J. Am. Chem. Soc. **1992**, *114*, 6195–6199

The "CUPID" Method for Calculating the Continuous Probability Distribution of Rotamers from NMR Data

Željko Džakula,[†] William M. Westler, Arthur S. Edison,[‡] and John L. Markley*,[‡]

Contribution from the Biochemistry Department, College of Agricultural and Life Sciences, University of Wisconsin—Madison, 420 Henry Mall, Madison, Wisconsin 53706.
Received September 30, 1991

Abstract: We present a new method, ContinUous ProbabIlity Distribution of rotamers (CUPID), for determining the distribution of rotamer probability $\rho(\chi)$ about a dihedral angle χ. This method utilizes measured vicinal homonuclear and heteronuclear spin–spin coupling constants (J) and nuclear Overhauser enhancements (NOEs) from NMR spectra and demands no prior assumption about the conformations or the degree of flexibility across the bond. To analyze the NMR measurements, it is necessary to know $J(\chi)$, the angular dependence of each coupling constant, and $r(\chi)$, the angular dependence of each interatomic distance giving rise to an NOE. With $2N$ or more measured coupling constants and/or NOEs and with known functions $J(\chi)$ and $r(\chi)$, one determines the probability distribution by solving for the coefficients of the Fourier expansion of $\rho(\chi)$ to order N. As a further refinement to $\rho(\chi)$, we make the assumption that the distribution functions are sums of Gaussians. The theory underlying CUPID is presented here; the accompanying article (Džakula, Ž.; Edison, A. S.; Westler, W. M.; Markley, J. L. J. Am. Chem. Soc., following paper in this issue) demonstrates an application of CUPID to the analysis of simulated data for an α-helix in a protein.

1. Introduction

The dihedral angular dependence of vicinal nuclear spin–spin coupling constants, $J(\theta)$, is well established.[1–4] This dependence was first described by Karplus[1] as $J(\theta) = a \cos^2(\theta) + b \cos(\theta) + c$, where a, b, and c are empirical parameters, and θ is the dihedral angle between two coupled nuclei. If the angle is rigid, θ can be determined from measurements of two different couplings, with some exceptions.[5] However, conformationally flexible molecules give rise to average coupling constants weighted by their probability distribution.[6] It is generally impossible to know, a priori, the conformational state of a molecule. To account for conformational averaging, it is usually assumed[7–13] that a set of three discrete values of θ exhausts all conformational possibilities. The angles are estimated by a "best guess", "chemical intuition" (leading most often to staggered conformers), or a calculation of rotational potentials. The problem is then expressed as

$$\langle J\rangle = \sum_{i=1}^{3} p_i J(\theta_i), \text{ with } \sum_{i=1}^{3} p_i = 1$$

where p_i (i = 1, 2, 3) are the unknown populations of the discrete conformers, and $\langle J\rangle$ is the measured averaged coupling constant. We refer to this as the "discrete model". Many difficulties are associated with this model. First, deviations from the commonly used staggered conformers are often found.[14–16] Second, even when nonstaggered angles are recognized, it is still necessary to estimate the actual values of θ before the problem can be solved. Third, the three conformers are assumed to cover the entire conformational space. Fourth, the discrete probability distribution, defined by delta (δ) functions, neglects contributions to $\langle J\rangle$ from dihedral angles with low probabilities but large couplings. Finally, due to the normalization, solutions for p_i are possible regardless of the assumed angles; thus the approach provides no check on their appropriateness. Clearly, this model is inadequate to describe many molecules (especially biological macromolecules) where torsional motion will broaden the assumed δ functions and where forces resulting from tertiary structural interactions, such as hydrogen bonds, can cause significant deviations from the assumed discrete distribution.[17]

Vicinal spin–spin couplings between α and β protons in amino acids have been used in conjunction with intra-residue NOEs to determine χ_1 rotamer populations in the side chains of peptides and proteins according to a set of rules formulated by Wagner et al. (Figure 4 of ref 18). However, the approach is valid only when the conformation is fixed, i.e., in the absence of rotational averaging. In addition, the approach assumes that the dihedral angle χ_1 takes only staggered values. When the NMR data are inconsistent with such predictions, as sometimes happens,[18] the model does not distinguish whether the discrepancy is due to several interconverting rotamers (staggered and/or nonstaggered) or one nonstaggered conformer.

Wyssbrod[5] proposed two methods for analyzing peptide coupling data: a intersection-of-sets method (based on direct geometrical analysis of Karplus curves) and conformation-locus diagrams for determining restricted χ_1 conformations of the side chains of amino acids with a β-methylene group. The conformation-locus method

---

[†] Permanent address: Institute of Physiology and Biochemistry, Faculty of Biology, School of Sciences, and Biophysical Laboratory, Center for Multidisciplinary Studies, University of Belgrade, Studentski trg 3, 11000 Belgrade, Serbia, Yugoslavia.

* To whom correspondence should be addressed.

[‡] Graduate Biophysics Program, University of Wisconsin—Madison.

(1) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11–18.

(2) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871.

(3) Barfield, M.; Grant, D. M. In *Advances in Magnetic Resonance*; Waugh, J. S., Ed.; Academic Press: New York, 1965; Vol. 1, pp 149–193.

(4) Kowalewski, J. In *Annual Report in NMR Spectroscopy*; Webb, G. A., Ed.; Academic Press: New York, 1982; Vol. 12, pp 81–176.

(5) Wyssbrod, H. R. In *Proceedings of the Seventh American Peptide Symposium*; Rich, D. H., Gross, E., Eds.; Pierce Chemical Co.: Rockford, IL, 1981; pp 261–265.

(6) Three distinct cases of averaging need to be considered:[7] (*a*) motion rapid compared with the reciprocal of the coupling constant which yields couplings that are rotationally averaged; (*b*) motion slower than the reciprocal of the coupling constant with only one conformer detected by *NMR*; (*c*) motion much slower than the reciprocal of the coupling constant such that each conformer exists long enough to be detected by *NMR*. In this article, we consider only case (*a*) which most often is the situation in liquids since vicinal coupling constants are generally quite small.

(7) Pople, J. A. *Mol. Phys.* **1958**, *1*, 3–8.

(8) Jardetzky, O. *Biochim. Biophys. Acta* **1980**, *621*, 227–232.

(9) Bystrov, V. L.; Portnova, S. L.; Balashova, T. A.; Tsetlin, V. I.; Ivanov, V. T.; Kostetzky, P. V.; Ovchinnikov, Yu. A. *Tetrahedron Lett.* **1969**, *59*, 5225–5228.

(10) Colucci, W. J.; Gandour, R. D.; Mooberry, E. A. *J. Am. Chem. Soc.* **1986**, *108*, 7141–7147.

(11) Cowburn, D.; Live, D. H.; Fischman, A. J.; Agosta, W. C. In *Intramolecular Dynamics*; Jortner, J., Pullman, B., Eds.; D. Reidel: New York, 1982; pp 473–480.

(12) Cowburn, D.; Live, D. H.; Fischman, A. J.; Agosta, W. C. *J. Am. Chem. Soc.* **1983**, *105*, 7435–7442.

(13) Fischman, A. J.; Wyssbrod, H. R.; Agosta, W. C.; Cowburn, D. *J. Am. Chem. Soc.* **1978**, *100*, 54–58.

(14) Nagayama, K.; Wüthrich, K. *Eur. J. Biochem.* **1981**, *115*, 653–657.

(15) Fujinaga, M.; Sielecki, A. R.; Read, R. J.; Ardelt, W.; Laskowski, M., Jr.; James, M. N. G. *J. Mol. Biol.* **1987**, *195*, 397–418.

(16) Langs, D. A.; Smith, G. D.; Stezowski, J. J.; Hughes, R. E. *Science* **1986**, 1240–1242.

(17) Fischman, A. J.; Live, D. H.; Wyssbrod, H. R.; Agosta, W. C.; Cowburn, D. *J. Am. Chem. Soc.* **1980**, *102*, 2533–2539.

(18) Wagner, G.; Braun, W.; Havel, T. F.; Schaumann, T.; Gō, N.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 611–639.

0002-7863/92/1514-6195$03.00/0

uses a correlation diagram of ${}^3J_{\alpha\beta_2}$ versus ${}^3J_{\alpha\beta_3}$ vicinal proton–proton couplings (see also ref 19). However, it does not take into account rotational averaging between distinct conformers. At about the same time, Nagayama and Wüthrich[14] used the same approach to study χ_1 conformers of BPTI. They accounted for rapid small-amplitude vibrations about a single rotamer by assuming simple uniform averaging within a limited interval of angles. Hyberts et al.[20] suggested an extension of that method, based on use of H^α–H^β COSY cross-peak multiplet structure in conjunction with H^N–H^β NOEs in order to assign stereospecifically the β-resonances and to determine concomitantly values of χ_1 and ϕ torsion angles. The weakness of that approach is that it attempts to determine simultaneously two torsion angles (χ_1 and ϕ) on the basis of a data set limited to only four measured numbers: two scalar couplings and two NOEs. All these methods are suitable only when there is a single conformer and when no large-amplitude rotations take place.

The literature contains references to at least two continuous models used in computing average values of coupling constants. The first, proposed by Bystrov et al.,[21] assumes no potential barrier and simply averages J over a complete rotation. The second assumes knowledge of the rotational potential energy and uses a Boltzmann distribution to describe the averaged coupling constant (see ref 22–24 and references therein). This approach is useful in principle, but suffers when knowledge of the rotational potential function is lacking (as is the case for most molecules, especially biological macromolecules).

As far as rotational averaging of NOEs is concerned, an attempt was made by Schirmer et al.[25] to reconstruct the continuous probability distribution of glycosyl rotamers in nucleosides from NOEs. They fitted the experimental data to a probability distribution made up of a sum of Gaussian curves. This approach has been difficult to implement and has not gained much popularity.

A uniform averaging model was used by Braun et al.[26] to estimate the upper limits of interproton distances from rotationally averaged NOE data. In that model the distance between the hydrogen atoms was allowed to vary uniformly between a minimum (equal to the van der Waals contact distance of 2.0 Å) and an unknown maximum value, which has to be determined from the experiment. Their simple approach formed a conceptual basis for the estimation of interproton distance constraints from NOE data, which then serve as input to distance–geometry analysis.

Recently, Torda et al.[27] recognized the importance of rotational averaging of NOEs and proposed the use of molecular dynamics in order to compute the average proton–proton distances and NOE intensities. This approach, however, is both cumbersome and computationally expensive.

As pointed out above, all these approaches have certain weaknesses and fail to identify multiple rotational isomers (staggered or nonstaggered), fail to quantify their probabilities, or do not estimate their mobilities. We present here a method for the analysis of NMR coupling constants and nuclear Overhauser enhancements called ContinUous ProbabIlity Distribution of rotamers (CUPID) that yields a continuous angular distribution of probability, $\rho(\chi)$. Our approach simultaneously solves the problems of determining the flexibility and the most probable angles.

2. Theory

2.1. Foundations. 2.1.1. Angular Probability Distribution. The probability distribution of rotamers, $\rho(\chi)$, can be expanded as a Fourier series by making use of the fact that it must be a periodic function of the dihedral angle χ

$$\rho(\chi) = \rho_0 + \sum_{n=1}^{\infty} \rho_n \cos(n\chi) + \sum_{n=1}^{\infty} \sigma_n \sin(n\chi) \tag{1}$$

where ρ_0, ρ_n, and σ_n are coefficients of the expansion. After taking into account the normalization condition that the distribution function must obey

$$\int_{-\pi}^{\pi} \rho(\chi)\,\mathrm{d}\chi = 1 \tag{2}$$

and the fact that integrals of functions cos ($n\chi$) and sin ($n\chi$) over the interval ($-\pi$, π) vanish, it is easily seen that

$$\rho_0 = 1/(2\pi) \tag{3}$$

In practice, $\rho(\chi)$ can be approximated by evaluating eq 1 as a limited sum:

$$\rho(\chi) \approx \frac{1}{2\pi} + \sum_{n=1}^{N} \rho_n \cos(n\chi) + \sum_{n=1}^{N} \sigma_n \sin(n\chi) \tag{4}$$

The main goal of this work is to reconstruct $\rho(\chi)$ by calculating the $2N$ coefficients ρ_n and σ_n ($n = 1, \ldots, N$) from a set of experimentally measured spin–spin coupling constants and NOEs.

2.1.2. Rotationally Averaged Spin–Spin Coupling Constants. The general relationship between vicinal nuclear spin–spin coupling constants and dihedral angles (Karplus equation) can be expressed as

$$J(\theta) = a\cos^2(\theta) + b\cos(\theta) + c = \frac{a}{2}\cos(2\theta) + b\cos(\theta) + c + \frac{a}{2} \tag{5}$$

where J is the coupling constant, θ is the dihedral angle between the coupled spins, and a, b, and c are empirically derived parameters. Note that the angle θ may differ in phase from the defined dihedral angle χ. For example, as shown in Figure 1 of the accompanying paper,[28] the standard definition of the dihedral angle about the C^α–C^β bond in amino acids, χ_1, is such that $\chi_1 = 0°$ when the atoms, N, C^α, C^β, and X^γ (the heteroatom in γ position) are coplanar.[29] However, when $\chi_1 = 0°$, the dihedral angle between the atoms H^α and $H^{\beta 2}$ (about the same C^α–C^β bond) is $\theta = -120°$. When the defined dihedral angle χ about a given bond differs from θ, it is convenient to express θ in terms of χ, i.e., as $\chi + \Omega$, where Ω is a constant that defines the stereochemistry for a particular pair of coupled atoms. Equation 5 then becomes

$$\begin{aligned} J(\theta) = J(\chi + \Omega) &= \frac{a}{2}\cos[2(\chi + \Omega)] + b\cos(\chi + \Omega) + c + \frac{a}{2} = \frac{a}{2}\cos(2\Omega)\cos(2\chi) - \frac{a}{2}\sin(2\Omega)\sin(2\chi) + \\ &\quad b\cos(\Omega)\cos(\chi) - b\sin(\Omega)\sin(\chi) + c + \frac{a}{2} \end{aligned} \tag{6}$$

Different pairs of atoms coupled across the same bond have different dihedral angles θ. However, the notation $\chi + \Omega$ allows one to express all of the couplings across that bond, as well as the probability distribution ρ, in terms of the same variable, χ.

If rotation occurs about χ at a rate much greater than the coupling constant, an NMR measurement yields an average value for the scalar coupling constant

$$\langle J \rangle = \int_{-\pi}^{\pi} J(\chi + \Omega)\rho(\chi)\,\mathrm{d}\chi \tag{7}$$

---

(19) Wagner, G.; Wüthrich, K. *Methods Enzymol.* **1986**, *131*, 307–326.

(20) Hyberts, S. G.; Märki, W.; Wagner, G. *Eur. J. Biochem.* **1987**, *164*, 625–635.

(21) Bystrov, V. F.; Portnova, S. L.; Tsetlin, V. I.; Ivanov, V. T.; Ovchinnikov, Yu. A. *Tetrahedron* **1969**, *25*, 493–515.

(22) Bystrov, V. F. *Prog. Nucl. Magn. Reson. Spectrosc.* **1976**, *10*, 41–81.

(23) Barfield, M.; Hruby, V. J.; Meraldi, J.-P. *J. Am. Chem. Soc.* **1976**, *98*, 1308–1314.

(24) de Leeuw, F. A. A. M.; Altona, C.; Kessler, H.; Bermel, W.; Friedrich, A.; Krack, G.; Hull, W. E. *J. Am. Chem. Soc.* **1983**, *105*, 2237–2246.

(25) Schirmer, R. E.; Davis, J. P.; Noggle, J. H.; Hart, P. A. *J. Am. Chem. Soc.* **1972**, *94*, 2561–2572.

(26) Braun, W.; Bösch, C.; Brown, L. R.; Gō, N.; Wüthrich, K. *Biochim. Biophys. Acta* **1981**, *667*, 377–396.

(27) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *21*, 223–235.

(28) Džakula, Ž.; Edison, A. S.; Westler, W. M.; Markley, J. L. *J. Am. Chem. Soc.*, following paper in this issue.

(29) IUPAC-IUB Commission of Biochemical Nomenclature. *J. Mol. Biol.* **1970**, *52*, 1–17.

Combination of eqs 4, 6, and 7 and use of the orthogonality relations between the functions cos ($n\chi$) and sin ($m\chi$), where n, $m = 0, 1, 2, \ldots$, gives

$$\langle J\rangle = \left(c + \frac{a}{2}\right)\rho_0\int_{-\pi}^{\pi} \mathrm{d}\chi + b\cos(\Omega)\rho_1\int_{-\pi}^{\pi}\cos^2(\chi)\,\mathrm{d}\chi - b\sin(\Omega)\sigma_1\int_{-\pi}^{\pi}\sin^2(\chi)\,\mathrm{d}\chi + \frac{a}{2}\cos(2\Omega)\rho_2\int_{-\pi}^{\pi}\cos^2(2\chi)\,\mathrm{d}\chi - \frac{a}{2}\sin(2\Omega)\sigma_2\int_{-\pi}^{\pi}\sin^2(2\chi)\,\mathrm{d}\chi = c + \frac{a}{2} + \pi b[\rho_1\cos(\Omega) - \sigma_1\sin(\Omega)] + \pi\frac{a}{2}[\rho_2\cos(2\Omega) - \sigma_2\sin(2\Omega)] \tag{8}$$

Note that the Karplus equation truncates eq 4 to the second order. With at least $2N$ measured coupling constants from different pairs of nuclei across a common chemical bond, a set of $2N$ equations similar to eq 8 can be used to determine values of the coefficients ρ_n and σ_n of the Fourier expansion of $\rho(\chi)$ to the order N ($N \leq 2$), provided that the coefficients of the corresponding Karplus equations are known.

2.1.3. Rotationally Averaged NOE. We shall assume here that the internal rotation is faster than longitudinal relaxation (i.e., the period of one complete internal rotation, τ_r, is much smaller than T_1) and that the internal rotation is fast on the chemical shift time scale (i.e., τ_r is smaller than the reciprocal of the difference of the chemical shifts of the nuclei that exhibit dipolar coupling). When the internal rotation is slower than overall molecular tumbling (i.e., τ_r is longer than the correlation time of the overall tumbling of the molecule, τ), the rotational averaging of an NOE between an atom fixed in the coordinate system bound to the molecule and an atom involved in the internal rotation can be expressed as[30-33]

$$\langle\mathrm{NOE}\rangle = f(\tau,\tau_r)\langle r^{-6}\rangle = f(\tau,\tau_r)\int_{-\pi}^{\pi}\rho(\chi)r(\chi)^{-6}\,\mathrm{d}\chi \tag{9}$$

where r is the distance between the two interacting nuclei (a function of the dihedral angle χ) and f is a function of the correlation times whose explicit form is not important for our study. In the case of fast internal rotation ($\tau_r \ll \tau$), averaging of the NOE occurs as[30-33] $\langle r^{-3}\rangle^2$

$$\langle\mathrm{NOE}\rangle = f(\tau,\tau_r)\langle r^{-3}\rangle^2 = f(\tau,\tau_r)\left(\int_{-\pi}^{\pi}\rho(\chi)r(\chi)^{-3}\,\mathrm{d}\chi\right)^2 \tag{10}$$

The integrals in eqs 9 and 10 can be solved easily when the functions $r(\chi)^{-6}$ and $r(\chi)^{-3}$ are expanded into Fourier series. The analogy to the averaging of the coupling constants is obvious, but there is one important difference: unlike averaged couplings, the averaged NOEs are not truncated after the second order. In principle, the averaged NOEs can establish as many Fourier coefficients of $\rho(\chi)$ as there are NOE data points, with experimental error being the only limitation. The Fourier series of the functions $r(\chi)^{-3}$ and $r(\chi)^{-6}$ have the forms

$$r(\chi)^{-3} = C_0 + \sum_{n=1}^{\infty} C_n\cos(n\chi) + \sum_{n=1}^{\infty} S_n\sin(n\chi) \tag{11a}$$

$$r(\chi)^{-6} = \mathbf{c}_0 + \sum_{n=1}^{\infty}\mathbf{c}_n\cos(n\chi) + \sum_{n=1}^{\infty}\mathbf{s}_n\sin(n\chi) \tag{11b}$$

where the Fourier coefficients C_0, C_n, S_n, $\mathbf{c}_0$, $\mathbf{c}_n$, and $\mathbf{s}_n$ depend on the coordinates of the fixed atom. They are analogous to Karplus coefficients, and their values must be known in order to be able to apply CUPID analysis. Upon introducing the Fourier expansions 11a and 11b into the expressions for the average NOE [eqs 9 and 10, respectively] and using the Fourier expansion of the probability distribution (1), one obtains

$$\sqrt{\langle\mathrm{NOE}\rangle} = \sqrt{f(\tau,\tau_r)}\left(C_0 + \pi\sum_{n=1}^{\infty}C_n\rho_n + \pi\sum_{n=1}^{\infty}S_n\sigma_n\right) \approx \sqrt{f(\tau,\tau_r)}\left(C_0 + \pi\sum_{n=1}^{N}C_n\rho_n + \pi\sum_{n=1}^{N}S_n\sigma_n\right) \text{ when } \tau \gg \tau_r \tag{12a}$$

$$\langle\mathrm{NOE}\rangle = f(\tau,\tau_r)\left(\mathbf{c}_0 + \pi\sum_{n=1}^{\infty}\mathbf{c}_n\rho_n + \pi\sum_{n=1}^{\infty}\mathbf{s}_n\sigma_n\right) \approx f(\tau,\tau_r)\left(\mathbf{c}_0 + \pi\sum_{n=1}^{N}\mathbf{c}_n\rho_n + \pi\sum_{n=1}^{N}\mathbf{s}_n\sigma_n\right) \text{ when } \tau \ll \tau_r \tag{12b}$$

where N is the order of the highest terms that are not rejected from $\rho(\chi)$ upon truncation.

It is convenient to normalize experimental NOE data by dividing the intensity of the given NOE by the intensity of a suitably chosen calibration peak, $\mathrm{NOE}_{\mathrm{calibr}}$, multiplied by the sixth power of the corresponding internuclear distance, r^6_{calibr}.

$$\xi\sqrt{\langle\mathrm{NOE}\rangle} = \langle r(\chi)^{-3}\rangle = C_0 + \pi\sum_{n=1}^{N}C_n\rho_n + \pi\sum_{n=1}^{N}S_n\sigma_n \tag{13a}$$

$$\xi^2\langle\mathrm{NOE}\rangle = \langle r(\chi)^{-6}\rangle = \mathbf{c}_0 + \pi\sum_{n=1}^{N}\mathbf{c}_n\rho_n + \pi\sum_{n=1}^{N}\mathbf{s}_n\sigma_n \tag{13b}$$

where

$$\xi = r^{-3}_{\mathrm{calibr}}\langle\mathrm{NOE}_{\mathrm{calibr}}\rangle^{-1/2} = \mathrm{const} \tag{14}$$

The aim of this normalization is the elimination of the term $f(\tau,\tau_r)$, which inserts a factor of r^{-6} into the NOE expression. It is assumed here that τ and τ_r are the same for all atom pairs considered, including the pair that gives rise to the calibration peak. This requirement needs more detailed explanation. As pointed out above, at least two types of motion are relevant to cross relaxation: overall tumbling of the molecule as a whole and internal rotations. These motions generally have different correlation times. In proteins, the correlation times for internal rotations about different bonds vary considerably. For analysis of $\rho(\chi_1)$, the NOE peak between the $\mathrm{H}^{\beta 2}$ and $\mathrm{H}^{\beta 3}$ protons of the amino acid under consideration provides a reliable normalization factor, since it has the same correlation time of internal rotation (neglecting other torsional degrees of freedom in the amino acid) as other NOE peaks used in the analysis of the particular dihedral angle (χ_1). By comparing the normalized and unnormalized NOE peaks corresponding to different amino acids, one can extract correlation times for overall tumbling and correlation times for internal rotations.

2.1.4. Linear Regression of Fourier Coefficients ρ_n and σ_n. Suppose that M ($M_j + M_n = M$) experimental constraints are available: M_j vicinal coupling constants across the bond for which the torsional rotation is defined, and M_n NOEs from fixed (in the frame bound to the molecule) to rotating atoms. A set of M linear equations can be written for $2N$ unknown Fourier coefficients ρ_n and σ_n ($n = 1, \ldots, N$) as

$$\underline{\underline{A}}\cdot\underline{\underline{U}} = \underline{\underline{E}} \tag{15}$$

The $M \times 2N$ matrix $\underline{\underline{A}}$ has the following form

$$\underline{\underline{A}} = \left[\begin{array}{cc} \boxed{\underline{\underline{K}}} & \boxed{\underline{\underline{O}}} \\ \boxed{\qquad\underline{\underline{F}}\qquad} & \end{array}\right] \tag{16}$$

The submatrix $\underline{\underline{K}}$ contains Karplus parameters. Its dimensions are $M_j \times 4$ and its elements are

$$K_{i1} = \pi b_i\cos(\Omega_i);\quad K_{i2} = -\pi b_i\sin(\Omega_i);\quad K_{i3} = \pi\frac{a_i}{2}\cos(2\Omega_i);\quad K_{i4} = -\pi\frac{a_i}{2}\sin(2\Omega_i) \quad \text{where } i = 1, \ldots, M_j \tag{17}$$

where a_i and b_i are Karplus parameters of the ith scalar coupling,

---

(30) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035–6043.

(31) Keepers, J. W.; James, T. L. *J. Am. Chem. Soc.* **1982**, *104*, 929–939.

(32) Keepers, J. W.; James, T. L. *J. Magn. Reson.* **1984**, *57*, 404–426.

(33) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; VCH: New York, 1989; pp 171–175, 503–514.

and Ω_i is the phase factor for the dihedral angle θ of the ith pair of coupled atoms ($\Omega_i = \theta - \chi_1$).

Submatrix $\underline{Q}$ contains $M_j \times (2N-4)$ zeros. Submatrix $\underline{F}$, of dimension $M_n \times 2N$, contains the Fourier coefficients of the function $r(\chi)^{-3}$ when the internal rotation is fast ($\tau \gg \tau_r$)

$$F_{i(2j\text{-}1)} = \pi C_{ij}; \qquad F_{i(2j)} = \pi S_{ij} \qquad \text{where i} = 1, ..., M_n;\ j = 1, ..., N \tag{18a}$$

or the Fourier coefficients of the function $r(\chi)^{-6}$ in the case of slow internal rotation ($\tau \ll \tau_r$)

$$F_{i(2j\text{-}1)} = \pi \mathbf{c}_{ij}; \qquad F_{i(2j)} = \pi \mathbf{s}_{ij} \qquad \text{where } i = 1, ..., M_n;\ j = 1, ..., N \tag{18b}$$

The column $\underline{U}$ contains $2N$ unknown Fourier coefficients of the probability distribution

$$U_{2i\text{-}1} = \rho_i; \qquad U_{2i} = \sigma_i \qquad \text{where } i = 1, ..., N \tag{19}$$

Finally, the column $\underline{E}$ contains the M experimental data. It has the following form:

$$\underline{\underline{E}} = \begin{bmatrix} \boxed{\underline{\underline{J}}} \\ \\ \boxed{\underline{\underline{N}}} \end{bmatrix} \tag{20}$$

where the subcolumn $\underline{J}$ contains M_j measured scalar couplings

$$\mathbf{j}_i = \langle J_i \rangle - c_i - a_i/2 \qquad \text{where } i = 1, ..., M_j \tag{21}$$

Karplus parameters c_i and a_i correspond to ith scalar coupling. Subcolumn $\underline{N}$ contains M_n measured NOEs. In the case of fast internal rotation ($\tau \gg \tau_r$), its elements are

$$\mathbf{n}_i = \xi\langle \mathrm{NOE}_i \rangle^{1/2} - C_{i0} \qquad \text{where } i = 1, ..., M_n \tag{22a}$$

where C_{i0} is the zeroth-order Fourier coefficient of the function $r(\chi)^{-3}$ corresponding to the ith pair of cross-relaxing nuclei. When the internal rotation is slow ($\tau \ll \tau_r$), the form of the elements N_i is slightly different

$$\mathbf{n}_i = \xi^2\langle \mathrm{NOE}_i \rangle - \mathbf{c}_{i0} \qquad \text{where } i = 1, ..., M_n \tag{22b}$$

where $\mathbf{c}_{i0}$ is the zeroth-order coefficient of the function $r(\chi)^{-6}$ corresponding to the ith NOE.

If the relation $M \geq 2N$ is reduced to the equality $M = 2N$, the unknown coefficients ρ_n and σ_n are found easily by solving the set of equations (15) by using Cramer's rule. Otherwise, if one chooses to retain a smaller number of terms from the Fourier series of $\rho(\chi)$ than the theoretical maximum allowed by the size of the set of experimental data, linear regression can be used. The latter case (with $M > 2N$) is preferable, because the additional data compensate for experimental errors in the measured data and increase the reliability of the higher-order coefficients ρ_n and σ_n. The best set of Fourier coefficients ρ_n and σ_n (when $M > 2N$) corresponds to the minimum of the sum of the squares of differences between experimental and theoretical couplings and NOEs, given by

$$\mathcal{F} = \sum_{i=1}^{M_j}(\langle J_i \rangle_{\text{expt}} - \langle J_i \rangle_{\text{theor}})^2 + \xi\sum_{i=1}^{M_n}(\langle \mathrm{NOE}_i \rangle_{\text{expt}}^{1/2} - \langle \mathrm{NOE}_i \rangle_{\text{theor}}^{1/2})^2, \quad \tau \gg \tau_r \tag{23a}$$

$$\mathcal{F} = \sum_{i=1}^{M_j}(\langle J_i \rangle_{\text{expt}} - \langle J_i \rangle_{\text{theor}})^2 + \xi^2\sum_{i=1}^{\mathrm{M}_n}(\langle \mathrm{NOE}_i \rangle_{\text{expt}} - \langle \mathrm{NOE}_i \rangle_{\text{theor}})^2, \quad \tau \ll \tau_r \tag{23b}$$

where ξ is defined by eq 14. $\mathcal{F}$ is a function of ρ_n and σ_n ($n = 1, ..., N$). The necessary condition for this function to have a minimum is given by the following set of equations:

$$(\partial\mathcal{F}/\partial\rho_n)_{\rho_{m\neq n},\sigma_k} = 0; \qquad (\partial\mathcal{F}/\partial\sigma_n)_{\rho_k,\sigma_{m\neq n}} = 0 \qquad \text{where } n, m, k = 1, ..., N \tag{24}$$

which can be rewritten as

$$\alpha\cdot\mathbf{U} = \beta \tag{25}$$

The elements of the $2N \times 2N$ matrix $\underline{\underline{\alpha}}$ are

$$\alpha_{ij} = \sum_{m=1}^{M} A_{mi}A_{mj} \qquad \text{where } i, j = 1, ..., 2N \tag{26}$$

Column $\underline{\beta}$ has the following elements:

$$\beta_i = \sum_{m=1}^{M_j} \mathbf{j}_m A_{mi} = \sum_{m=1}^{M_n} \mathbf{n}_m A_{mi} \qquad \text{where } i = 1, ..., 2N \tag{27}$$

The optimal set of values of the Fourier coefficients (ρ_n and σ_n) is found easily by solving the matrix equation (25) by applying Cramer's rule.

2.2. Nonnegativity and Higher-Order Terms. 2.2.1. Reconstruction of the Higher-Order Terms by Using Gaussian Curves. When the Fourier series of the probability distribution is truncated by the neglect of higher-order terms, it can exhibit oscillatory behavior. These oscillations are absent in the real probability distribution and are an artifact caused by the truncation. Although the oscillations often have a small amplitude, they can became troublesome by violating the nonnegativity condition at some values of χ. This problem can be circumvented by employing Gaussian curves centered at (or near) the peaks of the trigonometric distribution. The probabilities, exact positions, and widths of the Gaussians have to be adjusted so that the first $2N$ Fourier coefficients of the sum of Gaussians have the same values as those obtained directly from experimental data [eqs 15–27]. Fourier coefficients of a Gaussian curve with a maximum at φ and with width Δ are

$$\rho_0 = \frac{1}{2\pi}; \qquad \rho_n = \frac{1}{\Delta\pi^{3/2}}\int_{-\pi}^{\pi} e^{-(\chi-\varphi)^2/\Delta^2} \cos(n\chi)\mathrm{d}\chi = \frac{1}{\pi}\cos(\mathrm{n}\varphi)e^{-(n\Delta/2)^2} \tag{28}$$

$$\sigma_n = \frac{1}{\Delta\pi^{3/2}}\int_{-\pi}^{\pi} e^{-(\chi-\varphi)^2/\Delta^2} \sin(n\chi)\,\mathrm{d}\chi = \frac{1}{\pi}\sin(\mathrm{n}\varphi)e^{-(n\Delta/2)^2} \tag{29}$$

where $n = 1, ..., \infty$ (the integrals are estimated under the condition that $\Delta \ll 2\pi$). Let us assume that the probability distribution is a sum of three Gaussians centered at the positions φ_1, φ_2, and φ_3, having the probabilities p_1, p_2, and p_3 (where $\sum_{i=1}^{3} p_i = 1$ and $p_i \geq 0$), and having the same width $\Delta \ll 2\pi$. The condition of nonnegativity is automatically satisfied, but care must be taken to fulfill the condition of periodicity (replicas of the Gaussians at $\varphi_i \pm 2\pi$ have to be used, where $i = 1, 2, 3$). Expressions for the Fourier coefficients of the distribution involving these parameters are

$$\rho_n = \frac{1}{\pi}e^{-(n\Delta/2)^2}\sum_{i=1}^{3} p_i \cos(n\varphi_i) \tag{30}$$

$$\sigma_n = \frac{1}{\pi}e^{-(n\Delta/2)^2}\sum_{i=1}^{3} p_i \sin(n\varphi_i) \qquad \text{where } n = 1, ..., \infty \tag{31}$$

If the first six Fourier coefficients (ρ_1, σ_1, ρ_2, σ_2, ρ_3, and σ_3) of the probability distribution are known from experimental spin–spin couplings and cross-relaxation data, the six independent Gaussian parameters (φ_1, φ_2, φ_3, p_1, p_2, and Δ) can be found by solving the set of six nonlinear equations of the form of eqs 30 and 31. This can be done by using iterative numerical procedures. Initial values of the Gaussian parameters for these iterations can be taken from the truncated trigonometric series. Since Gaussians contain an infinite number of higher-order Fourier components, this procedure is equivalent to prediction of the higher-order Fourier coefficients, which cannot be found directly from experimental data. The predicted coefficients are model-dependent, and they are valid only under the assumptions that peaks in the probability distribution have Gaussian shapes and that their widths are similar.

2.2.2. Estimation of Higher-Order Fourier Coefficients and Nonlinear Fitting. Different approaches can be used to satisfy the nonnegativity condition when the available experimental data do not provide reliable values for ρ_3 and σ_3, for example, when only coupling constants (no NOEs) are measured or when NOEs contain large experimental error. We have investigated two ways of estimating the approximate range of third-order terms (and sometimes also fourth-order terms) in the Fourier expansion (eq 1); both exploit the condition that $\rho(\chi)$ cannot be negative for the whole range of values of χ. It is assumed that the addition of these higher- (third- and possibly fourth-) order terms only forces the curve to be positive or zero everywhere without changing the general features of the curve, and that still higher-order terms (ρ_{N+m} and σ_{N+m}, $m \geq 2$ or $m > 2$) practically vanish.

The first, and most straightforward, way to estimate ρ_{N+1} and σ_{N+1} is as follows. If the Fourier coefficients ρ_n and σ_n ($n = 1, ..., N$) are known, the relation

$$\frac{1}{2\pi} + \sum_{n=1}^{N} \rho_n \cos (n\chi) + \sum_{n=1}^{N} \sigma_n \sin (n\chi) + \rho_{N+1} \cos [(N+1)\chi] + \sigma_{N+1} \sin [(N+1)\chi] \geq 0, \forall \chi \quad (32)$$

defines an area in the $\rho_{N+1} - \sigma_{N+1}$ plane which corresponds to a positive sum of first $N + 1$ terms of the Fourier series of the probability distribution over the whole range of χ values (if this area exists). The set of straight lines that are tangent to the border of this area is given by the expression obtained by converting eq 32 into an equality; these correspond to a set of values of χ ranging from $-\pi$ to π. A systematic search in the four-dimensional cube in $\rho_{N+1} - \sigma_{N+1} - \rho_{N+2} - \sigma_{N+2}$ space is an extension of this approach.

Alternatively, the probability distribution $\rho(\chi)$ can be reconstructed by means of a nonlinear least-squares fit of the measured coupling constants and NOEs to equations having the form of eqs 8 and 12 with the condition $\rho(\chi) \geq 0$ incorporated into the fitting procedure.

The condition of nonnegativity in both cases is treated as a window which filters possible values for the selected higher order terms. We have found from simulated data that this approach often (but not always) improves the fit of the reproduced distribution function to the original function. Thus the use of nonnegativity in this way should be taken with caution.

3. Discussion

CUPID circumvents the major problems (see Introduction) of the discrete model. Most notably, no assumptions of the values of the dihedral angle are made; instead, the populations and angles are generated by the Fourier coefficients of the fitted distribution function. In addition, by integrating the product of the coupling constant and probability over all angles, CUPID avoids unreasonable approximations introduced by the discrete model's assumption of δ functions.

Although only ρ_1, σ_1, ρ_2, and σ_2 can be found by using coupling constants alone (eq 8), higher-order terms exist in the expansion of ρ (eq 4). For example, $\rho(\chi)$ for a 3-fold potential barrier contains the term $\rho_3 \cos (3\chi) + \sigma_3 \sin (3\chi)$, but this term, multiplied by $J(\theta)$ and integrated from $-\pi$ to π, makes a zero contribution to the average coupling. This means that the Karplus equation "selects" the components of the distribution function that have the same 2-fold symmetry as eq 5. The integral in eq 7 is a scalar product in the space of periodic, integrable functions in the interval $(-\pi, \pi)$, and one can detect only the projection of the function $\rho(\chi)$ into a subspace of these functions that has the same symmetry as $J(\theta)$. Components with different symmetry (3-fold, and higher) belong to orthogonal subspaces and cannot be seen from couplings since they average to zero. In the case $\rho(\chi) = \rho_0 + \rho_3 \cos (3\chi)$, for instance, the averaged J coupling constants would be the same as if the distribution were uniform $[\rho(\chi) = \rho_0]$. This distribution only holds for a rotor with 3-fold symmetry (for example, a methyl group). Any substituent other than a proton (in the case of a methyl group) will lower the rotor symmetry and introduce terms $\rho_1 \cos (\chi)$, $\sigma_1 \sin (\chi)$, $\rho_2 \cos (2\chi)$, and $\sigma_2 \sin (2\chi)$. NOEs have no truncation and are able to provide higher-order terms of $\rho(\chi)$ as well as improve the accuracy of the first- and second-order terms. The limitation associated with NOEs is signal-to-noise. We have found (accompanying article)[28] that third-order terms are often measurable but that fourth- and higher-order terms are generally difficult to obtain.

In addition to the truncation errors discussed above, the reliability of CUPID is dependent upon the accuracy of the coefficients in the underlying Karplus relationships. Further work is needed to refine these parameters,[34,35] especially for conformationally flexible systems.

CUPID provides equations for both $\langle r^{-6} \rangle$ and $\langle r^{-3} \rangle^2$ averaging of NOEs. Before choosing the more appropriate set of equations, one must determine what type of averaging of NOEs actually occurs. Therefore, it is necessary to know the correlation times for overall tumbling and for internal rotation. NOEs between atoms that do not participate in internal motion (fixed in the molecular frame, like H^{α} and H^{N}) depend only on the correlation time for overall tumbling. On the other hand, NOEs between β-methylene protons depend on both correlation times (overall tumbling and internal rotation). By combining measured intensities of NOEs with the corresponding interproton distances, which are fixed and known, it should be possible to estimate the correlation times of the internal rotation. Thus, in principle, one should be able to decide, solely on the basis of experimental data, whether $\langle r^{-3} \rangle^2$ or $\langle r^{-6} \rangle$ averaging is appropriate for a given residue.

The final step of CUPID, e.g., fitting of Gaussians, is similar to the method proposed by Schirmer et al.[25] Still, there are important differences between CUPID and Schirmer's method. First, CUPID is designed to utilize both spin–spin couplings and NOEs, while Schirmer's approach was applied only to the analysis of cross-relaxation data. Second, in CUPID, the integration representing the averaging is performed analytically, so that linear regression can be used, whereas Schirmer's method employed numerical integration. Finally, the initial guess for the fitted Gaussian parameters in CUPID is taken directly from the trigonometric series; Schirmer's approach provided no systematic way to estimate the initial values of the fitted parameters.

4. Conclusion

Nuclear magnetic resonance spin–spin coupling constants and NOEs have been known for many years to contain a wealth of structural information.[22] However, the utility of these NMR parameters has fallen far short of their potential because of the problem of conformational averaging. Many models based upon interconverting discrete structures have been proposed. However, the discrete models have the serious flaw of requiring a priori knowledge of the conformational angles. Additionally, models have been proposed that allow for any possible dihedral angle but do not allow for rotational averaging.[5,14,18-20] CUPID accounts for averaging without requiring preconceptions of the possible conformations. The only necessary "model" is the dihedral angular dependence of each coupling constant and the coordinates of the fixed atoms. The CUPID approach makes use of all relevant coupling and NOE data to provide an optimal distribution function for rotamer populations.

Acknowledgment. This work was supported by NIH Grants LM04958, GM35976, and RR02301 and by NEDO (Tokyo). Ž.Dž. is also supported by the Serbian Research Fund and by the Serbian Academy of Sciences. A.S.E. is a trainee of an NIH Molecular Biophysics Training Grant (GM08293). The authors thank Professors Slobodan Macura, Frank Weinhold, and Andrzej M. Krezel for helpful discussions.

---

(34) Pachler, K. G. R. *J. Chem. Soc., Perkin Trans.* **1972**, *2*, 1936–1940.

(35) Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783–2792.