Local knowledge helps determine protein structures

Michael R. Gryk and Jeffrey C. Hoch*

Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center, Farmington, CT 06030-3305

The nuclear Overhauser effect (NOE) has been the workhorse of structural studies of macromolecules in solution by NMR spectroscopy. The NOE is a measure of the rate of magnetization transfer between nuclei, induced by modulation of the magnetic dipole–dipole coupling between nuclear spins by overall tumbling of the molecule in solution (1). The efficiency of magnetization transfer via this mechanism is low; consequently, NOEs are typically only observed for protons within ≈5 Å, and lengthy experiments are required in order to detect the weak effect. NOE magnitudes provide a "spectroscopic ruler" through the approximate r^{-6} dependence on the distance between proton pairs. Although it takes surprisingly few such distance estimates to determine the overall fold of a protein (2)—provided the estimates are distributed evenly throughout the protein—even a complete set of NOEs yields far fewer constraints on the structure than are required to determine all $3N - 6$ degrees of freedom for an N-atom protein. The fold is even more highly underdetermined when protein flexibility is taken into account. By complementing NMR measurements with prior knowledge of protein structure, usually in the form of a potential energy function that describes the physical plausibility of a model structure, NMR has emerged as a valuable complement to x-ray crystallography for structure determination. The Protein Data Bank (PDB) (3) now contains >7,000 NMR protein structures, and >1,000 new structures are anticipated to be added in 2008. One way to further hasten the rate of protein structure determination by NMR is to develop alternatives that avoid the need to measure and assign NOEs. In this issue of PNAS, Shen *et al.* (4) describe a viable alternative to NOE-based structure determination for small proteins.

Nevertheless, reliance on NOEs for determination of biomolecular structure presents real challenges. A NOE must be assigned to two specific nuclei before it can be used to constrain a structural model, and assignment becomes more challenging as the size of the molecule increases. Even a modest-sized protein can yield many thousands of detectable NOEs. NOEs are frequently missing from flexible regions, and NOEs that can be measured are usually highly correlated with one another. The information content of all NOEs is not the same: NOEs between residues that are distant in the primary sequence (called "long-range NOEs") are more important for constraining the overall fold than are NOEs between neighboring residues ("short-range NOEs") or intraresidue NOEs that provide no additional structural information ("redundant NOEs"). The latter two types of NOEs are far more prevalent. The confidence level in an assignment is greater when a NOE is correlated with other assigned NOEs, but distinguishing a unique long-range NOE from an erroneous assignment can be challenging. Indeed, the small number of demonstrably incorrect NMR structures reported can nearly all be attributed to a handful of erroneous NOE assignments.

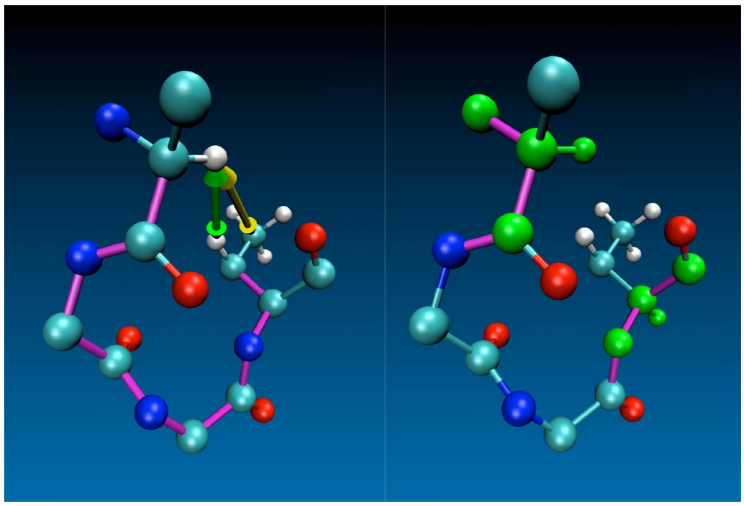

Fig. 1. Structural information derived from chemical shifts is more local than information derived from NOEs. (*Left*) NOEs observed between the same pair of residues, depicted as arrows between the H^{α} atom of one residue and the H^{β} atom (green) and methyl group (yellow) of a "distant" residue, are highly correlated and long-range in their impact: each provides approximately the same structural information, highly constraining the values of all 10 intervening dihedral angles (magenta). (*Right*) In contrast, chemical shift measurements for the atoms depicted in green provide more local structural constraints for the adjoining dihedral angles. The figure was prepared by using VMD (15).

NOEs are by no means the only NMR-observable parameters that are sensitive to structure. Vicinal couplings, residual dipolar couplings, and chemical shifts are all sensitive to structure and are all routinely used to supplement NOE-derived distance restraints in protein structure determination (5). Chemical shifts are the easiest to measure and are the first parameters to be determined in assigning resonances to specific nuclei. Chemical shift is a measure of the Larmour precession frequency of a magnetic spin under the influence of the local magnetic field (6). The field experienced by a nucleus is perturbed from the applied field strength of the instrument by the distribution of electrons near the nucleus, which can either diminish or augment the applied field. As the name implies, chemical shift is sensitive to the number and nature of substituents covalently connected to the nucleus, and to their conformation. The correlation between ^{1}H, ^{13}C, and ^{15}N chemical shifts of the polypeptide backbone and backbone dihedral angles has long been viewed as a potential alternative to NOEs for determining solution structure, but the imprecision of the correlation has proved limiting. Nevertheless, parallel progress in improving the correlation between chemical shifts and structure (both quantum mechanical and empirical) and *de novo* methods for protein structure prediction led Wishart and Case (7) to anticipate in 2001 that a combination of chemical shifts and structure prediction methods would be-

Author contributions: M.R.G. and J.C.H. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 4685.

*To whom correspondence should be addressed. E-mail: hoch@uchc.edu.

come a viable method for determining the solution structure of small proteins.

The article by Shen *et al.* (4) describes a confluence of precisely the two approaches foreseen by Wishart and Case (7). Baker and colleagues (8) are the developers of ROSETTA, a program that predicts protein structure from sequence data alone by using conformational ensembles for peptide fragments selected from the PDB on the basis of sequence homology, combined with an empirical cost function that quantifies the physical plausibility of conformations, to assemble the fragments into the intact protein. ROSETTA has consistently out-performed competing approaches in an annual competition [Critical Assessment of Techniques for Protein Structure Prediction (CASP); www.predictioncenter.org] designed to test the state of the art in protein structure prediction. Bax and colleagues (9) developed TALOS (torsion angle likelihood obtained from shift and sequence similarity), a widely used program for predicting torsion angles for peptide fragments on the basis of chemical shift measurements, again based on structures in the PDB and a corresponding database of chemical shifts, derived from the Biomagnetic Resonance Data Bank (10). To determine small protein structures from chemical shifts alone, Shen *et al.* employ an expanded TALOS-like database of chemical shifts to select ensembles of peptide fragment conformations, which are assembled into a final structure by using ROSETTA.

The results of this approach—applied to 16 proteins ranging up to 129 residues—are remarkable. The root-mean-square deviation (rmsd) for atoms making up the peptide backbone averages only 1.4 Å from the corresponding target structures, as determined by more-conventional NOE-based approaches. To explore the question of unintentional bias, the authors also performed "blind" structure computations by using chemical shifts for nine proteins whose structures were not yet deposited in the PDB, in collaboration with two structural genomics consortia. The rmsds from the target structures for these blind computations averaged 1.0 Å.

A very similar approach was recently reported by Cavalli *et al.* (11), also using a database of measured chemical shifts and known protein structures and using ROSETTA to assemble the final structure from ensembles of peptide fragment conformations. Of the five proteins in common to the two studies, only three used the same PDB entry as the target structure. For these three, the results by Shen *et al.* (4) average 1.1 Å from the target structures (backbone rmsd), whereas the results by Cavalli *et al.* differ by 1.5 Å. The significantly improved accuracy obtained by Shen *et al.* seems to be attributable to their use of $^{1}H^{N}$ and $^{13}C'$ chemical shifts, in addition to the ^{15}N, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{13}C^{\beta}$ shifts used by Cavalli *et al.*—a telling difference suggesting that the addition of more chemical shifts (for example $^{1}H^{\beta}$) may allow for future improvements. Different prior structure determinations of the same protein yielded backbone rmsds of 0.6 Å for ubiquitin (PDB entries 1D3Z vs. 1UBQ) and 0.8 Å for calbindin (1ICB vs. 4ICB), illustrating that the precision of the structures determined by chemical shift approaches that of traditional methods.

The impetus for using chemical shifts instead of NOEs for structural studies extends beyond the desire to avoid the time-consuming experiments needed to detect weak NOEs or the desire to reduce the number of assignments. The local nature of chemical shift structural restraints means that the effects of misassignment are similarly local (Fig. 1). In contrast, misassignment of even a single long-range NOE can have a profoundly negative impact on the accuracy of the computed structure. Indeed, all of the demonstrably incorrect NMR protein structures resulted from misassignment of a mere handful of NOEs. Conversely, because chemical shift structural restraints are by their very nature local, they are well suited for defining the structure of smaller proteins where all effects are local, while being less sufficient for larger structures. This may account for the increase in rmsd vs. number of protein residues observed by both Shen *et al.* (4) and Cavalli *et al.* (11). For these larger systems, supplementing the chemical shift data with a few longer-range constraints (NOEs, residual dipolar couplings, or spin labels) may prove effective in improving the accuracy of the determination. An interesting exercise would be to apply the method of Shen *et al.* to compute the structures of the incorrect NMR protein structures reported previously. We predict that, in many instances, the correct fold will be obtained despite the presence of misassignments. Another potential benefit of chemical shifts is that they can be measured for disordered or flexible regions of a protein that typically fail to yield detectable NOEs. Berjanskii and Wishart (12) have recently shown that backbone chemical shifts reflect local flexibility, as well as structure.

The relative independence of structural information derived from chemical shifts reflects the different contributions to the chemical shifts experienced by different nuclei along the peptide backbone (7). In contrast, the high degree of correlation among NOEs (Fig. 1) hinders the use of powerful resampling statistical methods (13) for error analysis or for detecting incorrect assignments. With chemical shift-based structure determination, these tools, which have proven so powerful for x-ray crystallography, could be returned to the armamentarium of NMR spectroscopists.

The useful protein size limit of chemical shifts for structure determination appears to be the result of cumulative error in the correlation between chemical shifts and structure. If this is so, then improvements in accuracy should lead to a concomitant increase in the size of the protein for which the method works. Possible improvements are straightforward in principle. More accurate and extensive correlation between chemical shift and local conformation can be anticipated, either through additional empirical data or from recent improvements in the quantum mechanical computation of chemical shifts (14). In the approach of Shen *et al.* (4), chemical shifts were used to select families of peptide conformations but not to refine the final structure. Adding chemical shift terms to the cost function used for final assembly and refinement should help extend the applicability of the method. Although clearly there is more work to be done, it can be said that 2008 is the year in which an alternative to NOE-based determination of protein structure in solution became truly viable for small proteins.

1. Noggle JH, Shirmer RE (1971) *The Nuclear Overhauser Effect: Chemical Applications* (Academic, New York).
2. Connolly PJ, Stern AS, Hoch JC (1994) *J Am Chem Soc* 116:2675–2676.
3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) *Nucleic Acids Res* 28:235–242.
4. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, *et al.* (2008) *Proc Natl Acad Sci USA* 105:4685–4690.
5. Clore GM, Gronenborn AM (1998) *Proc Natl Acad Sci USA* 95:5891–5898.
6. Abragam A (1961) *Principles of Nuclear Magnetism* (Oxford Univ Press, Oxford).
7. Wishart DS, Case DA (2001) *Methods Enzymol* 338:3–34.
8. Simons KT, Bonneau R, Ruczinski I, Baker D (1999) *Proteins Suppl* 3:171–176.
9. Cornilescu G, Delaglio F, Bax A (1999) *J Biomol NMR* 13:289–302.
10. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, *et al.* (2008) *Nucleic Acids Res* 36:D402–D408.
11. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) *Proc Natl Acad Sci USA* 104:9615–9620.
12. Berjanskii MV, Wishart DS (2008) *J Biomol NMR* 40:31–48.
13. Brünger A (1992) *Nature* 355:472–475.
14. Moon S, Case DA (2006) *J Comput Chem* 27:825–836.
15. Humphrey W, Dalke A, Schulten K (1996) *J Mol Graph* 14:33–38.