

A Dipolar Coupling Based Strategy for Simultaneous Resonance Assignment and Structure Determination of Protein Backbones

Fang Tian, Homayoun Valafar, and James H. Prestegard*

Contribution from the Southeast Collaboratory for Structural Genomics, University of Georgia, Athens, Georgia 30602-4712

Received July 24, 2001

Abstract: A new approach for simultaneous protein backbone resonance assignment and structure determination by NMR is introduced. This approach relies on recent advances in high-resolution NMR spectroscopy that allow observation of anisotropic interactions, such as dipolar couplings, from proteins partially aligned in field ordered media. Residual dipolar couplings are used for both geometric information and a filter in the assembly of residues in a sequential manner. Experimental data were collected in less than one week on a small redox protein, rubredoxin, that was ^{15}N enriched but not enriched above 1% natural abundance in ^{13}C . Given the acceleration possible with partial ^{13}C enrichment, the protocol described should provide a very rapid route to protein structure determination. This is critical for the structural genomics initiative where protein expression and structural determination in a high-throughput manner will be needed.

Introduction

Various genome sequencing projects are producing large amounts of information about genes and gene products. Of the genes sequenced, many will code for proteins of unknown structure and function. It is believed that the elucidation of structure in these cases will provide a stepping stone to the necessary functional characterization.^{1–5} However, determination of structure for proteins on a genomic scale demands rapid, high-throughput, analysis. X-ray crystallographers are in a good position to meet this demand, but NMR spectroscopists are far less prepared; protein structure determination by conventional NMR methods often takes months. A major portion of the time devoted to NMR-based analysis involves resonance assignment, the first in a two-step process that leads to structure determination.⁶ Here, a new approach is described that can save major amounts of time, by merging assignment and structure determination steps and by focusing on backbone structures instead of structures that include both backbone and side chain atomic positions.

It is not assignment, per se, that slows the NMR structure determination process, but the fact that assignment has to be nearly complete. Assignment of both backbone and side chain resonances is a prerequisite for protein structure determination by traditional NOE-based methods.⁶ This is because sequential connectivity is done through the backbone, but the majority of structure-rich, long-range, NOEs come from side chain–side chain contacts. For many structural genomics applications, such

as the development of a fold library and prediction of function, it may not be necessary to experimentally place side chains in structures; backbone structures may be adequate.^{7–9} Side chain resonances are also the most difficult to assign, and a significant amount of time is required for data collection on their behalf. Backbone assignments are, on the other hand, efficient and even subject to automation when ^{13}C , ^{15}N labeled proteins are available.^{10–12} Thus, consideration of determining protein backbone structures directly seems justified.

NOE data are unfortunately sparse when restricted to backbone atoms. Hence, it is important to consider alternate sources of information such as the orientational information that comes from residual dipolar interactions.^{13–16} Unlike constraints derived from NOE data, constraints derived from residual dipolar coupling data do not require close proximity of the atoms; rather than constraining distances of separation, the relative orientation of various bond vectors, or molecular fragments, is constrained.

In partially ordered systems, residual dipolar couplings are given by eq 1, where $\gamma_{i,j}$ are the gyromagnetic ratios of nuclei

$$D_{ij}^{\text{res}} = - \left(\frac{\mu_0}{4\pi} \right) \frac{\gamma_i \gamma_j h}{2\pi^2 r_{ij}^3} \sum_{kl} S_{kl} \cos(\alpha_k) \cos(\alpha_l) \quad (1)$$

i and j , r_{ij} is the internuclear distance between these nuclei, $\cos(\alpha_{k,l})$ are direction cosines describing the orientation of the

* To whom correspondence should be addressed: (phone) (706)542-6281; (e-mail) jpresteg@ccrc.uga.edu.

(1) Sali, A.; Kuriyan, J. *Trends Biochem. Sci.* **1999**, *24*, M20–M24.
 (2) Blundell, T. L.; Mizuguchi, K. *Prog. Biophys. Mol. Biol.* **2000**, *73*, 289–295.
 (3) Burley, S. K.; Almo, S. C.; Bonanno, J. B.; Capel, M.; Chance, M. R.; Gaasterland, T.; Lin, D. W.; Sali, A.; Studier, F. W.; Swaminathan, S. *Nat. Genet.* **1999**, *23*, 151–157.
 (4) Montelione, G. T.; Anderson, S. *Nat. Struct. Biol.* **1999**, *6*, 11–12.
 (5) Zarembinski, T. I.; Hung, L. W.; Mueller-Dieckmann, H. J.; Kim, K. K.; Yokota, H.; Kim, R.; Kim, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 15189–15193.
 (6) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

(7) Skolnick, J.; Fetrow, J. S. *Trends Biotechnol.* **2000**, *18*, 34–39.
 (8) Fetrow, J. S.; Skolnick, J. *J. Mol. Biol.* **1998**, *281*, 949–968.
 (9) Holm, L.; Sander, C. *J. Mol. Biol.* **1993**, *233*, 123–138.
 (10) Zimmerman, D. E.; Kulikowski, C. A.; Huang, Y. P.; Feng, W. Q.; Tashiro, M.; Shimotakahara, S.; Chien, C. Y.; Powers, R.; Montelione, G. T. *J. Mol. Biol.* **1997**, *269*, 592–610.
 (11) Gronwald, W.; Willard, L.; Jellard, T.; Boyko, R. E.; Rajarathnam, K.; Wishart, D. S.; Sonnichsen, F. D.; Sykes, B. D. *J. Biol. NMR* **1998**, *12*, 395–405.
 (12) Bartels, C.; Guntert, P.; Billeter, M.; Wüthrich, K. *J. Comput. Chem.* **1997**, *18*, 139–149.
 (13) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1697–1697.
 (14) Hansen, M. R.; Rance, M.; Pardi, A. *J. Am. Chem. Soc.* **1998**, *120*, 11210–11211.
 (15) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279–9283.

internuclear vector relative to the principal axis system of an order tensor, and the S_{kl} are elements of this order tensor. Clearly expressions such as this can be inverted to give constraints on the orientation of internuclear vectors. There are many recent examples of using residual dipolar coupling as a source of information for the determination of protein structure.^{17–22}

So far all applications of residual dipolar couplings rely on prior assignment of the resonances from which anisotropic interactions are measured. But this may not be essential. Structure determination can be placed ahead of assignment and orientational constraints can be used at a later point to resolve chemical shift ambiguities that inhibit prior resonance assignment. An example of something similar comes from solid-state NMR where dipolar coupling and chemical shift anisotropy patterns have been used to directly assess structure.^{23,24}

Given the possibility of abandoning prior assignment, a reevaluation of the need for isotope enrichment is also in order. Most protocols for backbone resonance assignment require high levels of ^{13}C enrichment in addition to ^{15}N enrichment. This is primarily needed for experiments that include ^{13}C – ^{13}C transfers to side chain atoms to facilitate amino acid type identification, HNCACB and CACB(CO)NH experiments, for example. These are in fact low sensitivity, time-consuming, experiments compared to those that do not involve ^{13}C – ^{13}C transfers.²⁵ ^{13}C uniform labeling at high levels is also expensive, particularly for proteins that do not express well in bacterial culture. Thus, release from a requirement of prior assignment may allow savings in both time and resources.

The work presented here uses dipolar couplings to define structural features of fragments ahead of resonance assignments and then uses the same couplings directly in combination with backbone chemical shifts to connect these fragments in a sequence-specific way. Since structural constraints in addition to chemical shift make the connection, the process is robust, and reliance on identification of amino acid type for placement in the protein sequence is reduced. The procedure is illustrated with an application to structure determination of a well-characterized, highly soluble, small protein, the Zn-substituted, triple mutant (W3Y, I23V, and L32I) rubredoxin from *Pyrococcus furiosus*. There are several crystal structures of closely related forms for structural comparison and validation of the method proposed here.²⁶ The protein is small enough (54 amino acids) and soluble enough (4.5 mM) to explore the potential of simultaneous backbone structure determination and resonance assignment with a high level of ^{15}N enrichment, but ^{13}C at natural abundance.

(16) Al-Hashimi, H. M.; Valafar, H.; Terrell, M.; Zartler, E. R.; Eidsness, M. K.; Prestegard, J. H. *J. Magn. Reson.* **2000**, *143*, 402–406.

(17) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, in press.

(18) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142–2143.

(19) Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283–294.

(20) Andrec, M.; Du, P. C.; Levy, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 1222–1229.

(21) Hus, J. C.; Marion, D.; Blackledge, M. *J. Am. Chem. Soc.* **2001**, *123*, 1541–1542.

(22) Fowler, C. A.; Tian, F.; Al-Hashimi, H. M.; Prestegard, J. H. *J. Mol. Biol.* **2000**, *304*, 447–460.

(23) Wang, J.; Denny, J.; Tian, C.; Kim, S.; Mo, Y.; Kovacs, F.; Song, Z.; Nishimura, K.; Gan, Z.; Fu, R.; Quine, J. R.; Cross, T. A. *J. Magn. Reson.* **2000**, *144*, 162–167.

(24) Marassi, F. M.; Opella, S. J. *J. Magn. Reson.* **2000**, *144*, 150–155.

(25) Sattler, M.; Schleucher, J.; Griesinger, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.

(26) Bau, R.; Rees, D. C.; Kurtz, D. M.; Scott, R. A.; Huang, H. S.; Adams, M. W. W.; Eidsness, M. K. *J. Biol. Inorg. Chem.* **1998**, *3*, 484–493.

Experimental Section

Materials. Rubredoxin was expressed with ^{15}N labeling and purified as described previously.²⁷ A sample with bacteriophage as an alignment medium was prepared to be 4.5 mM rubredoxin in 50 mM sodium phosphate and 100 mM NaCl at pH 6.3 with enough phage to give a 14 Hz splitting of the ^2H NMR signal from 10% $^2\text{H}_2\text{O}$ in the sample. A sample with bicelles as an alignment medium was prepared to be 2.5 mM rubredoxin in 50 mM sodium phosphate and 50 mM NaCl at pH 6.3. The 7% (weight/volume) DMPC/DHPC at a 3:1 molar ratio doped with CTAB (4% relative to DMPC) gave a 12 Hz splitting of the ^2H NMR signal from 10% $^2\text{H}_2\text{O}$ in the sample.

Spectroscopy. Spectra were acquired on Varian Inova 500 and 800 MHz spectrometers using gradient triple resonance probes (Varian Inc, Palo Alto, CA). For the collection of residual dipolar coupling data a phase-modulated HSQC experiment,²⁸ a soft HNCA-E. COSY experiment,²⁹ and a 2D IP-HSQC experiment³⁰ were selected. Both the original HNCA-E. COSY pulse sequence and the IP-HSQC pulse sequence were modified for acquiring dipolar couplings from natural abundance ^{13}C protein samples (the pulse sequences are provided in the Supporting Information). Since there are no ^{13}C – ^{13}C couplings in these samples, selective excitation and decoupling of carbon resonances are no longer needed. Also, a filter to select only ^{15}N magnetization with an adjacent ^{13}C was added to the original 2D IP-HSQC sequence.

Since most splittings of resonances in the spectra acquired are the sum of scalar and residual dipolar coupling, the samples were run under both isotropic and partially oriented conditions, and differences in splittings were used to extract residual dipolar contributions. A duplicate sample without phage was used as an isotropic reference. Data collection for the HNCA experiments typically included 12 t1 points, 64 t2 points, and 1024 t3 points over a period of 30 h with a repetition rate of 1 s. Data collection for the HSQC experiments included 96 t1 points and 1024 t2 points over 2 h with a 1 s repetition rate. Data collection for the IP-HSQC experiments included 320 t1 points and 1024 t2 points over 10 h with a 1 s repetition rate. All data processing, including extraction of frequency differences for coupling constant measurements, was done with the FELIX98 package (Accelrys, San Diego, CA).

Data Analysis. Analysis of residual dipolar couplings was carried out by using a combination of previously described or distributed software (Orderten_SVD³¹ and BACKCALC_RDC, available from our website, <http://tesla.ccrcc.uga.edu>) and software written explicitly for this application. The additional software is in the form of Perl scripts and C programs. Searches for allowed geometries and allowed residue connectivities were carried out on a Silicon Graphics O2 workstation. Computation time was highly dependent on levels of ambiguity in the data, but it took approximately 2 min per residue for the extension of starting fragments.

Results

Data Acquired. Obtaining sufficient dipolar couplings for simultaneous structure determination and assignment could be a challenge, especially when dealing with proteins labeled with only ^{15}N . With recent improvements in the sensitivity of NMR instruments, however, it is feasible to do the most sensitive triple resonance experiments such as HNCO and HNCA on ^{15}N only labeled proteins with ^{13}C at natural abundance. Losses in moving these experiments from a 90% labeled sample to a natural abundance sample are not as great as the factor of 90 that one might anticipate. When HNCA type experiments are run on a natural abundance ^{13}C sample, the magnetization of an amide proton is transferred to either inter or intra C^α spins, but not

(27) Tian, F.; Fowler, C. A.; Zartler, E. R.; Jenney, F. A.; Adams, M. W.; Prestegard, J. H. *J. Biol. NMR* **2000**, *18*, 23–31.

(28) Tolman, J. R.; Prestegard, J. H. *J. Magn. Reson.* **1996**, *112*, 245–252.

(29) Weisemann, R.; Ruterjans, H.; Schwalbe, H.; Schleucher, J.; Bermel, W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 231–240.

(30) Wang, Y. X.; Marquardt, J. L.; Wingfield, P.; Stahl, S. J.; Huang, S.; Torchia, D.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 7385–7386.

(31) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334–342.

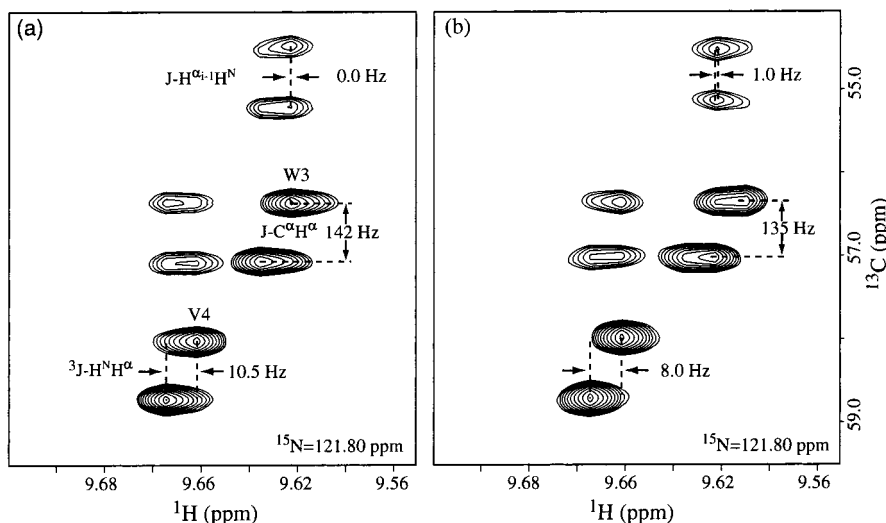


Figure 1. Soft HNCA-E. COSY spectra of rubredoxin under (a) isotropic and (b) phage aligned conditions. The C^α chemical shifts, C_i^α to C_{i-1}^α connectivities, and three-bond $H^N H^\alpha$ scalar coupling constants can be obtained from spectrum (a), and $C^\alpha-H^\alpha$ dipolar couplings and both inter- and intraresidue $H^N H^\alpha$ dipolar couplings can be obtained from differences between the measured couplings in these two spectra.

both, while on ^{13}C uniformly labeled samples, the magnetization of an amide proton is split between inter and intra C^α spins, reducing transfer to any one connectivity by an average factor of 2. In addition, there are carbon-carbon couplings and additional spin relaxation pathways in 90% ^{13}C samples that result in further losses. As a result, the sensitivity loss for an HNCA experiment on moving from 90% to 1% natural abundance ^{13}C sample is less than a factor of 40. We would also like to acquire several pieces of dipolar data in a single experiment to reduce the data acquisition time. The soft HNCA-E. COSY experiment²⁹ was chosen for collecting experimental data because it has many of the features we desire. Likewise the IP-HNCO experiment provides multiple couplings and allows compensation for reduced levels of ^{13}C labeling.

Figure 1 shows two spectra from soft HNCA-E. COSY experiments on ^{15}N -labeled rubredoxin. The spectrum on the left was acquired under isotropic conditions, and the spectrum on the right was acquired under aligned conditions using a phage medium. The information that can be extracted from these two spectra includes C^α chemical shifts, C_i^α to C_{i-1}^α connectivities, three-bond $H^N H^\alpha$ scalar coupling constants, $C^\alpha-H^\alpha$ dipolar coupling constants, and both inter- and intraresidue $H^N H^\alpha$ dipolar coupling constants.

Two-bond $^{13}C'-^1H$ and one-bond $^{13}C'-^{15}N$ dipolar coupling constants were measured from the same natural abundance ^{13}C protein samples by using the modified 2D IP-HSQC experiment described in the Experimental Section. The one-bond $^{15}N-^1H$ couplings were measured by an intensity-based, phase-modulated 2D HSQC experiment.²⁸ The measured dipolar couplings are listed and an example of data from the IP-HSQC experiment is given in the Supporting Information. There are a few residues in the rubredoxin sequence, including the last three, which show significant chemical shift changes in the bicelle aligned medium. Their couplings were excluded from the following analysis. In all, about 400 dipolar couplings, 37 three-bond $H^N H^\alpha$ scalar couplings, and 44 C^α chemical shifts were used in the analysis.

Structures of Starting Points. The dipolar couplings mentioned above can be assembled into groups, connected to a single amide $N-H$ pair. Potentially, there will be six dipolar couplings ($C^\alpha-H^\alpha_i$, $N_i-H^N_i$, $H^N_i-H^\alpha_i$, $H^N_i-H^\alpha_{i-1}$, $C'_{i-1}-H^N_i$, $N_i-C'_i$) associated with each residue, for each alignment medium used. In principle five dipolar couplings should be adequate for the

determination of the orientation of a peptide segment independent of other information (there are five independent order tensor elements in eq 1). This simplistic view fails in our case because the peptide segments include variable torsion angles; relative to the amide peptide plane containing the N_i-H_i vector, two of the couplings, the intraresidue $C^\alpha-H^\alpha$ coupling and the intraresidue $H^N H^\alpha$ coupling, depend on ϕ and a third, the interresidue $H^N H^\alpha$ coupling, depends on both ϕ_{i-1} and φ_{i-1} . Even with scalar couplings to restrict ϕ values, this leaves us with marginal data. This problem can be resolved by dividing the structure determination process into two steps, using data from two or more connected residues as a starting point and then extending the starting point one residue at a time after some global parameters are determined. Connecting two sequential residues provides six additional couplings while contributing only two more variables (ϕ_{i-1} , φ_{i-1}). Once started, the order parameters are defined and each successive addition involves identifying only the two variables, ϕ and φ . It is of course better if several segments of more than two residues can be connected as a starting point.

A good way to connect two residues and their associated dipolar information is to look for coincidence of an interresidue C^α chemical shift of the first residue with an intraresidue C^α chemical shift as seen from another residue. Normally chemical shift identities are not adequate to uniquely connect very many residues because of chemical shift degeneracies, but we need only a few starting points. In rubredoxin one fragment of 5 residues could be connected together from C_i^α to C_{i-1}^α chemical shift information alone. This fragment corresponds to residues 2-6 in the sequence. For the convenience of description we use residue numbers to label this fragment, but we do not need to map the connected residues to the primary sequence at this point.

About 40 dipolar couplings were measured for residues in this first peptide fragment. The backbone structure of this fragment was constructed from the best fit to these measurements by using a χ^2 function defined as follows:²¹

$$\chi^2 = \sum \left(\frac{D_i^{\text{exp}} - D_i^{\text{cal}}}{\sigma_i} \right)^2$$

Here D_i^{exp} and D_i^{cal} are the calculated and measured dipolar couplings respectively and σ_i is an estimate of error. D_i^{cal} was

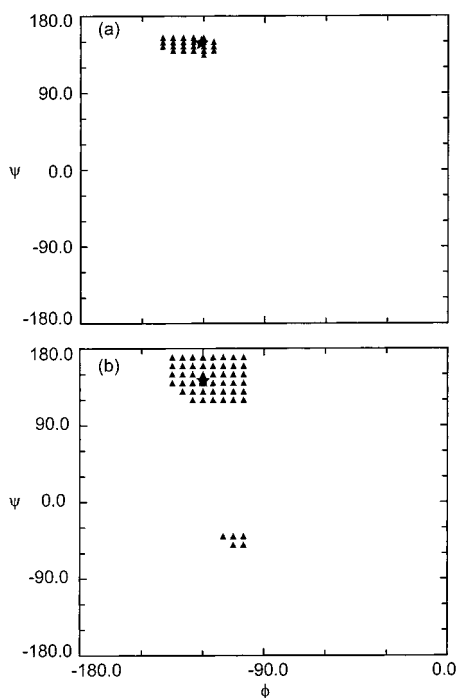


Figure 2. Solutions for ϕ and φ with χ^2 below 20% of the maximum: (a) residue 2 and (b) residue 3. The stars indicate values taken from the X-ray structure (1BRF).

calculated by using eq 1 while assuming a standard peptide plane geometry. For σ_i , 5% of the largest dipolar coupling measured from each aligned medium is added to the experimental uncertainty to account for variability in peptide plane geometry. The S_{zz} , S_{xx} , and S_{yy} of eq 1 written in the principal axis frame were initially determined from distributions of the dipolar couplings.³² The χ^2 function was then minimized for three residues at a time by searching the two pairs of torsion angles (ϕ, φ) defining the local structure, and the two sets of Euler angles (α, β, γ) defining the orientations of the principal axes for the order tensors in two media. Solutions for ϕ and φ of this fragment from initial searches are plotted in Figure 2. All solutions with a χ^2 less than 20% of the maximum are shown. The stars indicate where ϕ and φ values from the X-ray structure lie. While a single cluster of solutions is found for residue 2, two clusters of solutions were found for residue 3. Note that the two solutions are related by a 180° rotation about φ . This is an example of the 4-fold orientational degeneracy normally associated with structures from residual dipolar coupling data. While the data from the second alignment medium reduces the degeneracy, the extent to which this occurs depends on the particular orientations of dipolar vectors, the relative orientations of principal axis systems, and the experimental errors. Distance constraints from NOEs, if available, can be very useful in further reducing such degeneracies, but here the correct solution was selected based on its lower χ^2 value. All five residues were then combined, and all ϕ and φ angles, and Euler angles, were adjusted to minimize the χ^2 between all measured and calculated dipolar couplings. The magnitudes of order parameters were not refined in the present case due to the high accuracy of the initial estimates.

From the structure of the starting fragment, the relative orientations of order frames in phage and bicelle can be defined. The Euler rotations relating the alignment tensor for rubredoxin in bicelles to that for rubredoxin in phage were determined to

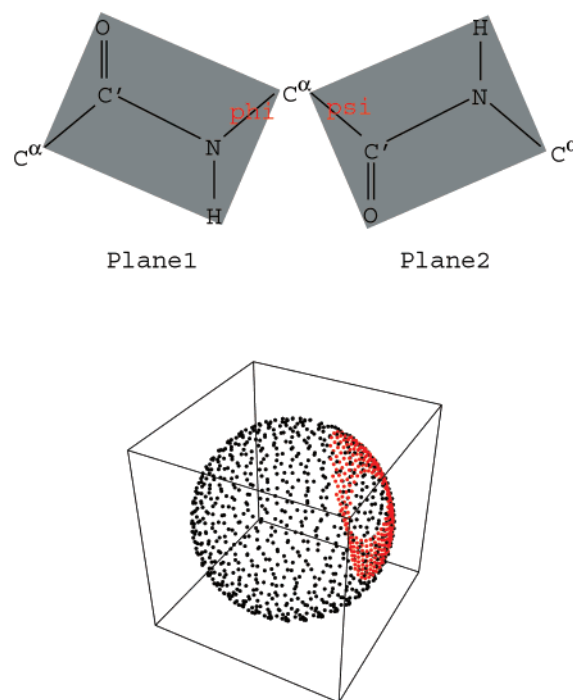


Figure 3. Illustration of the restrictions on allowed orientations for a sequential residue in a dipeptide fragment. Starting from the known orientation of peptide plane 1, possible orientations of plane 2 can be calculated as a function of ϕ and φ . Red dots on the sphere represent the allowed orientations of the normal to peptide plane 2 as ϕ rotates from -20 to 20° and φ rotates from -180 to 180° . The orientations sampled by plane 2 are very limited.

be 22° , 92° , and 125° (plus three degenerate sets). These, along with the principal order parameters, can be used as globally fixed parameters in all subsequent fragment constructions and fragment extensions. While the longest cluster was used to establish global parameters, there will usually be several clusters of two or more residues that can be connected based on D_i^{exp} and D_i^{cal} chemical shift information; these can also be used as starting points. For rubredoxin, there were 11 such starting points. For fragments with more than one structural solution, the angles with the minimum χ^2 of each fragment were saved and each was used as a potential starting point. Usually, only a correct solution will allow extension and connection with other fragments.

Extension of Starting Fragments. The utility of dipolar coupling based structural information in selecting appropriate residues for extension of fragments is illustrated in Figure 3. Beginning with the orientation of a single fragment, which can be as small as a peptide plane (plane 1), the possible orientations that a sequential peptide plane (plane 2) can adopt are restricted to those accessible by rotations about ϕ and φ . If the rotation about ϕ can be constrained to $\pm 20^\circ$ by using the relationship of measured $^3\text{J}-\text{H}^{\text{N}}\text{H}^\alpha$ scalar couplings to the ϕ torsion angle, the allowed orientations are further restricted. The red portion of the sphere in Figure 3 represents allowed orientations of the normal to peptide plane 2 as a function of ϕ and φ rotations. Obviously, the orientations sampled by plane 2 are very limited.

From the order parameters for a starting fragment (plane 1) and each allowed orientation of plane 2 (a certain ϕ and φ combination), one can use eq 1 to calculate the corresponding sets of dipolar couplings for a connected residue. These values can be screened against experimental sets. Normally several acceptable sets are found. Other experimental criteria, such as coincidence of chemical shifts for α carbons as measured from

(32) Clore, C. M.; Gronenborn, A. M.; Bax, A. *J. Magn. Reson.* **1998**, *133*, 216–221.

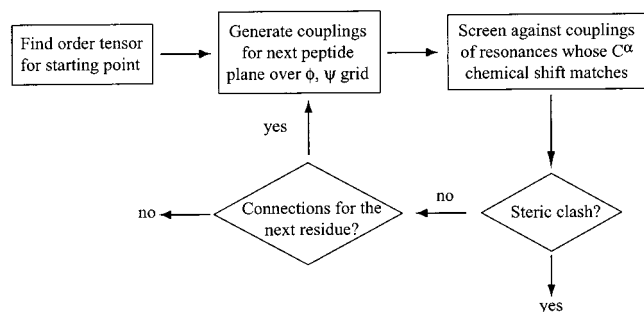


Figure 4. Flowchart illustrating a strategy for simultaneous protein backbone resonance assignment and structure determination.

i and $i - 1$ N-H amide pairs, can then be used to further reduce the options for selection of a connected residue. Steric clash and an inability to find additional sequential matches can be used to further reduce selections if necessary. As a bonus, the match may exist for only certain particular ϕ and φ combinations, fixing the structural relationship between two residues at the same time.

As in the case of our starting fragments, there are some additional ambiguities that are due to the fact that absolute orientations of the fragments are 4-fold degenerate because dipolar and other NMR data on anisotropic parameters are insensitive to 180° rotations about the principal axes of the order tensor. These ambiguities can be resolved in several ways such as distance constraints from NOEs or relative orientations of bond vectors from cross-correlation effects,³³ but here heavy reliance was placed on dipolar couplings from a second alignment medium.¹⁶ The general procedures used are summarized in the flowchart shown in Figure 4.

A 17-residue stretch of rubredoxin, which turns out to be that from residue 2 to residue 18, can be used as an illustration of structure determination by extension of starting fragments. Two of the 11 possible starting points were used in the initial construction of this stretch, fragment I and fragment III (latter we will see that fragment I corresponds to residues 2–6 and fragment III corresponds to residues 11–14). Extension of fragment III in the amino terminus direction was first attempted. The C_{i-1}^α chemical shift of the first residue in fragment III (60.45 ppm) matches within experimental error (± 0.08 ppm) the C_i^α chemical shifts in two unconnected groups of data, 60.44 and 60.49 ppm (these will turn out to be data associated with residues 10 and 45). ϕ and φ angles which would connect these residues to fragment III were varied and dipolar couplings were calculated at each grid point. Comparing calculated couplings to experimental values, it is not possible to find a match with values in the second group. Therefore, a unique connection can be made to, and the related torsion angles (ϕ , φ) can be defined for, the residue associated with the first group. The C_{i-1}^α chemical shift for the residue associated with the first group of couplings has an unusual value (45.88 ppm). This can only come from a glycine residue. Applying a similar process to extension in the C-terminal direction, connections to groups of couplings representing three additional residues can be made. Here an additional glycine is encountered. Two glycines separated by 7 residues are unique to the 10–17 segment of rubredoxin. Hence, both the structure and the sequential assignment for this segment are determined.

For fragment I, the C_i^α chemical shift of last resonance, 59.08 ppm matches the C_{i-1}^α chemical shifts in three groups of data, 59.07, 59.10, and 59.04 ppm (corresponding to residues

7, 9, and 42). Comparing the calculated dipolar couplings with measurements for these groups, connection to one group, that with a C_i^α chemical shift of 59.10, is not orientationally allowed. Although the group with a C_i^α chemical shift of 59.04 remains a possible choice for connection, a subsequent connection cannot be found. Therefore, the group with a C_i^α chemical shift of 59.07 ppm is uniquely connected to the last residue of fragment I. After two more residues are added to this fragment, a unique chemical shift for a glycine residue occurs and the fragment is mapped to sequence positions 2–8.

The connection between residue 8 of fragment I and residue 10 of fragment III also cannot be made simply based on chemical shifts of C^α s. Two groups of data, which have very similar C_i^α and C_{i-1}^α chemical shifts (45.92, 59.10 and 45.88, 59.04 ppm) are potential candidates for this position. The choice here is made based on a combination of allowed dipolar couplings and the now restricted orientations of fragments I and III. Once the choice is made the structure of, and the assignments for, the first 18 residues of rubredoxin are determined. The local structure of the 2–18 fragment is shown in Figure 5a (light blue), which has root-mean-square deviation of 1.2 Å for backbone atoms with respect to the corresponding atoms in the high-resolution X-ray structure (1BRF, white).

Assembly of Fragments. The remaining groups of data could be assembled in a similar fashion into five separate fragments. Each fragment terminates with a connection to a C_{i-1}^α chemical shift that is unique and appears nowhere in the list of N-H connected C_i^α shifts. Examination of the rubredoxin sequence reveals 5 prolines. Proline has a secondary amide nitrogen that does not provide $\text{H}^{\text{N}}-C_i^\alpha$ connections. Hence each fragment must be separated by a proline. Based simply on this fact, and the length of the unassigned fragments, each fragment could be inserted into the sequence. If torsion angles (ϕ , φ) of proline and its previous residue can be specified, a complete backbone structure can be determined. As in the case of connecting residue 8 to residue 10 above, a proper connection across proline should maintain the local structures and orientations of the two fragments it connects. Also proline adopts a very limited range of ϕ angles. Initially assuming all prolines to have a trans amide bond configuration, the torsion angle, ϕ , was searched between -40 and -80° and φ was searched from -180 to 180° . The torsion angle, ϕ , of the previous residue is constrained by its own $^3\text{J}-\text{H}^{\text{N}}\text{H}^\alpha$ scalar coupling as well as $\text{H}^{\text{N}}\text{H}^\alpha$ intraresidue and $\text{C}^\alpha\text{H}^\alpha$ intraresidue dipolar couplings. In all cases, a single cluster of solutions for these torsion angles was found. A complete structure of rubredoxin (residue 2–50) is shown in Figure 5b (light blue). With respect to the corresponding atoms in the high-resolution X-ray structure (white) it has a root-mean-square deviation of 1.8 Å over all backbone atoms of residues 2–50.

Discussion and Conclusion

The results presented above demonstrate that it is possible to determine an accurate backbone structure of a protein without the necessity of prior resonance assignment. The method employed to accomplish this task utilizes recent advances in high-resolution NMR spectroscopy that exploit measurement of residual dipolar couplings in biomolecules weakly aligned by magnetic field oriented media. The dipolar couplings provide not only orientational constraints on the overall structure but a filter for assembly of residues in a sequential manner. The dipolar couplings are not sufficient in themselves for resonance assignment, but when combined with limited information from chemical shifts and backbone scalar couplings, an assignment and structure emerge simultaneously.

(33) Reif, B.; Hennig, M.; Griesinger, C. *Science* **1997**, *276*, 1230–1233.

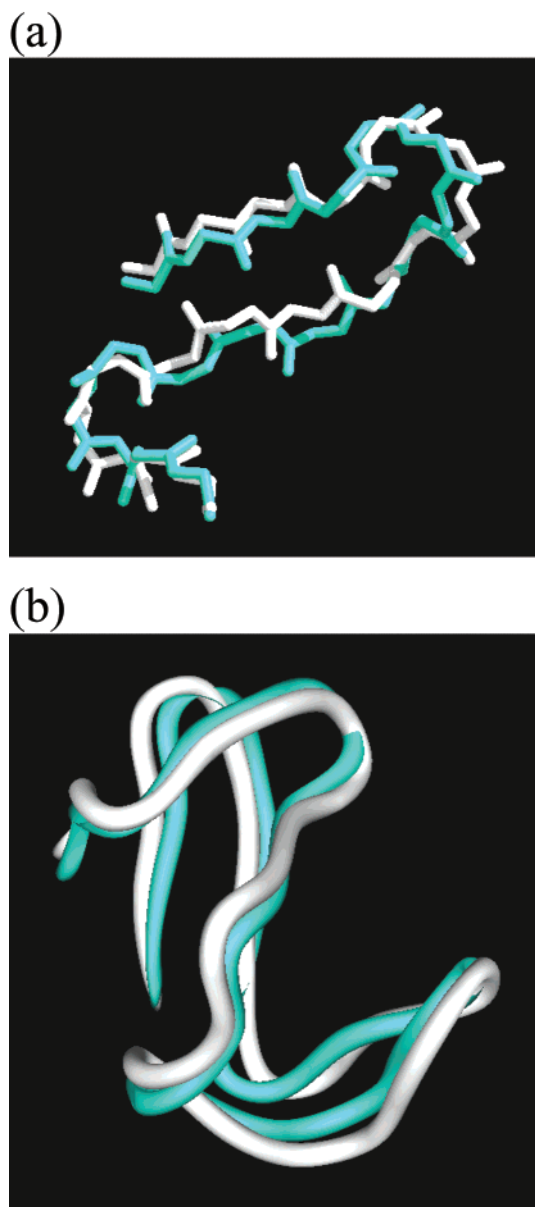


Figure 5. Comparison of the structure of rubredoxin from this study (light blue) and a high-resolution X-ray structure (white) (1BRF): (a) residue 2-18 and (b) residue 2-50.

Experimental data used in this study were collected by using two 2D and one 3D experiments in one isotropic and two aligned media. Total acquisition time was less than one week with use of a protein that was ^{15}N enriched but not enriched above 1% natural abundance in carbon. The experiments employed are sufficiently sensitive to give data involving couplings to ^{13}C sites at natural abundance, but substantial reductions in data acquisition times would be possible with partial ^{13}C enrichment.

Other improvements in efficiency may come from concatenation of experiments. Only a small percentage of the magnetization associated with amide protons (1% for natural abundance ^{13}C samples) is used during any one experiment involving transfers to ^{13}C , and different parts of the magnetization are used in experiments involving transfer to different carbons.

These advantages, as well as advantages associated with improved relaxation properties, can be retained with 10–15% ^{13}C enrichment. Between concatenation and partial enrichment, data acquisition times on small soluble proteins will be reduced to less than a day. This is critical for the structural genomics initiative, where protein expression and structural determination in a high-throughput manner will be needed. Beyond these applications, the ability to determine structure independent of complete residue assignment is also useful where segments of proteins are of interest. For example, it is possible to look at protein backbone structures near the active site in the presence of various ligands without determining full structures.

The backbone structure determined above is of course not a complete structure. It is nevertheless a high-resolution structure. This may be important for computational approaches to the addition of side chains. Several groups have now showed that it is possible to computationally determine at least internal side chain conformations if accurate backbone atom positions are available.^{34–36} In addition, a well-defined backbone structure can facilitate the assignment of NOESY spectra, which is a prerequisite for the full structure determination by traditional NOE-based methods.

The protein to which we have applied this simultaneous structure determination and assignment approach is small. For larger proteins, the C_i^α chemical shifts and orientations of the peptide planes are likely to be more degenerate as a result of the increased number of resonances. Also, precision of measurement of small dipolar couplings will decrease. However, we anticipate that these losses can be compensated by addition of other experiments. ^{15}N edited NOESY, for example, can be added to the list of experiments to provide additional information at a modest cost in time. Although NOE experiments, in general, are not considered to be especially sensitive, a focus on short- and medium-range NOEs for backbone resonance assignments will use NOEs that are among the most easily obtained.³⁷ Furthermore, the sequential backbone NOEs provide additional constraints on torsion angles (ϕ , φ). The NOESY experiment might also be concatenated with the experiments described above to minimize acquisition time. Applications to test protein size limits and improve the efficiency of the approach are in progress.

Acknowledgment. This work is supported by grants from the National Institutes of Health, GM 062407, and the National Science Foundation, MCB 9726341. We would like to thank Dr. F. A. Jenny, Jr., and Dr. M. W. Adams for providing the rubredoxin sample.

Supporting Information Available: Three figures show the modified pulse sequences for the soft HNCA-E. COSY and IP-HSQC experiments, and an example of IP-HSQC data, and two tables list the dipolar couplings for rubredoxin in aligned media (PDF). This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA011806H

(34) Looger, L. L.; Hellinga, H. W. *J. Mol. Biol.* **2001**, *307*, 429–445.
 (35) Shenkin, P. S.; Farid, H.; Fetrow, J. S. *Proteins* **1996**, *26*, 323–352.

(36) Wilson, C.; Gregoret, L. M.; Agard, D. A. *J. Mol. Biol.* **1993**, *229*, 996–1006.

(37) Doreleijers, J. F.; Ravés, M. L.; Rullmann, T.; Kaptein, R. *J. Biomol. NMR* **1999**, *14*, 123–132.