DOI 10.5012/bkcs.2010.31.9.2717

Unambiguous Determination of Intermolecular Hydrogen Bond of NMR Structure by Molecular Dynamics Refinement Using All-Atom Force Field and Implicit Solvent Model

JunGoo Jee

Center for Priority Areas, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan
E-mail: jee-jungoo@tmu.ac.jp
Received June 7, 2010, Accepted July 29, 2010

Key Words: NMR structure, Intermolecular hydrogen bond, Ubiquitin

The structure determination of a biomolecule by NMR depends heavily on the distance restraints derived by the NOE cross peaks that are observed between two protons within 6 Å through space.[1] Therefore, the existence of the NOE peaks and their correct assignments to two corresponding protons are essential for an accurate and precise structure determination. Recent developments of NOE assignment and calculation algorithms have enabled the determination of protein 3D structures without any manual interpretation, provided chemical shifts are assigned in most atoms and sufficient NOE peaks exist.[2,3] Along with these advances, the necessity of determining complicated structures such as complexes is increasing. In order to calculate protein-protein or protein-DNA complex structures by NMR, it is necessary to have sufficient intermolecular distance restraints. However, if a system undergoes fast exchange on the NMR time scale, it is often problematic to gain sufficient NOE-based intermolecular distance restraints. In particular, collecting distance information for the regions where hydrophilic or electrostatic interactions are dominant is more difficult, occasionally resulting in inaccurate and imprecise structures. Many hydrophilic interactions are formed by side-chain moieties that contain labile protons, from which it is often difficult to observe NOE signals.

Molecular dynamics refinement (hereafter, MD refinement), which employs an all-atom force field and the generalized Born implicit solvent model, has improved the quality of NMR structures, by being applied in the final stage of calculation.[4] MD refinement is more effective for a region where the experimental restraints are insufficient. Only with the limited number of restraints, MD refinement enabled NMR structures to get close to the native-like folds.[5,6] However, besides the apparent improvement revealed by the parameters for the accuracy and precision of structures, cases where the refinement leads to a direct understanding of confusing roles are limited in number. In this paper, I report the unambiguous determination of an intermolecular hydrogen bond in the complex structure between ubiquitin and ubiquitin-interacting motif (UIM) by MD refinement.

Ubiquitin (Ub), a protein consisting of 76 amino acids, is a reversible tag for post-translational modification. It can be attached to the side-chain of a lysine residue on a protein surface by forming an isopeptide bond with its C-terminal glycine through a series of enzyme reactions. Its surface-exposed lysines can also be linked with another Ub, producing poly-Ub.[7,8] Ubiquitination plays crucial roles in various cellular signaling pathways, including the well-known proteasome-dependent degradation pathway. The tagged Ub is recognized by various downstream regulators containing ubiquitin-binding domains (UBDs).[9,10] The surface around Ile-44 of Ub is the main site for interacting with the UBDs, whereas each UBD has a unique motif for binding to Ub. A noticeable and general feature of the complex between mono-Ub and UBDs is the weak binding affinity. Ub and UBD complexes display the equilibrium dissociation constant, Kd, in the range of μM to mM.[11] Intriguingly, some proteins contain a ubiquitin-like domain (UBL) inside their sequences as a domain, which has a sequence composition and properties that are very similar to Ub, although it cannot generate poly-Ub due to the lack of C-terminal-free glycine.

The UIM and ubiquitin-associated (UBA) domain are the two most abundant UBDs. The UIM contains the -ϕ-x-x-Ala-x-x-x-Ser-x-x-Ac- motif, where ϕ and Ac denote large hydrophobic and acidic residues, respectively. Here the serine is irreplaceable. The substitution of alanine for the serine leads to a substantial decrement in the affinity for Ub,[12] indicating the critical role of the side-chain hydroxyl group. On the basis of the mutation data and the electrostatic property of the hydroxyl group, the existence of a hydrogen bond, where the hydroxyl group of the serine takes part, can reasonably be presumed. The proper definition of a hydrogen bond requires the coordinates of a proton as well as heavy atoms. However, the explicit determination of the proton position in X-ray and NMR structures is not straightforward. Protons are invisible with X-ray diffraction. The use of NMR to observe signals from labile side-chain protons is often ineffective, because the exchange rate with a solvent can unfavorably influence the NMR line-shape.

NMR has been applied to determine various complex structures between Ub and UBD, including two mono-Ub/UIM complexes − Ub/Vps20-UIM (PDB ID: 1Q0W) and Ub/S5a-UIM2 (1YX6).[13,14] The deposited ensembles consist of 20 and 18 structures for Ub/Vps20-UIM and Ub/S5a-UIM2, respectively. These structures have good precisions, revealing respective root mean square deviation (RMSD) values of 0.49 and 0.18 Å for the 1 - 70 residues of Ub. Despite the finenesses of their structures, the intermolecular interaction *via* the conserved serine in the UIM region is not conclusive. In Ub/Vps20-UIM, six of the 20 structures display a hydrogen bond between the

Table 1. Statistics of ensemble structures

		PDB	CNS	CNW	AMD
Ub/Vps20-UIM (1Q0W)	Å	0.49 (0.98)	0.51 (0.98)	0.56 (1.02)	0.34 (0.80)
	%	87.3/9.8/2.0/0.9	85.3/13.0/1.0/0.7	88.2/10.2/0.3/1.2	94.4/5.1/0.2/0.4
Ub/S5a-UIM2 (1YX6)	Å	0.18 (0.61)	0.49 (0.87)	0.44 (0.85)	0.26 (0.62)
	%	80.1/18.6/1.2/0.1	66.8/28.6/3.0/1.5	78.8/18.2/1.4/1.5	91.3/7.3/0.8/0.7

PDB indicates the ensemble in the RCSB database, CNS represents the structures calculated by CNS "anneal.inp" protocol, CNW is those refined by the ARIA water refinement, and AMD stands for those by AMBER refinement. The upper row contains the values of backbone (heavy atoms) RMSD in the 1 - 70 residues of Ub. The values in the bottom row are the percentages in the most favored, favored, allowed, and disallowed regions of the Ramachandran analysis for all the regions. The numbers of structures in the PDB ensembles are 20 and 18 for Ub/Vps20-UIM and Ub/S5a-UIM2, respectively, and 30 for all the others.

Oγ of Ser-270 in UIM and the amide H of Gly-47 of Ub. In Ub/S5a-UIM2, only three of the 18 structures show the hydrogen bond between the Oγ of Ser-294 in UIM and the H of Gly-47 in Ub. These low numbers of structures with hydrogen bonds in an ensemble prevent us from drawing conclusions about the existence as well as the role of the hydroxyl group from the serine.

Motivated by previous studies in which MD refinement has improved the NMR structures,[15] I recalculated the complex structures of Ub/Vps20-UIM and Ub/S5a-UIM2 with the deposited and the artificially generated restraints, respectively. Three types of calculations were performed for each. The first included conventional simulated annealing in the torsion angle space using the "Crystallography & NMR System" (CNS) package.[16] The structures were further refined with the ARIA water refinement (CNW)[17] and AMBER MD (AMD) refinement methods,[18] respectively. In the AMD calculation, of the 300 structures that were calculated by a CNS run, 100 structures were further refined. Finally, the 30 structures that showed the lowest overall energies and did not reveal significant violations against input restraints (< 0.5 Å for distance and < 5° for torsion angle restraints) were selected as a final ensemble in each calculation. For the Ub/S5a-UIM2 structure, where experimental restraints were not deposited together, synthetic distance restraints were generated. The synthetic restraints were prepared for two protons that were located within 5 Å in more than 75 % of the original structures, and were separated by more than one residue in primary sequence ($|i - j| > 1$). The upper distance limits were loosened by adding 1 Å to the mean values of the distances.

Table 1 reveals the accuracy and precision parameters in the final structures. It should first be noted that the structures by the synthetic restraints did not deviate from the original ones. The precision that is revealed by the backbone RMSD value became worse (~0.3 Å) than the original one in the CNS calculation of Ub/S5a-UIM2. However, the pairwise RMSD value between two Ub regions in the deposited and the CNS structures of Ub/S5a-UIM2 was 0.69 ± 0.07 Å, which was smaller than 0.97 ± 0.14 Å for Ub/Vps20-UIM, indicating the structural closeness in Ub/S5a-UIM2 cases. Second, as reported previously,[17] CNW and AMD refinements improved the accuracies of the structures, which were representatively reflected by the percentile of the most favored region in a Ramachandran analysis. Third, the AMD results also had better precision. Interestingly, it showed better RMSD values than the deposited structure in the case of Ub/Vps20-UIM, which demonstrated the usefulness of MD refinement. Fourth, the most dramatic improvement was found in the hydrogen bond *via* the hydroxyl moiety of the serine residues in UIM (Table 2). Both the CNS and CNW results revealed a hydrogen bond only in the small portion of the final ensemble. Moreover, the CNS structures in Ub/Vps20-UIM showed different hydrogen bond partners compared with other cases. However, the AMD structures revealed 27 and 19 hydrogen bonds between the Oγ of the serine and the H of Gly-47 for Ub/Vps20-UIM and Ub/S5a-UIM2, respectively. Ratios of 0.90 (27/30) and 0.63 (19/30) would probably be the acceptable levels to clarify the existence of a hydrogen bond in NMR structures.

Table 2. Number of hydrogen bonds between Oγ of serine in UIM and amide proton of Gly-47 in Ub.

	PDB	CNS	CNW	AMD
Ub/Vps20-UIM	6	5[a]	5	27
Ub/S5a-UIM2	3	3	2	19

[a]Instead of the amide proton Gly-47 in Ub, the hydrogen bond through the amide proton of Ala-46 was identified.

Visual inspection also revealed the better outcomes by AMD refinement well. AMD structures presented the superior convergence in Ub/Vps20-UIM (Fig. 1). While the backbone convergence was comparable to the deposited coordinates, AMD structures of Ub/S5a-UIM2 revealed the better geometries in the side-chains (Fig. 1). We previously reported the NMR structure between the UBL domain of hHR23B and S5a-UIM that was refined by AMD.[19] It revealed the identical pattern − a hydrogen bond between the amide proton of Gly-50 that corresponds to the Gly-47 in Ub and the Oγ of the serine in UIM − implying a considerable similarity between the Ub and UBL of hHR23B in recognizing UIM. Most importantly, the hydrogen bond patterns in Ub/Vps20-UIM and Ub/S5a-UIM2 were in total agreement with the high-resolution X-ray structure (PDB ID: 2D3G) that was determined with a resolution of 1.7Å.[20]

The refined structures may make it possible to explain the pH dependence in the interaction between Ub and UIM. We have reported that an acidic condition weakens the binding affinity.[19] Compared with the Kd (~10 μM) at pH 8, the value of Kd at pH 5 increases about four-fold (~40 μM) in the complex between Ub and the UIM of S5a. We have suggested that His-68

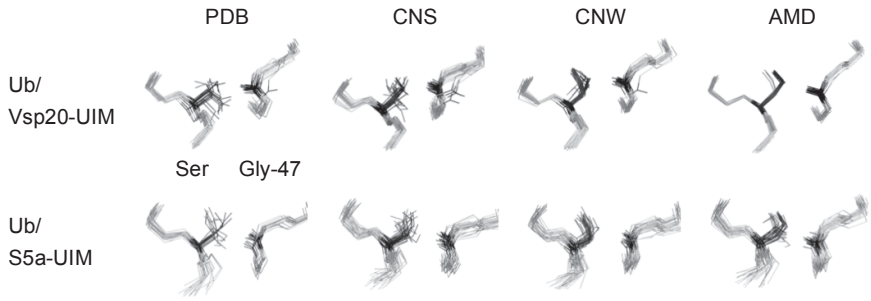

Figure 1. Overlaid structures of Ub/Vps20-UIM and Ub/S5a-UIM2. For the overlays, all the heavy atoms in the 46 - 48 residues for Ub, and the serine and its ±1 residue for UIM, the residues 269 - 271 and 293 - 295 for Vps20-UIM and S5a-UIM2, respectively, were used. The backbone atoms of these six residues are shown in gray. The side-chain heavy and Hγ atoms of Ser-270 and Ser-294 in Vps20-UIM and S5a-UIM2, respectively, and the amide atoms in Gly-47 of Ub are shown in black. All the ensembles are aligned to have the same directions. All the figures were generated by PyMOL (http://www.pymol.org).

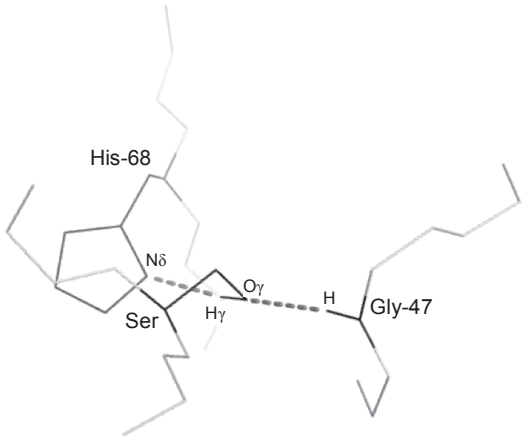

Figure 2. Hydrogen bond network between Gly-47 and His-68 in Ub and the serine in UIM. The first structure in the AMD ensemble of Ub/S5a-UIM2 was used. This coordinate has the same orientation as those in Fig. 1. The observed hydrogen bonds between Gly-47 and His-68 in Ub and Ser-294 in UIM are shown using dashed lines. For clarity, only heavy atoms are drawn except two protons, H of Gly-47 in Ub and Hγ of Ser-294 in S5a-UIM.

of Ub is involved in the interaction based on the following two observations. First, the pattern of Kd changes was similar to that of the Hε1 chemical shift changes in His-68 with varying pH. Second, the H68V mutant did not show any pH dependence. Nonetheless, the details of the dependence have been little understood. Remarkably, the Ub/S5a-UIM2 structures by AMD reveal a hydrogen bond between the Nδ1 of His-68 in Ub and the Hγ of Ser-294 in UIM2 in 12 (of 30) structures. To efficiently function as a hydrogen bond acceptor, the Nδ1 of the imidazole group should not be protonated, implying that a basic condition is more favorable. This is fairly consistent with our previous observations (Fig. 2). But even so, it will not lead to the unanimous explanation for the pH dependence under currently available data, because no hydrogen bond between His-68 in Ub and Ser-270 in UIM is found in the refined Ub/Vps20-UIM structures. Several NOE restraints from His-68 in Ub/Vps20-UIM complexes restrict the local geometry, which was unalterable by AMD. The values of the χ2 angle in His-68 of the AMD refined Ub/Vps20-UIM structure is -99.9 ± 6.5^{o}, which is in contrast to the value of 90.1 ± 7.9^{o} in His-68 of Ub/S5a-UIM2. Despite the discrepancy, however, considering the current situation that there is no appealing model, the hydrogen bond *via* His-68 in the refined structures will be the meaningful data that guides future studies with experimental and computational tools.

In conclusion, it has been shown that AMD refinement is very useful for defining an intermolecular hydrogen bond in NMR structure calculation. The refined structure also provides a clue for explaining the pH dependence in Ub and UIM complexes. As reported by Choi *et al*.,[21] serine-mediated hydrogen bonds are the third most populated hydrogen bonds found in protein-protein intermolecular interactions, after the backbone-backbone and backbone-aspartate ones. The abundance imposes the requirement of an method to determine the interface of protein-protein complexes. The precise geometry is particularly important in the complex structures between Ub and UBDs. Ub recognizes various targets with the same surface, where both hydrophobic and hydrophobic interactions are involved. Hence, the details of the hydrophilic interactions are necessary to find the common binding modes.

Materials and Methods

The standard protocol, "anneal.inp", was used for the CNS calculation. The ARIA water refinement protocol was adapted for the CNW run. The AMBER software package (version 9) with an ff99SB force field and generalized Born implicit model was applied for AMD.[22] The AMBER calculation consists of three stages: 1500 steps of energy minimization, 20-ps molecular dynamics, and 1500 steps of energy minimization. The force constants for distance and torsion-angle restraints were 50 kcal·mol^{-1}·Å^{-2} and 200 kcal·mol^{-1}·rad^{-2}, respectively. The distance restraints for the structure calculations of Ub/Vps20-UIM consisted of 1044, 394, 258, 364, and 75 restraints for intra ($|i - j| = 0$), sequential ($|i - j| = 1$), medium ($1 < |i - j| < 5$), long ($|i - j| > 4$), and inter-molecular ranges, respectively. Torsion angles were restrained in 51 ϕ and 51 φ angles. Ub/S5a-UIM2 calculations used only distance restraints, which comprised 652 medium ($1 < |i - j| < 5$), 1181 long ($|i - j| > 4$), and 176 inter-molecular restraints. The RMSD values, torsion angles, and hydrogen bond patterns were analyzed using MOMOL.[23]

Acknowledgments. I thank Prof. Masahiro Shirakawa for giving me the opportunity to participate in his project.

References

1. Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
2. Altieri, A. S.; Byrd, R. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 547.
3. Jee, J.; Güntert, P. *J. Struct. Funct. Genomics* **2003**, *4*, 179.
4. Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317.
5. Chen, J.; Won, H. S.; Im, W.; Dyson, H. J.; Brooks, C. L., III *J. Biomol. NMR* **2005**, *31*, 59.
6. Chen, J.; Im, W.; Brooks, C. L., III *J. Am. Chem. Soc.* **2004**, *126*, 16038.
7. Pickart, C. M. *Annu. Rev. Biochem.* **2001**, *70*, 503.
8. Hershko, A.; Ciechanover, A. *Annu. Rev. Biochem.* **1998**, *67*, 425.
9. Dikic, I.; Wakatsuki, S.; Walters, K. J. *Nat. Rev. Mol. Cell. Biol.* **2009**, *10*, 659.
10. Hurley, J. H.; Lee, S.; Prag, G. *Biochem. J.* **2006**, *399*, 361.
11. Lee, S.; Tsai, Y. C.; Mattera, R.; Smith, W. J.; Kostelansky, M. S.; Weissman, A. M.; Bonifacino, J. S.; Hurley, J. H. *Nat. Struct. Mol. Biol.* **2006**, *13*, 264.
12. Young, P.; Deveraux, Q.; Beal, R. E.; Pickart, C. M.; Rechsteiner, M. *J. Biol. Chem.* **1998**, *273*, 5461.
13. Swanson, K. A.; Kang, R. S.; Stamenova, S. D.; Hicke, L.; Radhakrishnan, I. *EMBO J.* **2003**, *22*, 4597.
14. Mueller, T. D.; Feigon, J. *EMBO J.* **2003**, *22*, 4634.
15. Jee, J.; Ahn, H. C. *Bull. Korean Chem. Soc.* **2009**, *30*, 1139.
16. Brunger, 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.
17. Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. *Proteins* **2003**, *50*, 496.
18. Case, D. A.; Cheatham, T. E., III.; 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.
19. Fujiwara, K.; Tenno, T.; Sugasawa, K.; Jee, J. G.; Ohki, I.; Kojima, C.; Tochio, H.; Hiroaki, H.; Hanaoka, F.; Shirakawa, M. *J. Biol. Chem.* **2004**, *279*, 4760.
20. Hirano, S.; Kawasaki, M.; Ura, H.; Kato, R.; Raiborg, C.; Stenmark, H.; Wakatsuki, S. *Nat. Struct. Mol. Biol.* **2006**, *13*, 272.
21. Choi, H.; Kang, H.; Park, H. *J. Comput. Chem.* **2010**, *31*, 897.
22. Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.
23. Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph.* **1996**, *14*, 51.