

Relationships between the Precision of High-Resolution Protein NMR Structures, Solution-Order Parameters, and Crystallographic B Factors

ROBERT POWERS, G. MARIUS CLORE,* DANIEL S. GARRETT, AND ANGELA M. GRONENBORN*

Laboratory of Chemical Physics, Building 2, National Institutes of Diabetes and Digestive and Kidney Diseases,
National Institutes of Health, Bethesda, Maryland 20892

Received February 12, 1993

One of the principal motivations for studying proteins by nuclear magnetic resonance stems from the desire to describe the *solution* structure of these molecules as compared to the generally perceived static picture obtained by X-ray crystallography. Indeed, it is one of the unique features of NMR spectroscopy that in addition to structural data, dynamic properties can be probed and characterized by measuring relaxation parameters. Furthermore, any mobility of the protein in solution will necessarily modulate the measured NMR parameters and should influence the resulting structure. It has been argued that regions of a protein that are highly mobile would be expected to be defined to a lesser degree of precision than regions that are rigid (1, 2).

The structures of proteins determined by NMR are based on the calculation of a large ensemble of structures, where each structure is compatible with the experimental NMR data, comprising principally short (<5 Å) approximate interproton distance restraints. Thus, each individual structure in the ensemble represents an equally good description of the "true mean" structure. Because the nuclear Overhauser effect at short mixing times is proportional to r^{-6} , the resulting interproton distance restraints are $\langle r^{-6} \rangle^{-1/6}$ averages. Hence, the mean structure that is probed by NMR is not an arithmetic mean of all the conformations present in solution but an $\langle r^{-6} \rangle^{-1/6}$ mean. The precision of the structure determination is dependent on the number and distribution of the interproton distance restraints (3) and is simply given by the average atomic rms difference, $\langle \text{rmsd}_{x-m} \rangle$, between the individual structures and their mean coordinate positions. For high-resolution NMR structures (4) which are characterized by a backbone $\langle \text{rmsd}_{x-m} \rangle$ of ≤ 0.4 Å and are based on an average of more than ~ 15 restraints per residue, including stereospecific assignments of the β -methylene protons and methyl groups of all side chains that are not conformationally disordered, one would expect an empirical correlation on a residue-by-residue basis between precision and mobility. To test this hypothesis we examined data for three proteins, interleukin-8 (IL-8), interleukin-1 β (IL-1 β),

and interleukin-4 (IL-4), for which high-resolution NMR structures (5–7), generalized-order parameters for the NH vectors from ^{15}N relaxation measurements (8–10), and refined crystal structures (11–16) are available.

The overall order parameters S^2 for the individual backbone NH vectors are plotted against the corresponding values of the backbone $\langle \text{rmsd}_{x-m} \rangle$ in Fig. 1A. The data comprise order parameters for 64 out of 72 residues for IL-8, 127 out of 153 residues for IL-1 β , and 113 out of 133 residues for IL-4. (Note that IL-8 is a symmetric homodimer, each subunit having 72 residues). The data reveal a large scatter (for values of $\langle \text{rmsd}_{x-m} \rangle \leq 0.8$ Å, S^2 spans a width of ~ 0.3 units), and an inverse nonlinear relationship between S^2 and $\langle \text{rmsd}_{x-m} \rangle$ can be observed. The apparent nonlinearity in the correlation arises from the fact that the maximum value of S^2 cannot exceed 1.0 (i.e., no motion). For values of $\langle \text{rmsd}_{x-m} \rangle \leq 0.4$ Å, S^2 reaches an average plateau value of 0.85 ± 0.15 , reflecting the fact that small-magnitude thermal motions are always present. For $\langle \text{rmsd}_{x-m} \rangle \geq 1.2$ Å (data not shown), S^2 appears to reach a limiting value of ~ 0.2 , reflecting the fact that a tethered fragment of polypeptide chain cannot exhibit completely random motion. An approximately linear correlation is obtained by plotting S^2 versus $(1 + \langle \text{rmsd}_{x-m} \rangle)^{-1}$, a function which, like S^2 , is limited to values between 0 and 1, as shown in Fig. 1B.

The large scatter observed in the S^2 vs $\langle \text{rmsd}_{x-m} \rangle$ and $(1 + \langle \text{rmsd}_{x-m} \rangle)^{-1}$ plots may be attributed to the fact that the correlation between S^2 and $\langle \text{rmsd}_{x-m} \rangle$ arises only via an *indirect* relationship. While S^2 is directly dependent on motions faster than the overall correlation time (τ_c) of the molecule, the $\langle \text{rmsd}_{x-m} \rangle$ is dependent on the number and distribution of NOE interproton distance restraints per residue (3). Clearly both the mobility of a residue and the number of observable NOEs (i.e., short interproton distance contacts) are linked to the packing density of interresidue interactions, thereby establishing an indirect correlation between S^2 and $\langle \text{rmsd}_{x-m} \rangle$. To further complicate this correlation, a correspondence between S^2 and $\langle \text{rmsd}_{x-m} \rangle$ need only exist for small values of S^2 . Any increase in the mobility of a residue associated with a small S^2 will cause the intensity of inter-

* To whom correspondence should be addressed.

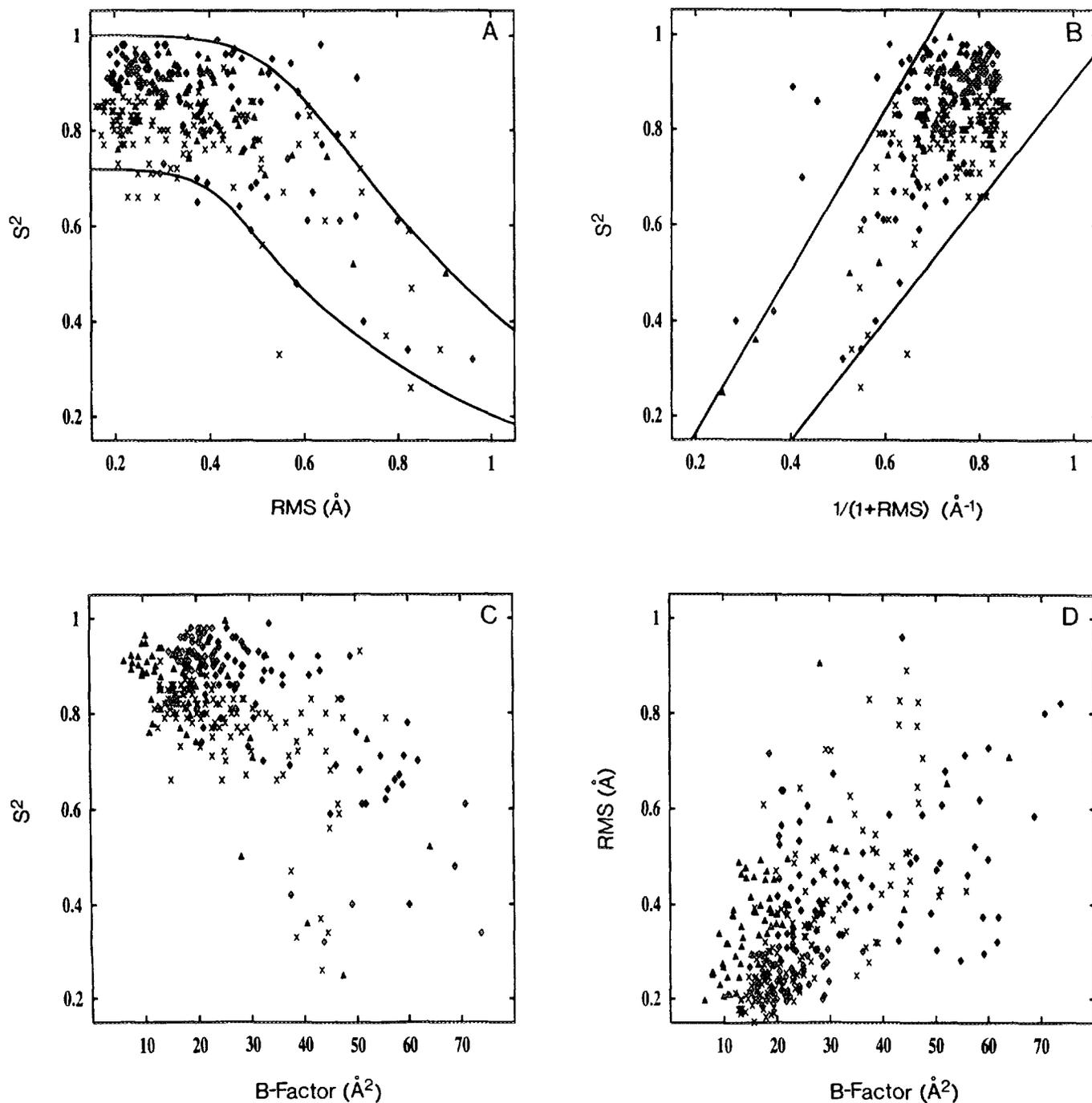


FIG. 1. Empirical relationships between backbone coordinate precision of NMR structures, overall solution-order parameter (S^2) for the backbone NH vectors, and backbone crystallographic B factors for IL-8 (Δ), IL-1 β (\times), and IL-4 (\diamond). The coordinate precision of an NMR structure determination is given by the average atomic rms difference, $\langle \text{rmsd}_{x-m} \rangle$, between the individual structures of an ensemble and their mean coordinate positions (denoted as rms in the figure). The lines in (A) and (B) serve to guide the eye.

residue NOEs to be attenuated resulting in fewer interproton distance restraints with larger upper bounds and correspondingly larger $\langle \text{rmsd}_{x-m} \rangle$ values. Motion, however, that is slower than the correlation time has no effect on S^2 , but may increase the value of $\langle \text{rmsd}_{x-m} \rangle$, reflecting multiple

conformations within particular regions of the molecule. Taking the above into account, it would be inappropriate to expect a perfect correlation between S^2 and $\langle \text{rmsd}_{x-m} \rangle$; indeed, the finding of a general trend is a pleasing result. It is also important to note that an exhaustive and accurate anal-

ysis of the NMR data is a necessary prerequisite for the above relationship to be of general validity. Incomplete analysis of the NMR spectra, misassignments of NOEs, and inappropriate boundary limits will result in relatively large $\langle \text{rmsd}_{x-m} \rangle$ values. Therefore, even though the mobility of a protein region may possibly be inferred from the relative values of $\langle \text{rmsd}_{x-m} \rangle$ in high-resolution NMR structures, it is only the values of S^2 that can accurately determine local mobility.

A plot of backbone S^2 versus crystallographic B factors is presented in Fig. 1C. An inverse correlation would intuitively be expected since high mobility in solution as evidenced by a low S^2 would manifest itself by random thermal motions or static conformational disorder in the crystal lattice, resulting in large B factors. Although a trend in this regard may be inferred from the data, the large degree of scatter indicates that additional factors heavily influence this simple relationship. Indeed, a small S^2 value does not necessarily result in a large B factor since regions involved in crystal contacts will be restricted in their thermal motions in the lattice but frequently exhibit a considerable degree of mobility in solution. Conversely, large S^2 values in regions involved in slow exchange between different conformations could result in large B factors instead of the predicted small values because of the inability to differentiate between multiple conformers in the crystal (i.e., static disorder). Likewise, the relationship between the $\langle \text{rmsd}_{x-m} \rangle$ and B factors is a complex one. Although, in general, regions which are well defined in the solution structure will correspond to regions exhibiting small B factors in the X-ray structure, the reverse is not necessarily true and detailed comparisons must be carried out for each individual case.

Given the above relationships, one can only conclude that the precision of the backbone coordinates on a residue-by-residue basis observed in high-resolution protein NMR structures is approximately correlated to backbone mobility in solution (S^2). A similar type of approximate correlation is also observed between B factors and coordinate precision in X-ray structures (17). This observation is reassuring since it indicates that in regions of high mobility the precision of the NMR solution coordinates is correspondingly reduced. Indeed, the observation of overly precise coordinates in regions of high mobility (either $S^2 \leq 0.4$ or significant conformational heterogeneity as evidenced by T_2 exchange line broadening) can be taken as indicative of the presence of errors in the interproton distance restraints in such regions (e.g., misassignments, upper bounds that are too low, distance ranges that are too restrictive). Conversely, reduced precision in regions that are not particularly mobile is indicative of a lack of an appropriate number of distance re-

straints, for example, due to incomplete assignments of the NOE cross peaks in multidimensional spectra. It cannot, however, be emphasized enough that the $\langle \text{rmsd}_{x-m} \rangle$ does not provide a measure of the magnitude of the conformational space sampled by a protein in solution. Rather it simply reflects the precision with which the mean $\langle r^{-6} \rangle^{-1/6}$ solution coordinates have been determined. Thus, broadly speaking, S^2 , together with chemical-exchange line broadening, is the solution equivalent of the crystallographic B factor, and the $\langle \text{rmsd}_{x-m} \rangle$ for an ensemble of NMR structures is equivalent to the precision of the crystallographic coordinates, as obtained from independent X-ray structure determinations of the same crystal form.

ACKNOWLEDGMENT

This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health.

REFERENCES

1. S. V. Hyberts, M. S. Goldberg, T. F. Havel, and G. Wagner, *Protein Sci.* **1**, 736 (1992).
2. J. W. Peng and G. Wagner, *Biochemistry* **31**, 8571 (1992).
3. G. M. Clore, M. A. Robien, and A. M. Gronenborn, *J. Mol. Biol.*, in press.
4. G. M. Clore and A. M. Gronenborn, *Science* **252**, 1390 (1991).
5. G. M. Clore, E. Appella, M. Yamada, K. Matsushima, and Gronenborn, *Biochemistry* **29**, 1689 (1990).
6. G. M. Clore, P. T. Wingfield, and A. M. Gronenborn, *Biochemistry* **30**, 2315 (1991).
7. R. Powers, D. S. Garrett, C. J. March, E. A. Frieden, A. M. Gronenborn, and G. M. Clore, *Biochemistry*, in press (1993).
8. B. L. Grasberger, A. M. Gronenborn, and G. M. Clore, *J. Mol. Biol.*, in press.
9. G. M. Clore, P. C. Driscoll, P. T. Wingfield, and A. M. Gronenborn, *Biochemistry* **29**, 7387 (1990).
10. C. Redfield, J. Boyd, L. J. Smith, R. A. G. Smith, and C. M. Dobson, *Biochemistry* **31**, 10431 (1992).
11. E. T. Baldwin, I. T. Weber, R. St. Charles, J.-G. Xuan, E. Appella, M. Yamada, K. Matsushima, B. F. P. Edwards, G. M. Clore, A. M. Gronenborn, and A. Wlodawer, *Proc. Natl. Acad. Sci. USA* **88**, 5026 (1991).
12. B. C. Finzel, L. L. Clancy, D. R. Holland, S. W. Muchmore, K. D. Watenpugh, and H. M. Einspahr, *J. Mol. Biol.* **209**, 779 (1989).
13. J. P. Priestle, H. P. Schär, and M. G. Grütter, *Proc. Natl. Acad. Sci. USA* **86**, 9667 (1989).
14. B. Veerapandian, G. L. Gilliland, R. Raag, A. L. Svensson, Y. Masui, Y. Hirai, and T. L. Poulos, *Protein Struct. Funct. Genet.* **12**, 10 (1992).
15. B. Shaanan, A. M. Gronenborn, G. H. Cohen, G. L. Gilliland, B. Veerapandian, D. R. Davies, and G. M. Clore, *Science* **257**, 961 (1992).
16. A. Wlodawer, A. Pavlosky, and A. Gutschina, *FEBS Lett.* **309**, 59 (1992).
17. S. E. V. Phillips, *J. Mol. Biol.* **142**, 531 (1980).