Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

21

Protein Structure Calculation from NMR Data

Tapas K. Mal, Stefan Bagby, and Mitsuhiko Ikura

1. Introduction

Until 1984, structural information of biomolecules at atomic resolution could only be determined by X-ray diffraction techniques with protein single crystals ***(1)***. In the mid- 1980s, Wüthrich and co-workers demonstrated that nuclear magnetic resonance (NMR) spectroscopy ***(2)*** could be used as a technique for protein structure determination ***(3)***. This permits biomolecular structure determination with comparable accuracy to X-ray diffraction, but in a solution environment that is much closer to the physiological milieu than the single crystals required for protein crystallography. Today, many if not most, NMR measurements with proteins are performed with the ultimate aim of determining their three-dimensional (3D) structure. NMR is not a "microscope with atomic resolution" that directly produces an image of a protein. Rather, NMR yields a wealth of indirect structural information from which the 3D structure can only be elucidated by extensive calculations. The first structure determinations of peptides and proteins in solution ***(4–8)*** were fascinating yet tedious and lengthy struggles because of the lack of established NMR techniques and numerical methods for structure calculation. In the early days of structure calculations from NMR data (NOE-derived distance restraints and ^3J-derived torsion angle restraints), mostly two types of distance geometry algorithms were used: (1) DISGEO, which operated in distance space and used the metric matrix method to convert distance restraints into Cartesian coordinates ***(9)***, and (2) DISMAN, which operated in torsion angle space and used restrained minimization of a variable target function ***(10)***. Subsequent developments in both NMR methodologies and structure calculation methods have made NMR an indispensable tool for determining biomolecular structures. Here we focus on the most commonly used software packages for 3D structure calculations from

From: *Methods in Molecular Biology*, vol. 173:
Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques
Edited by: H. J. Vogel

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

NMR data, X-PLOR ***(11)***, ARIA (ambiguous restraints for iterative assignment) ***(12,13)***, and DYANA (dynamical algorithm for NMR applications ***(14,15)***. X-PLOR and ARIA use the distance space, whereas DYANA works in the torsion angle space.

2. Materials

Certain pieces of hardware and software are essential for structure calculation from NMR data. Hardware options include SUN (http://www.sun.com), Silicon Graphics Inc. (SGI) (http://www.sgi.com), and Hewlett-Packard (HP) (http://www.hp.com) platforms and PC systems using Linux (http://www.linux.com). The most commonly used software packages are X-PLOR ***(11)*** (http://pauli.csb.yale.edu/xplor-info), ARIA ***(12,13)*** (http://www.nmr.embl-heidelberg.de/nmr/nilges/aria), and DYANA ***(14,15)*** (www.mol.biol.ethz.ch/wuthrich/software/dyana).

3. Methods

Structure determination from NMR data can be divided into two steps: (1) collection of structural restraints, and (2) calculation of structures using these restraints. The first step is common to all three structure calculation methods under discussion in this chapter. Two types of restraints are commonly used in structure determination from NMR data:

1. Distance restraints derived from NOE (nuclear Overhauser effect) measurements.
2. Dihedral angle restraints derived from the measurement of vicinal coupling constants (^3J).

Collection of structural restraints typically involves the following steps:

1. Assign as many NOE cross peaks from NOESY (nuclear Overhauser enhancement spectroscopy) spectra as possible.
2. Measure cross peak intensities of all assigned NOEs (*see* **Subheading 4.**).
3. Classify NOEs into three different categories, for example "short range" ($d(i, i+1)$), "medium range" ($dij(1<|i–j|\leq 4)$), and "long range" ($dij(|i–j|\geq 5)$).
4. Translate NOE intensities into different classes of upper distance bounds (typically, 1.8–2.8, 1.8–3.5, and 1.8–5.0 Å) (*see* **Subheading 4.**).
5. Measure coupling constants such as $^3J_{HN-H\alpha}$, $^3J_{H\alpha-H\beta}$, $^3J_{H\alpha-Ni+1}$ (*see* **Subheading 4.**).
6. Delineate secondary structure elements using NOE, ^3J, and chemical shift data (*see* **Subheading 4.**).
7. Where possible, assign stereo-specific geometry for diastereo or prochiral groups utilizing both NOE and ^3J data (*see* **Subheading 4.**).
8. Assign hydrogen bond restraints from exchange, NOE and ^3J data (*see* **Subheading 4.** for details — it is advisable not to use these restraints in preliminary structure calculations but to add them in the structure refinement protocols).

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

3.1. Structure Calculation Methods

3.1.1. X-PLOR

NMR data alone are not sufficient to determine the positions of all atoms in a biological macromolecule, but must be supplemented by information about the covalent structure of the protein — the amino acid sequence, bond lengths, bond angles, chiralities, and planar groups — as well as by the steric repulsion between nonbonded atom pairs. Details of structure calculation by X-PLOR are given:

1. Build a protein structure file (psf) from the protein sequence using the standard X-PLOR topology (topallhdg.pro) and peptide bond linkage (toph19.pep). This psf will contain the following information: atom names, types, charges, and masses; residue names and segment names and a list of bond terms, angle terms, dihedral terms, improper terms, explicit hydrogen-bonding terms, explicit nonbonded exclusions, and nonbonded group partitions. It does not contain atomic coordinates, parameters, or restraints.
2. Generate a template structure using standard X-PLOR parameter (parallhdg.pro) and molecular structure (psf) files. The template structure will have an extended conformation with good local geometry and no nonbonded contacts.
3. Create NOE restraints table (*see* **Subheading 4.**).
4. Create dihedral angle restraints table (*see* **Subheading 4.**).
5. Select an appropriate potential function for NOE restraints. The most commonly used potential function is a flat bottom (square well) potential with a soft asymptote ***(11)***.
6. Produce structure from psf with randomized ϕ and ψ angles. χ^i-angles are not affected.
7. Energy minimization using Powell gradient function (50 steps) to remove nonbonded interaction.
8. Molecular dynamics — based simulated annealing (MDSA). Typically, MDSA from an extended template structure with randomized ϕ and ψ angles works in four stages: a high-temperature search phase, two cooling phases, and a final-energy minimization step. During the high-temperature molecular dynamics (MD) search phase, a low repulsion energy value is used to allow the atoms to pass through each other and to increase the convergence rate. The temperature is reduced from 2000 K to 1000 K in the first cooling phase, and all weights on the different energy terms are gradually brought to their final values (*see* **Table 1**). The second cooling phase comprises a slow cooling from 1000 K to 100 K. The final stage comprises 1000*g* radient energy minimization steps using the final weighting values of the various energy terms (*see* **Table 1**).

Table 1

3.1.1.1. Analysis of Structures Derived from MDSA

1. Restraints violations: Following calculations, the structures must be analyzed to determine whether they fulfil the given experimental restraints. Usually a viola-

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

Table 1
Simulated Annealing Protocol

	Stage			
	1	2	3	4
Temperature[a] (K)	2000	2000 → 1000	1000 → 100	100
Number of steps	6500	3500	3000	1000
Parameters and force constants				
k_{NOE} (kcal mol^{-1} Å^{-2})	10 → 50	50	50	50
k_{repel} (kcal mol^{-1} Å^{-4})	0.002	0.01 → 4.0	4.0	4.0
$k_{dihedral}$ (kcal mol^{-1} Å^{-2})	5	200	200	200
repel	0.9	0.9 → 0.75	0.75	0.80

[a]The temperature is maintained by coupling to a bath ***(16)*** with a coupling constant of 10 ps^{-1}.

tion greater than approx 1 Å is an indication that there is a serious problem, especially if any such violation occurs repeatedly across the ensemble of structures. It is generally useful to identify the region of structures where these restraints are violated and critically inspect them using interactive computer graphics and check the corresponding NOE assignment and volume integration.

2. Atomic root-mean-square deviations (RMSD): The RMSD of the ensemble from the mean of the ensemble is a test that is commonly used to determine the "precision" of the structures, or how close the calculated structures are to each other.
3. Torsion angle distributions: Plots of the ϕ, ψ, χ^i-torsion angles vs amino acid sequence of the protein and Ramachandran plots are useful to analyze local conformations and this allows assessment of the local geometry of the structures.
4. Quality assessment: Programs are available to perform a "quality check" of the structures derived from NMR data, for example, PROCHECK-NMR ***(17)*** and WHATIF ***(18)***. These software packages assess the quality of structures by comparing structural parameters with their values in databases derived from high resolution X-ray structures.

3.1.1.2. Additional Restraints for Structure Refinement

Additional structural restraints can be added to the distance and torsion angle restraints for structure calculation in the final refinement of the structures. For example, ^{13}C secondary chemical shifts ***(19)***, which are related to backbone ϕ and ψ angles, ^{1}H chemical shifts ***(20,21)***, which are influenced for example by short-range ring-current effects from aromatic groups, magnetic anisotropy of C = O and C–N bonds ***(22)*** and residual dipolar couplings that give information on angles between covalent bonds and globally defined axes in the molecule

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

(23). Although the inclusion of these restraints has little impact on precision, the accuracy of the structures is improved ***(24)***.

3.1.2. ARIA

ARIA has recently been developed to automate the structure calculation from NMR data using X-PLOR following assignments of unambiguous NOEs ***(12,13)***. In general, ARIA is an extended version of X-PLOR. It is a combination of FORTRAN subroutines linked to X-PLOR and scripts that convert data formats and control the overall flow of iterative structure calculations.

The most time-consuming step in NMR structure determination is assignment of NOEs. Often, several protons have the same chemical shift. An NOE crosspeak involving such degenerate protons cannot be converted directly into a distance restraint and used in structure calculation. In normal practice, NOEs that can be assigned unambiguously are used to calculate preliminary 3D structures and then additional NOE cross peaks can be assigned on the basis of these structures ***(25)***. The additional restraints are used to calculate a second generation of structures, which, in turn, is then used to obtain more NOE assignments. This is highly time consuming and laborious. ARIA can automate this process and also identify NOEs that do not provide valid structural information (such as spectral artifacts) and reject them from the structure calculation. A typical ARIA protocol involves the following steps:

1. Assign all possible unambiguous NOEs from NOESY spectra.
2. Determine distance restraints from unambiguous NOEs (*see* **Subheading 3.1.1.**).
3. Make peaklists that contain both unambiguous and ambiguous NOE information.
4. Structure calculation with an extended version of X-PLOR. The following steps are automated within ARIA:
   a. Calculation of a set of structures from unambiguous restraints.
   b. Selection of a subset of lowest energy structures for iterative assignment of ambiguous NOEs.
   c. Calibration of the ambiguous NOEs to distance restraints called ambiguous distance restraints (ADRs) on the basis of the subset of lowest energy structures and selection of the NOEs that make the greatest contribution to the ambiguous NOE cross peaks.
   d. Calculation of another set of structures with unambiguous NOE restraints and assigned ADRs.
   e. Analysis of the violations from the calculated structures and rejection of those NOEs that are constantly violated.
   f. Repetition of **steps 2–5** until completion of the user-defined number of iterations or until no significant changes in structures and data sets are detected.
5. Analysis of the structures obtained from ARIA (*see* **Subheading 3.1.1.1.**).

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

3.1.2.1. Ambiguous distance restraints

An ambiguous NOE cross peak at the chemical shift coordinates *F1* and *F2* contains contributions from all proton pairs with the same chemical shifts. On the basis of the isolated spin pair approximation (ISPA), an ambiguous NOE can be treated as the sum of the inverse sixth powers of individual proton-proton distances assuming that the proportionality factor is identical for all protons:

$$\mathrm{NOE}_{\mathrm{F1,F2}} \propto \sum_{a=1}^{N_\delta} d_a^{-6}$$

where a runs through all N_δ contributions to a cross peak at frequencies *F1* and *F2*, and d_a is the distance between two protons corresponding to the ath contribution. The ambiguous NOE corresponds thus to a "d^{-6} summed distance" D:

$$\bar{D} = \left(\sum_{a=1}^{N_\delta} d_a^{-6}\right)^{-1/6}$$

This distance is determined from preliminary structures calculated from unambiguous restraints.

3.1.2.2. Distance Target Functions

During the structure calculation the distances in the structure are typically held to upper and lower bounds according to distance restraints by a gradient-bound flat bottom potential with a soft asymptote that takes care of any large violation ***(26,27)***. The energy of a single distance restraint is

$$E_{\mathrm{NOE}} = k_{\mathrm{NOE}} \begin{cases} (L-D)^2 & \text{if } D<L \\ 0 & \\ (D-U)^2 & \text{if } U<D\leq U+\sigma \\ \alpha + \beta(D-U)^{-1} + \gamma(D-U) & \\ & \text{if } D>U+\sigma \end{cases}$$

where D is the distance measured in the current structure model or a $(\sum d_a^{-6})^{-1/6}$ distance, k_{NOE} is the energy constant, and U and L are the upper and lower bounds of the interproton distances, respectively.

$$U = (d_{\mathrm{ref}}^{-6} \times V/V_{\mathrm{ref}})^{-1/6} + \Delta^+$$

$$L = (d_{\mathrm{ref}}^{-6} \times V/V_{\mathrm{ref}})^{-1/6} + \Delta^-$$

where Δ^+ and Δ^- account for errors as a result of exchange, motion and spin diffusion; d_{ref} is the distance from the iteratively calculated structures as $\langle d^{-6}\rangle^{-1/6}$ averaged over all values for which the distance is smaller than a cutoff (3–6

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

Å); V_{ref} is evaluated as the arithmetic average over all corresponding volumes. For a good data set, the best estimated value for Δ^+ is suggested to be max(0.15, 0.15+0.08[D-2.6]) and Δ^- is 0.15D ***(28)***. The parameter σ determines the distance at which the potential switches to asymptotic behavior, γ is the asymptotic slope of the potential, the coefficients α and β are determined such that E_{NOE} is continuous and differentiable at the point $U+\sigma$. If D is between *L* and *U*, the energy and gradient are zero. For large restraint violations, the force approaches a maximum value or can be decreased depending on α and β. This makes the optimization more stable and improves convergence by permitting transient large violations during calculation and thus allows the structures to escape deep local minima. This is important for structure calculation with ADRs.Assignment of ADRsThe criterion used in ARIA is based on the estimation of the relative size of contributions of different assignment possibilities to the peak volume. For each contribution k to the ambiguous NOE, the minimum or average distance (D^k_{min} or $D^k_{av)}$ is determined from the calculated structure. The contribution C^k of the assignment *k* to the cross peak is estimated as:

$$C^k = (D^k)^{-6} / \sum_i^{N\delta} (D_a^i)^{-6}$$

The C^k are then sorted according to size, and the largest contribution (*Np*) is chosen such that

$$\sum_i^{N\delta} C^i > p$$

The cutoff parameter *p* can be varied for different iterations, in general starting from values close to 1.0 for the first iteration to 0.8 in the last iteration. The smaller the final value of *p* chosen, the fewer peaks remain ambiguous.

3.1.2.4. Advantages of ARIA Over General X-PLOR

1. Ambiguous data can be used from the beginning of the structure calculation.
2. Hydrogen bonds are very difficult to assign, especially at the termini of secondary structure elements, and in irregular structures. Hydrogen bond restraints can be used as ambiguous restraints in ARIA.
3. Sometimes it is difficult to know the disulphide bond pattern in a protein. Disulphide bond restraints can be input as ambiguous restraints in ARIA.

3.1.3. DYANA

DYANA ***(14,15)*** calculates solution 3D structures of biomolecules from distance restraints and torsion angle restraints collected from NMR experiments by performing simulated annealing and molecular torsion angle dynamics (TAD) using variable target functions. Both X-PLOR and ARIA also use molecular dynamics and simulated annealing with variable target functions but work in Cartesian coordinates. The principal differences of TAD from simulated annealing in Cartesian coordinates are:

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

1. It works with internal coordinates rather than Cartesian coordinates.
2. The number of degrees of freedom in TAD is almost 10 times smaller as the covalent structure parameters such as bond lengths, bond angles, chiralities, and planarities remain fixed at their optimal values during structure calculation.
3. Strong potentials are required to preserve the covalent structure and geometry in conventional Cartesian space molecular dynamics whereas a soft potential function is used in TAD, as the concomitant high frequency motions are absent.
4. TAD gives higher efficiency structure calculation as it uses longer time steps for the numerical integration of motions.

In DYANA, the molecules are treated as a tree structure consisting of a base rigid body that is fixed in space and n rigid bodies, which are connected by rotatable bonds ***(29)***. The degrees of freedom are exclusively torsion angles, i.e., rotation about single bonds. Each rigid body is made up of one or several mass points (atoms) for which the relative positions are invariable. The tree structure starts from a "base," typically at the N-terminus of the polypeptide chain, and terminates with "leaves" at the ends of side chains and at the C-terminus. The rigid bodies are numbered from 0 to n and the base has the number 0. A typical DYANA protocol involves the use of CALIBA and gridsearch to create a starting conformation with all torsional angles as independent uniformly distributed random variables (discussed later). This is followed by simulated annealing and energy minimization as follows:

1. Perform a short minimization to reduce high energy interactions: 100 conjugate gradient minimization steps are performed at target level 3, i.e., including only distance restraints between atoms up to three residues apart along the sequence, followed by a further 100 minimization steps including all restraints.
2. Exclude all hydrogen atoms from the check for steric overlap, and increase the repulsive core radii of heavy atoms that are covalently bound to hydrogen atoms by 0.15 Å with respect to their standard values. Set the weighting factor for upper and lower distance bounds to 1, for steric lower bounds to 0.5, and for torsion angle constraints to 5 $Å^2$.
3. Perform a TAD calculation at constant high temperature (typically $T_{high} \approx 10{,}000$ K). One fifth of all N torsion angle dynamic steps are performed at T_{high} (typical value of N is 4000 to 8000).
4. Perform the remaining 4N/5 torsion angle dynamic steps with slow cooling to zero.
5. Incorporate all hydrogen atoms to check for steric overlap. Reset the repulsive core radii to their standard values, increase the weighting factor for steric restraints to 2, and perform 100 conjugate gradient minimization steps with inclusion of all restraints.
6. Perform 200 TAD steps at zero reference temperature.
7. Perform 1000 conjugate gradient minimization steps including all restraints.

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

During the TAD calculation, the list of van der Waals lower distance bounds is updated every 50 steps, or, during minimization, whenever a torsion angle has changed by more than 10^0 since the last update, or after 100 minimization steps.

DYANA is an integrated program, which includes CALIBA (calibration of NOE intensity vs distance restraints) and a versatile multidimensional gridsearch algorithm. CALIBA calibrates NOE intensities into distance restraints. It has different calibration functions for backbone, side-chains and methyl groups. The calibration functions are $V = A/r^6$, $V = B/r^4$, $V = C/r^4$ where V is the peak volume and r is the corresponding distance. The parameters A, B, and C are either user-defined or calculated automatically (typically, $B = A/d_{\min}^2$ and $C = B/3$, $d$min = 2.4 Å). The multidimensional gridsearch algorithm analyzes the local conformation of an arbitrary molecular fragment of a protein involving the three torsional angles ϕ, ψ, and χ^1 of an amino acid residue, determines the stereospecific assignments of β-protons and generates torsion angle restraints.

4. Notes

1. NOEs are the essential NMR data for defining the secondary and tertiary structures of a protein because they permit connection of pairs of hydrogen atoms in amino acid residues that may be far apart in the protein sequence, but close in space (less than about 5 Å apart). The NOE arises from the transfer of magnetization between spins coupled by the dipole–dipole interaction in a molecule undergoing Brownian motion in a liquid ***(30–32)***. The intensity of an NOE, i.e., the volume of the corresponding cross peak in a NOESY spectrum ***(31,33,34)***, is related to the distance r between the two interacting spins by

   $$V = (r^{-6} > f(\tau_c)$$

   where r^{-6} is averaged since the distance r may vary in molecules with inherent flexibility. The remaining dependence of the magnetization transfer on motion enters through the function $f(\tau_c)$ that includes effects of global and internal motions of the molecule.

   The NOE is quantified by the volume or intensity of the corresponding cross peak in the NOESY spectrum ***(35)***. Because the linewidths can vary appreciably for different resonances, cross peaks should be quantified by integration over the peak area rather than by measuring peak heights.
2. Measured crosspeak (NOE) volumes are translated to distance ranges. The lower bound is determined from the sum of the van der Waals' radii and the upper bound from the NOE intensity. NOEs are usually translated into upper bounds on interatomic distances rather than precise distance restraints because the presence of internal motions, spin diffusion and, possibly, chemical exchange may affect the intensity of an NOE ***(35)***.

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

Assuming a rigid body, upper distance bounds (u) are calibrated using the equation $V = k/u^6$, where k is a constant that can be determined from known distances, for example the sequential distances $d(H_{\alpha i}, HN_{i+1})$ and $d(HN_i, HN_{i+1})$ in a regular secondary structure element ***(36)*** or by reference to a preliminary structure ***(37)***. The value of u obtained from the above equation may either be used directly as an upper distance bound or NOEs may be calibrated into different classes according to their volume, using the same upper-bound u for all NOEs in a given class. The upper distance bounds are typically put into three classes according to the measured volume of the corresponding NOE cross-peak, for example 2.8 Å (strong), 3.5 Å (medium), and 5.0 Å (weak) ***(7,38)***. This calibration usually yields good results provided that there is a large number of restraints. However, if greater accuracy is required, for example when ligand-binding sites are being studied, a means of obtaining tighter distance restraints from NOE peak intensities is necessary. The full relaxation matrix is commonly used to achieve this ***(39,40)***.

3. Two lines from a typical NOE distance restraint file:

```
assign (resid 1 and name HA)     (resid 2 and name HN)   2.0 0.2 0.8
assign (resid 1 and name HG1#)   (resid 31 and name HA)  2.5 1.3 5.5
```

   The first statement selects the first atom of residue number 1 and the second statement selects the second atom of residue number 2 or 31. The interpretation of the real numbers is dependent on the particular restraining function used for NOE restraints. Here, the first number is deduced from the NOE intensity and the third number is the error value (to account for exchange, spin diffusion, chemical exchange, error in integration of peaks, and so on).

4. NOEs that involve groups of protons with degenerate chemical shifts, in particular methyl groups, are commonly referred to pseudoatoms located at the geometric center of the protons that they represent, and the upper bound is increased by a pseudoatom correction equal to the proton–pseudoatom distance ***(41)***. Programs for automated pseudoatom distance corrections in NOE tables include that written in FORTRAN by M. Nilges (EMBL, Heidelberg) with a C version by M. Osawa (OCI, Toronto) (available from our website — http://diana.oci.utoronto.ca/ikura/datasoft.html).

   In X-PLOR, the setup of pseudoatoms is accomplished by using the NOE assign statement with multiple protons in either atom selection. For example, a medium-range NOE from an Ala methyl group of residue number 1 to the HN proton of the residue number 12 can be written as:

```
assign (resid 1 and name HB#) (resid 12 and name HN) 3.0 1.8 3.1
```

   This assign statement sets the lower bound to 1.2 Å and the upper bound to 6.1 Å. The additional pseudoatom correction (1.1 Å) ***(41)*** is added to upper distance bounds. Pseudoatoms (multiple atom selections) should be used primarily for unresolved NOE cross peaks like those of methyl groups, prochiral centers, and aromatic rings. In the case of stereospecific assignments, the distances should be specified explicitly.

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

Table 2
Pseudoatom Representation for Some Amino Acids Used in the Structure Determination of Proteins from NMR Data

Residue	Pseudoatom representation		^{1}H atoms represented
	X-PLOR/ARIA	DYANA	
Gly	HA#	QA	α-methylene
Ala	HB#	QB	β-methylene
Val	HG1#, HG2#	QG1, QG2	γ1-, γ2-methyl
	HG#	QQG	all six g-methyl
Ile	HG1#, HG2#	QG1, QG2	γ1-methylene,
	HD#	QD	γ2-methyl
Leu	HD1#, HD2#	QD1, QD2	δ1-, δ2-methyl
	HD#	QD	all six δ-methyl
Pro	HB#, HG#, HD#	QB, QG, QD	β-, γ-, δ-methylene
Ser, Asp, Cys, His, Trp	HB#	QB	β-methylene
Thr	HG#	QG	γ-methylene
Asn	HD2#	QD2	δ2-amido
Glu	HB#, HG#	QB, QG	β-, γ-methylene
Gln	HB#, HG#	QB, QG	β-, γ-methylene
	HE2#	QE2	ε2-amido
Phe, Tyr	HD#, HE#	QD, QE	δ1- and δ2-ring, ε1- and ε2-ring

Table 2

The pseudoatom nomenclature used in X-PLOR and ARIA is different from DYANA and is listed in **Table 2**. A universal nomenclature of pseudoatoms for NMR structure calculation and representation has been recommended ***(42)***.

5. Calcium-binding proteins require an additional distance restraints list to account for Ca^{2+}–protein interactions. For example, calmodulin binds to four Ca^{2+}, which are located in EF-hand loop regions ***(43,44)***. The distance list shown below was used in calculation of a calmodulin–peptide complex structure ***(45)***. In the list, sites I, II, III, and IV indicate the four Ca^{2+}-binding sites in calmodulin. Calcium atoms are assigned the residue numbers 149–152.

```
! site I
assign (segid A and resid 20 and name OD2)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 22 and name OD1)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 24 and name OD2)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 26 and name O)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 31 and name OE1)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 31 and name OE2)(resid 149 and name CA2) 2.5 0.8 0.3
! site II
assign (segid A and resid 56 and name OD2)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 58 and name OD2)(resid 150 and name CA2) 2.5 0.8 0.3
```

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging


```
assign (segid A and resid 60 and name OD1)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 62 and name O)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 67 and name OE1)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 67 and name OE2)(resid 150 and name CA2) 2.5 0.8 0.3
! site III
assign (segid A and resid 93 and name OD2)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 95 and name OD2)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 97 and name OD1)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 99 and name O)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 104 and name OE1)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 104 and name OE2)(resid 151 and name CA2) 2.5 0.8 0.3
! site IV
assign (segid A and resid 129 and name OD2)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 131 and name OD2)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 133 and name OD2)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 135 and name O)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 140 and name OE1)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 140 and name OE2)(resid 152 and name CA2) 2.5 0.8 0.3
```

6. The vicinal scalar coupling constant, ^3J, between atoms separated by three covalent bonds can provide useful geometric information that is complementary to that from the NOE data. In contrast to NOEs, the coupling constants give information only on the local conformation. They are nevertheless important to define accurately the local conformation, to obtain stereo-specific assignments for diastereotopic protons (usually β protons), and to detect torsion angles (usually c^i) that occur in multiple states. Vicinal scalar coupling constants can be translated into dihedral angles by a Karplus-type equation ***(46)***:

$$^3J(\theta) = A\cos^2\theta + B\cos\theta + C$$

where the parameters *A*, *B*, and *C* are constants and are determined for various types of coupling constant by a best fit of the measured ^3J values to the corresponding values calculated with the above equation from known structures. The most commonly used Karplus relations in proteins are given in **Table 3**.

7. The following is an example of a dihedral angle restraint table.

```
assign   (resid 2 and name C) (resid 3 and name CA)
         (resid 3 and name N)(resid 3 and name C)    1–120 40.0 2 {* 9 Hz *}
assign   (resid 3 and name C) (resid 4 and name CA)
         (resid 4 and name N)(resid 4 and name C)    1–120 50.0 2 {* 8 Hz *}
```

The four selections of each assign statement specify the particular dihedral angle. The first number after the selections specifies the energy constant in kcal/mol/rad, the second number specifies degrees to which the dihedral angle is restrained, the third number specifies the range around the restrained angle, and the last number specifies the exponent of the restraining function ***(11)***.

8. Secondary structures in proteins have characteriztic NOE patterns and ^3J coupling constants ***(3)***. These two parameters have extensively been used to assign secondary structures (α-helix, 3_{10}-helix, β-sheet and coil) in proteins, the details of which have been documented elsewhere ***(3)***. Recently, the chemical shifts of

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

Table 3
The Most Commonly Used Karplus Relations, $^3J(\theta) = A\cos^2\theta + B\cos\theta + C$, for Proteins to Obtain a Torsion Angle θ from the Corresponding ^3J Coupling Constant

Angle	Coupling	A(Hz)	B(Hz)	Offset C(Hz)	(degree)[a]	Ref.
ϕ	H^N–H^α	6.98	–1.38	1.72	–60	***47***
	H^N–C'	4.32	0.84	0.00	180	***47***
	H^N–Cb	3.39	–0.94	0.07	60	***47***
ψ	H^α–N_{i+1}	–0.88	–0.61	–0.27	–120	***48***
χi	H^α–H^β	9.5	–160	1.80	–120/0	***49***
	N–H^β	–4.40	1.20	0.10	120/–120	***50***
	C'–H^β	7.20	–2.04	0.60	0/120	***51***

[a]Difference between θ and the standard torsion angle ϕ, ψ, and χ^I.

C^α, C^β, C´, and H^α are also being routinely used for identifying local backbone conformation in proteins ***(52,53)***. The C^α and C´ nuclei show an upfield shift in β-strand and a downfield shift in helical structures relative to random coil shifts. Both C^β and H^α nuclei exhibit the opposite correlation of a downfield shift in β-strands and an upfield shift in helices. Various methods are available for identifying secondary structure elements from the chemical shifts, such as the chemical shift index (CSI) ***(54)***. We routinely employ Metzler's method ***(55)*** that uses a combination of C^α and C^β chemical shifts of *i*-1, *i*, and *i*+1 residues.

Information about the secondary structure elements in a protein helps with the structural determination process in two ways. First, it allows deduction of dihedral angle restraints based on regular secondary structures. To this end, a new method called TALOS has recently been developed to extract ϕ and ψ angle restraints by searching a database for chemical shift and sequence homology ***(56)***. Second, hydrogen-bonding restraints may be added to regions assigned to a regular secondary structure, although caution must be taken to ensure that the amide exchange rate data concur with the secondary structure already deduced from the NOEs, ^3J-coupling constants, and chemical shifts. Any discrepancy implies a distortion of the regular structure or the presence of flexible regions.

Acknowledgments

This work is supported by a grant from the Medical Research Council of Canada (MRCC) and National Cancer Institute of Canada (NCIC) to M. I. and by OCI/Amgen Fellowship and NCIC to T. K. Mai and M. Ikura are MRCC Scientist and Howard Hughes Medical Institute of International Research Scholars. We thank to Tao Yuan, Hong Qian, and Antonio Pineda-Lucena for useful discussions.

Au: OK?

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

References

1. Drenth, J. (1994) *Principles of Protein X-ray Crystallography*. Springer, New York.
2. Abragam, A. (1961) *Principles of Nuclear Magnetism*. Clarendon, Oxford.
3. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
4. Arseniev, A. S., Kondakov, V. I., Maiorov, V. N., and Bystrov, V. F. (1984) The solution spatial structure of short scorpion insectotoxin I5A. *FEBS Lett.* **165,** 57–62.
5. Braun, W., Bosch, C., Brown, L. R., Gö, N., and Wüthrich, K. (1981) Combined use of proton-proton Overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon. *Biochim. Biophys. Acta* **667,** 377–396.
6. Clore, G. M., Brünger, A. T., Karplus, M., and Gronenborn, A. M. (1986) Application of molecular dynamics with interproton distance restraints to three dimensional protein structure determination: a model study of crambin. *J. Mol. Biol.* **191,** 523–551.
7. Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985) Solution conformation of proteinase inhibitor IIa from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182,** 295–315.
8. Zuiderweg, E. R. P., Billeter, M., Boelens, R., Scheek, R. M., Wüthrich, K., and Kaptein, R. (1984) Spatial arrangement of the three a helices in a solution structure of E. coli lac repressor DNA-binding domain. *FEBS Lett.* **174,** 243–247.
9. Havel, T. F. and Wüthrich, K. (1984) A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^{1}H–^{1}H proximities in solution. *Bull. Math. Biol.* **46,** 673–698.
10. Braun, W. and Gö, N. (1985) Calculation of protein conformation by proton-proton distance constraints: a new efficient algorithm. *J. Mol. Biol.* **186,** 611–626.
11. Brünger, A. T. (1992) *X-PLOR, Version 3. 1. A System for X-Ray Crystallography and NMR*. Yale University Press, New Haven, CT.
12. Nilges, M. (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245,** 645–660.
13. Nilges, M., Macias, M., O'Donoghue, S. I., and Oschkinat, H. (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β spectrin. *J. Mol. Biol.* **269,** 408–422.
14. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273,** 283–298.
15. Güntert P. (1998) Structure calculation of biological macromolecules from NMR data. *Quart. Rev. Biophys.* **31,** 145–237.
16. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3680–3690.

17. Laskowski, R. A., Rullman, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** 477–486.
18. Vriend, G. and Sander, C. (1993) Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallogr.* **26,** 47–60.
19. Kuszewski, J, Qin, J., Gronenborn, A. M., and Clore, G. M. (1995) The impact of direct refinement against $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **106,** 92–96.
20. Kuszewski, J., Gronenborn, A. M., and Clore, G. M. (1995) The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **107,** 293–297.
21. Kuszewski, J., Gronenborn, A. M., and Clore, G. M. (1996) A potential involving multiple proton chemical shift restraints for nonstereospecifically assigned methyl and methylene protons. *J. Magn. Reson. Ser. B* **112,** 79–81.
22. Brüschweiler, R., Liao, X., and Wright, P. E. (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268,** 886–889.
23. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., and Bax, A. (1997) Use of dipolar $^{1}H–^{15}N$ and $^{1}H–^{13}C$ couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* **4,** 732–738.
24. Clore, G. M. and Gronenborn, A. M. (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.* **16,** 22–34.
25. Güntert, P., Berndt, K. D., and Wüthrich, K. (1993) The program ASNO for computer-supported collection of NOE upper distance restraints as input for protein structure determination. *J. Biol. NMR* **3,** 601–606.
26. Nilges, M., Gronenborn, A. M., Brünger, A. T., and Clore, G. M. (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2,** 27–38.
27. Hanggi, G. and Braun, W. (1994) Pattern recognition and self-correcting distance geometry calculations applied to myohemerythrin. *FEBS Lett.* **344,** 147–153.
28. Folmer, R. H. A., Nilges, M., Papavoine, C. H. M., Harmsen, B. J. M., Konings, R. N. H., and Hilbers, C. W. (1997) Refined structure, DNA binding studies, and dynamics of the bacteriophage Pf3 encoded single-stranded DNA binding protein. *Biochemistry* **36,** 9120–9135.
29. Abe, H., Braun, W., Noguti, T., and Gö, N. (1983) Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles in proteins. General recurrent equations. *Comput. Chem.* **8,** 239–247.
30. Solomon, I. (1955) Relaxation processes in a system of two spins. *Phys. Rev.* **99,** 559–565.
31. Macura, S. and Ernst, R. R. (1980) Elucidation of cross relaxation in liquids by 2D NMR spectroscopy. *Mol. Phys.* **41,** 95–117.
32. Neuhaus, D. and Williamson, M. P. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis.* VCH, New York.

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

33. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71,** 4546–4556.
34. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95,** 1–6.
35. Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *The Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Clarendon, Oxford.
36. Billeter, M., Braun, W., and Wüthrich, K. (1982) Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations. *J. Mol. Biol.* **155,** 321–346.
37. Güntert, P., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. J., and Wüthrich, K. (1991) Structure determination of the Antp(C39(S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217,** 531–540.
38. Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., and Gronenborn, A. M. (1986) The three-dimensional structure of aI-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrainted molecular dynamics. *EMBO J.* **5,** 2729–2735.
39. Borgias, B. A. and James, T. L. (1989) Two-dimensional nuclear Overhauser effect: complete relaxation matrix analysis. *Methods Enzymol.* **176,** 169–183.
40. Bovin, A. M., Rullmann, J. A., Lamerichs, R. M., Boelens, R., and Kaptein, R. (1993) Ensemble iterative relaxation matrix approach: a new NMR refinement protocol applied to the solution structure of Crambin. *Proteins* **15,** 385–400.
41. Wüthrich, K., Billeter, M., and Braun, W. (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance restraints with nuclear magnetic resonance. *J. Mol. Biol.* **169,** 949–961.
42. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., et al. (1998) Recommendation for the presentation of NMR structures of proteins and nucleic acids. *Pure Appl. Chem.* **70,** 117–142.
43. Babu, Y. S., Sack, J. S., Greenhough, J. J., Bugg, C. E., Means, A. R., and Cook, W. J. (1985) Three-dimensional structure of calmodulin. *Nature* **315,** 37–40.
44. Wylie, D. C. and Vanaman, T. C. (1988) Structure and evolution of the calmodulin family of calcium regulatory protein, in *Calmodulin* (Cohen, P. and Klee, C. B., eds.), ELSEVIER Science Publishers, Amsterdam, pp. 1–15.
45. Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256,** 632–638.
46. Karplus, M. (1963) Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85,** 2870–2871.

Job: Vogel 2 689-8 (MiMB)
Chapter: 21 (*Mal et al.*)
Pub Date: March 2001
Operator: Sean Murdock
Date: Tuesday, January 23, 2001
Revision: 1st Galleys/Paging

47. Wang, A. C. and Bax, A. (1996) Determination of the backbone dihedral angles (in human ubiquitin from reparametrized empirical Karplus equations. *J. Am. Chem. Soc.* **118,** 2483–2494.
48. Wang, A. C. and Bax, A. (1995) Reparametrization of the Karplus relation for $^{3}J(H_a$-N) in peptides from uniformly $^{13}C/^{15}N$ enriched human ubiquitin. *J. Am. Chem. Soc.* **117,** 1810–1813.
49. De Marco, A. C., Llinas, M., and Wüthrich, K. (1978) Analysis of the ^{1}H-NMR spectra of ferrichrome peptides. I. The non-amide protons. *Biopolymers* **17,** 617–636.
50. De Marco, A. C., Llinas, M., and Wüthrich, K. (1978) $^{1}H–^{15}N$ spin-spin couplings in alumichrome. *Biopolymers* **17,** 2127–2742.
51. Fischman, A. J., Live, D. H., Wyssbrod, H. R., Agosta, W. C., and Cowburn, D. (1980) Torsion angles in the cystine bridge of oxytocin in aqueous solution. Measurements of circumjacent vicinal couplings between 1H, 13C, and ^{15}N. *J. Am. Chem. Soc.* **102,** 2533–2539.
52. Wishart, D. S. and Nip, A. M. (1998) Protein chemical shift analysis: a practical guide. *Biochem. Cell Biol.* **76,** 153–163.
53. Venters, R. A., FarmerII, B. T., Fierke, C. A., and Spicer, L. D. (1996) Characterizing the use of perdeuteration in NMR studies of large proteins: 13C, ^{15}N and ^{1}H assignments of human carbonic anhydrase II. *J. Mol. Biol.* **264,** 1101–1116.
54. Wishart, D. S. and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239,** 363–392.
55. Metzler, W. J., Constantine, K. L., Friedrichs, M. S., Bell, A. J., and Ernst, E. G. (1993) Characterization of the three-dimensional solution structure of human profilin: 1H, ^{13}C and ^{15}N assignments and global folding pattern. *Biochemistry* **32,** 13,818–13,829.
56. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biolmol. *NMR* **13,** 289–302.