

Improving the Packing and Accuracy of NMR Structures with a Pseudopotential for the Radius of Gyration

John Kuszewski, Angela M. Gronenborn, and G. Marius Clore*

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases National Institutes of Health, Bethesda, Maryland 20892-0520

Received December 21, 1998

NMR structures tend to be poorly packed and somewhat expanded relative to X-ray structures.¹ This is not a reflection of differences between the solution and crystal states, but rather of the nature of the experimental data and the computational methods employed to determine solution NMR structures. Indeed, the experimental NMR observables agree better with those calculated from high-resolution crystal structures than those calculated from the corresponding NMR structures.² NMR structures are mainly based on NOE-derived short interproton distance restraints which are generally classified into ranges. The NOE restraints determine the protein fold and pull the various structural elements into close spatial proximity, but the repulsive forces generated by the van der Waals term used to prevent atomic overlap tend to expand the structure. Because there are many more possible expanded structures than tightly packed ones that are compatible with the NOE data, the expansion can be regarded as an entropic effect. For globular proteins where there are numerous correlated NOE distance restraints and where the distance between structural elements is limited by covalent geometry, the problem of expansion may not be too severe, providing the structure is relatively well restrained by the NOE data. Intermolecular interfaces, on the other hand, will be particularly expanded since the density of intermolecular NOEs is usually limited and there are no intermolecular covalent restraints. Thus, there is a need for a structural restraint that can counteract the tendency toward expansion. In this paper, we demonstrate that the incorporation of a pseudopotential for the radius of gyration, R_{gyr} , a parameter that provides information regarding the global conformation of a molecule, can readily be incorporated into NMR structure calculations and results in significant increases in accuracy, particularly for protein–protein complexes.

The R_{gyr} of a group of atoms is defined as the rms distance from each atom of the molecule to their centroid

$$R_{\text{gyr}} = \left\{ \sum_{j=1}^N [r_j - (\sum_{i=1}^N r_i/N)]^2 / N \right\}^{1/2}$$

where r_i and r_j are the position vectors of atoms i and j , and N is the number of atoms. Previous work has shown that the value of R_{gyr} can be predicted with reasonable accuracy on the basis of the number of residues in a globular protein or protein domain using the relationship $R_{\text{gyr}}(\text{pred}) = 2.2N_{\text{residues}}^{0.38}$.³ R_{gyr} can also be measured experimentally using small-angle X-ray scattering with excellent agreement between observed values and those calculated from crystal structures.⁴

* To whom correspondence should be addressed. E-mail: clore@speck.niddk.nih.gov. Telephone: (301) 496-0782. Fax: (301) 496-0825.

(1) Gronenborn, A. M.; Clore, G. M. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *30*, 351–385. Abagayan, R. A.; Totrov, M. M. *J. Mol. Biol.* **1997**, *267*, 678–685.

(2) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891–5898.

(3) Skolnick, J.; Kolinski, A.; Ortiz, A. R. *J. Mol. Biol.* **1997**, *265*, 217–241.

R_{gyr} is not a geometrically specific quantity. Thus, restraining a protein structure to a particular target value of R_{gyr} does not constrain it to have any specific structure, but ensures that the structure is overall about as tightly packed as the majority of proteins whose structures have been determined by X-ray crystallography.

The pseudopotential E_{rgyr} which has been incorporated into CNS,⁵ is defined as

$$E_{\text{rgyr}} = k_{\text{rgyr}} \sum_{m=1}^{N_{\text{restraints}}} [R_{\text{gyr-calcd}}(\text{atoms}_m) - R_{\text{gyr-target}}(\text{atoms}_m)]^2$$

where k_{rgyr} is a force constant, $R_{\text{gyr-calcd}}$ and $R_{\text{gyr-target}}$ are the calculated and target values of R_{gyr} , respectively, and atoms_m represents the set of atoms used by restraint m . In our implementation different groups of atoms can have independent R_{gyr} restraints, thus facilitating applications to nonglobular proteins which can be subdivided into overlapping globular components. $R_{\text{gyr-calcd}}$ depends on the values of the coordinates of all atoms in a particular R_{gyr} restraint simultaneously, so the forces in relation to E_{rgyr} are applied to each of these atoms in order to move them closer to the target value. Since there is only one restraint for the selected group of atoms, the value of k_{rgyr} must be set to a relatively high value, typically $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

We have applied the R_{gyr} restraint to a number of cases where both NMR and X-ray structures are available: the B1 domain of streptococcal protein G (GB1) (6 kDa),^{6a,7a} the dimeric barrier-to-autointegration factor (BAF) (21 kDa),^{6b,7b} the trimeric ectodomain of gp41 (e-gp41) (44 kDa),^{6c,7c} and the tetramerization domain of p53 (20 kDa).^{6d,7d} A conventional simulated annealing protocol⁸ was employed and the force constant for E_{rgyr} was set at a constant value throughout the calculation.⁹ In making use of the E_{rgyr} potential, two factors need to be taken into consideration. If an experimental value of R_{gyr} is not available, and the value is estimated on the basis of the number of residues, appropriate account must be taken of disordered residues and the extent to which the structure deviates from a globular shape. Thus, for example, the p53 tetramerization domain consists of residues 319–360, but the N- and C-terminal residues are disordered. Residues 325–357 form an approximately globular structure, so that the target value of R_{gyr} is only estimated for these residues and the potential is only applied to these residues. In the case of e-gp41, the structure is highly anisotropic, approximately 34 \AA in width and $\sim 110 \text{ \AA}$ in height. In this case, the helical core of

(4) Bu, Z.; Koide, S.; Engelman, D. M. *Protein Sci.* **1998**, *7*, 2681–2683.

(5) Brünger, 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.* **1998**, *D54*, 905–921.

(6) (a) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657–661. (b) Cai, M.; Huang, Y.; Zheng, R.; Wei, S.-Q.; Guirlando, R.; Lee, M. S.; Craigie, R.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1998**, *5*, 903–909. (c) Caffrey, M.; Kaufman, J.; Stahl, S. J.; Wingfield, P. T.; Covell, D. G.; Gronenborn, A. M.; Clore, G. M. *EMBO J.* **1998**, *17*, 4572–4584. (d) Clore, G. M.; Ernst, J.; Clubb, R. T.; Omichinski, J. G.; Kennedy, W. M. P.; Sakaguchi, K.; Appella, E.; Gronenborn, A. M. *Nat. Struct. Biol.* **1995**, *2*, 321–332.

(7) (a) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. *Biochemistry* **1994**, *33*, 4721–4729. (b) Umland, T. C.; Davies, D. R., manuscript in preparation. (c) Malashkevich, V. N.; Chan, D. C.; Chutkowski, C. T.; Kim, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9134–9139. (d) Jeffrey, P. D.; Gorina, S.; Pavletich, N. P. *Science* **1995**, *265*, 1498–1502.

(8) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1987**, *2*, 27–38.

(9) The optimal value of k_{rgyr} lies in the $50\text{--}100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ range. Values in this range do not affect the conformation of mobile surface side chains. Higher values result in violations with respect to the experimental restraints and deterioration in the covalent geometry. Lower values progressively reduce the impact of the R_{gyr} restraint.

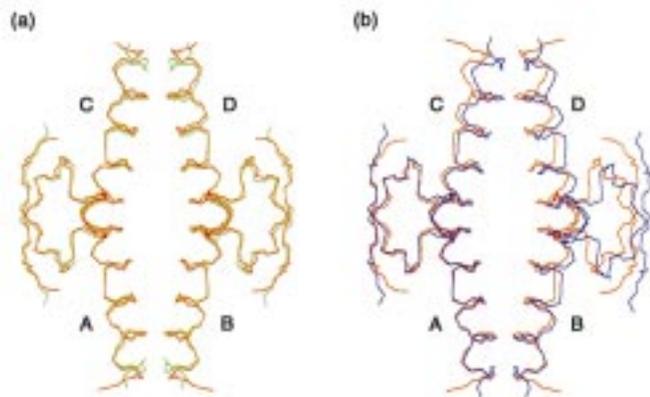


Figure 1. Best-fit backbone (N, C α , C) superpositions of the X-ray structure (residues 325–356; red) of the p53 tetramerization domain onto the mean NMR structures calculated with (green, panel a) and without (blue, panel b) the incorporation of the R_{gyr} restraint. The four subunits are labeled A–D. The coordinates are best-fitted to residues 326–354 (the ordered core) of subunit A only to clearly illustrate the effect of the R_{gyr} restraint on the packing of the tetramer. The expansion of the structure and the lateral translational displacement of the BD dimer in the case of the structure calculated without the R_{gyr} restraint is evident.

e-gp41 (which is ~ 96 Å in length and excludes residues 81–106 which form a flexible loop at the apex of the structure) was divided into four overlapping globular segments of approximately $34 \times 34 \times 34$ Å, and four separate R_{gyr} restraints were employed simultaneously, one for each segment.

The results are summarized in Table 1. The improvement in accuracy ranges from 9 to 83% and is smallest for monomeric proteins or individual subunits (9–39%) and largest for complexes or multimeric proteins (31–83%). Figure 1 illustrates the improvement in the accuracy (83%) of the global structure of the p53 tetramerization domain upon refinement against R_{gyr} . The agreement with the experimental restraints, and the precision and quality of the structures (as measured by the percentage of residues in the most favored regions of the Ramachandran plot) is unaffected by refinement against R_{gyr} . In addition, inspection of the surface side chains indicates that these are not perturbed by refinement against R_{gyr} , and no folding back of surface side chains is observed.

We have also investigated the use of the R_{gyr} restraint to obviate the need for classification of NOEs into distance ranges. We carried out a series of calculations on GB1 in which all the NOE distance restraints were set to an upper bound of 5.0 Å. In the absence of the R_{gyr} restraint, the backbone rmsd to the X-ray structure increased from 0.69 Å with a classified set of NOE restraints (cf. Table 1) to 0.81 Å. The inclusion of the R_{gyr} restraint brought the backbone rmsd to the X-ray structure back down to 0.58 Å, which is essentially identical to that of the structure obtained using the classified NOE restraints supplemented by the R_{gyr} restraint (cf. Table 1).

Since the majority of NMR restraints are restricted to atoms in close spatial proximity, there are two major problems relating to long-range order in any NMR structure determination: angular orientation and translational. In this paper we have demonstrated that the latter is readily resolved by the addition of a simple pseudopotential describing R_{gyr} . The target values of R_{gyr} are easily computed from the number of residues in the protein, and appropriate measures can always be taken to divide the protein or protein complex into overlapping approximately globular segments. Thus, the combined use of residual dipolar coupling restraints which provide direct information on angular orientations,¹⁰ and the R_{gyr} restraint which provides translational information, resolve many of the outstanding problems related to long-range structure in NMR structure determination. The R_{gyr} restraint produces significant improvements in accuracy, particularly for

Table 1. Effect of Incorporation of the R_{gyr} Restraint on Accuracy^a

protein ^b	backbone (N, C α , C) rmsd to X-ray structure (Å) ^c		% improvement in accuracy	
	without R_{gyr} restraint	with R_{gyr} restraint		
GB1	0.69	0.57	21%	
BAF	monomer	0.82	0.73	12%
	dimer	1.18	0.85	39%
gp41	monomer	1.58	1.14	39%
	trimer	1.71	1.15	49%
p53	monomer	0.37	0.34	9%
	AC dimer	0.47	0.36	31%
	tetramer	0.66	0.36	83%

^a The number of experimental NMR restraints used in calculating the structures is as follows. GB1 (56 residues):^{6a} 1232 restraints comprising 735 interproton distance restraints, 145 torsion angle restraints, 53 3J coupling restraints, and 300 dipolar coupling restraints (152 measured in a liquid crystalline medium of phage *fd*^{11a} and 148 in a liquid crystalline medium of bicelles^{11b}). BAF (89 residues per subunit):^{6b} 1655 restraints per monomer comprising 864 distance restraints including 48 intermolecular restraints, 257 torsion angle restraints, 66 3J coupling restraints, 165 $^{13}\text{C}\alpha/\beta$ secondary chemical shift restraints, 44 ^1H methyl shift restraints and 259 dipolar coupling restraints. e-gp41 (123 residues per subunit; residues 27–149):^{6c} 2160 restraints per monomer comprising 1500 distance restraints, including 232 intermolecular restraints, 360 torsion angle restraints, 35 3J coupling restraints, 26 $^3\text{DC}\alpha(\text{ND})$ isotope shift restraints, and 239 $^{13}\text{C}\alpha/\beta$ chemical shift restraints. Tetramerization domain of p53 (42 residues per subunit; residues 319–360):^{6d} 1118 restraints per monomer comprising 938 distance restraints including 309 intersubunit restraints, 71 torsion angle restraints, 36 3J coupling restraints and 73 $^{13}\text{C}\alpha/\beta$ chemical shift restraints. The conformational database potential² was employed in all calculations. The target values of R_{gyr} given by $2.2N_{\text{residues}}^{0.38}$ are as follows. GB1: 10.16 Å for residues 1–56. BAF: 15.76 Å for residues 1–89 of both subunits taken together. gp41: 14.76 Å for residues 3–26 and 97–122, 14.54 Å for residues 10–33 and 90–114, 14.88 Å for residues 20–43 and 78–104, and 14.18 Å for residues 30–53 and 72–92, for the three subunits taken together; the four overlapping segments are globular. p53: 13.57 Å for residues 325–355 of the four subunits taken together. The precision of the backbone coordinates is as follows: GB1, 0.20–0.26 Å; BAF dimer, 0.32–0.34 Å; gp41 trimer, 0.60–0.75 Å, p53 tetramer (residues 326–354) 0.31–0.36 Å. ^b The X-ray structures of GB1, BAF, gp41, and p53 are listed in ref 7a–d, respectively. ^c The values reported are the rmsd values between the X-ray structure and the mean NMR structure. Between 20 and 30 simulated annealing structures were calculated in each case. The rmsd relates to residues 1–56 of GB1, residues 3–89 of BAF, residues 8–43, and 86–119 of SIV gp41 (the fragment of the core corresponding to the crystal structure^{7c}), and residues 326–354 of p53 (i.e., the ordered core).

protein complexes, and may also be used to eliminate the need for distance classification of NOEs without any loss in overall accuracy, a feature which could be particularly useful in the structure determination of larger proteins.

Acknowledgment. We thank T. C. Umland and D. R. Davies for giving us the X-ray coordinates of BAF prior to publication, and A. Bax and M. Caffrey for useful discussions.

Supporting Information Available: One table giving a comparison of the target values of R_{gyr} with the calculated values of R_{gyr} for the NMR structures obtained with and without the R_{gyr} restraint, and for the X-ray structures (PDF). This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA9843730

(10) Tjandra, N.; Omichinski, G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732–728.

(11) (a) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571–10572. (b) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.