Journal of the Korean Magnetic Resonance Society **2015**, *19*, 1-10

DOI 10.6564/JKMRS.2015.19.1.001

Comparison of NMR structures refined under implicit and explicit solvents

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 May 2, 2015; Revised May 21, 2015; Accepted May 28, 2015

Abstract Refinements with atomistic molecular dynamics (MD) simulation have contributed to improving the qualities of NMR structures. In most cases, the calculations with atomistic MD simulation for NMR structures employ generalized-Born implicit solvent model (GBIS) to take into accounts solvation effects. Developments in algorithms and computational capacities have ameliorated GBIS to approximate solvation effects that explicit solvents bring about. However, the quantitative comparison of NMR structures in the latest GBIS and explicit solvents is lacking. In this study, we report the direct comparison of NMR structures that atomistic MD simulation coupled with GBIS and water molecules refined. Two model proteins, GB1 and ubiquitin, were recalculated with experimental distance and torsion angle restraints, under a series of simulated annealing time steps. Whereas the root mean square deviations of the resulting structures were apparently similar, AMBER energies, the most favored regions in Ramachandran plot, and MolProbity clash scores witnessed that GBIS-refined structures had the better geometries. The outperformance by GBIS was distinct in the structure calculations with sparse experimental restraints. We show that the superiority stemmed, at least in parts, from the inclusion of all the pairs of non-bonded interactions. The shorter computational times with GBIS than those for explicit solvents makes GBIS a powerful method for improving structural qualities particularly under the conditions that experimental restraints are insufficient. We also propose a method to separate the native-like folds from non-violating diverged structures.

Introduction

Structure determination of biomolecules using NMR data relies on an iterative process that tightly couples NOE assignments and structure calculation.[1] Experimental restraints guide the conformational searches. In turn, the calculated structures help with gathering additional experimental restraints. The advanced algorithm accelerated by the increased computational capacity can automate the procedure for NMR structure calculation, provided the assigned chemical shifts are available for most atoms and the NOESY data are qualified to obtain sufficient structural restraints.[2] While the researches have established a standard protocol, the experimental restraints are still limited in number compared with that by X-ray crystallography. It often hinders the generation of accurate and precise structures. One of the computational approaches to overcoming the difficulty and improving the structural quality is to use sophisticated force fields. Traditional software for NMR structure calculation use simplified force fields compared to atomistic molecular dynamics

* 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

(MD) simulation. In details, Lennard-Jones potential and solvation energies are not included. The simplification enables the efficient traverse of the conformational spaces through high-temperature annealing. However, the geometries of the regions that lack structural restraints, often diverge and are occasionally inaccurate. Atomistic MD calculations are suitable for characterizing the areas 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.[3,4] GBIS is useful in attaining the structural quality expected by the calculation under explicit solvents, without spending massive computational times. We have reported on the successful applications of GBIS into refining protein, protein–protein complex and membrane protein NMR structures.[5-16]

Despite the improvements of GBIS, the comparison of NMR structures by GBIS and explicit solvents is lacking in number. Xia *et al*. pioneered the application of GBIS into the refinements of NMR structures and compared the results with those by explicit solvents.[17] However, the comparison was limited to a single MD duration time due to the computation capacity at the time. In this study, we report the direct comparison of NMR structures refined under GBIS and water molecules in two model proteins, GB1 and ubiquitin (UBQ), under various simulation times. We first calculated the structures by CYANA[18] with experimental distance and torsion angle restraints. Subsequently, AMBER software[19] refined the structures with implicit and explicit solvents under a series of simulated annealing MD time steps: 50, 100, 150, 300, and 500 ps. In addition to the root mean square deviations of the backbone atoms, we compared the qualities from the viewpoints of Cα chemical shifts and residual dipolar couplings (RDC) that were obtained by experiments but not used for structure calculations. We also extracted geometry-related parameters such as AMBER energies,[19] the most favored regions in Ramachandran plot,[20] and MolProbity clash scores.[21]

Experimental Methods

Restraints for NMR structure calculation—We used only the distance and backbone torsion angle restraints for calculating the structures of GB1 and UBQ. The PDB database contains the data as the ID of 3GB1 for GB1 and 1D3Z for UBQ. The numbers for the distance restraints are 582 and 1,446 for GB1 and UBQ, respectively. The distance restraints for intra ($|i-j|=0$), sequential ($|i-j|=1$), medium ($1<|i-j|<5$), and long ($|i-j|>4$) ranges were 122/288, 122/294, 82/236, and 256/628 for GB1/UBQ, respectively. The torsion angle restraints consist of 52 phi and 49 psi angles for GB1, whereas UBQ has 62 phi angles. The sparse distance restraints for GB1 and UBQ were 62 (32 sequential, 15 medium, and 10 long-range restraints) and 89 (2 intra, 36 sequential, 28, medium, and 23 long-range), respectively.

NMR structure calculations—Structure calculations consist of CYANA run and AMBER-based refinement with the CYANA results. We first calculated 300 structures of GB1 and UBQ with experimental distance and torsion angle restraints with CYANA. The top 100 CYANA structures that did not show significant violations against the experimental inputs were chosen for AMBER refinements. We used ff14SB all-atom force field for GBIS and explicit solvents by AMBER package (ver. 14). GBIS is set to have *igb*=8. The thickness of TIP3P water molecules was 10 Å. The cut-off for non-bonded interactions was set to infinity for GBIS and 10 Å for explicit solvents. As a conformational search method in MD refinements, we applied a restrained simulated annealing with PMEMD module. Here, the temperature was increased to 1,000 K for the first quarter duration. It stayed at 1,000 K for the second quarter, followed by a stepwise cooling to 0 K for the latter half. The force constants for distance and torsion angle restraints were 50 kcal/(mol·Å^2) and 200 kcal/(mol·rad^2), respectively. The integration time step for restrained simulated annealing was 2 fs with SHAKE restraints. Of 100 structures, the best 40 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 analyses.

Quantitative analyses of NMR structures—The resulting structures were compared from the viewpoints of two backbone RMSDs, one between the resulting structures (eRMSD; mean root-mean-square deviation in an ensemble to mean structure) and the other between the resulting structures and reference X-ray structures, 2QMT for GB1 and 1UBQ for UBQ (rRMSD; mean root-mean-square deviation in an ensemble to reference structure). We chose the ranges of 1-56 and 1–70 residues in GB1 and UBQ, respectively, for the calculation of RMSDs. In addition to AMBER energies, we calculated the portions of the most favored regions in the Ramachandran plot and MolProbity packing scores. The parameters were obtained by PROCHECK-NMR and MolProbity software packages,[20,21] respectively. It is noted that AMBER energies were calculated with identical parameters excluding solvent molecules to fairly compare the energies by implicit and explicit solvents. The R-factor for RDC means the Pearson correlation *r*. The Q-factors for RDC and Cα chemical shifts were calculated with the following formula,

$$Q = \frac{\sqrt{\sum (P_{\text{calc}} - P_{\text{exp}})^2}}{\sqrt{\sum P_{\text{exp}}^2}}.$$

Results and Discussion

Overall geometries little improved by lengthening the duration of molecular dynamics simulation- Increments of time steps in simulated annealing of MD simulation slightly decreased AMBER energies (Tables 1-6). The decrement of energy is indicative of the geometrical improvements. Tables 1, 2, 4 and 5, however, indicate that the improvements by the increased MD steps were insignificant both in GB1 and UBQ. In details, the parameters of R- and Q-factors for RDC and Q-factor for Cα chemical shifts, that reflect the differences between the experimental inputs and the simulated values from the resulting structures, were almost unvaried. The values of eRMSD and rRMSD fluctuated, not exhibiting a clear correlation with the MD durations. The data suggest that the spaces for geometry to get better were restricted in GB1 and UBQ owing to the sufficiency of experimental restraints. However, it should be noted that the geometric qualities by any atomistic MD refinements are markedly superior to those by CYANA.[5] It again supports the needs for employing atomistic MD refinements. Our data are largely consistent with what Xia *et al*. reported.[17]

Refinements with GBIS outperformed those with explicit solvents- It is clear that the GBIS runs outperformed those by explicit solvents (hereafter WAT) (Tables 1, 2, 4 and 5). AMBER energies, the most favored regions in Ramachandran plot, and MolProbity clash scores revealed that GBIS-refined structures had the improved qualities in both GB1 and UBQ. The extents of improvements were more distinct in GB1 than in UBQ. For instance, the values of rRMSD at 500 ps were 0.54 Å by GBIS and 0.66 Å by WAT in GB1, whereas UBQ showed 0.66 Å for GBIS and 0.71 Å for WAT. GBIS refinements are performed by calculating the non-bonded interactions between all the pairwise atoms. In other words, the cut-off for the non-bonded interaction is set to infinity.

On the other hand, WAT in this study included the non-bonded interactions between the two atoms located within 10.0 Å. We should stress that the calculation times by GBIS are shorter than those by WAT by three- to five-fold, despite the use of all the non-bonded interactions. Our data prove the obvious merits to employ GBIS than WAT in refining NMR structures. We guessed that the outperformance in GBIS would originate from the unlimited accounts of the non-bonded interactions.

To know the effects caused by the omission of the long-range non-bonded interactions, we set the cut-off to 10.0 Å and recalculated the structures by GBIS (hereafter $\text{GBIS}^{\text{cutoff}}$). However, the truncation impeded little the performances of GBIS (Tables 3 and 6). The differences of the results in GBIS and

GBIS^{cutoff} (Tables 1 & 3 and 4 & 6) lie within uncertainties both in GB1 and UBQ. It may mean that the effects caused by the long-range non-bonded interactions are minor at least in the cases where experimental restraints suffice.

Refinements with GBIS outperformed those with explicit solvents more under sparse restraints-
We further hypothesized that the inclusion of the long-range non-bonded energies might produce notable differences if experimental restraints were insufficient.

Table 1. Statistics from the top 40 GB1 structures by GBIS

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama* (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα√ Chemical Shifts
50	-1963	0.53 ± 0.11	0.55 ± 0.11	94.6	0.0	0.93 ± 0.06 0.29 ± 0.15	0.019
100	-1967	0.61 ± 0.18	0.49 ± 0.10	94.4	0.0	0.94 ± 0.05 0.27 ± 0.12	0.021
150	-1975	0.39 ± 0.16	0.58 ± 0.10	96.1	0.0	0.93 ± 0.06 0.29 ± 0.13	0.018
300	-1976	0.40 ± 0.10	0.55 ± 0.10	95.7	0.0	0.94 ± 0.05 0.26 ± 0.12	0.018
500	-1978	0.37 ± 0.11	0.54 ± 0.09	95.4	0.0	0.92 ± 0.06 0.30 ± 0.13	0.018

* "Rama" means the portion of the most favored region in Ramachandran plot
√ Back-calculated values of Cα chemical shifts

Table 2. Statistics from the top 40 GB1 structures by explicit solvents

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-1867	0.52 ± 0.14	0.61 ± 0.11	91.9	0.56	0.92 ± 0.05 0.33 ± 0.13	0.019
100	-1881	0.45 ± 0.11	0.60 ± 0.11	91.2	0.47	0.92 ± 0.05 0.31 ± 0.13	0.018
150	-1888	0.48 ± 0.13	0.64 ± 0.11	91.5	0.56	0.93 ± 0.05 0.30 ± 0.12	0.019
300	-1899	0.44 ± 0.10	0.64 ± 0.11	90.3	0.32	0.92 ± 0.04 0.31 ± 0.12	0.018
500	-1903	0.49 ± 0.13	0.66 ± 0.11	90.7	0.29	0.91 ± 0.04 0.36 ± 0.11	0.018

Table 3. Statistics from the top 40 GB1 structures by GBIS with cut-off for non-bonded interactions

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-1959	0.39 ± 0.15	0.51 ± 0.10	95.0	0.0	0.95 ± 0.04 0.23 ± 0.10	0.019
100	-1964	0.33 ± 0.12	0.49 ± 0.09	94.2	0.0	0.96 ± 0.04 0.23 ± 0.11	0.019
150	-1969	0.45 ± 0.14	0.55 ± 0.10	94.8	0.0	0.93 ± 0.06 0.29 ± 0.14	0.019
300	-1976	0.34 ± 0.10	0.52 ± 0.09	96.2	0.0	0.96 ± 0.04 0.22 ± 0.10	0.016
500	-1976	0.40 ± 0.13	0.54 ± 0.09	95.3	0.0	0.94 ± 0.05 0.27 ± 0.13	0.019

To test the idea, we performed structure calculations with sparse distance restraints. We extracted only the backbone-backbone distance restraints but kept all the experimental torsion angle restraints.

The numbers of the distance restraints corresponded to 10.1 and 6.2 % of original GB1 and UBQ restraints. The atomistic MD calculations with the sparse restraints were also executed under a series of durations in the three conditions: GBIS, WAT, and GBIS^{cutoff} (Tables 5-8).

Table 4. Statistics from the top 40 UBQ structures by GBIS

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-3046	0.29 ± 0.08	0.63 ± 0.04	89.8	0.41	0.90 ± 0.01 0.29 ± 0.02	0.013
100	-3052	0.33 ± 0.09	0.65 ± 0.05	89.8	0.39	0.89 ± 0.01 0.30 ± 0.03	0.013
150	-3056	0.32 ± 0.11	0.66 ± 0.05	89.0	0.53	0.89 ± 0.01 0.30 ± 0.02	0.014
300	-3063	0.30 ± 0.11	0.64 ± 0.04	89.2	0.39	0.89 ± 0.01 0.30 ± 0.02	0.014
500	-3065	0.27 ± 0.10	0.66 ± 0.04	90.6	0.49	0.89 ± 0.01 0.30 ± 0.03	0.014

Table 5. Statistics from the top 40 UBQ structures by explicit solvents

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-2905	0.37 ± 0.08	0.69 ± 0.04	89.4	1.22	0.91 ± 0.03 0.28 ± 0.04	0.013
100	-2932	0.36 ± 0.08	0.69 ± 0.05	88.3	1.14	0.91 ± 0.03 0.28 ± 0.04	0.015
150	-2944	0.36 ± 0.08	0.69 ± 0.04	88.7	0.89	0.90 ± 0.03 0.29 ± 0.04	0.015
300	-2961	0.37 ± 0.07	0.70 ± 0.05	88.7	0.93	0.91 ± 0.03 0.28 ± 0.04	0.014
500	-2969	0.34 ± 0.09	0.71 ± 0.04	87.2	1.03	0.91 ± 0.03 0.28 ± 0.04	0.014

Table 6. Statistics from the top 40 UBQ structures by GBIS with cut-off for non-bonded interactions

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-3041	0.26 ± 0.07	0.60 ± 0.04	90.4	0.35	0.89 ± 0.01 0.30 ± 0.02	0.013
100	-3049	0.26 ± 0.07	0.61 ± 0.03	90.1	0.37	0.90 ± 0.01 0.30 ± 0.02	0.014
150	-3053	0.24 ± 0.08	0.62 ± 0.03	89.5	0.53	0.89 ± 0.01 0.30 ± 0.02	0.015
300	-3061	0.26 ± 0.07	0.62 ± 0.04	89.2	0.67	0.89 ± 0.01 0.31 ± 0.02	0.014
500	-3065	0.27 ± 0.07	0.63 ± 0.03	90.8	0.53	0.89 ± 0.01 0.30 ± 0.03	0.015

The mean rRMSD values of starting CYANA structures were 4.57 and 3.08 Å for GB1 and UBQ, respectively, exhibiting the inaccuracy of the ensembles.

Nevertheless, GBIS refinements at 500 ps could produce the native-like folds, reflected by rRMSD values of 0.79 Å for GB1 and 0.60 Å for UBQ (Tables 7 &10).

Table 7. Statistics from the top 40 GB1 structures by GBIS with sparse restraints

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-1878	3.81 ± 1.14	4.23 ± 1.10	90.2	0.09	0.69 ± 0.14 0.67 ± 0.23	0.024
100	-1913	2.94 ± 1.08	3.15 ± 1.05	91.1	0.06	0.75 ± 0.11 0.61 ± 0.19	0.023
150	-1914	2.74 ± 0.92	3.33 ± 0.97	91.2	0.09	0.72 ± 0.12 0.65 ± 0.22	0.025
300	-1940	1.91 ± 1.16	1.91 ± 1.07	92.2	0.12	0.85 ± 0.10 0.44 ± 0.17	0.023
500	-1949	1.69 ± 1.01	1.71 ± 1.01	93.0	0.03	0.85 ± 0.11 0.43 ± 0.19	0.026

Table 8. Statistics from the top 40 GB1 structures by explicit solvents with sparse restraints

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-1775	2.74 ± 0.78	3.06 ± 0.95	87.5	0.59	0.68 ± 0.15 0.73 ± 0.32	0.029
100	-1803	2.84 ± 0.91	2.90 ± 0.89	87.2	0.85	0.75 ± 0.11 0.61 ± 0.21	0.019
150	-1814	2.74 ± 0.90	2.73 ± 0.85	86.8	0.76	0.76 ± 0.12 0.59 ± 0.23	0.026
300	-1833	2.69 ± 0.85	2.52 ± 0.91	87.3	0.82	0.77 ± 0.11 0.59 ± 0.21	0.027
500	-1838	2.82 ± 0.89	2.47 ± 0.85	87.5	0.73	0.73 ± 0.13 0.65 ± 0.24	0.030

Table 9. Statistics from the top 40 GB1 structures by GBIS with sparse restraints and cut-off for non-bonded interactions

Run (ps)	AMBER Energy (kcal/mol)	eRMSD (Å)	rRMSD (Å)	Rama (%)	MolProbity Clash Score	RDC R-factor Q-factor	Cα Chemical Shifts
50	-1884	3.92 ± 1.00	4.22 ± 0.59	90.8	0.03	0.72 ± 0.12 0.63 ± 0.20	0.023
100	-1900	3.13 ± 0.93	3.75 ± 0.87	91.2	0.0	0.73 ± 0.14 0.61 ± 0.25	0.024
150	-1912	2.49 ± 1.07	3.24 ± 0.91	92.9	0.0	0.79 ± 0.10 0.51 ± 0.15	0.024
300	-1922	2.70 ± 1.02	3.06 ± 1.23	92.5	0.03	0.79 ± 0.14 0.55 ± 0.25	0.027
500	-1942	2.18 ± 1.21	2.34 ± 1.08	92.7	0.06	0.83 ± 0.11 0.48 ± 0.20	0.025

However, three GBIS runs in GB1 resulted in the incorrect folds. The rRMSD values at 50, 100 and 150 ps were 4.23, 3.15 and 3.33 Å, respectively (Table 7). All the runs using WAT and GBIS^{cutoff} produced inaccurate structures under the sparse restraints in GB1 (Tables 8 & 9). The results in UBQ were markedly different from GB1. Except the run at 50 ps in GBIS^{cutoff}, all the other runs by WAT and GBIS^{cutoff} generated the accurate folds having rRMSD smaller than 2.0 Å (Tables 11 & 12). Taken together, our data supports two ideas; first, the truncation of non-bonded interactions impedes the performance of GBIS when the input data is insufficient. Second, the longer MD duration is helpful with finding accurate structure under sparse restraints. Of the experimental data, RDC correlated well with the accuracy of the resulting ensemble, whereas Cα chemical shifts were less sensitive to 3D folds. It is consistent with the observation that the chemical shifts reflect the local geometries.[22]

Calculating energies of each clustered ensemble could separate native-like fold in GB1- The systematic evaluations enabled to uncover the apparent correlations between rRMSD and geometric parameters. We found the strongest correlation between eRMSD and rRMSD (Fig. 1), which is entirely consistent with our previous results.[5] On the other hand, the correlation between AMBER energies and rRMSDs was indistinct (Fig. 1). However, we thought that the elevated AMBER energy can be a criterion to filter out wrong folds if combined with

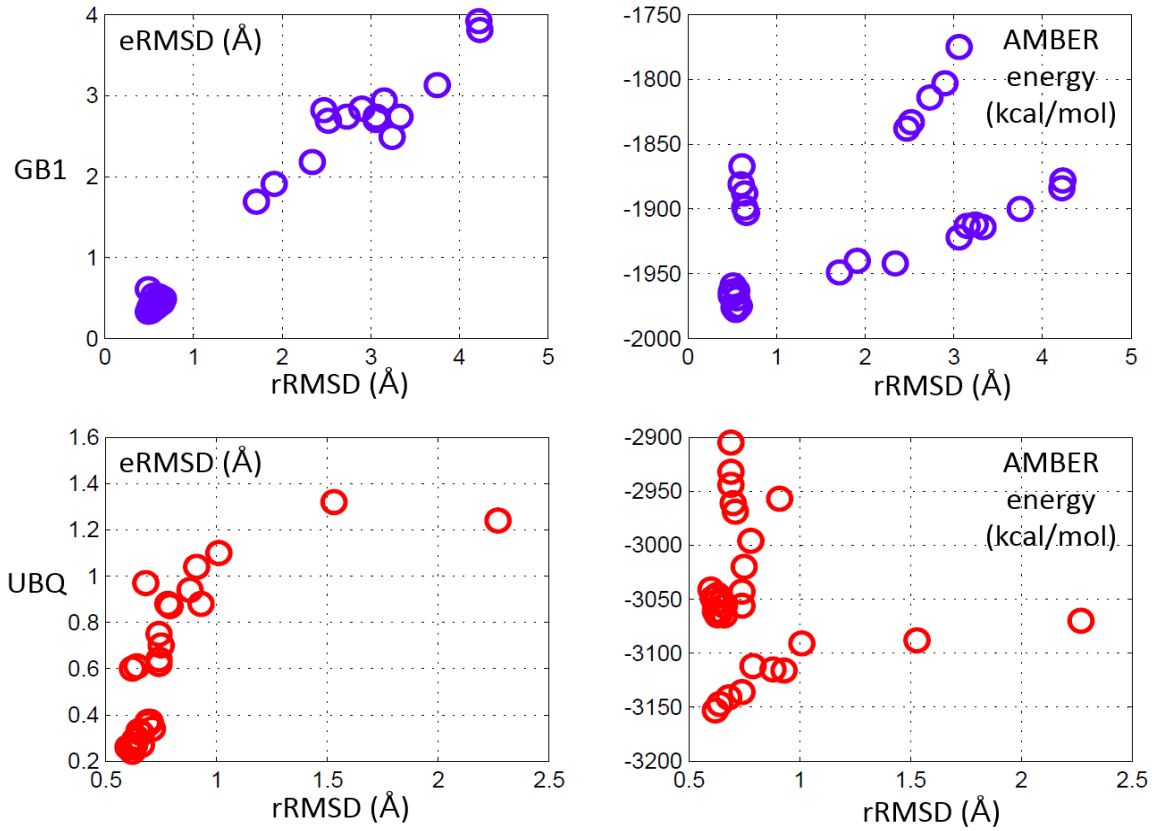

Figure 1. Scatter plots between rRMSD and two geometric parameters, eRMSD and AMBER energy, using the data from all the runs in Tables 1-12. GB1 and UBQ were plotted in blue and red. The correlation coefficients (*r*) between rRMSD and eRMSD were 0.99 and 0.71 for GB1 and UBQ, respectively. AMBER energy and rRMSD revealed 0.55 and -0.11 as correlation coefficients for GB1 and UBQ.

eRMSD. When non-violating structures are clustered with sufficient structures (> 5 %) and reasonable eRMSD (< 2 Å), the cluster of the lowest energy will have a high probability as an accurate ensemble. If there is no cluster, no structure will be selected as a representative. To confirm our idea, we focused on the run of GB1 at 50 ps by GBIS (Table 7), because the structures diverged, revealing 3.81 and 4.23 Å as eRMSD and rRMSD, respectively. We generated 500 structures with the sparse restraints by GBIS. Of 500, 228 structures did not show a significant violation of input restraints. We clustered the 228 structures based on the structural similarities by k-means method with 3.0 Å as a radius cut. Only the clusters that contain more than ten structures were considered, leading to two populated clusters. The cluster-1 consisted of 28 structures while 26 structures belonged to the cluster-2. The AMBER energy of the cluster-1 was -1901 ± 16 (kcal/mol) and that of cluster-2 was -1882 ± 16. The eRMSD values of cluters-1 and -2 were 1.65 ± 0.44 and 1.65 ± 0.48 Å, respectively. The cluster-1 was selected as a representative ensemble because its energy is lower than that of the cluster-2. Remarkably, the rRMSD values validated that the cluster-1 contains the native-like structures. The value by the cluster-1 was 2.31 ± 0.54 Å, whereas that of the cluster-2 was 4.62 ± 0.53. Fig. 2 showed the overlaid structures from the cluster-1 and -2. The cluster-2 included inaccurate folds. In details, the position of α-helix that backs the β-sheet was wrong, due to the total lack of long-range restraints. It is noteworthy that the structures in the two clusters did not violate the input restraints. Our data propose the way to collect the native-like folds of non-violating structures. Many algorithms for *de novo* protein folding employ the clustering algorithm to find native-like folds of decoys.[23-25] In most cases, the structure having the lowest energy is used as a reference to digitize the structural similarity in all the structures. However, it cannot discriminate two structures that have the similar RMSDs to the reference structure but show totally incorrect folds each other. Our method instead groups the structures and then calculates the mean energy of each cluster.

Conclusion

We performed extensive MD refinements with two model proteins, GB1 and UBQ, by changing restraints, simulation times and solvent models, and analyzed the data in the quantitative way. The total

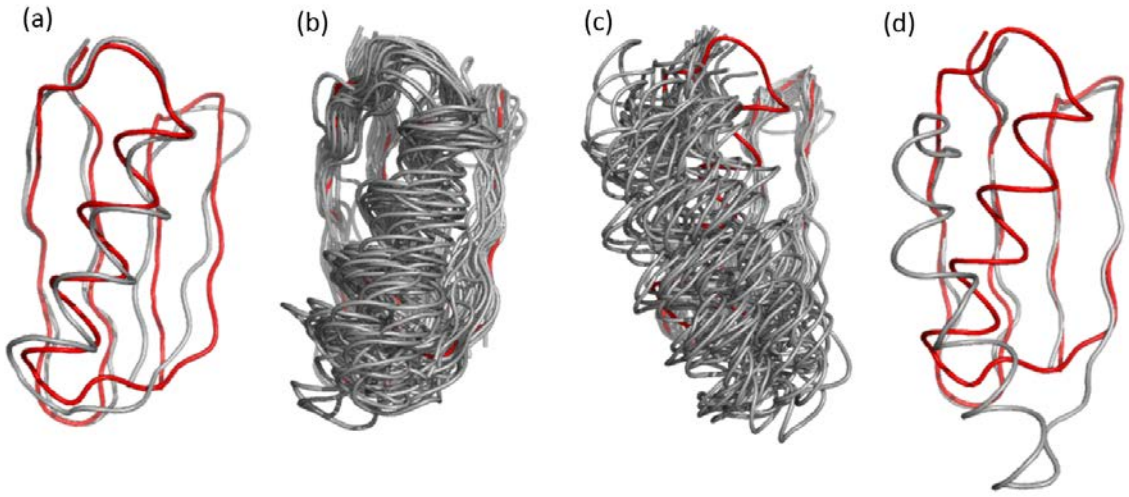

Figure 2. Two clusters (b, c) of GB1 structures from the run of 50 ps MD time steps with the sparse restraints by GBIS. Overlaid figures of the representative structure (gray) and the reference (PDB ID: 2QMT, red) in (b) and (c) were drawn in (a) and (d), respectively. The figures were generated by PyMOL (https://www.pymol.org).

structures refined by atomistic MD calculations reach 6,000 (12×5×100 = 6000) in number. It is a huge amount that one cannot access with ease. We made it feasible by automating runs using a cluster-based system. The results further enabled the study of the correlations between geometric parameters, suggesting a method to separate the native-like folds. Many biologically important proteins including membrane proteins and protein-protein complexes undergo intrinsic dynamics, which in turn cause the unfavorable line-broadenings of NMR signals for structure calculation. It often results in the lack of experimental restraints, hindering the determination of accurate and precise 3D structures. GBIS is a method to overcome the difficulty. One can use orthogonal computational methods including Rosetta and I-TASSER for the similar purpose.[23-25] These fragments assemble methods have become one of the primary options to predict the accurate structures of proteins. The recent data suggest the complementary use of atomistic MD simulation and Rosetta for improving the qualities of modeled structures.[26,27] The papers indicate the atomistic MD simulations are helpful with providing starting structures for Rosetta-refinements. Our data will be an excellent addition in this direction, guiding the combined approach for NMR structure calculation with sparse data.

Acknowledgements

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

References

1. K. Wüthrich, *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
2. P. Güntert, *Eur. Biophys. J.* **38**, 129 (2009)
3. A. Onufriev, D.A. Case, and D. Bashford, *J. Comput. Chem.* **23**, 1297 (2002)
4. V. Tsui, and D.A. Case, *Biopolymers* **56**, 275 (2000)
5. J.G. Jee, *Bull. Korean Chem. Soc.* **35**, 1944 (2014)
6. J.G. Jee, *J. Kor. Mag. Res. Soc*. **18**, 24 (2014)
7. J.G. Jee, *J. Kor. Mag. Res. Soc*. **17**, 11 (2013)
8. N. Sekiyama, J.G. Jee, S. Isogai, K. Akagi, T.H. Huang, M. Ariyoshi, H. Tochio, and M. Shirakawa, *J. Biomol. NMR* **52**, 339 (2012)
9. J.G. Jee, T. Mizuno, K. Kamada, H. Tochio, Y. Chiba, K. Yanagi, G. Yasuda, H. Hiroaki, F. Hanaoka, and M. Shirakawa, *J. Biol. Chem.* **285**, 15931 (2010)
10. J.G. Jee, *Bull. Korean Chem. Soc.* **31**, 2717 (2010)
11. K. Furuita, J.G. Jee, H. Fukada, M. Mishima, and C. Kojima, *J. Biol. Chem.* **285**, 12961 (2010)
12. J.G. Jee, and H.C. Ahn, *Bull. Korean Chem. Soc.* **30**, 1139 (2009)
13. M. Takeda, N. Sugimori, T. Torizawa, T. Terauchi, A.M. Ono, H. Yagi, Y. Yamaguchi, K. Kato, T. Ikeya, J.G. Jee, P. Güntert, D.J. Aceti, J.L. Markley, and M. Kainosho, *FEBS J.* **275**, 5873 (2008)
14. J.G. Jee, I.J. Byeon, J.M. Louis, and A.M. Gronenborn, *Proteins* **71**, 1420 (2008)
15. A. Ohno, J.G. Jee, K. Fujiwara, T. Tenno, N. Goda, H. Tochio, H. Kobayashi, H. Hiroaki, and M. Shirakawa, *Structure* **13**, 521 (2005)
16. F. Fujiwara, T. Tenno, K. Sugasawa, J.G. Jee, I. Ohki, C. Kojima, H. Tochio, H. Hiroaki, F. Hanaoka, and M. Shirakawa, *J. Biol. Chem.* **279**, 4760 (2004)

17. B. Xia, V. Tsui, D.A. Case, H.J. Dyson, and P.E. Wright, *J. Biomol. NMR* **22**, 317 (2002)
18. P. Güntert, C. Mumenthaler, and K. Wüthrich, *J. Mol. Biol.* **273**, 283 (1997)
19. D.A. Case, T.E. Cheatham 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R.J. Woods, *J. Comput. Chem.* **26**, 1668 (2005)
20. R.A. Laskowski, J.A. Rullmannn, M.W. MacArthur, R. Kaptein, and J.M. Thornton, *J. Biomol. NMR* **8**, 477 (1996)
21. I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall 3rd, J. Snoeyink, J.S. Richardson, and D.C. Richardson, *Nucleic Acids Res.* **35**, W375 (2007)
22. D.A. Case, *Curr. Opin. Struct. Biol.* **23**, 172 (2013)
23. A. Roy, A. Kucukural, and Y. Zhang, *Nat. Protoc.* **5**, 725 (2010)
24. S. Raman, O.F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, G. Liu, T.A. Ramelot, A. Eletsky, T. Szyperski, M.A. Kennedy, J. Prestegard, G.T. Montelione, and D. Baker, *Science* **327**, 1014. 2010
25. Y. Shen, O.F. Lange, F. Delaglio, P. Rossi, J.M. Aramini, G. Liu, A. Eletsky, Y. Wu, K.K. Singarapu, A. Lemak, A. Ignatchenko, C.H. Arrowsmith, T. Szyperski, G.T. Montelione, D. Baker, and A. Bax, *Proc. Natl. Acad. Sci. U S A* **105**, 4685 (2008)
26. S. Lindert, and J.A. McCammon, *J. Chem. Theory Comput.* **11**, 1337 (2015)
27. S. Lindert, J. Meiler, and J.A. McCammon, *J. Chem. Theory Comput*. **9**, 3843 (2013)