Journal of the Korean Magnetic Resonance Society **2016**, *20*, 102-108
DOI 10.6564/JKMRS.2016.20.4.102

Strategy for Determining the Structures of Large Biomolecules using the Torsion Angle Dynamics of CYANA

Jun-Goo Jee*

Research Institute of Pharmaceutical Sciences, College of Pharmacy, Kyungpook National University,
80 Daehak-ro, Buk-gu, Daegu 41566, Republic of Korea

Received Nov 15, 2016; Revised Nov 21, 2016; Accepted Dec 2, 2016

Abstract With the rapid increase of data on protein-protein interactions, the need for delineating the 3D structures of huge protein complexes has increased. The protocols for determining nuclear magnetic resonance (NMR) structure can be applied to modeling complex structures coupled with sparse experimental restraints. In this report, I suggest the use of multiple rigid bodies for improving the efficiency of NMR-assisted structure modeling of huge complexes using CYANA. By preparing a region of known structure as a new type of residue that has no torsion angle, one can facilitate the search of the conformational spaces. This method has a distinct advantage over the rigidification of a region with synthetic distance restraints, particularly for the calculation of huge molecules. I have demonstrated the idea with calculations of decaubiquitins that are linked via Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, or Lys63, or head to tail. Here, the ubiquitin region consisting of residues 1–70 was treated as a rigid body with a new residue. The efficiency of the calculation was further demonstrated in Lys48-linked decaubiquitin with ambiguous distance restraints. The approach can be readily extended to either protein-protein complexes or large proteins consisting of several domains.

Keywords NMR; Structure calculation; Complex structure; Torsion angle dynamics

Introduction

The determination of biomolecular structures using nuclear magnetic resonance (NMR) data is an iterative process that couples the nuclear Overhauser effect (NOE) assignment with the structure calculation.[1] Experimental restraints confine the conformational spaces. In turn the calculated structures facilitate the obtainment of additional restraints. Accumulated endeavors have successfully resulted in standard protocols for calculating the structure of biomolecules in a fully automatic way, provided the chemical shifts for most NOE-generating atoms can be assigned.[2] However, the completeness of assignments is not a sufficient condition for calculating accurate and precise structures. NMR signals should provide sufficient restraints in both quality and quantity.[3] The restraints for conventional NMR structure calculations are limited in number compared with those for X-ray crystallography. The obtainable experimental

*Correspondence to: **Jun-Goo Jee,** College of Pharmacy, Kyungpook National University, 80 Daehak-ro, Buk-gu, D aegu 41566, Republic of Korea, Tel: 82-53-950-8568; E-mail: jjee@knu.ac.kr

restraints decrease in systems with unfavorable motion that causes line broadening of the NMR signals to occur, occasionally leading to inaccurate and imprecise structures. In addition, because of the iterative features of automatic structure calculation, the soundness of the template structures in the initial stages is important, as the wrong structures in the first cycle can lead to inaccurate results in the CYANA calculation.[3]

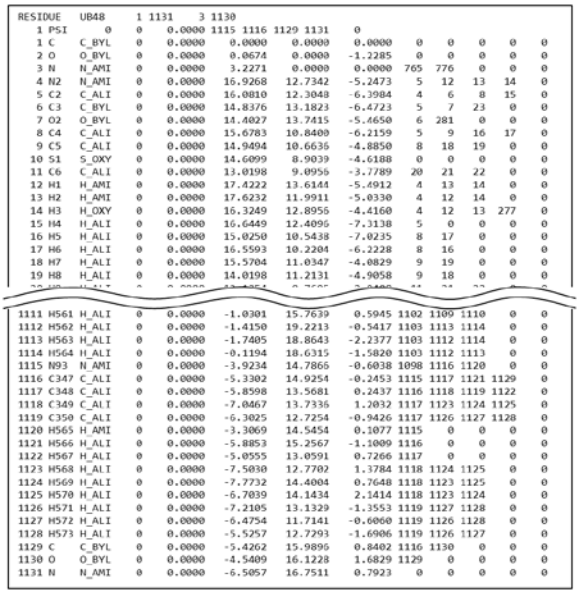
Simulations of biomolecular docking into a complex can employ the protocol for NMR structure calculation, provided the 3D coordinates of the comprising molecules are available. Besides NMR-derived data such as 2D chemical shift perturbations, the restraints from other experiments, including residue-specific mutations, small-angle X-ray scattering (SAXS), and cryo-electron microscopy (cryo-EM), can be combined to simulate the docked conformation accurately. The popular application for this purpose is the High Ambiguity Driven biomolecular DOCKing (HADDOCK) approach.[4] From its beginnings of simulating docking by ambiguous restraints with 2D NMR data, HADDOCK has evolved to integrate information (even originating from coevolution) for visualizing 3D complexes.[5] The information-driven docking simulation treats each unit as a rigid body once it reads structures independently. The concavo-convex geometry is the primary factor for generating the proper structure, assisted by electrostatic interactions.

The data acquired from proteomics techniques have accumulated information on the proximity of proteins in cells. Subsequently, the need for shaping huge complexes by related proteins has increased. As an approach to satisfying this need, in this paper, I suggest a method that handles a particular area in a protein as a rigid body, using CYANA,[6] which is the most popular software used for calculating structures from NMR data.[7] In addition to its ease of use, CYANA is able to calculate structures automatically using NOE data. Therefore, the incorporation of rigid body handling into the automatic NOE assignment is expected to be useful in a broad range of relevant areas. I have applied the new approach for generating polyubiquitin (polyUb) molecules. Ubiquitin (Ub), which consists of 76 amino acids, can be attached to a specific lysine residue of a target protein, where the isopeptide bond between the carboxyl (C) terminus of Gly76 and the ε-amine of the lysine side chain is formed by a series of enzymes (E1, E2, and E3). A distinct feature of ubiquitylation is that another Ub molecule can be added to the attached Ub. The second isopeptide bond is formed between the C terminus of the distal Ub, and the seven lysine residues (6, 11, 27, 29, 33, 48, and 63) or the α-amino terminus (Met1) of the proximal Ub molecule.[8] The connection of another Ub molecule can continue, thereby forming polyUb. The topology of the linkage in polyUb acts as a signal in defining the destiny of the protein that carries polyUb. For instance, the Lys48-linked polyUb is a tag for proteasome to degrade the tagged protein. Intriguingly, recent research studies have reported the existence of branched chains and mixed linkages of polyUb.[8] Therefore, to understand its precise roles and control its function, the studies on the structure and dynamics of polyUb are indispensable. In particular, a biophysical technique that can reflect the flexibility of the biomolecules will be a more appropriate tool.

Experimental Methods

Preparation of CYANA library for polyubiquitin- In-house-written Python scripts prepared the new residues compatible for CYANA calculation from a known structure. The scripts used the functions of PyMOL and CYANA to add additional atoms for connection, after aligning each residue in the same direction. For preparing a unit for polyUb, the first structure of the NMR ensemble (PDB ID: 1D3Z) was used. Eight new residues (UB01, UB06, UB11, UB27, UB29, UB33, UB48, and UB63) were prepared for the linkages with the α-amine group of Met1, and the ε-amine groups of Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63, respectively. The new residue did not contain any torsion angles. Once the coordinates comprising residues 1–70 were extracted from 1D3Z, two atoms (C and O) were added for connection with

Table 1. CYANA library for preparing Lys48-linked polyubiquitin



```
RESIDUE   UB48     1 1131     3 1130
   1 PSI      0    0    0.0000 1115 1116 1129 1131    0
   1 C    C_BYL    0    0.0000    0.0000    0.0000    0.0000    0    0    0    0    0
   2 O    O_BYL    0    0.0000    0.0674    0.0000   -1.2285    0    0    0    0    0
   3 N    N_AMI    0    0.0000    3.2271    0.0000    0.0000  765  776    0    0    0
   4 N2   N_AMI    0    0.0000   16.9268   12.7342   -5.2473    5   12   13   14    0
   5 C2   C_ALI    0    0.0000   16.0810   12.3048   -6.3984    4    6    8   15    0
   6 C3   C_BYL    0    0.0000   14.8376   13.1823   -6.4723    5    7   23    0    0
   7 O2   O_BYL    0    0.0000   14.4027   13.7415   -5.4650    6  281    0    0    0
   8 C4   C_ALI    0    0.0000   15.6783   10.8400   -6.2159    5    9   16   17    0
   9 C5   C_ALI    0    0.0000   14.9494   10.6636   -4.8850    8   18   19    0    0
  10 S1   S_OXY    0    0.0000   14.6099    8.9039   -4.6188    0    0    0    0    0
  11 C6   C_ALI    0    0.0000   13.0198    9.0956   -3.7789   20   21   22    0    0
  12 H1   H_AMI    0    0.0000   17.4222   13.6144   -5.4912    4   13   14    0    0
  13 H2   H_AMI    0    0.0000   17.6232   11.9911   -5.0330    4   12   14    0    0
  14 H3   H_OXY    0    0.0000   16.3249   12.8955   -4.4160    4   12   13  277    0
  15 H4   H_ALI    0    0.0000   16.6449   12.4095   -7.3138    5    0    0    0    0
  16 H5   H_ALI    0    0.0000   15.0250   10.5438   -7.0235    8   17    0    0    0
  17 H6   H_ALI    0    0.0000   16.5593   10.2204   -6.2228    8   16    0    0    0
  18 H7   H_ALI    0    0.0000   15.5704   11.0347   -4.0829    9   19    0    0    0
  19 H8   H_ALI    0    0.0000   14.0198   11.2131   -4.9058    9   18    0    0    0
 ...
1111 H561 H_ALI    0    0.0000   -1.0301   15.7639    0.5945 1102 1109 1110    0    0
1112 H562 H_ALI    0    0.0000   -1.4150   19.2213   -0.5417 1103 1113 1114    0    0
1113 H563 H_ALI    0    0.0000   -1.7405   18.8643   -2.2377 1103 1112 1114    0    0
1114 H564 H_ALI    0    0.0000   -0.1194   18.6315   -1.5820 1103 1112 1113    0    0
1115 N93  N_AMI    0    0.0000   -3.9234   14.7865   -0.6038 1098 1116 1120    0    0
1116 C347 C_ALI    0    0.0000   -5.3302   14.9254   -0.2453 1115 1117 1121 1129    0
1117 C348 C_ALI    0    0.0000   -5.8598   13.5681    0.2437 1116 1118 1119 1122    0
1118 C349 C_ALI    0    0.0000   -7.0467   13.7336    1.2032 1117 1123 1124 1125    0
1119 C350 C_ALI    0    0.0000   -6.3025   12.7254   -0.9426 1117 1126 1127 1128    0
1120 H565 H_AMI    0    0.0000   -3.3069   14.5454    0.1077 1115    0    0    0    0
1121 H566 H_ALI    0    0.0000   -5.8853   15.2567   -1.1009 1116    0    0    0    0
1122 H567 H_ALI    0    0.0000   -5.0555   13.0591    0.7266 1117    0    0    0    0
1123 H568 H_ALI    0    0.0000   -7.5030   12.7702    1.3784 1118 1124 1125    0    0
1124 H569 H_ALI    0    0.0000   -7.7732   14.4004    0.7648 1118 1123 1125    0    0
1125 H570 H_ALI    0    0.0000   -6.7039   14.1434    2.1414 1118 1123 1124    0    0
1126 H571 H_ALI    0    0.0000   -7.2105   13.1329   -1.3553 1119 1127 1128    0    0
1127 H572 H_ALI    0    0.0000   -6.4754   11.7141   -0.6060 1119 1126 1128    0    0
1128 H573 H_ALI    0    0.0000   -5.5257   12.7293   -1.6906 1119 1126 1127    0    0
1129 C    C_BYL    0    0.0000   -5.4262   15.9896    0.8402 1116 1130    0    0    0
1130 O    O_BYL    0    0.0000   -4.5409   16.1228    1.6829 1129    0    0    0    0
1131 N    N_AMI    0    0.0000   -6.5057   16.7511    0.7923    0    0    0    0    0
```

Table 2. CYANA sequence for preparing Lys48-linked decaubiquitin

```
UB48 1
LEU  71
ARG  72
LEU  73
ARG  74
GLY  75
GLY  76

UB48 101
LEU  171
ARG  172
LEU  173
ARG  174
GLY  175
GLY  176

UB48 201
LEU  271
ARG  272
LEU  273
ARG  274
GLY  275
GLY  276

.....
```

the previous residue. The coordinates of C and O were adjusted to form an isopeptide bond with the NZ atom of a specific lysine of the rigidified Ub molecule. In UB01, the coordinates of C and O were added to form typical peptide bond with the α-amine group of Met1. For the sequential connection with residues 71–76 in Ub, the coordinate of the N atom of residue 71 was added to the rigid bodies.

Calculation of polyubiquitins- To test whether the approach works, the eight decaUbs consisting of UB01, UB06, UB11, UB27, UB29, UB33, UB48, and UB63 were calculated. For each polyUb, 1000 structures were generated under 20,000 steps of the torsion angle dynamics by CYANA (ver. 3.0) with or without distance restraints, and 100 structures showing the lowest target functions were selected. ALIS-NMR (Automated pLatform for Integrative Structure calculation for NMR) automated the procedure by distributing the runs into a Linux cluster for parallel computation. Structures of Lys48-linked decaUb with artificial distance restraints were also calculated. The distance restraints consisted of ambiguous distances between the Cα-Cα pairs of Ile44 and Val70, with 7 Å upper limits.

Results

Generation of polyubiquitins with rigid bodies- Table 1 lists the CYANA library of Ub-derived residues for the Lys48-linked polyUb as an example. Table 2 shows the CYANA sequence for generating Lys48-linked decaUb. Figure 1 displays the representative structures for head-to-tail-, Lys48-, and Lys63-linked decaUb molecules, demonstrating the soundness of the approach. In conventional nomenclature of polyUb, the units are named as “distal” and “proximal” according to their distances from the tagged protein. In the current approach, a smaller number corresponds to a more distal unit. The linkage with the target protein occurs at unit 10.

Distribution of the radius of gyration in each polyubiquitin- To determine the distribution in each polymer, we calculated the radius of gyration (R_{gyr}) and carried out histogram analyses to visualize its distributions (figure 2). One may argue that it would be less meaningful to calculate the spatial distribution of decaUb molecules, because the CYANA target function does not contain sophisticated non-bond energies between atoms. Nevertheless, it is noteworthy that the distribution of R_{gyr} varied depending on chain types. The most deviated pattern was found in Lys6-linked decaUb, where both the

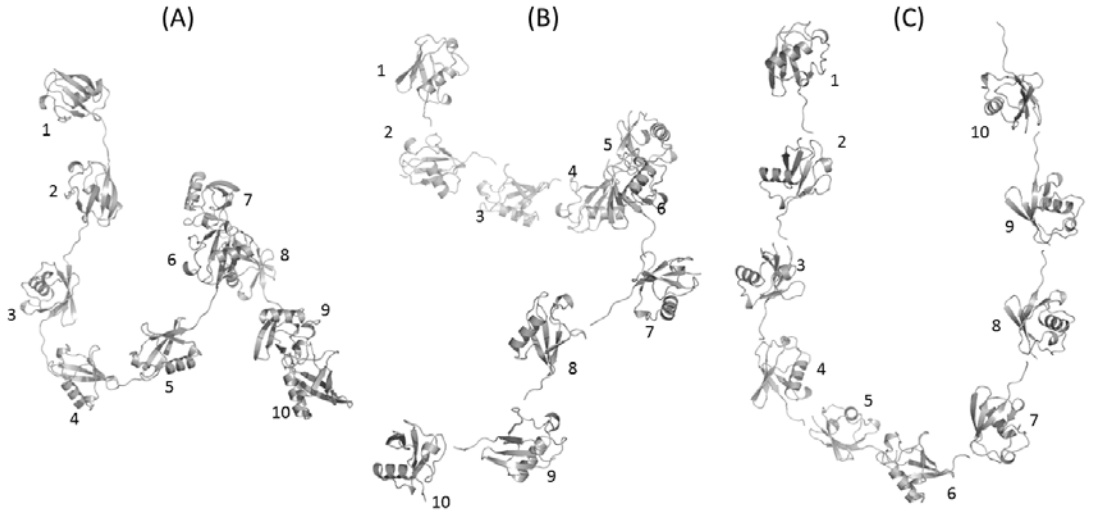


Figure 1. Representative structures of decaubiquitins. (A) Head-to-tail-, (B) Lys48-, and (C) Lys63-linked decaubiquitin molecules are drawn. A smaller unit number corresponds to a more distal unit. The linkage with the target protein is assumed to occur at unit 10.

values of the averaged R_{gyr} and the standard deviation (50.5 ± 8.4 Å) were the lowest compared with the other polymers. Several studies have indicated that head-to-tail-linked and Lys63-linked polyUbs behave in similar manners for recognizing Ub-interacting motifs. Interestingly, these decaUb polymers shared a similar R_{gyr} distribution with each other, of 81.5 ± 15.8 and 84.0 ± 14.6 for head-to-tail- and Lys63-linked decaUb, respectively.

Restrained calculation of Lys48-linked decaubiquitin- The conformation of Lys48-linked decaUb with artificial distance restraints was also simulated. Lys48-linked polyUb is a signal that leads the tagged protein into the degradation pathway. Compared with other polyUbs, the conformation of Lys48-linked polyUb has been more studied by biophysical methods, including X-ray crystallography and NMR.[9-12] Several X-ray structures of tetraUb have been reported, although disagreement about the relative positions between units remains.[9,12] The pH is reported to influence the conformation.[11] In all cases, however, the results have repeatedly indicated that the Ile44 residue in one unit is proximal in space to the Val70 residue of another unit of polyUb. Because a quaternary structure comprising more than tetraUb has not been reported, it is meaningful to shape the overall fold of decaUb. I confined the inter-unit distance in an ambiguous way, where the Cα atom of Ile44 in one unit was located within 7 Å of the Cα atom of Val70 in any of the other 9 units. The resulting 100 structures revealed very compact quaternary structures, with 34.5 ± 3.7 as the R_{gyr} value (figure 3). Individual inspection of the structures allowed a rough classification of the conformers into two groups. In one group, the very first unit is close to the fourth unit (figure 3A). In the other group, the first and the second units are close (figure 3B). The latter group is the one observed in the X-ray crystal structure of tetraUb in neutral pH, although the relative positions to the third and the fourth units are somewhat deviated.

Efficiency of calculation in head-to-tail-linked polyubiquitin- The efficiency in structure calculation declines as the size of a protein grows. The torsion angle dynamics that CYANA employs is more advantageous in searching conformational space than conventional Cartesian dynamics, because the

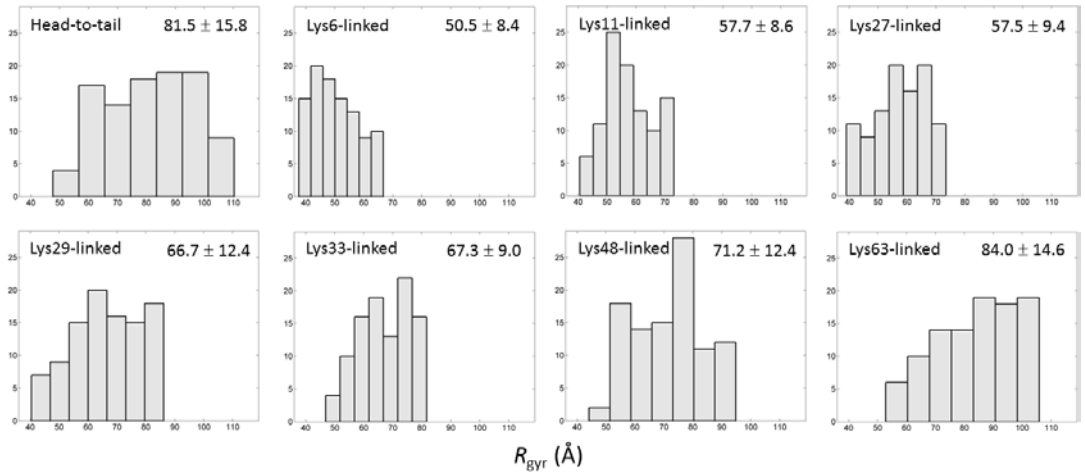


Figure 2. Histogram analysis of the radius of gyration (R_{gyr}) in 8 decaubiquitin molecules. Each histogram was prepared with 100 decaubiquitin structures. The structure types and the values of the mean and standard deviation for R_{gyr} are written in each panel. All graphs are arranged to have the identical ranges in both *x*- and *y*-axes for clear comparison.

degrees of freedom to consider during a time step are reduced. My approach can decrease the degrees of freedom even more by introducing rigid bodies. To test whether this approach is indeed efficient compared with typical CYANA calculation, the number of Ub units was increased by 5 in head-to-tail-linked polyUb. In typical CYANA calculation, to rigidify the Ub, experimental distance restraints that were deposited along with the NMR structure of Ub (PDB ID: 1D3Z) were used. Although typical CYANA run could generate up to 25 polyUb successfully, it could not manage to calculate 30 polyUb, reporting errors related with the memory capacity. On the other hands, the rigid body method could simulate the structures of 30 Ubs, despite the total number of amino acids being as large as 2280 (= 76 × 30).

Generation of mixed chain tetraubiquitin- Next, structures of polyUb consisting of mixed chains were generated. Recent studies have indicated that mixed chains of polyUb play an important role in nuclear factor kappa B signaling or protein trafficking. Although the detailed biological roles of the chains are still unknown, the structural image will be helpful for understanding their functions. TetraUb comprising UB01-UB11-UB63-UB48 was prepared, because the combination has been identified by experiments. The resulting 100 conformers were more diverse compared with those of Lys48-liked polyUb, having an R_{gyr} value of 32.9 ± 3.5. Because there is no information on the spatial proximity, only the initial conformers were simulated (figure 4). The studies on the inter-unit contacts or overall shape can be combined to refine the structures for reflecting the real structure.

Discussion

Conventional structural biology has focused on elucidating the structures of monomers or complexes that do not include flexible linker regions. The linker region can hinder crystallization, and the electron density in the linker part is often invisible in the diffraction. In NMR spectroscopy, the inclusion of the regions can increase the overall tumbling time, leading to line broadening of the NMR signals. The linker regions are nevertheless important for the function of a protein. The length and flexibility of the linker can confine interdomain positions and motions. Post-translational modification at the linker parts

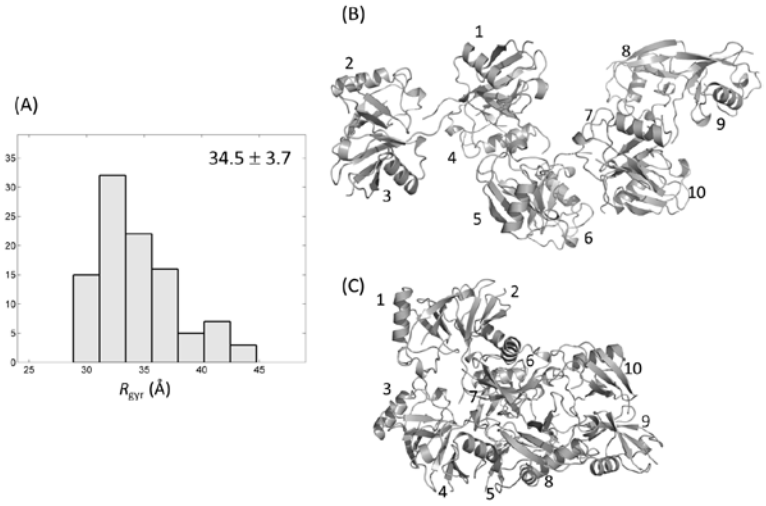


Figure 3. Ambiguous distance-restrained Lys48-linked decaubiquitin. (A) Histogram analysis of the radius of gyration (R_{gyr}). Conformations with proximity for 1–4 units (B) and 1–2 units (C) are represented.

occasionally endows the protein with new properties that are closely coupled with the cellular signaling cascade. Therefore, visualization of the structural change of a whole protein will be more beneficial for understanding biological phenomena. My approach can readily be extended to shaping the whole structure of a protein, with less hindered by the large size.

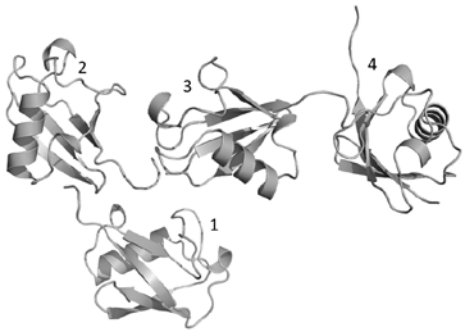


Figure 4. Representative structure of mixed chain tetraubiquitin. The numbers 1 through 4 correspond to the UB01, UB11, UB63, and UB48 units, respectively.

It should be noted that the structures in the current approach are the starting points for employing sophisticated methods. The conformers in this study were calculated only under the constraints of intermolecular repulsions without applying van der Waals and electrostatic potentials. Refinements with atomistic force fields in implicit or explicit solvents can enhance the structural accuracy as well as precision.[13-15] Compatible codes of the latest graphics processing units will accelerate the calculation synergistically.[16-18] Here, the experimental restraints from cryo-EM, SAXS, and NMR can be employed as well. In any case, a reliable starting structure is important for amplifying relevant conformations. Instead of producing trajectories from a single conformation, more prevalent tactics for modeling structures would be to start from multiple structures and to find clustered structures with lower energies as shown in the Rosetta algorithm. This study will be a meaningful addition for such purposes.

Acknowledgements

This work was supported by a National Research Foundation grant funded by the Korean government (2014R1A1A2059460).

References

1. K. Wüthrich, *NMR of Proteins and Nucleic Acids*; Wiley: New York (1986)
2. T. Herrmann, P. Guntert, and K. Wuthrich, *J. Mol. Biol*. **319**, 209 (2002)
3. J. G. Jee, and P. Guntert, *J. Struct. Funct. Genomics* **4**, 179 (2003)
4. C. Dominguez, R. Boelens, and A. M. Bonvin, *J. Am. Chem. Soc.* **125**, 1731 (2003)
5. T. A. Hopf, C. P. Scharfe, J. P. Rodrigues, A. G. Green, O. Kohlbacher, C. Sander, A. M. Bonvin, and D. S. Marks, *Elife* **3**:e03430 (2014)
6. P. Güntert, C. Mumenthaler, and K. Wüthrich, *J. Mol. Biol*. **273**, 283 (1997)
7. M. P. Williamson, and C. J. Craven, *J. Biomol. NMR* **43**, 131 (2009)
8. D. Komander, and M. Rape, *Annu. Rev. Biochem*. **81**, 203 (2012)
9. E. J. Eddins, R. Varadan, D. Fushman, C. M. Pickart, and C. Wolberger, *J. Mol. Biol.* **367**, 204 (2007)
10. T. Tenno, K. Fujiwara, H. Tochio, K. Iwai, E. H. Morita, H. Hayashi, S. Murata, H. Hiroaki, M. Sato, K. Tanaka, and M. Shirakawa, *Genes Cells* **9**, 865 (2004)
11. R. Varadan, O. Walker, C. Pickart, and D. Fushman, *J. Mol. Biol.* **324**, 637 (2002)
12. W. J. Cook, L.C. Jeffrey, E. Kasperek, and C. M. Pickart, *J. Mol. Biol*. **236**, 601 (1994)
13. J. G. Jee, *Bull. Korean Chem. Soc*. **35**, 1944 (2014)
14. J. G. Jee, *J. Kor. Magn. Reson. Soc*. **18**, 24 (2014)
15. J. G. Jee, *J. Kor. Magn. Reson. Soc*. **17**, 11 (2013)
16. A. W. Gotz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, *J. Chem. Theory. Comput*. **8**, 1542 (2012)
17. R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, and R.C. Walker, *J. Chem. Theory. Comput*. **9**, 3878 (2013)
18. J. G. Jee, *J. Kor. Magn. Reson. Soc*. **18**, 69 (2014)