, ff12SB, and ff13—with experimental restraints. The root mean square deviations of backbone atoms and packing scores that reflect the apparent structural qualities were almost indistinguishable except ff13. Qualitative comparison of parameters, however, indicates that ff99SB-ILDN is more recommendable, at least in the cases of ubiquitin and GB1.

Keywords Molecular dynamics, GBIS, AMBER force field, NMR, Structure

Introduction

Improvements of algorithms have automated the determination of 3D NMR structures of biomolecules under the conditions that the assignments of chemical shifts for most protons are available, and NOE peaks suffice.[1] Several papers have reported fully automatic structure calculations with processed NMR data without any manual interpretation,[2-4] although the methods are not yet popular to end users. Despite the technical advances, however, the improvements of NMR structures through refinements mainly rely on the expertise of the researcher who performs the structure calculations. It is notable that there is no straightforward validation parameter to judge the soundness of a structure.[5] One of tactics to improve structural qualities is to employ additional computational calculations. The assorted strategies are two: the use of an empirical structure database and the application of sophisticated calculation methods stemming from molecular dynamics (MD) simulations.

Up-to-date algorithms and computational capacities have permitted MD simulation faithfully to reflect the behaviors of biomolecules. Here sophisticated atomistic force field plays an important role. A force field decides the energies between atoms, angles and torsion angles. When structures are trapped in unfavorable geometries, the raised energy enables

* 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, Republic of Korea; E-mail: jjee@knu.ac.kr

atoms to relocate toward optimal positions. Atomistic MD simulation restrained with NMR data can improve NMR structures. In most cases, atomistic MD simulations accompany generalized-Born implicit solvent model (GBIS). The implicit solvent model approximates the solvation effects with continuous medium, compensating computation times needed for calculating large amounts of explicit solvents. GBIS is particularly effective for improving the geometries in the regions where NMR restraints are insufficient to fix geometries. Several applications have reported the potency of GBIS. For instance, GBIS was helpful with unambiguous positioning of the donors and acceptors of hydrogen bonds, allowing further insights into the pH dependence of binding affinity in the complex between UIM and ubiquitin.[6] In addition to improving the local geometry, GBIS was capable of determining the global fold with limited experimental restraints. Brooks and his colleagues showed that GBIS could yield accurate 3D folds with less than 10 % of the original NOE data.[7]

On the other hand, the small changes of parameters can lead to marginal differences in atomistic MD simulation. For example, our results with two model proteins of ubiquitin (UBQ) and GB1 demonstrated that GBIS generates different structures dependent on implicit solvent models.[8] Our data also indicated that a change of surface tension value could lead to the structures in a better agreement with X-ray structure of crambin.[9] The results may emphasize the need to optimize parameters of GBIS. The case that structure determination with conventional methods is difficult will necessitate GBIS more. In the case, the numbers of experimental restraints are often insufficient, making structures more sensitive to changes of GBIS parameters.[10] As a step for finding optimal GBIS protocol, in this study, we compared the qualities of NMR structures of two model proteins, UBQ and GB1, under a series of AMBER force fields—ff99SB, ff99SB-ILDN, ff99SB-NMR, ff12SB, and ff13—with experimental NMR data.

Experimental Methods

Experimental restraints for calculating the structures of UBQ and GB1 were extracted from the PDB database (http://www.rcsb.org), where they are deposited as 1D3Z and 3GB1, respectively. The restraints for calculations included only the distance and backbone dihedral angle restraints. The distance restraints were 1,446 and 584 in number for UBQ and GB1, respectively. The torsion angle restraints consisted of 62 angles for UBQ, whereas those for and in GB1 were 52 and 49. Structure calculations by a set of restraints consisted of two steps: CYANA calculation[11] and AMBER-based refinement [12] from the CYANA results. We first calculated 300 structures of UBQ and GB1 with experimental distance and torsion angle restraints using CYANA. GBIS with AMBER package (ver. 12) further refined the best 100 CYANA structures that did not show significant violations against the experimental inputs. The option for generalized-Born implicit solvent model was igb=5. As a conformational search method by GBIS, we applied a restrained simulated annealing of 20 ps. 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 best 20 structures that showed the lowest energies with no significant violation against the distance (< 0.5 Å) and torsion angle restraints (< 5°) were selected as an ensemble for further analyses. Linux-cluster machines consisting of 120 cores made all the calculations by in-house written automation protocol, ALIS-NMR (Automatic pLatform for Iterative Structure calculation by NMR data).

For quantitative analyses, we compared the resulting structures in terms of two backbone RMSDs: eRMSD for the root-mean-square deviation in an ensemble to the mean structure, and rRMSD for the root-mean-square deviation in an ensemble to the reference X-ray structure. The reference structures were 1UBQ for UBQ and 2QMT for GB1. We chose the ranges of 1–70 and 1–56 residues in UBQ and GB1, respectively, for the RMSDs calculation. In addition, PROCHECK-NMR[13] and MolProbity[14] software packages calculated the most favored region in the Ramachandran plot and MolProbity packing

score, respectively. MolProbity generated two overall parameters of Clash score (C-score) and MolProbity score (M-score), which reflect the qualities of all-atom contacts and protein geometries, respectively. In both scores, the lower values indicate the better qualities. We also fitted the experimental residual dipolar coupling (RDC) data against GBIS-refined structures to determine R- and Q-factors. The values indicate how much the resulting structures satisfy the protein ensembles with which experimental restraints are gathered. Please note that it is necessary to use several parameters simultaneously for validation, since there is no single parameter to reflect the overall quality of resulting structures.[5]

Results & Discussion

Detailed description and comparison of force fields is beyond the scope of the current study, but it will be adequate to explain the force fields in this study for better understanding. The ancestor of modern force fields for MD simulation with AMBER package is ff94.[15] Ab initio quantum mechanical calculations have derived parameters of ff94. One of the main successors that ameliorate ff94 is ff99. The ff99SB is a modified version from ff99.[16] It improves the discrepancy of ϕ/φ dihedral angles by fitting the energies from multiple conformations of glycine and alanine tetrapeptides. Shaw and his colleagues reported that the side-chain torsion potentials of ff99SB in the residues of Ile, Leu, Asp and Asn deviate from expectations based on Protein Data Bank statistics, and developed ff99SB-ILDN.[17]
The ff99SB-NMR emanates from ff99SB by modifying backbone dihedral angles as to satisfy NMR observations.[18] Because ff99SB has revealed a tendency to stabilize less helical conformations of transiently folded peptides, Simmerling et al. devised ff12SB.[19] On the other hand, ff13 is different from ff99 and its related AMBER force fields with new charge models as well as van der Waals parameters.[19] Tables-1 and 2 list the parameters regarding global folds by five force fields. No force field dominantly outdid the others. The deviations in each criterion were not huge considering standard errors. Besides R- and Q-factors,[20] visual comparison of backbone ^{1}H-^{15}N RDC data between experimental and fitted values revealed the improve qualities and the

Table 1. GBIS-refined ubiquitin structures*

Model	eRMSD (Å)	rRMSD (Å)	Energy (kcal/mol)	Rama-chandran & (%)	(MolProbity) C-score M-score	(RDC) R-factor Q-factor
X-ray	n.a.#	n.a.	n.a.	95.5	1.690 1.640	0.975 0.133
CYANA	0.42	0.91	n.a.	73.0	0.486 2.195	0.762 0.495
ff99SB	*0.34*	0.65	-3,143	**89.2**	0.243 1.360	0.901 0.288
ff99SB-ILDN	**0.29**	**0.62**	-3,092	87.7	0.203 **1.295**	0.897 0.299
ff99SB-NMR	0.30	0.63	-3,033	88.4	**0.041** 1.338	0.907 0.285
ff12SB	**0.29**	**0.62**	-2,548	87.9	0.243 1.441	*0.886* *0.314*
ff13	0.31	*0.69*	-3,696	*83.2*	1.257 *2.042*	**0.920** **0.254**

* Of the ensemble by GBIS, the criteria showed the best values are written in bold and underlined. The worst values are italicized. Emphases omitted AMBER energies, because they vary by different parameterization.

"n.a." means "not available".

& For Ramachan analysis, only most favored regions are considered.

apparent similarities in backbone geometries well (Fig.1 and Table-1).

In order to inspect the geometrical quality of side-chains, we then extracted $\chi 1$ angles of aromatic

Table 2. GBIS-refined GB1 structures

Model	eRMSD (Å)	rRMSD (Å)	Energy (kcal/mol)	Rama-chandran (%)	(MolProbity) C-score M-score	(RDC) R-factor Q-factor
X-ray	n.a.	n.a.	n.a.	96.0	0.000 1.119	0.987 0.132
CYANA	0.34	0.72	n.a.	85.5	0.000 1.789	0.927 0.316
ff99SB	0.50	0.59	-2,078	**96.6**	0.058 **0.654**	0.947 0.243
ff99SB-ILDN	*0.54*	**0.55**	-2,025	94.7	0.058 0.674	**0.950** **0.238**
ff99SB-NMR	0.41	*0.61*	-1,997	95.9	**0.000** 0.732	0.936 0.272
ff12SB	**0.39**	0.57	-1,594	94.2	**0.000** 1.269	0.922 0.310
ff13	0.46	0.55	-2,091	*88.2*	*0.292* *1.270*	*0.914* *0.339*

Note that structural restraints did not include RDC data. It is clear that ff13 resulted in somewhat worse values than the others in the most favored region of Ramachandran plot. One may argue that ff13 needs more optimization due to its recent advent. However, it will require more data to regard ff13 as a premature one, because ff13 revealed the best result from the viewpoint of RDC in UBQ (Table 1 and Fig. 1).

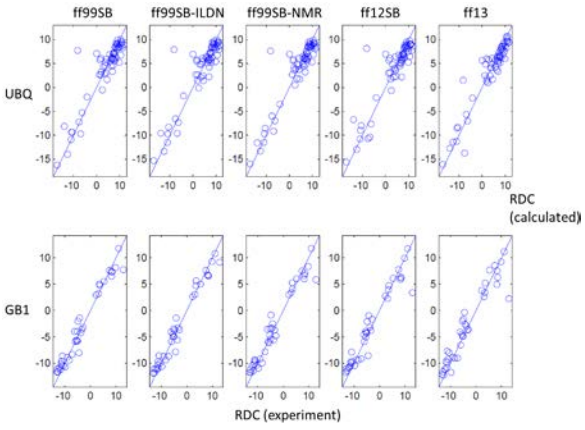

Figure 1. Residual dipolar coupling data fitted into GBIS-refined structures PALES software package[23] calculated R- and Q-factors by RDC.

Table 3. $\chi 1$ angles of GBIS-refined ubiquitin structures**

Model	Phe-4	Phe-45	Tyr-59	His-68
X-ray	-60.7	178.0	-63.3	-69.1
CYANA	-57.6 ± 12.5	176.9 ± 7.3	-85.9 ± 6.7	-72.2 ± 12.4
ff99SB	-68.0 ± 3.1	171.8 ± 5.7	-72.5 ± 4.6	-54.3 ± 4.1
ff99SB-ILDN	**-66.2** ± 3.0	**175.2** ± 4.1	**-71.3** ± 3.3	-52.7 ± 2.8
ff99SB-NMR	-67.8 ± 5.0	172.8 ± 4.6	-72.9 ± 2.8	-55.9 ± 3.2
ff12SB	*-73.0* ± 3.3	*164.2* ± 4.9	*-74.0* ± 2.5	*-49.6* ± 8.2
ff13	-66.8 ± 2.8	173.9 ± 3.1	-73.7 ± 4.7	**-56.4** ± 5.8

** All the values indicate degrees with mean ± standard deviation. The closest values to those in X-ray structure are written in bold and underlined. The worst values are italicized.

residues and calculated their statistics. The $\chi 1$ angles of aromatic residues in an NMR ensemble is a good indicator for the accuracy and precision of NMR structure.[21] Due to the practical difficulty of obtaining distance restraints from aromatic atoms, the parts in an ensemble are frequently divergent and occasionally wrong. UBQ contains four (Phe-4, Phe-45, Tyr-59, His-68), and GB1 has six (Tyr-3, Phe-30, Tyr-33, Trp-45, Tyr-45, Phe-52) aromatic residues. All the GBIS-refined structures had well converged geometries, not deviating much from the values in X-ray structures (Tables 3 and 4). The aromatic residue that showed the most deviated mean value was His-68 of UBQ by ff12SB. The standard error of the residue was the biggest. Yet there was no force field that surpassed the others in $\chi 1$ angles of aromatic residues either. We could conclude that ff99SB-ILDN is more recommendable, nevertheless, because it did generate worse results less than any other force fields, at least in the cases of UBQ and GB1.

Table 4. χ_1 angles of GBIS-refined GB1 structures

Model	Tyr-3	Phe-30	Tyr-33	Trp-43	Tyr-45	Phe-52
X-ray	-66.0	-71.2	167.5	-73.0	173.4	-64.4
CYANA	-57.4 ± 6.4	-74.9 ± 2.8	178.4 ± 4.4	-72.1 ± 2.9	173.4 ± 1.5	-100.1 ± 3.8
ff99SB	-57.1 ± 2.8	-76.7 ± 2.8	***178.7*** ± 3.4	**-71.4** ± 2.6	**175.6** ± 2.3	-72.5 ± 2.8
ff99SB -ILDN	-57.0 ± 1.9	***-76.8*** ± 2.1	178.3 ± 2.3	***-69.4*** ± 2.3	176.3 ± 1.5	**-71.6** ± 2.7
ff99SB -NMR	***-55.7*** ± 2.2	-75.6 ± 2.7	177.8 ± 2.9	-71.3 ± 1.9	176.9 ± 1.6	-72.1 ± 4.1
ff12SB	**-61.9** ± 3.3	-73.9 ± 1.4	**168.2** ± 1.9	-70.9 ± 1.5	***167***.0 ± 2.1	-73.0 ± 2.7
ff13	-58.1 ± 3.8	**-72.9** ± 1.6	173.4 ± 3.4	-70.0 ± 2.1	167.2 ± 3.6	***-77.4*** ± 4.3

In conclusion, we compared the performances of GBIS under popular AMBER force fields. There was no significant deviation of the results, probably owing to the sufficiency of experimental restraints. The difference may increase, if the number of restraints becomes much smaller. Therefore, it may necessitate studying the influences of force fields under the combinations of various restraints. It will moreover be meaningful to extend the study into other force fields such as CHARMM,[22] because it has somewhat different philosophy for defining forces. Our results will be a useful guidance on calculating NMR structures that are difficult with conventional methods.

Acknowledgements

This work was supported by the National Research Foundation (NRF) grant funded by the Korean government (MSIP) (NRF-2012R1A1A2007246).

References

1. Güntert, P. *European biophysics journal : EBJ* **38**, 129. (2009).
2. Lopez-Mendez, B.; Güntert, P. *Journal of the American Chemical Society* **128**, 13112. (2006).
3. Ikeya, T.; Takeda, M.; Yoshida, H.; Terauchi, T.; Jee, J. G.; Kainosho, M.; Güntert, P. *Journal of biomolecular NMR* **44**, 261. (2009).
4. Schmidt, E.; Güntert, P. *Journal of the American Chemical Society* **134**, 12817. (2012).
5. Montelione, G. T.; Nilges, M.; Bax, A.; Güntert, P.; Herrmann, T.; Richardson, J. S.; Schwieters, C. D.; Vranken, W. F.; Vuister, G. W.; Wishart, D. S.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L. *Structure 21*, 1563. (2013).
6. Jee, J. G. *Bull Korean Chem Soc* **31**, 2717. (2010).
7. Chen, J.; Im, W.; Brooks, C. L., 3rd *Journal of the American Chemical Society* 2004, **126**, 16038. (2004).
8. Jee, J. G. J *Kor Mag Res Soc* **17**, 11. (2013).
9. Jee, J. G.; Ahn, H. C. *Bull Korean Chem Soc* **30**, 1139. (2009).
10. Jee, J. G. *Bull Korean Chem Soc* **34**. (2014).
11. Güntert, P.; Mumenthaler, C.; Wüthrich, K. *Journal of molecular biology* **273**, 283. (1997).
12. 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* **26**, 1668. (2005).
13. Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *Journal of biomolecular NMR* **8**, 477. (1996).
14. 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 research* **35**, W375. (2007).
15. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **117**, 5179. (1995).
16. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **65**, 712. (2006).
17. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **78**, 1950. (2010).
18. Li, D. W.; Bruschweiler, R. *Angew Chem Int Ed Engl* **49**, 6778. (2010).
19. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, I., C.L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *University of California, San Francisco.* 2012.
20. Bax, A.; Grishaev, A. *Curr Opin Struct Biol* **15**, 563. (2005).
21. Jee, J. G.; Ikegami, T.; Hashimoto, M.; Kawabata, T.; Ikeguchi, M.; Watanabe, T.; Shirakawa, M. The *Journal of biological chemistry* **277**, 1388. (2002).
22. MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *The Journal of Physical Chemistry B* **102**, 3586. (1998).
23. Zweckstetter, M. *Nature protocols* **3**, 679. (2008).