JMB

Rapid Determination of Protein Folds Using Residual Dipolar Couplings

C. Andrew Fowler[1,2], Fang Tian[1], Hashim M. Al-Hashimi[1,2] and James H. Prestegard[1]*

[1]*Complex Carbohydrate Research Center, The University of Georgia, 220 Riverbend Road, Athens GA 30602-4712, USA*

[2]*Department of Chemistry, Yale University, PO Box 208107 New Haven, CT 06520-8107, USA*

*Corresponding author

Over the next few years, various genome projects will sequence many new genes and yield many new gene products. Many of these products will have no known function and little, if any, sequence homology to existing proteins. There is reason to believe that a rapid determination of a protein fold, even at low resolution, can aid in the identification of function and expedite the determination of structure at higher resolution. Recently devised NMR methods of measuring residual dipolar couplings provide one route to the determination of a fold. They do this by allowing the alignment of previously identified secondary structural elements with respect to each other. When combined with constraints involving loops connecting elements or other short-range experimental distance information, a fold is produced. We illustrate this approach to protein fold determination on ^{15}N-labeled *Eschericia coli* acyl carrier protein using a limited set of ^{15}N-^{1}H and ^{1}H-^{1}H dipolar couplings. We also illustrate an approach using a more extended set of heteronuclear couplings on a related protein, ^{13}C,^{15}N-labeled NodF protein from *Rhizobium leguminosarum*.

© 2000 Academic Press

Keywords: structural genomics; NMR; ACP; NodF; protein structure

Introduction

Current efforts directed toward sequencing the complete genomes of various species are producing results at an astonishing rate. Many of the identified genes code for proteins with no known function, and have a low level of sequence homology to previously characterized proteins.[1,2] Structural genomics provides a route to functional characterization by firstly obtaining the three dimensional structures of large numbers of representative proteins.[1,3,4] Various computational methods would then be used to predict specific activities or other properties of biological

Present address: H. M. Al-Hashimi, Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA.

Abbreviations used: NOE, nuclear Overhauser effect; ACP, acyl carrier protein; FFFs, fuzzy functional forms; IPTG, isopropyl β-D-thiogalactopyranoside; SDS-PAGE, sodium dodecyl sulfate - polyacrylamide gel electrophoresis; HPLC, high performance liquid chromatography; LB, Luria-Bertani broth; POSE, positioning of structural elements.

E-mail address of the corresponding author: jpresteg@ccrc.uga.edu

interest.[5–9] The challenge arises in the shear number of structure determinations which must be undertaken. Here we discuss one approach that can help meet this challenge.

Nuclear magnetic resonance spectroscopy (NMR) provides a means of obtaining structural information that complements X-ray crystallography by being applicable in solution to proteins that may be resistant to a crystallographic approach. Traditional NMR techniques for the determination of macromolecular structure, however, are not well suited to genomics applications because of the vast amount of time required to produce a structure. These methods rely heavily upon refinement against distance constraints obtained from NOE measurements. Because of the steep distance dependence ($1/r^6$), and resultant short-range character of NOEs, critical fold determining constraints often come from side-chain to side-chain contacts in the hydrophobic core of proteins. This results in the requirement that the vast majority of all observed resonances in the NMR spectrum are assigned, including those at the very extremes of amino acid side-chains. Much of the time involved in determining a structure is consumed during the resonance assignment and NOE identification

stage. Software is currently available that can at least partially automate these stages;[10–16] however, the minimum set of experiments that must be performed in order for these routines to work reliably can often require well over a month of data acquisition. With the number of samples expected in any sort of high throughput genomics application, a more rapid means of extracting structural information is required.

Recently developed NMR methodology based on the measurement of residual dipolar couplings in partially ordered media can provide an alternative means of rapidly obtaining structural information.[17, 18] Residual dipolar couplings complement NOE data in that they provide angular restraints that can be used to orient remote structural elements in a protein, even when the intervening distance is too great to allow a determination using NOE-based methods.[19–22] One can think of any reasonably rigid piece of a protein as a structural element; in particular, regular elements of secondary structure, such as α or 3_{10}-helices, β-sheets, and reverse turns, can serve as these elements. The identification of these elements can be done relatively easily from short and intermediate-range NOEs involving backbone protons, or directly from the experiments needed to make sequential backbone resonance assignments.[23,24] Once the secondary structural elements are identified and backbone resonance assignments are made, residual dipolar couplings adequate for fold determination can be measured using backbone resonances alone. Here, we focus entirely on those residual dipolar couplings which exist between amide protons and directly bonded amide nitrogen atoms ($^1D_{NH}$), alpha carbon atoms and directly bonded alpha protons ($^1D_{CH}$), intra or inter-residue proton-proton pairs involving amide and alpha protons (D_{HH}), and a given carbonyl carbon atom and the amide nitrogen atom or proton of a sequential residue ($^1D_{NC}$, $^2D_{HC}$).

The magnitude of an observed dipolar coupling is given by:

$$D_{ij} = \frac{-\mu_0 \gamma_i \gamma_j \hbar}{4\pi^2 r^3} \left\langle \frac{3\cos^2 \theta - 1}{2} \right\rangle \qquad (1)$$

where γ_i and γ_j are the gyromagnetic ratios of the ith and jth nuclei, r is the distance between nuclei i and j, and θ_i is the angle between the internuclear vector and the magnetic field (see Figure 1). The angle brackets denote a time average; in isotropic solution the angular term averages to zero and dipolar couplings are not observed, but if the sample is partially oriented we can measure a residual dipolar coupling which manifests itself in a manner similar to scalar couplings. Provided the measurement of at least five independent residual dipolar couplings per structural element can be be made, the preferred orientation of all elements can be determined.

An order matrix approach proves convenient for determination of element orientation.[25] The angular
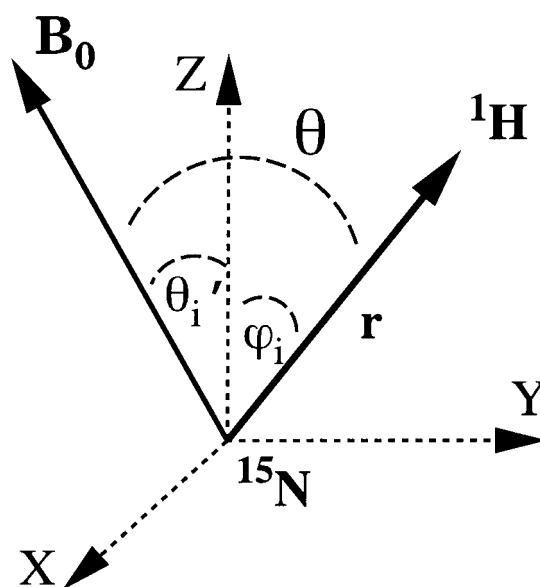

Figure 1. Diagram illustrating the parameters which are required to define a dipolar coupling and the elements of the Saupe order matrix.

term in equation (1) can be rewritten as:

$$\left\langle \frac{3\cos^2 \theta_k - 1}{2} \right\rangle = \sum_{i,j=\{x,y,z\}} S_{ij} \cos \phi_{ik} \cos \phi_{jk} \qquad (2)$$

where ϕ_{ik} is the angle between the kth internuclear vector and the ith axis of an arbitrarily chosen coordinate frame and S_{ij} are Saupe order parameters. These are given by:

$$S_{ij} = \left\langle \frac{3\cos^2 \theta_i' \cos^2 \theta_j' - \delta_{ij}}{2} \right\rangle \qquad (3)$$

where θ_i' is the angle between the ith axis of our arbitrarily chosen frame and the magnetic field, and δ_{ij} is the Kronecker delta function (Figure 1).

The order matrix parameters, only five of which are independent, can be determined from a set of measured couplings and the assembled order matrix diagonalized to give three Euler angles relating a secondary structure element to its principal alignment frame and two principal order parameters describing the extent and asymmetry of order (S_{zz} and $\eta = (S_{yy} - S_{xx})/S_{zz}$, respectively). For a rigid molecule, all structural elements must see the same molecular alignment frame; thus, we can orient all rigid elements of a molecule correctly with respect to each other by superimposing their alignment frames. Subsequent translation of elements to satisfy a minimum set of distance constraints produces a backbone structure, or protein fold.

We demonstrate the above approach on ^{15}N labeled *Escherichia coli* acyl carrier protein (ACP), a

77-amino acid protein that carries a phosphopan-
tetheine prosthetic group at Ser36. ACP is one of
the most abundant soluble proteins in *E. coli*, shut-
tling growing acyl chains during fatty acid biosyn-
thesis, as well as serving as an acyl chain donor for
a variety of enzymes and cellular processes.[26,27]
We also illustrate a variation of our method which
makes use of a much larger number of dipolar
couplings on ^{13}C, ^{15}N doubly labeled NodF protein
from *Rhizobium leguminosarum*, a 92-amino acid
protein. NodF, combined with the NodE gene pro-
duct, is required for the biosynthesis of host-
specific, polyunsaturated fatty acids which are
used in the production of lipochitin oligosaccharide
signals for symbiotic root nodulation on host
plants.[28,29] NodF also has a relatively high level of
homology with ACP, especially around the pros-
thetic group attachment site (S36 in ACP, S45 in
NodF), and, in fact, chimeras of the two proteins
which retain acyl donor activity have been
produced.[30] Thus, structural comparison of these
two proteins is of some interest.

Results

For ACP, both sequential and NOE assignments
have been previously reported based upon both
homonuclear and heteronuclear experiments.[31–33]
In a test of an ability to make assignments and
identify secondary structural elements using a
minimum of data sets, we examined ^{15}N-NOESY-
HSQC[34] and HNHA[35] scalar coupling data sets.
Only ^{1}H-^{15}N amide pairs that could be identified
as being sequentially connected and belonging to a
particular type of secondary structure (α-helix)
were used in subsequent analyses. In general, the
appearance of strong H^N-H^N connectivities in com-
bination with i to $i+3$ and i to $i+4$ H_α-H^N corre-
lations or a small value of $^{3}J_{HNHA}$ were deemed
adequate for this conclusion. Connected H^N reson-
ances were assigned to sequence specific positions
based on restriction to particular amino acid types
using previously reported RELAYED-COSY data.[31]
In a modern application the RELAYED-COSY data
would be replaced with data from a ^{15}N-TOCSY-
HSQC experiment. As an example of the procee-
dure, a stretch of five connected amino acid resi-
dues Gly-X-X-X-Gly was unambiguously assigned
to residues 12 through 16 of ACP, as this is the
only stretch of the primary sequence which has
two glycine residues separated by three interven-
ing amino acid residues. Similarly, the two con-
secutive alanine residues at residues 67 and 68 in
helix 3 are unique in the primary sequence. The
three helices identified and used for the final fold
determination ran from Ile3 through Glu13, Leu37
through Asp51, and Val65 through His75. Except
for truncation of the first helix by two residues,
these correspond to helices identified in the orig-
inal structural analysis.[32,36]

For NodF, initial assignments and identification
of secondary structural elements had been pre-

viously reported using ^{15}N-edited NOESY, TOCSY,
and HNHA experiments.[37,38] This actually corre-
sponds to a modern minimal set of experiments for
a ^{15}N-labeled protein, and only these experiments
were used in the initial backbone assignment. Total
data acquisition time was approximately five days
on a 4.5 mM sample using a 500 MHz spec-
trometer for this original set of experiments.
Because we had a doubly ^{13}C, ^{15}N labeled sample
of NodF and wished to illustrate our approach
using a broader range of dipolar couplings, we
also assigned the backbone resonances using stan-
dard triple resonance methods.[39,40] HNCACB,
CBCA(CO)HN, and HNCO experiments were run
on a 2.5 mM sample at 600 MHz, requiring a total
of roughly two days of instrument time in this
recent application. A small number of inconsisten-
cies were found when compared with the original
assignments, nearly all of which were at the termi-
ni of the protein or in extended loop regions. The
identification of helical regions was unaffected and
these are used as rigid structural elements for the
determination of a fold. The complete set of cor-
rected backbone resonance assignments is given in
Table S1 of Supplementary Material. The helices in
NodF used for fold determination ran from Leu5
through Val17, Leu46 through Leu58, and Asn76
through Gly86.

Spectra containing dipolar coupling information
were obtained for both ACP and NodF in phage-
containing media.[41] Details of the spectral acqui-
sition and the extraction of dipolar couplings are
given in Materials and Methods. Examples of some
of the spectra recorded can be found in Sup-
plementary Material. The residual dipolar coup-
lings used in the orientation of helical segments of
ACP and NodF, along with those NOEs sub-
sequently used in fold determination, are listed in
Tables 1 and 2, respectively. Errors in the measure-
ment of dipolar couplings are approximately
± 0.5 Hz for $^{1}D_{HN}$ and $^{1}D_{CH}$ couplings measured
using phase encoded experiments, and ± 1.0 Hz for
all other couplings. Larger errors were used in the
determination of the molecular alignment frame,
2-3 Hz for $^{1}D_{NH}$ and correspondingly larger errors
for other couplings ($^{1}D_{CH}$ of 3-5 Hz, D_{HH} of
4-6 Hz), mainly to account for deviations between
the actual structure of the molecular fragments and
our assumption of an idealized rigid geometry.

Analysis of dipolar couplings in terms of a struc-
ture was pursued using an order matrix-based pro-
gram, Orderten_SVD.[20] The program finds a set of
allowable solutions (typically 20,000, requiring less
than one minute) by selecting couplings from a
normal distribution about the input values, solving
a set of equations as given in equations (1) and (2),
generating and diagonalizing an order matrix, and
keeping any solutions consistent with the original
data. Solving the set of linear equations resulting
from equations (1) and (2) also requires knowledge
of a local structural element so that angles ϕ_i can
be calculated. Local structures used were idealized

Table 1. Data used to determine the backbone fold of ACP

	Couplings (Hz)					
	Helix 1		Helix 2		Helix 3	
A. *Dipolar couplings in α-helices*						
Amide couplings $(N_i\text{-}H_i^N)$	I3	1.4	L37	−2.6	Q66	8.2
	E4	0.4	D38	1.6	A67	7.7
	E5	3.4	T39	−0.3	I69	6.8
	V7	−1.0	V40	−2.6	D70	6.1
	K8	0.8			N73	7.4
	I10	2.0			G74	5.5
	I11	−0.3			H75	7.7
	G12	2.1				
	E13	2.1				
	Q14	0.0				
	L15	−1.9				
Amide-alpha couplings $(H_i^N\text{-}H_{\alpha i})$	R6	3.0	L37	0.4	Q66	0.0
	V7	−3.5	L42	0.0	A68	−8.5
	K9	4.5	V43	3.5	H75	−9.0
	L15	0.0	M44	2.0		
			L46	−2.0		
$(H_i^N\text{-}H_{\alpha i \pm 1})$			V43$_N$-L42$_\alpha$	2.0		
			M44$_N$-A45$_\alpha$	−2.0		
Amide-amide couplings $(H_i^N\text{-}H_{i+1}^N)$			D38	2.0	N73	2.0
			T39	2.0		
			V40	2.0		

B. *NOEs and distances used to position helices in POSE*
I3 H^N -F50 H$^\alpha$, V7 H^N-F50 H$_\zeta$, V7 H$^\alpha$-F50 H$_\zeta$, V7 methyl-A68 H^N, I3 methyl-N73 H^N
  6 Å   8 Å   8 Å   8 Å   8 Å

polyalanine α-helices generated using standard backbone torsional angles.[42]

The set of allowable solutions coming from Orderten-SVD is plotted as a distribution of points on Sauson-Flamsteed plots, as shown for the ACP data in Figure 2. The red, black, and blue spots depict the direction of the axes (z, y, and x, respectively) of the principle alignment frame as seen from each helical element. The spread of solutions gives an approximate picture of the precision with which the alignment frame can be determined. Since dipolar coupling data from a single aligning medium cannot detect inversion of axes, two clusters of solutions appear 180° apart for each of the axes. This is commonly known as the inversion problem;[43] four order frames which differ only in the signs of pairs of axes are allowed. The correct frame can be selected either by using a second aligning medium,[43] or by seeking consistency with other experimental distance information. Since the same set of coordinates was used as input into Orderten-SVD for each helix, the Sauson-Flamsteed projection of the order tensor frame for each helix qualitatively depicts the relative orientation of each
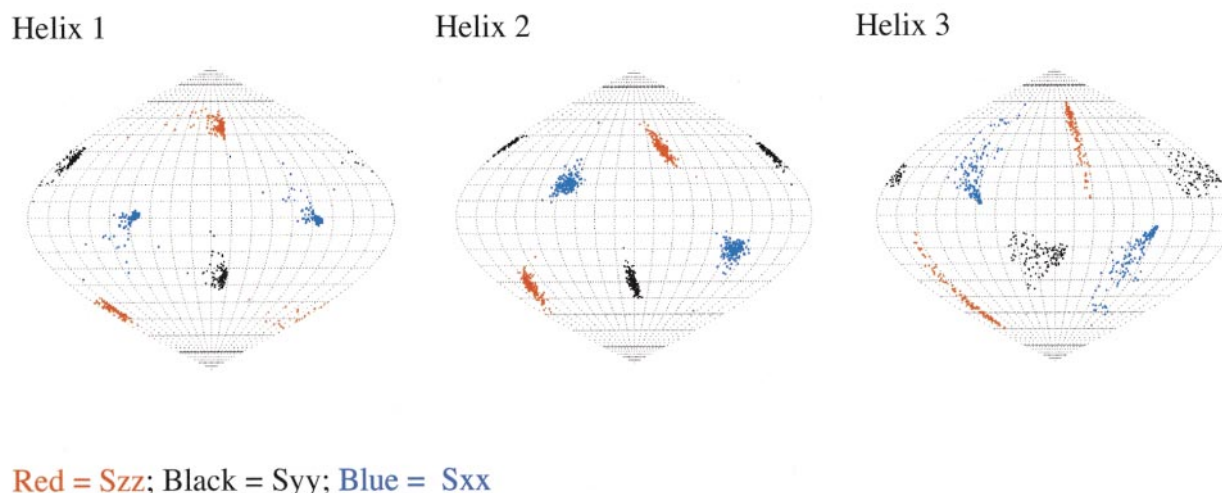

Helix 1   Helix 2   Helix 3

Red = Szz; Black = Syy; Blue = Sxx

Figure 2. Output from the Orderten-SVD program for the helices of ACP. The various spots correspond to the axes of the principle axis system which is defined by the molecular alignment frame for each helix and are displayed as Sauson-Flamsteed projections.

Table 2. Data used to determine the backbone fold of NodF

	Couplings (Hz)											
	Helix 1				Helix 2				Helix 3			
A. Dipolar couplings in α-helices												
Amide (^{15}N-^{1}H)	T6	−0.5	L7	7.4	L46	7.4	G47	4.0	N76	11.0	I77	8.5
	S11	7.6	N14	5.0	D50	6.2	W53	6.8	G78	7.8	E82	9.2
	K15	11.8			L58	6.0			G86	9.2		
Amide-alpha (^{1}H-^{1}H)	T6	−0.6	N14	5.0	L46	−3.8	D54	1.8	I77	−3.3	R85	−4.7
									I77-N76			1.0
Amide-amide (^{1}H-^{1}H)									I77-G78			0.5
									E82-V81			1.0
Alpha couplings (^{13}C-^{1}H)	L7	−4.2	I13	7.3	L48	−7.7	V51	4.7	I77	4.6	V81	0.2
	K15	−3.1			L52	−0.5	W53	−6.3	R85	1.4		
					D54	0.4	Q57	−3.7				
					L58	0.0						
Carbonyl (^{13}C′-^{15}N)	T6	−3.0	L7	4.9	A49	5.2	V51	−0.7	G78	0.1	V80	−0.3
	S11	1.5	A12	1.9	W53	−0.4	L58	3.7	V81	2.3	E82	−0.3
	I13	−1.7	N14	0.8					R85	2.3	G86	3.4
	K15	0.2										
Carbonyl (^{13}C′-^{1}H^{N})	T6	0.4	L7	1.3	A49	2.2	V51	−4.6	G78	−2.3	V80	−3.6
	S11	3.0	A12	0.0	W53	−3.4	L58	−0.9	V81	−3.7	E82	−1.2
	I13	1.1	N14	−0.3					R85	0.3	G86	2.3
	K15	−5.2										

B. NOEs and distances used to position helices in POSE

V7 H^{α}-E82 H^{α}, N14 H^{α}-G78 H^{α}, V17 H^{α}-A49 H^{α}, D54 H^{α}-I77 H^{N}, L58 H^{α}-I77 H^{N}

   8 Å   10 Å   7 Å   10 Å   7 Å

fragment. In the case presented, positions of axes, and in particular the z axis, which corresponds to the direction of highest order, do not appear to be very different when viewed from these standard helices. Therefore, the helices must be nearly parallel, and only small rotations are needed to superimpose order frames for molecular structure assembly.

Before rotating the individual helices into the correct order tensor frame as determined by Orderten-SVD, solutions were filtered for a consistent set of principal order parameters among the three helices. These turned out to be S_{zz} between -0.001 and -0.0015 and η between 0.65 and 0.9. The correctly aligned helices were then used as input to the program POSE for the determination of a backbone fold. This program was written to translate each rigid structural element through a set of grid points in order to find positions which satisfy a small number of experimental distance constraints that include NOE-derived distances, an approximate maximum extension of intervening loop regions, and the minimum distances for van der Waals contact. All NOE constraints are set to a distance slightly larger than that normally used in more traditional methods (6-7 Å *versus* 5 Å maximum), mainly to account for any deviations between true local structures and the idealized polyalanine helices used in our fold determination protocol. For NOEs which involve side-chain protons (in the case of ACP we use some NOEs to the ring of Phe50), the position of the NOE is set to that of the corresponding methyl group in the idealized helices, and an additional 2-3 Å is added to compensate for the difference in side-chain extension. A complete list of the NOE distance constraints used for ACP is given in Table 1B. Loop distance constraints were set explicitly to a maximum of 2.5 Å times the number of intervening residues. The program repeats the translational search for each of the four possible orientations of each structural element. All solutions which satisfy the data are recorded by the program for analysis.

The final backbone fold for ACP is depicted in Figure 3(a); for comparison, Figure 3(b) shows a previously determined NOE structure of ACP (PDB ID: 1ACP) in approximately the same orientation.[36] The fold is the average of an ensemble of approximately 900 structures with an RMSD of 2.5 Å. A second set of solutions was also produced, but it did not correspond to a closely packed structure and was discarded. In addition to positioning the three helices, we have managed to collect enough dipolar coupling data to orient a type I reverse β-turn, from residues Thr23 to Ala26, located in the very long loop region between helices 1 and 2 in ACP. We were only able to observe a single NOE, which was insufficient to position this small (four-residue) polypeptide unambiguously with respect to the helices in three-dimensional space; however, one of the four allowed orientations is within approximately
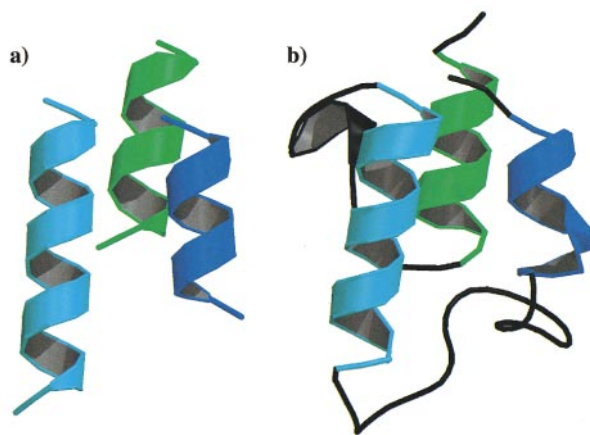

Figure 3. Comparison of ACP structures. (a) Backbone fold as determined from residual dipolar couplings and five long-range NOEs. (b) High-resolution NOE structure.[32]

15° of the orientation of this turn in the original structure.[36]

The backbone fold of NodF can be seen in Figure 4. In this case, each helix was treated as three individual, smaller helices of roughly equal length. These were oriented individually using Orderten-SVD and then reassembled back into three longer helices. Splitting the helices into multiple smaller segments was possible for NodF because of the inclusion of $^1D_{CH}$, $^1D_{NC'}$, and $^2D_{HC'}$ couplings in addition to the $^1D_{NH}$ and D_{HH} couplings used in ACP. This has an advantage in that it allows for a better fit when the larger structural element deviates from ideality; in fact, splitting the helices was required in this case to get solutions out of Orderten-SVD using reasonable error estimates. The inversion problem did not present any difficulties when recombining the shorter fragments back into full-length helices; in all cases, only one of the four allowed orientations maintains the correct helical twist. This method does, however, lead deviations from regular helix geometry to concentrate at the reattachment sites. Solutions from Orderten-SVD were checked for a consistent set of order parameters, and all three helices had a set of solutions with S_{zz} ranging from -0.001 to -0.0015 and η from 0.5 to 0.65. The concatenated helices were then positioned in the program POSE, as previously described for ACP, using the NOE distance constraints shown in Table 2B and loop maximum distance constraints of 2.5 Å times the number of amino acid residues in the loop. This yielded a single ensemble of 303 allowable structures with an RMSD for the backbone heavy atoms of 2.3 Å. A high-resolution structure of NodF using NOEs in addition to dipolar couplings is currently being investigated but is not yet available for comparison.
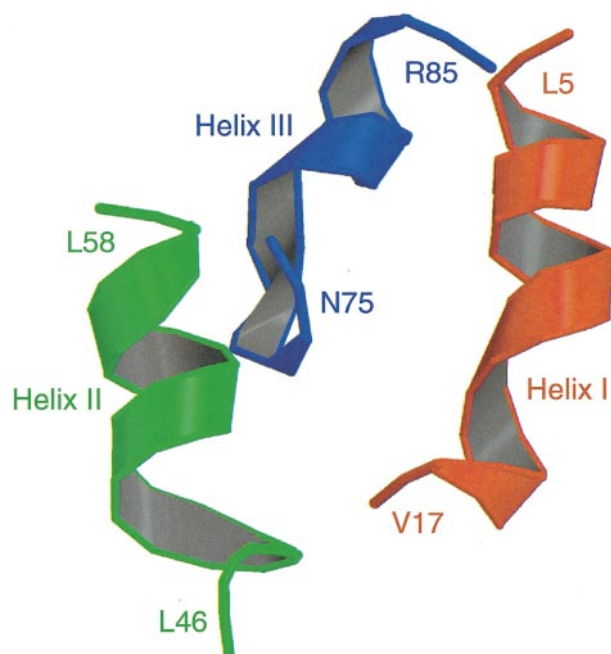
Figure 4. Backbone fold of NodF as determined from residual dipolar couplings and five long-range NOEs.

Discussion

In traditional, NOE-based structure determination using NMR, the vast majority of time is spent at the stage of data acquisition and analysis, with the goal being the ability to obtain assignments for resonances that are as complete as possible. This often requires identification of poorly resolved resonances for side-chain protons in addition to sequential backbone connectivities. As a result, it is not atypical to spend more than one month collecting and interpreting data before structure calculations can begin.

In contrast, the backbone assignments needed for the determination of a backbone fold can be done relatively quickly. When a doubly ^{13}C, ^{15}N-labeled sample is available, a backbone walk using HNCACB and CBCA(CO)NH experiments can be performed.[39,40] Our experiments on a 2.5 mM NodF sample required only two days. When only a singly ^{15}N-labeled sample is available, assignments come from a pair of ^{15}N filtered experiments, a TOCSY experiment to classify resonances into spin systems and a NOESY experiment to obtain sequential (i, $i + 1$) connectivities, as well as a few long-range connectivities.[23,44] An HNHA-style experiment is also run to identify secondary structural motifs from $^{3}J_{HNHA}$ coupling constants.[23] Although we used previous assignments in our ACP application, the equivalent experiments on NodF required just over four days.

Most experiments for the actual collection of dipolar coupling data are also straightforward to run and relatively efficient in terms of time.[45–49]

We actually acquired a number of redundant data sets in these initial applications. However, we can estimate times corresponding to a minimal set of experiments. For ACP, an oriented phase CT-COSY experiment plus the phase encoded ^{15}N-^{1}H HSQC would have required approximately two days. For NodF, phase encoded ^{15}N-^{1}H HSQC, CT-COSY, in-phase/anti-phase HSQC, and C$^{\alpha}$-H$^{\alpha}$ HSQCs with four different mixing times would have required only three days. Hence, total data acquisition times can be kept to less than one week with current instrumentation; times should be even shorter using recent advances in NMR hardware.

In order to use our dipolar coupling data to produce a backbone fold, we did have to make a number of assumptions. The first and most important of these is that secondary structural elements conform to an idealized structure. For short polypeptide segments this assumption is good; as shown for NodF, larger elements that deviate from ideality can be broken down into smaller fragments, and then the intact elements can be reassembled. We also assumed that the individual elements are rigid, i.e. that there are no large internal motions. Structure determination by aligning order tensor frames is actually quite tolerant to moderate motions, perhaps even more so than NOE-based methods. In the systems investigated, neither assumption seemed to introduce any severe problems with the final folds.

The accuracy of the backbone folds could be a concern. Figure 3(a) illustrates the average positioning of the three helices of *E. coli* ACP determined using dipolar couplings and POSE. For comparison, the previously determined NOE structure of ACP is shown in Figure 3(b). The overall geometry is apparently quite similar. In fact, the RMSD between the backbone atoms in the helices of both structures is slightly less than 3 Å, a resolution similar to that of a low-resolution crystal structure. This level of error between structures is remarkably good considering that the C terminus of helix 1 in the original NMR structure (on the right in Figure 3) is quite badly kinked and deviates significantly from our assumption of an ideal helix. Furthermore, helix 3 (the central helix in the figure) is twisted by approximately 40° about the helical axis. This contributes to the RMSD but does not produce easily visualized departures between the two structures. Hence, a fold description is very well reproduced, while actual positions of backbone atoms may suffer moderate displacements.

We can not make a direct comparison to a structure determined by alternate methods for the NodF protein from *R. leguminosarum*. However, the structure of NodF can be compared to ACP. NodF yields a bundle of three helices in which helix 1 is anti-parallel to the other two helices, as are the helices in ACP. The similarity is not surprising, as NodF has a reasonably high level of both sequence and functional homology with ACP, especially around the point where the

phosphopantetheine prosthetic group is attached (Ser36 in ACP, Ser45 in NodF). In each case this site is in a conserved Asp-Ser-Leu motif and the leucine residue begins helix 2. Furthermore, ACP has a hydrophobic cleft between helices 2 and 3 in which an acyl chain can lie.[50] In both structures a number of hydrophobic residues line the contact area between these helices (F50, I54, A59, V65, A68 and Y71 in ACP; A49, W53, I77, V81, and V84 in NodF), implying that this same motif may exist in NodF.

We can also compare *E. coli* ACP and NodF to structures of other acyl carrier proteins. One of these is the actinorhodin acyl carrier protein (*act* ACP) from the polyketide synthase of *Streptomyces coelicolor*, for which the structure of the apoprotein has been solved using NOE-based methods.[51] This structure also has three helices (residues 7-16, 42-53, and 72-85) arranged in a similar three-helix bundle. Another comparison is with the fatty acid synthase ACP from spinach. The structure for spinach ACP was derived using homology constraints from *E. coli* ACP combined with a limited set of NOEs.[52] The structure again indicates a three-helix bundle with a number of hydrophobic residues that could provide a site for acyl chain stabilization in the area between the second and third helices. Very recently a crystal structure of the holo-ACP synthetase from *Bacillus subtilis* appeared which contains an ACP molecule in complex.[53] Here too the general fold of the three major helices seems to be preserved.

In the examples presented, we have dealt only with helical proteins. Application to a wider range of orientable secondary structural elements is of interest. The orienting of a reverse turn in ACP was mentioned in the results, and while there were too few observed distance constraints to correctly position this fragment, it provides some support for application of our method to non-helical structures.

One last issue that remains to be discussed is how useful such a fold can be. Classification of proteins using homology searches has a long history and continues to play an important role today. At low levels of sequence homology, models of backbone structures are the primary result of these methods. Experimental backbone structures can supplement and replace models generated in this way. They can add to, or be searched against in, databases of structural motifs for proteins such as SCOP or CATH.[54, 55] Several tools also exist for comparing features of backbone structures or for extending backbone structures to more complete models; some examples include DALI, WHAT IF, and TOP.[56, 57, 58] One recent tool of note employs "fuzzy functional forms" (FFFs).[6, 59] FFFs basically define a conserved geometric placement of residues within a set of related enzymes, typically represented by the positions of backbone atoms, which are not necessarily sequential, but tend to lie in or near the active site. An attempt to identify function is then made by searching for this

pattern. The target for the search is a backbone structure, usually produced by sequence homology modeling with an accuracy of 4 to 6 Å. Our approach produces more accurate structures without the need for homology. Thus, we are optimistic about the potential utility of experimental backbone folds in structural genomics applications.

Materials and Methods

Protein purification and sample preparation

[15]N-labeled *E. coli* acyl carrier protein (ACP) was expressed and isolated using a procedure modified from previously reported methods.[60–62] Briefly, ACP was overexpressed in *E. coli* BL-21(DE3), and the IPTG-induced cells were harvested and frozen. Thawed and resuspended cells were disrupted by sonication, centrifuged, and most other proteins were removed from the resulting supernatant by the addition of ammonium sulfate to a final concentration of 0.37 g/ml. The denatured proteins were pelleted, and ACP was precipitated from this supernatant by adjusting the pH to 3.5 by the dropwise addition of concentrated HCl. ACP was resuspended and treated with 0.2 M hydroxylamine-HCl at pH 9.0 to strip off any acyl chains. The pH was lowered to 6.0 and the protein purified over a Q-Sepharose column (Pharmacia, Piscataway, NJ) using a linear gradient from 0.15 M to 1.0 M NaCl; ACP containing fractions were identified by SDS-PAGE. At this point, *E. coli* holo-acyl carrier protein synthase was expressed and partially purified by batch treatment with DEAE-cellulose (DE-52, Whatman, Clifton, NJ) followed by cation exchange chromatography of the supernatant over an SP-Sepharose column (Pharmacia) eluted with a linear gradient of NaCl from 0 M to 1 M.[63,64] The resulting crude synthase was used to convert mixed ACP entirely to the *holo* form using coenzyme A (Sigma, St. Louis, MO) as the phosphopantetheine source.[63,64] Following incubation of both proteins with coenzyme A, conversion to *holo*-ACP was verified by SDS-PAGE. Conversion to *holo*-protein was judged greater than 95%, and product was recovered from the reaction mixture using a Q-Sepharose column eluted with 0.5 M NaCl. Finally, the *holo*-ACP containing fractions, as determined by SDS-PAGE, were purified by gel filtration HPLC using a TSK-750 column (Bio-Rad, Hercules, CA).

[13]C,[15]N NodF was prepared by Dr Otto Geiger as outlined.[37] In short, IPTG-induced cells were harvested, frozen, resuspended, and lysed by passage through a French pressure cell. Ethanol was added dropwise while stirring at 4 °C to a final concentration of 50% (v/v) and the solution incubated for an additional hour. Precipitated proteins were removed by centrifugation and the supernatant dialyzed to remove ethanol. The NodF from the dialysate was further purified using a DEAE-cellulose column (DE-52) eluted with a linear gradient from 0.1 M to 1.0 M NaCl. NodF containing fractions were determined using SDS-PAGE.

Pf1 bacteriophage, used as an orienting medium for NMR experiments, was prepared as outlined.[41,65] *Pseudomonas aeruginosa* was grown on LB to an A_{600} of ~1.0, infected with Pf1, and grown for an additional 16 hours. Host cells were removed by centrifugation and the Pf1 precipitated by adding 60 g/l of NaCl and 20 g/l of PEG-8000 while stirring on ice. The phage was collected by centifugation. Phage was then purified by ultracentrifugation using a KBr density gradient; a step gradient

from 20 % to 50 % KBr (w/v) was used, with the phage loaded in the 20 % layer. After centrifugation, the visible faint blue band of phage near the middle of the gradient was carefully extracted by syringe and then extensively dialyzed against TE buffer (10 mM Tris, 1 mM EDTA (pH 7.0)) to remove KBr. Finally, Pf1 was pelleted by ultracentrifugation and the pellet stored at 4 °C until needed.

Final NMR samples of ACP and NodF were obtained by concentrating and exchanging buffer using Centricon-3 devices (Amicon, Beverly, MA). ^{15}N *holo*-ACP samples were 25 mM bis-Tris in 10 % ^{2}H$_2$O, 90 % H$_2$O (pH 6.0), containing 10 mM CaCl$_2$ for protein stability (protein concentrations were 2 mM isotropic and 1 mM aligned). ^{13}C,^{15}N NodF samples were 200 mM potassium phosphate in 10 % ^{2}H$_2$O, 90 % H$_2$O (pH 6.1) with no other salts present (protein concentrations were 2.5 mM isotropic and 1 mM aligned). Trace amounts of sodium azide were added to all samples as a preservative. Alignment of samples for the observation of dipolar couplings was achieved using Pf1 filamentous bacteriophage.[41] Samples with the correct volume and buffer conditions were lyophilized and then rehydrated with enough aqueous phage to give a ^{2}H$_2$O quadrupolar splitting of 12-18 Hz (approximately 20 mg/ml phage).

NMR spectroscopy

All NMR spectra were acquired on Varian Inova spectrometers operating at 500, 600, or 800 MHz and equipped with 5 mm triple resonance gradient probes. Experiments run on each sample for both assignment and the observation of couplings, including experimental conditions, are listed in Table 3; examples of some of the spectra can be found in Supplementary Material. ACP side-chain assignments and ^{3}J$_{HNHA}$ scalar couplings were previously reported by Prestegard and co-workers using different, COSY based methods and are not reported again here.[31, 32, 36] Peak positions were determined and intensities or volumes were measured using the routines built into the software package Felix (MSI, San Diego, CA).

Measurement of scalar and residual dipolar couplings

A large number of different couplings were measured for the determination of a backbone fold; we briefly present here the methods by which couplings can be extracted from the various NMR spectra used. ^{1}J$_{NH}$ and ^{1}D$_{NH}$ couplings were measured in two ways. A coupling-enhanced HSQC experiment was used,[66] in which the couplings can simply be measured by dividing the difference of two peak centers by two. For our purposes, measuring the couplings from the positions of peak maxima, as interpolated by Felix, has a high enough level of precision (on the order of ±1.0 Hz). A phase-encoded HSQC experiment was also used.[45] In this case, two different spectra are acquired in an interleaved fashion with the coupling modulated by either a sine or cosine function during the evolution time, T. The coupling can be extracted using the following formula:

$$\Delta J = \frac{\arctan\left(\frac{V_{\sin}}{V_{\cos}}\right)}{\pi T} \tag{4}$$

where ΔJ is the difference between the true scalar (or

dipolar) coupling and an experimental tune frequency (typically 93-95 Hz for ^{15}N-^{1}H couplings) and $V_{\sin}$ and $V_{\cos}$ are the volumes of the peaks in the sine and cosine experiments, respectively. Errors on the measurement of couplings using this method were typically less than ±0.5 Hz.

^{3}J$_{HNHA}$ couplings were measured using either an HNHA experiment[35] or CT-COSY[46] experiments. For the HNHA experiment, the couplings can be extracted from the intensities of the auto- and cross-peaks using the following relation[35]:

$$\frac{I_{\text{cross}}}{I_{\text{auto}}} = -\tan^2(\pi J_{HH} t_{\text{mix}}) \tag{5}$$

in which I_{cross} and I_{auto} are the cross-peak and auto-peak intensities, J_{HH} is the HNHA coupling of interest, and t_{mix} the mixing time of the experiment.

The following equation can be used to extract ^{3}J$_{HNHA}$, as well as ^{3}D$_{HNHA}$ and any other observable ^{1}H-^{1}H dipolar couplings, from CT-COSY experiments[46,67]:

$$J = \frac{\arctan\left(\frac{I_{\text{cross}}}{I_{\text{auto}}}\right)}{\pi T} \tag{6}$$

Here J is once again the scalar or dipolar coupling being measured, I_{cross} and I_{auto} are the cross-peak and auto-peak intensities, and T is constant time during which the coupling evolves. Because intensities are difficult to measure as volumes for the antiphase cross-peaks, two values of T were used, which allowed a scaling factor to be introduced in equation (6) so that peak height ratios could accurately represent intensity ratios.

In the case of doubly ^{13}C,^{15}N labeled NodF, we measured an additional set of couplings involving ^{13}C. C$^\alpha$H$^\alpha$ couplings are measured using a *J*-modulated constant-time HSQC experiment.[47] In this experiment, intensities of the observed peaks are proportional to a cosine function:

$$I \propto \cos[2\pi^1 J_{CH}(T - \Delta)] \tag{7}$$

in which T is the constant time required to refocus undesired carbon-carbon couplings and Δ is a slight change to this delay which modulates the intensities of the observed peaks. In actuality the coupling is extracted by optimizing the fit of the cosine function to a number of intensities from a set of experiments run with different values for Δ. Finally, an in-phase, anti-phase experiment was run to obtain values for ^{1}D$_{NC'}$ and ^{2}D$_{HC'}$.[48] These sets of couplings are extracted simply by measuring the differences between the centers of two peaks in a multiplet; the one bond ^{15}N-^{13}C' coupling is measured in the indirect dimension while the displacement in the direct dimension yields the two bond ^{1}H^N-^{13}C' coupling.

Fold determination

For the determination of a protein backbone fold, the molecular orientation frame (also called the principal alignment frame) for each rigid fragment was determined by singular value decomposition using the program Orderten-SVD.[20] In the case of ACP, each helix was treated as a single rigid entity; in determining the fold of NodF, each helix was split into three pieces with roughly equivalent numbers of couplings for each smaller fragment, followed by reassembly of the fragments back into full-length helices. The starting coordinates for

Table 3. Experiments run on ACP and NodF for backbone assignments and the measurement of residual dipolar couplings

Experiment	Reference		Time (h)	t_{mix} (ms)	Field (MHz)	No. of points	Sweep width (Hz)	Final size (pts)
A. ^{15}N-ACP experiments								
NOESY-^{15}N-HSQC	34	Isotropic	38	150	600	1024 × 158 × 18	6000 × 6000 × 2000	512 × 512 × 64
Intensity modulated	45	Isotropic	1.5	43	600	512 × 128	6000 × 2200	1024 × 512
^{15}N-HSQC		Aligned	3.0	43	600	512 × 128	6000 × 2200	1024 × 512
^{15}N-HMQC-CT-COSY	46	Isotropic	20	25	800	1024 × 48 × 18	10,650 × 5600 × 3200	1024 × 128 × 64
			20	35	800	1024 × 48 × 18	10,650 × 5600 × 3200	1024 × 128 × 64
		Aligned	20	25	800	1024 × 48 × 18	10,650 × 5600 × 3200	1024 × 128 × 64
			20	35	800	1024 × 48 × 18	10,650 × 5600 × 3200	1024 × 128 × 64
B. ^{13}C,^{15}N-NodF experiments								
NOESY-^{15}N-HSQC	34	Isotropic	48	150	800	1024 × 148 × 24	9600 × 9600 × 2920	1024 × 512 × 64
TOCSY-^{15}N-HSQC	69	Isotropic	32	70	500	512 × 128 × 30	6000 × 1330 × 3000	512 × 256 × 64
HNHA	35	Isotropic	20	26	800	512 × 32 × 24	9200 × 9200 × 3200	512 × 128 × 64
CBCA(CO)NH	70, 71	Isotropic	16		600	512 × 52 × 24	7200 × 9050 × 2400	512 × 128 × 64
HNCACB	71, 72	Isotropic	16		600	512 × 52 × 24	7200 × 9050 × 2400	512 × 128 × 64
HNCO	71, 73	Isotropic	12		600	512 × 48 × 24	7200 × 2420 × 2400	512 × 128 × 64
Phase encoded	45	Isotropic	1.5	43	600	512 × 128	6000 × 2200	1024 × 512
^{15}N-HSQC		Aligned	2.0	22	800	1024 × 96	9620 × 2800	1024 × 512
Coupling enhanced	66	Isotropic	3.0		600	512 × 256	7200 × 2400	1024 × 512
^{15}N-HSQC		Aligned	1.5		600	512 × 128	7200 × 2400	1024 × 512
^{15}N-HMQC-CT-COSY	46	Aligned	18	25	800	512 × 48 × 32	9620 × 5600 × 2800	1024 × 128 × 64
			18	40	800	512 × 48 × 32	9620 × 5600 × 2800	1024 × 128 × 64
Intensity modulated	47	Isotropic	6 ea.	see legend	800	512 × 256	8000 × 20000	2048 × 1024
^{13}C-HSQC		Aligned	8 ea.	see legend	500	1024 × 80	6500 × 12000	2048 × 1024
IPAP ^{15}N-^{13}C',	48	Isotropic	6		800	1024 × 512	9600 × 2920	1024 × 1024
^{1}H^{N}-^{13}C'		Aligned	12		800	1024 × 480	9600 × 2920	1024 × 1024

For the intensity modulated ^{13}C experiment on NodF the mixing times used were 0.0, 0.5, 1.0, 1.5, 2.0, and 2.4 ms for the isotropic sample and 0.5, 1.0, 1.6, 1.8, 2.0, 2.6, 3.1, and 3.8 ms for the aligned sample. For ^{15}N-edited experiments, the final matrix size indicated is that after the upfield half of the data was discarded in the direct dimension.

each fragment came from an idealized polyalanine starting structure generated using standard values for the backbone ϕ, ψ, and ω torsion angles.[42] The molecular fragments were rotated such that the principle axis systems of each of their orientation frames were coincident.

The oriented fragments were translated with respect to each other in three-dimensional space using a new program, POSE (positioning of structural elements) and a minimal set of NOEs (two to three distance constraints per rigid element, see Tables 1 and 2). The program will be made available on our website: *http://tesla.ccrc.uga.edu*. A general outline of how the program works is given in the flowchart in Figure 5. Briefly, any two of the correctly oriented structural elements are provided as input to the program along with a list of NOEs and other distance constraints (for example, a short three or four-residue loop would force two elements which pack closely together into an anti-parallel orientation). One of the atoms involved in an NOE is chosen as a starting point, and a spherical grid of possible solutions to that NOE is generated about it. The second element is then translated so that the atom connected by the NOE superimposes on the first point of the grid. A check is made to assure that any remaining NOEs and other distance constraints are satisfied, and then an additional check is made for van der Waals collisions between every possible pair of atoms. If there are any violations of distance constraints or close contacts, the solution is rejected; if all conditions are satisfied, the solution is allowed and written to a file. The program runs until all points in the grid are tested, appending each possible solution to the output file. If the inversion problem that arises from our inability to distinguish the positive and negative directions for each axis of the alignment frame has not previously been solved using a second aligning medium,[43] the process is repeated for all four allowed inversions of the starting structure of the second element. Up to this point, the entire process required very little computational time (less than two minutes using an SGI Indy with a 150 MHz R4400 processor). In order to position the third helix, the first two helices were combined into a new "first" element, the third helix was added as the "second" element, and these were used along with an appropriate set of distance constraints as inputs to POSE. In order to assure that we did not throw away any valid solutions, we repeated the process using every allowed solution for the first pair of elements. Additional elements can be added by repeating the entire process in an iterative fashion.

Coordinates for backbone atoms of secondary structure elements in NodF are being deposited in the RCSB Protein Data Bank[68] with accession number 1FH1.
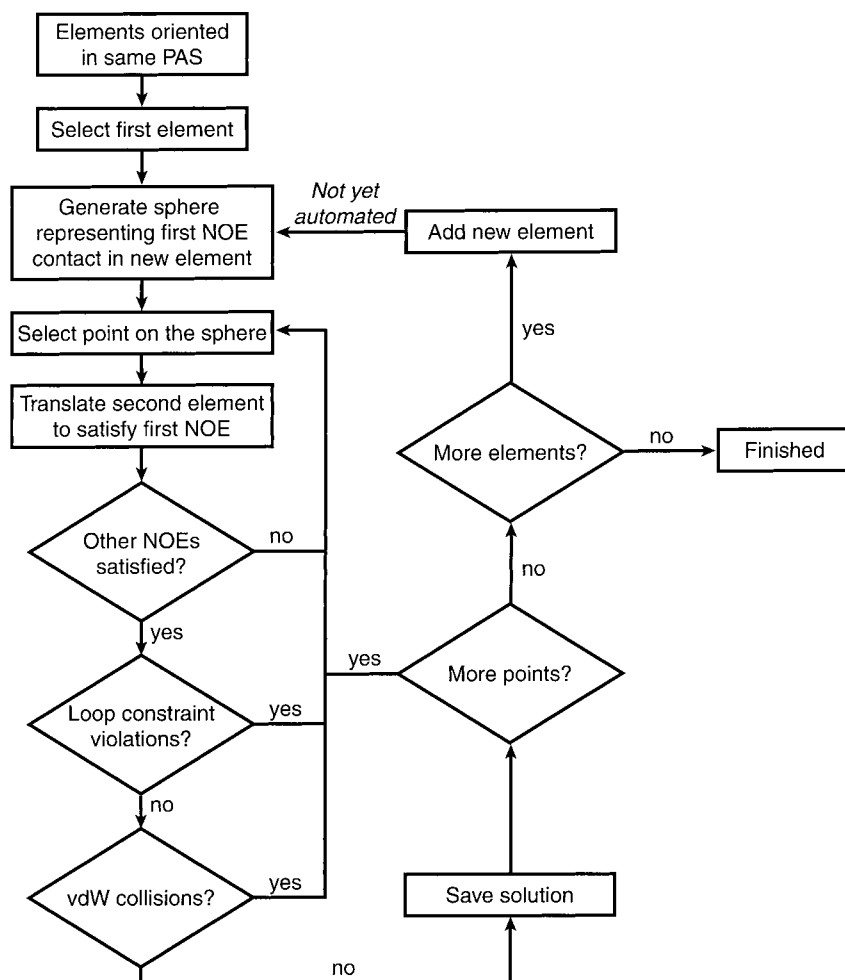

Figure 5. Flow chart illustrating the operation of the program POSE.

Acknowledgments

References

1. Sali, A. & Kuriyan, J. (1999). Challenges at the frontiers of structural biology. *Trends Biol. Sci.* **24**, M20-M24.

2. Zarembinski, T. I., Hung, L. W., Mueller-Dieckman, H. J., Kim, K. K., Yokota, H., Kim, R. & Kim, S. H. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl Acad. Sci. USA,* **95**, 15189-15193.

3. Sali, A. (1998). 100,000 protein structures for the biologist. *Nature Struct. Biol.* **5**, 1029-1032.

4. Montelione, G. T. & Anderson, S. (1999). Structural genomics: keystone for a Human Proteome Project. *Nature Struct. Biol.* **6**, 11-12.

5. Skolnick, J., Fetrow, J. S. & Kolinski, A. (2000). Structural genomics and its importance for gene function analysis. *Nature Biotechnol.* **18**, 283-287.

6. Skolnick, J. & Fetrow, J. S. (2000). From genes to protein structure and function: novel applications of computational approaches in the genomic era. *Trends Biotechnol.* **18**, 34-39.

7. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science,* **273**, 595-602.

8. Holm, L. & Sander, C. (1994). Searching protein structure databases has come of age. *Proteins: Struct. Funct. Genet.* **19**, 165-173.

9. Gibrat, J. F., Madej, T. & Bryant, S. H. (1996). Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **6**, 377-385.

10. Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A common sense approach to peak picking in two, three, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214-220.

11. Koradi, R., Billeter, M., Engeli, M., Güntert, P. & Wüthrich, K. (1998). Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. *J. Magn. Reson.* **135**, 288-297.

12. Zimmerman, D. E., Kulikowski, C. A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C. Y., Powers, R. & Montelione, G. T. (1997). Automated analysis of protein NMR assignments using methods from artificial intelligence. *J. Mol. Biol.* **269**, 592-610.

13. Gronwald, W., Willard, L., Jellard, T., Boyko, R. F., Krishna, R., Wishart, D. S., Sönnichsen, F. D. & Sykes, B. D. (1998). CAMRA: chemical shift based computer aided protein NMR assignments. *J. Biomol. NMR,* **12**, 395-405.

14. Bartels, C., Güntert, P., Billeter, M. & Wüthrich, K. (1997). GARANT - A general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. *J. Comput. Chem.* **18**, 139-149.

15. Nilges, M., Macias, M. J., O'Donoghue, S. I. & Oschkinat, H. (1997). Automated NOESY interpretaion with ambiguous distance restraints: the refined NMR solution structures of the pleckstring hom-

16. Nilges, M. & O'Donoghue, S. I. (1998). Ambiguous NOEs and automated NOE assignment. *Prog. Nucl. Magn. Reson. Spect.* **32**, 107-139.

17. Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science,* **278**, 1111-1114.

18. Prestegard, J. H., Al-Hashimi, H. M. & Tolman, J. R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quart. Rev. Biophys.* In the press.

19. Fischer, M. W. F., Losonczi, J. A., Weaver, J. L. & Prestegard, J. H. (1998). Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry,* **38**, 9013-9022.

20. Losonczi, J. A., Andrec, M., Fischer, M. W. F. & Prestegard, J. H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138**, 334-342.

21. Clore, G. M. & Gronenborn, A. M. (1998). New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl Acad. Sci. USA,* **95**, 5891-5898.

22. Delaglio, F., Kontaxis, G. & Bax, A. (2000). Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.* **122**, 2142-2143.

23. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.

24. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* **31**, 1647-1651.

25. Saupe, A. (1968). Recent results in the field of liquid crystals. *Angewandte Chemie Int. (Edit. English),* **7**, 97-112.

26. Magnuson, K., Jackowski, S., Rock, C. O. & Cronan, J. E., Jr (1993). Regulation of fatty acid biosynthesis in *Eschericia coli. Microbiol. Rev.* **57**, 522-542.

27. Vanden, Boom T. & Cronan, J. E., Jr (1989). Genetics and regulation of bacterial lipid metabolism. *Annu. Rev. Microbiol.* **43**, 317-343.

28. Spaink, H. P., Sheeley, D. M., van Brussel, A. A. N., Glushka, J., York, W. S., Tak, T., Geiger, O., Kennedy, E. P., Reinhold, V. N. & Lugtenberg, B. J. J. (1991). A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium. Nature,* **354**, 125-130.

29. Geiger, O., Spaink, H. P. & Kennedy, E. P. (1991). Isolation of the *Rhizobium leguminosarum* NodF nodulation protein: NodF carries a 4'-phosphopantetheine prosthetic group. *J. Bacteriol.* **173**, 2872-2878.

30. Ritsema, T., Gehring, A. M., Stuitje, A. R., van der Drift, K. M. G. M., Dandal, I., Lambalot, R. H., Walsh, C. T., Thomas-Oates, J. E., Lugtenberg, B. J. J. & Spaink, H. P. (1998). Functional analysis of an interspecies chimera of acyl carrier proteins indicates a specialized domain for protein recognition. *Mol. Gen. Genet.* **257**, 641-648.

31. Holak, T. A. & Prestegard, J. H. (1986). Secondary structure of acyl carrier protein as derived from two-dimensional ^{1}H NMR spectroscopy. *Biochemistry,* **25**, 5766-5774.

32. Kim, Y. & Prestegard, J. H. (1989). A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry,* **28**, 8792-8797.

ology domain from β-spectrin. *J. Mol. Biol.* **269**, 408-422.

33. Andrec, M. R., Hill, R. B. & Prestegard, J. H. (1995). Amide exchange rates in *Eschericia coli* acyl carrier protein: correlation with protein structure and dynamics. *Protein Sci.* **4**, 983-993.

34. Jahnke, W., Baur, M., Gemmecker, G. & Kessler, H. (1995). Improved accuracy of NMR structures by a modified NOESY-HSQC experiment. *J. Magn. Reson.* **106**, 86-88.

35. Vuister, G. W. & Bax, A. (1993). Quantitative *J* correlation: a new approach for measuring homonuclear three-bond $J(H^N H^\alpha)$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772-7777.

36. Kim, Y. & Prestegard, J. H. (1990). Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins: Struct. Funct. Genet.* **8**, 377-385.

37. Ghose, R., Geiger, O. & Prestegard, J. H. (1996). NMR investigations of the structural properties of the nodulation protein, NodF, from *Rhizobium leguminosarum* and its homology with *Escherichia coli* acyl carrier protein. *FEBS Letters,* **388**, 66-72.

38. Huang, K., Andrec, M., Heald, S., Blake, P. & Prestegard, J. H. (1997). Performance of a neural-network-based determination of amino acid class and secondary structure from ^1H-^{15}N NMR data. *J. Biomol. NMR,* **10**, 45-52.

39. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. I. & Skelton, N. J. (1996). *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego.

40. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectr.* **34**, 93-158.

41. Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065-1074.

42. Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*, 2nd edit., W.H. Freeman and Co., New York.

43. Al-Hashimi, H. M., Valafar, H., Terrell, M., Zartler, E. R., Eidsness, M. K. & Prestegard, J. H. (2000). Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J. Magn. Reson.* **143**, 402-406.

44. Clore, G. M. & Gronenborn, A. M. (1991). Applications of three and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Prog. Nucl. Magn. Reson. Spectr.* **23**, 43-92.

45. Tolman, J. R. & Prestegard, J. H. (1996). A quantitative *J*-correlation experiment for the accurate measurement of one bond amide ^{15}N-^1H couplings in proteins. *J. Magn. Reson.* **112 B**, 245-252.

46. Tian, F., Fowler, C. A., Zartler, E. R., Jenney, F. A. J., Adams, M. W. & Prestegard, J. H. (2000). Direct measurement of ^1H-^1H dipolar couplings in proteins: a complement to traditional NOE measurements. *J. Biomol. NMR,* **18**, 23-31.

47. Tjandra, N. & Bax, A. (1997). Measurement of dipolar contributions to $^1J_{CH}$ splittings from magnetic-field dependence of *J* modulation in two-dimensional NMR spectra. *J. Magn. Reson.* **124**, 512-515.

48. Wang, Y., Marquardt, J. L., Wingfield, P., Stahl, S. J., Lee-Huang, S., Torchia, D. & Bax, A. (1998). Simultaneous measurement of ^1H-^{15}N, ^1H-$^{13}C'$, and ^{15}N-$^{13}C'$ dipolar couplings in a perdeuterated 30 kDa protein dissolved in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 7385-7386.

49. Ottigen, M., Delaglio, F. & Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Mang. Reson.* **131**, 373-378.

50. Jones, P. J., Holak, T. A. & Prestegard, J. H. (1987). Structural comparison of acyl carrier protein in acylated and sulfhydryl forms by two-dimensional ^1H NMR spectroscopy. *Biochemistry,* **26**, 3493-3500.

51. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A., Murray, M., Hopwood, D. A. & Simpson, T. J. (1997). Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry,* **36**, 6000-6008.

52. Oswood, M. C., Kim, Y., Ohlrogge, J. B. & Prestegard, J. H. (1997). Structural homology of spinach acyl carrier protein and *Escherichia coli* acyl carrier protein based on NMR data. *Proteins: Struct. Funct. Genet.* **27**, 131-143.

53. Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C. C., Seehra, J. & Somers, W. S. (2000). Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure,* **8**, 883-895.

54. Lo Conte, L., Ailey, B., Hubbard, T. J. P., Brenner, S. E., Murzin, A. G. & Chothia, C. (2000). SCOP: a structural classification of proteins database. *Nucl. Acids Res.* **28**, 257-259.

55. Pearl, F. M. G., Lee, D., Bray, J. E., Sillitoe, I., Todd, A. E., Harrison, A. P., Thornton, J. M. & Orengo, C. A. (2000). Assigning genomic sequences to CATH. *Nucl. Acids Res.* **28**, 277-282.

56. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

57. Vriend, G. & Sander, C. (1991). Detection of common three-dimensional substructures in proteins. *Proteins: Struct. Funct. Genet.* **11**, 52-58.

58. Lu, W. (2000). TOP: a new method for protein structure comparisons and similarity searches. *J. Appl. Crystallog.* **33**, 176-183.

59. Fetrow, J. S. & Skolnick, J. (1998). Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and T_1 ribonucleases. *J. Mol. Biol.* **281**, 949-968.

60. Rock, C. O. & Cronan, J. E., Jr (1980). Improved purification of acyl carrier protein. *Anal. Biochem.* **102**, 362-364.

61. Morris, S. A., Revill, W. P., Staunton, J. & Leadlay, P. F. (1993). Purification and separation of *holo* and *apo* forms of *Saccharopolyspora erythraea* acyl-carrier protein released from recombinant *Eschericia coli* by freezing and thawing. *Biochem. J.* **294**, 521-527.

62. Hill, R. B., MacKenzie, K. R., Flanagan, J. M., Cronan, J. E., Jr & Prestegard, J. H. (1995). Overexpression, purification, and characterization of *Eschericia coli* acyl carrier protein and two mutant proteins. *Protein Express. Purif.* **6**, 394-400.

63. Lambalot, R. H. & Walsh, C. T. (1995). Cloning, overproduction, and characterization of the *Eschericia coli* holo-acyl carrier protein synthase. *J. Biol. Chem.* **270**, 24658-24661.

64. Lambalot, R. H. & Walsh, C. T. (1997). Holo-[acyl-carrier-protein] synthase of *Eschericia coli*. *Methods Enzymol.* **279**, 254-262.

65. Hansen, M. R., Hanson, P. & Pardi, A. (1999). Filamentous bacteriophage as a versatile method for aligning RNA, DNA, and proteins for measurement of NMR dipolar coupling interactions. *Methods Enzymol.* **317**, 220-240.

66. Tolman, J. R. & Prestegard, J. H. (1996). Measurement of amide ^{15}N-^{1}H one-bond couplings in proteins using accordion heteronuclear-shift-correlation experiments. *J. Magn. Reson.* **112**, 269-274.

67. Tian, F., Bolon, P. & Prestegard, J. (1999). Intensity-based measurement of homonuclear residual dipolar couplings from CT-COSY. *J. Am. Chem. Soc.* **121**, 7712-7713.

68. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissi, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235-242.

69. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1β. *Biochemistry,* **28**, 6150-6156.

70. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291-6293.

71. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson.* **103**, 203-216.

72. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha and beta-carbon resonances in proteins. *J. Magn. Reson.* 101B, 214-217.

73. Ikura, M., Kay, L. E. & Bax, A. (1990). A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry,* **29**, 4659-4667.

IDE𝐀L®

http://www.academicpress.com/jmb

Supplementary material for this paper is available on IDEAL.