Validity of Using the Radius of Gyration as a Restraint in NMR Protein Structure Determination

Xuemei Huang and Robert Powers*

Department of Biological Chemistry
Wyeth Research, 85 Bolton Street
Cambridge, Massachusetts 02140

Received November 8, 2000
Revised Manuscript Received February 5, 2001

NMR has been established as a powerful method for determining the structure of proteins and protein−ligand complexes.[1] Substitution of nonexchangeable protons with deuterons has effectively increased the size limit of a protein amenable for analysis by NMR.[2−4] However, the resulting precision and accuracy of a structure determined by NMR for a deuterated protein is relatively low due to the decrease in proton derived distance information.[5] Measurement of residual dipolar couplings in partially oriented proteins dissolved in a lipid bicelle solution[6] has demonstrated a tremendous impact on the accuracy and precision of structures calculated with minimal NOE restraints.[7,8] Other approaches have been implemented to increase the number of non-NOE based restraints in an effort to improve the quality of NMR structures of large molecular weight proteins where NOE information is minimal.[9] These methods have included the direct refinement against chemical shifts,[9,10] coupling constants,[11] and a conformational database potential.[12]

A recent study has shown that incorporating the radius of gyration (Rg) as a structure restraint target function can improve the packing and accuracy of a structure calculated by NMR.[13] It has also been proposed that NMR structures tend to be poorly packed relative to X-ray structures.[13] Since the overall fold of a protein determined by NMR is primarily a result of the experimental NOE distance restraints, the resulting protein packing is presumably determined from the balance of the NOE distance restraints and the repulsive forces from the dynamics force field function. Thus, if the number of repulsive interactions from the dynamics force field overwhelms the number of experimental restraints the resulting structure may be biased toward an expanded structure. The radius of gyration (Rg) of a group of atoms is defined as the root-mean-square distance from each atom of the molecule to their centroid

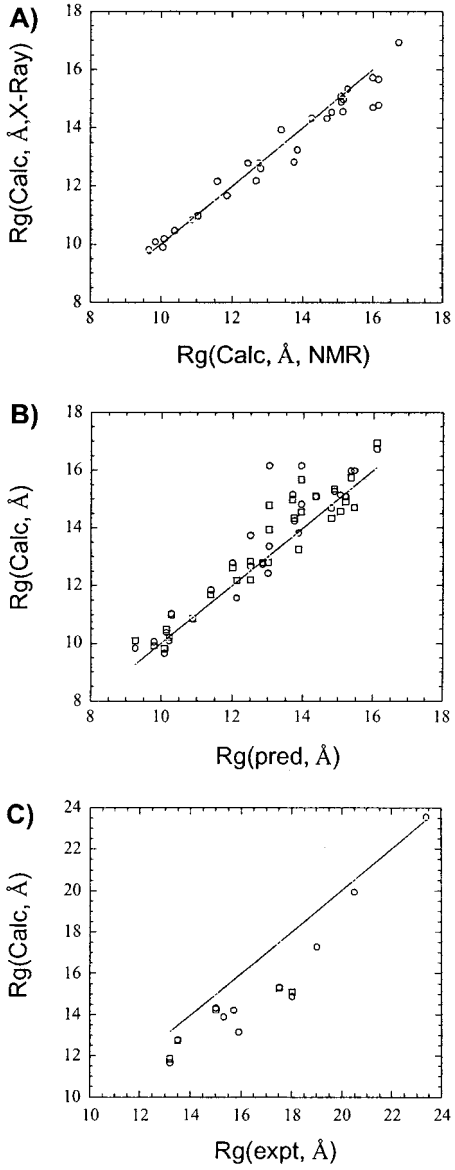

Figure 1. (A) Comparison of the radius of gyration calculated from the X-ray structural coordinates with the radius of gyration calculated from the corresponding NMR structure. (B) Comparison of the radius of gyration calculated from the X-ray (□) and NMR (○) structural coordinates with the predicted radius of gyration based on the number of residues in the protein. (C) Comparison of the radius of gyration calculated from the X-ray (○) and NMR (□) structural coordinates with the experimental radius of gyration reported in the literature. The four proteins with extremely large deviations between the structural and experimental Rg were excluded from the graph for clarity. The line corresponding to $y = x$ is included in each graph.

$$\mathrm{Rg(calc)} = \{ \sum_{j=1}^{N} [r_j - (\sum_{i=1}^{N} r_i/N)]^2/N \}^{1/2}$$

where r_i and r_j are the position vectors of atoms i and j, and N is the number of atoms. Thus, the radius of gyration target function may provide an effective global long-range restraint to counteract the tendency of protein structures to expand during a dynamic simulation.

In a globular protein, Rg can be predicted with reasonable accuracy on the basis of the number of residues by using the relationship $\mathrm{Rg(pred)} = 2.2N^{0.38}$ determined empirically from

(1) Clore, G. M.; Gronenborn, A. M. *Trends Biotechnol.* **1998**, *16*, 22−34.
(2) Gardner, K. H.; Kay, L. E. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357−406.
(3) Shan, X.; Gardner, K. H.; Muhandiram, D. R.; Rao, N. S.; Arrowsmith, C. H.; Kay, L. E. *J. Am. Chem. Soc.* **1996**, *118*, 6570−6579.
(4) Shan, X.; Gardner, K. H.; Muhandiram, D. R.; Kay, L. E.; Arrowsmith, C. H. *J. Biomol. NMR* **1998**, *11*, 307−318.
(5) Gardner, K. H.; Rosen, M. K.; Kay, L. E. *Biochemistry* **1997**, *36*, 1389−1401.
(6) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−4.
(7) Clore, G. M.; Starich, M. R.; Bewley, C. A.; Cai, M.; Kuszewski, J. *J. Am. Chem. Soc.* **1999**, *121*, 6513−6514.
(8) Huang, X.; Franklin, M.; Powers, R. *Biochemistry* **2000**, *39*, 13365−75.
(9) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891−8.
(10) Kuszewski, J.; Qin, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. B* **1995**, *106*, 92−6.
(11) Garrett, D. S.; Kuszewski, J.; Hancock, T. J.; Lodi, P. J.; Vuister, G. W.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. B* **1994**, *104*, 99−103.
(12) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1997**, *125*, 171−177.
(13) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337−2338.

10.1021/ja005770p CCC: $20.00
Published on Web 03/28/2001

analysis of high-resolution X-ray structures.[14] In an effort to address the appropriateness of the routine utility of the radius of gyration target function in structures determined by NMR, we have compiled a list of proteins where both NMR and X-ray structures have been solved. The goal is to provide insight into whether the use of the radius of gyration may inadvertently bias the calculated NMR protein structure and to further explore the generally accepted belief that X-ray structures are more compact relative to NMR structures.

For each protein, where the size ranges from 6 to 45 kDa, a theoretical Rg(pred) based on the number of residues has been calculated. Similarly, a calculated value for the radius of gyration, Rg(calc), was determined for each protein based on both the X-ray and NMR structure coordinates, where all hydrogens have been removed, using XPLOR.[15] In the case of NMR structures where an ensemble of protein coordinates were available, the Rg(calc) was calculated for each individual structure and an average Rg(calc) was used for the analysis. A total of 29 protein structures were identified for this analysis. Figure 1A compares Rg(calc) between the NMR and X-ray structure for each protein and clearly illustrates that in general both the NMR and X-ray structures for a given protein have a similar radius of gyration value ($r = 0.98$). There is no evidence for any systematic deviation between the radius of gyration calculated for the X-ray and NMR structure coordinates. This analysis implies that protein NMR structures, in general, have a similar compactness as those determined by X-ray crystallography. This differs from the previous observation where it was shown that NMR structures tend to be poorly packed and somewhat expanded relative to X-ray structures.[13] Of course, individual comparisons will vary based on the particulars of the X-ray and NMR structures. It is also plausible that other measures of compactness may yield results distinct from the Rg analysis reported here.

The comparison between the radius of gyration calculated from the NMR and X-ray structures, Rg(calc), with the predicted Rg values, Rg(pred), is shown in Figure 1B. Rg(pred) values are consistent with the Rg values calculated from both the experimental X-ray and NMR structures ($r = 0.95, 0.94$). This implies that the utilization of the radius of gyration as a target function should not bias the resulting structure.

However, deviation from a globular shape or regions of high mobility and disorder will skew Rg(calc) relative to Rg(pred). The appropriate use of the Rg target function has to account for these issues by either excluding the disordered residues from the Rg target function or using multiple Rg values to describe an elongated structure.[13] An alternative is to use a Rg value based on either a homologue or X-ray structure. Of course, experimentally determined Rg values would appear to be the preferred mechanism to obtain a target Rg value for a structure calculation.

Many studies have been done to determine the compactness of proteins by measuring Rg values using small-angle X-ray scattering (SAXS).[16] Experimentally determined Rg values are available in the literature for 11 proteins that have been structurally determined by NMR and X-ray. Figure 1C shows the comparison of the experimental Rg values with the Rg(calc) values. The experimental Rg values follow the general trend observed with Rg(pred) and Rg(calc), but are systematically larger. A presumed source of the discrepancy is the contribution of the protein's hydration shell to the experimental Rg values. The extent of the discrepancy may depend on the size of the outer hydration shell, shape of the protein, and the experimental buffer conditions. Additionally, for four of the proteins, the difference between the Rg values is mainly the result in a difference in the oligomeric state of the protein, where the Rg calculations are based on the monomeric structural coordinates for the protein.

On the basis of our analysis, the utilization of the radius of gyration target function as a routine component of an NMR structure determination protocol appears to be a valid approach and does not appear to bias the resulting structure. This is supported by the observation that a number of NMR protein structures exhibit very similar Rg values compared to the corresponding X-ray structure and Rg(pred). Also, this result provides support for the validity of using Rg(pred) in a structure calculation, which is critical given the observation that experimental Rg values over-estimate the structural radius of gyration. The comparison of the Rg values calculated from X-ray and NMR protein structures suggests that NMR structures, in general, are not expanded relative to X-ray structures. The use of the radius of gyration could prove to be a valuable addition to the structure determination of large molecular weight proteins where minimal restraint information may be available.

Supporting Information Available: Table of radius of gyration values calculated from the number of residues and their protein structure coordinates as well as the experimental values from literature (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA005770P

(14) Skolnick, J.; Kolinski, A.; Ortiz, A. R. *J. Mol. Biol.* **1997**, *265*, 217−41.

(15) Brunger, A. T. *X-PLOR*, Version 3.1 Manual; Yale University: New Haven, CT, 1993.

(16) Glatter, O.; Kratky, O.; Academic Press: London, 1982; pp 1−515.