Conformational Analysis of Molecules with Five-Membered Rings through NMR Determination of the Continuous Probability Distribution (CUPID) for Pseudorotation

Željko Džakula,[†] Michele L. DeRider,[‡] and John L. Markley*,[†]

Contribution from the National Magnetic Resonance Facility at Madison, Department of Biochemistry, and Department of Chemistry, University of Wisconsin−Madison, 420 Henry Mall, Madison, Wisconsin 53706

Received June 26, 1996[⊗]

Abstract: The continuous probability distribution (CUPID) method [Džakula, Ž.; Westler, W. M.; Edison, A. S.; Markley, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 6195] for conformational analysis of molecules from nuclear magnetic resonance (NMR) data has been extended to the determination of molecular conformations of five-membered rings. This approach, which should be particularly useful for studies of molecules containing pyrrolidine or furanose rings, is illustrated by the analysis of NMR data from the literature for small peptides containing L- and D-prolines, hydroxyprolines, and fluoroprolines. The CUPID approach to the analysis of five-membered rings, which takes advantage of linear regression, generally yields better fits to experimental data in a shorter time than the conventional discrete approach, which utilizes nonlinear fitting procedures. The new method proved successful in a few cases in which the conventional approach failed to produce satisfactory analysis of the data. Built-in error-propagation analysis in the CUPID software provides a direct assessment of the reliability of the calculated results.

1. Introduction

There has been considerable interest in determining the conformational states of biologically important five-membered rings such as those of proline residues in peptides and proteins and of ribose rings in nucleic acids. For example, the conformations of the five-membered pyrrolidine rings of proline and hydroxyproline are thought to control the stability and physiological function of collagen fibrils.[1] Similarly, changes in the conformations of the constituent ribose and deoxyribose rings are associated with major structural alterations in nucleic acids,[2] which in turn can have an impact on their biological functions.[3−5]

NMR spectroscopy provides rich information on the conformations of five-membered ring systems (for example, various cross-relaxation rates and multiple vicinal spin−spin couplings across most of the torsion angles of the five-membered ring). An analysis based on NMR data thus has the potential of specifying the (multiple-state) conformational probability in considerable detail. The analysis is greatly simplified by the fact that a single conformation of a five-membered ring can be approximated by specifying only two independent geometric descriptors: the pucker amplitude χ_{max} and the pseudorotation angle P.[2,6] The conventional approach used in analyzing ring pucker from NMR data normally assumes a conformational equilibrium between two discrete states, each characterized by its own pseudorotation value and pucker amplitude.[7,8] In the case of proline and its derivatives, nonlinear fitting procedures typically have been utilized to determine five parameters (two pseudorotation angles, two pucker amplitudes, and an equilibrium constant) from 6−10 experimentally measured NMR proton−proton couplings.[7,8] Despite its noteworthy successes,[7,8] this approach has limitations: (i) the five fitted parameters often exhibit strong correlation, and their "best-fit values" are therefore unreliable; (ii) the ratio of the size of the input experimental data set to the number of fitted parameters is sometimes unfavorable (*e.g.*, 6:5 in substituted Pro rings); and (iii) in some cases, the nonlinear fitting procedures are unable to identify best-fit requirements within reasonable time limits.

With the aim of overcoming these problems, we have investigated ways of adapting the continuous angular probability distribution (CUPID) method to the specialized case of five-membered rings. In earlier implementations of the CUPID approach,[9−11] no assumptions were made concerning the values of torsion angles or their interdependence. A version of the program (CUPID-5) has been developed that incorporates the general principles of pseudorotation in five-membered rings while maintaining the basic approach of using linear regression to obtain the best fit to the Fourier coefficients of the probability distribution. Further simplification is afforded by the fact that the probabilities of the pseudorotations are much more interesting than the distributions of the amplitudes, χ_{max}, which are usually unimodal, with maxima at about 40° and widths less than 10°.[2,6] Therefore, CUPID-5 was developed for optimal analysis of pseudorotation angles.

Two approaches were evaluated. In one, it was assumed that χ_{max} is fixed at its most likely value; the resulting one-dimensional distribution of pseudorotations, $\rho(P)$, is a *section*

[†] Department of Biochemistry.
[‡] Department of Chemistry.
[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1996.

(1) Stryer, L. *Biochemistry*; Freeman: New York, 1988.
(2) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205.
(3) Langridge, R.; Marvin, D. A.; Seeds, W. E.; Wilson, H. R.; Harper, C. W.; Wilkins, M. H. F.; Hamilton, L. D. *J. Mol. Biol.* **1960**, *2*, 38.
(4) Arnott, S.; Wilkins, M. H. F.; Fuller, W.; Langridge, L. D. *J. Mol. Biol.* **1967**, *27*, 535.
(5) Arnott, S. *Progr. Biophys. Mol. Biol.* **1970**, *21*, 265.
(6) DeTar, D. F.; Luthra, N. P. *J. Am. Chem. Soc.* **1977**, *99*, 1232.
(7) 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.
(8) Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; de Leeuw, H. P. M.; Altona, C. *Biopolymers* **1981**, *20*, 1211.
(9) Džakula, Ž.; Westler, W. M.; Edison, A. S.; Markley, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 6195.
(10) Džakula, Ž.; Edison, A. S.; Westler, W. M.; Markley, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 6200.
(11) Džakula, Ž.; Westler, W. M.; Markley, J. L. *J. Magn. Reson.* **1996**, *111B*, 109.

S0002-7863(96)02160-9 CCC: $12.00

of the two-dimensional probability distribution $\rho(\chi_{max},P)$.[12] In the other, a *projection* of $\rho(\chi_{max},P)$, *i.e.*, its integral over all possible values of χ_{max}, was evaluated.[12] Both approaches were applied to the analysis of conformations of a series of molecules containing five-membered rings. The results show that the CUPID-5 approach[13] overcomes many of the deficiencies (noted above) of the previous methods used to analyze these data.

2. Theory

This paper presents a method for determining the populations of five-membered rings from experimental NMR *J*-couplings, which are known to depend on torsion angles according to the well-known Karplus equations:[14]

$$J_j(\chi) = a_j \cos^2(\chi + \Omega_j) + b_j \cos(\chi + \Omega_j) + c_j \qquad (1)$$

where the index j enumerates the experimentally measured couplings, a_j, b_j, and c_j are the corresponding Karplus parameters, and Ω_j is the stereochemical factor defined by the geometry of the rotating segment. Equation 1 can be used interchangeably for homonuclear and heteronuclear couplings, but to achieve a more appropriate description of the homonuclear proton−proton scalar couplings, the extended Karplus equation[15] must be used:

$$J_j(\chi) = P_1 \cos^2(\chi + \Omega_j) + P_2 \cos(\chi + \Omega_j) + P_3 + \sum_i \Delta E_{ij}\{P_4 + P_5 \cos^2[(\chi + \Omega_j)sign_{ij} + P_6|\Delta E_{ij}|]\} \qquad (2)$$

where $sign_{ij} = \pm 1$, depending on the stereochemistry of the rotating segment, and parameters P_i are available in the literature.[15] The term ΔE_{ij} reflects the electronegativities of α-substituents (E_{ij}) and β-substituents (E_{kj})

$$\Delta E_{ij} = E_{ij} - E_H - P_7 \sum_k (E_{kj} - E_H) \qquad (3)$$

where E_H is the electronegativity of hydrogen. The extended Karplus equation (eq 2) can be rewritten so that it resembles eq 1:

$$J_j(\chi_{max},P) = a_j \cos^2(\chi + \Omega_{2j}) + b_j \cos(\chi + \Omega_{1j}) + c_j \qquad (4)$$

where

$$\Omega_{1j} = \Omega_j \qquad \Omega_{2j} = \arctan(\beta_j/\alpha_j)$$

$$a_j = 2\alpha_j/\cos(2\Omega_{2j}) \qquad b_j = P_2 \qquad c_j = \gamma_j - \alpha_j/2 \qquad (5)$$

and the parameters α, β, and γ are defined as follows

$$\alpha_j = \frac{P_1}{2}\cos(2\Omega_j) + \frac{P_5}{2}\sum_i \Delta E_{ij} \cos(2P_6|\Delta E_{ij}|) \times \cos(2\Omega_j) - \frac{P_5}{2}\sum_i sign_{ij}\Delta E_{ij} \sin(2P_6|\Delta E_{ij}|) \sin(2\Omega_j)$$

$$\gamma_j = P_1/2 + P_3 + (P_4 + P_5/2)\sum_i \Delta E_{ij} \qquad (6)$$

$$\beta_j = \frac{P_1}{2}\sin(2\Omega_j) + \frac{P_5}{2}\sum_i \Delta E_{ij} \cos(2P_6|\Delta E_{ij}|) \sin(2\Omega_j) + \frac{P_5}{2}\sum_i sign_{ij}\Delta E_{ij} \sin(2P_6|\Delta E_{ij}|) \cos(2\Omega_j)$$

The torsion angles χ_m in five-membered rings (m = 0, ..., 4) are dependent on one another and can be expressed as functions of two independent variables, according to the following expression:[2]

$$\chi_m = \chi_{max} \cos[P + 4\pi(m - M)/5] = \chi_{max} \cos(P + \phi_m) \qquad (7)$$

where the phase $\phi_m = 4\pi(m - M)/5$ is a constant for a given torsion angle, and the integer constant M reflects the naming convention for the torsion angles in the ring. The Karplus equations (eq 1 or 2) can be combined with eq 7 to show explicitly the dependence of couplings on the parameters χ_{max} and P

$$J_j(\chi_m) = a_j \cos^2[\chi_{max} \cos(P + \phi_m) + \Omega_{2j}] + b_j \cos[\chi_{max} \cos(P + \phi_m) + \Omega_{1j}] + c_j \qquad (8)$$

Two different indexes, m and j, are used because several different couplings j can be measured across the same torsion angle m. To extract $\rho(P)$ from the experimentally measured couplings by applying the usual CUPID procedures,[9−11] eq 8 has to be transformed into

$$J_j(P) = C_{0j} + \sum_{n=1}^{N_j}[C_{nj} \cos(nP) + S_{nj} \sin(nP)] \qquad (9)$$

where N_j is the order of truncation of the Fourier expansion. The analytical expression for the coefficients $\{C_{nj},S_{nj}\}$ of the Fourier series in eq 9 are obtained on using the following expressions:

$$\cos[\chi_{max} \cos(P + \phi_m) + \Omega_j] = \frac{1}{2}\{\exp(i\Omega_j)\exp(i\chi_{max} \cos\phi_m \cos P)\exp(-i\chi_{max} \sin\phi_m \sin P) + \exp(-i\Omega_j)\exp(-i\chi_{max} \cos\phi_m \cos P)\exp(i\chi_{max} \sin\phi_m \sin P)\} \qquad (10)$$

$$\exp[A \cos(nP)] = J_0(iA) - 2iJ_1(iA) \cos(nP) - 2J_2(iA) \cos(2nP) + 2iJ_3(iA) \cos(3nP) + 2J_4(iA) \cos(4nP) + ... \qquad (11)$$

$$\exp[A \sin(nP)] = J_0(iA) - 2iJ_1(iA) \sin(nP) + 2J_2(iA) \cos(2nP) - 2iJ_3(iA) \sin(3nP) + 2J_4(iA) \cos(4nP) + ... \qquad (12)$$

where n = 1, 2, ..., and $J_m(x)$ are the Bessel functions. Straightforward rearrangements lead to the following expressions, which are equally applicable to heteronuclear couplings (eq 1, standard Karplus equation with $\Omega_{2j} = \Omega_{1j} = \Omega_j$) and to proton−proton homonuclear couplings (eq 4):

$$C_{0j} = c_j + b_j \cos(\Omega_{1j})[J_0(\chi_{max} \cos\phi_m)J_0(\chi_{max} \sin\phi_m) - 2J_2(\chi_{max} \cos\phi_m)J_2(\chi_{max} \sin\phi_m)] + (a_j/2)\{1 + \cos(2\Omega_{2j})[J_0(2\chi_{max} \cos\phi_m)J_0(2\chi_{max} \sin\phi_m) - 2J_2(2\chi_{max} \cos\phi_m)J_2(2\chi_{max} \sin\phi_m)]\} + ... \qquad (13)$$

$$C_{1j} = -2b_j \sin(\Omega_{1j})J_1(\chi_{max} \cos\phi_m)[J_0(\chi_{max} \sin\phi_m) + J_2(\chi_{max} \sin\phi_m)] - a_j \sin(2\Omega_{2j})[J_0(2\chi_{max} \sin\phi_m) \times J_1(2\chi_{max} \cos\phi_m) + J_1(2\chi_{max} \cos\phi_m)J_2(2\chi_{max} \sin\phi_m) - J_2(2\chi_{max} \sin\phi_m)J_3(2\chi_{max} \cos\phi_m)] + ... \qquad (14)$$

$$S_{1j} = 2b_j \sin(\Omega_{1j})J_1(\chi_{max} \sin\phi_m)[J_0(\chi_{max} \cos\phi_m) + J_2(\chi_{max} \cos\phi_m)] + a_j \sin(2\Omega_{2j})[J_0(2\chi_{max} \cos\phi_m) \times J_1(2\chi_{max} \sin\phi_m) + J_1(2\chi_{max} \sin\phi_m)J_2(2\chi_{max} \cos\phi_m) - J_2(2\chi_{max} \cos\phi_m)J_3(2\chi_{max} \sin\phi_m)] + ... \qquad (15)$$

(12) Taupin, D. *Probabilities, Data Reduction, and Error Analysis in the Physical Sciences*; Les Editions de Physique: Les Ulis Cedex, France 1988.

(13) The CUPID software package including CUPID-5 is available from the authors; information may be obtained from the World Wide Web (http://www.nmrfam.wisc.edu).

(14) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11.

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

(16) Huggins, M. L. *J. Am. Chem. Soc.* **1953**, *75*, 4123.

Table 1. Karplus Parameters, Ω, and ϕ-Phases in L-Pro, Hydroxyproline (Hyp), and Fluoroproline (F-Pro)[a,b]

	$H^{\alpha}-H^{\beta 2}$	$H^{\alpha}-H^{\beta 3}$	$H^{\beta 2}-H^{\gamma 2}$	$H^{\beta 2}-H^{\gamma 3}$	$H^{\beta 3}-H^{\gamma 2}$	$H^{\beta 3}-H^{\gamma 3}$	$H^{\gamma 2}-H^{\delta 2}$	$H^{\gamma 2}-H^{\delta 3}$	$H^{\gamma 3}-H^{\delta 2}$	$H^{\gamma 3}-H^{\delta 3}$
a (Hz) in Pro	9.66	9.61	12.46	12.47	12.47	12.46	11.29	11.31	11.31	11.29
b (Hz) in Pro	−0.99	−0.99	−0.73	−0.73	−0.73	−0.73	−0.73	−0.73	−0.73	−0.73
c (Hz) in Pro	1.19	1.21	0.28	0.28	0.28	0.28	0.48	0.47	0.47	0.48
Ω_1 (deg) in Pro	−122.9	−1.7	−0.6	121.5	−121.7	0.5	0.7	122.2	−121.5	0.1
Ω_2 (deg) in Pro	−126.4	−3.6	−0.7	121.9	−122.1	0.6	−1.3	124.7	−124.0	2.1
a (Hz) in Hyp	9.66	9.61	9.68		9.55		9.21	8.62		
b (Hz) in Hyp	−0.99	−0.99	−0.99		−0.99		−0.99	−0.99		
c (Hz) in Hyp	1.19	1.21	1.01		1.08		1.09	1.38		
Ω_1 (deg) in Hyp	−122.9	−1.7	−1		−122.1		1.1	122.7		
Ω_2 (deg) in Hyp	−126.4	−3.6	6.5		−116.2		−9.8	119.1		
a (Hz) in F-Pro	9.66	9.61	10.48		10.29		10.1	9.27		
b (Hz) in F-Pro	−0.99	−0.99	−0.99		−0.99		−0.99	−0.99		
c (Hz) in F-Pro	1.19	1.21	0.47		0.56		0.48	0.91		
Ω_1 (deg) in F-Pro	−122.9	−1.7	−1.0		−122.1		1.1	122.7		
Ω_2 (deg) in F-Pro	−126.4	−3.6	9.8		−112.7		−12.9	115.1		
ϕ (deg)	−144	−144	0	0	0	0	144	144	144	144

[a] In D-Pro, the parameters *a*, *b*, *c*, and ϕ are the same as those in L-Pro, whereas the phase factors Ω_1 and Ω_2 have the same magnitudes and opposite sign when the 180° shift of the pseudorotation is not applied. Equations 2−6 were used with the following P_i parameter values for all the couplings in substituted prolines and for the $H^{\alpha}-H^{\beta}$ couplings in unsubstituted prolines:[8] $P_1 = 13.22$ Hz, $P_2 = -0.99$ Hz, $P_3 = 0$ Hz, $P_4 = 0.87$ Hz, $P_5 = -2.46$ Hz, $P_6 = 19.9°$, and $P_7 = 0$. The parameter values used for all other couplings in unsubstituted proline rings were[8] $P_1 = 13.70$ Hz, $P_2 = -0.73$ Hz, $P_3 = 0$ Hz, $P_4 = 0.56$ Hz, $P_5 = -2.47$ Hz, $P_6 = 16.9°$, and $P_7 = 0.14$. The values for the electronegativities were[19] $E(H) = 2.20$, $E(C) = 2.60$, $E(O) = 3.50$, $E(N) = 3.05$, and $E(F) = 3.90$.

$$C_{2j} = 2b_j \cos(\Omega_{1j})[J_0(\chi_{max}\cos\phi_m)J_2(\chi_{max}\sin\phi_m) - J_0(\chi_{max}\sin\phi_m)J_2(\chi_{max}\cos\phi_m)] + a_j \cos(2\Omega_{2j})[J_0(2\chi_{max}\cos\phi_m)J_2(2\chi_{max}\sin\phi_m) - J_0(2\chi_{max}\sin\phi_m)J_2(2\chi_{max}\cos\phi_m)] + \ldots \quad (16)$$

$$S_{2j} = 2b_j \cos(\Omega_{1j})[J_1(\chi_{max}\cos\phi_m)J_1(\chi_{max}\sin\phi_m) + a_j \cos(2\Omega_{2j})[J_1(2\chi_{max}\sin\phi_m)J_1(2\chi_{max}\cos\phi_m)] + J_1(2\chi_{max}\sin\phi_m)J_3(2\chi_{max}\cos\phi_m) + J_1(2\chi_{max}\cos\phi_m)J_3(2\chi_{max}\sin\phi_m)] + \ldots \quad (17)$$

These are the Fourier coefficients of spin−spin couplings under the assumption that the amplitude χ_{max} is fixed. A more realistic approach averages the Karplus equation over all possible values of χ_{max}. Its neglect of the correlation between χ_{max} and *P* can be justified on the basis of results from a statistical analysis of 60 X-ray structures of five-membered rings in sugars.[2] Upon assuming a Gaussian distribution of χ_{max} values of width *D* centered at χ^0_{max}, the Karplus equation assumes the following form:

$$\begin{aligned} J(P) &= \int_0^{\infty} J(P,\chi_{max}) \frac{\exp[-(\chi_{max}-\chi^0_{max})^2/(2D^2)]}{D\sqrt{2\pi}}\, d\chi_{max} \\ &= \int_{-\infty}^{\infty} \{a_j \cos^2[\chi_{max}\cos(P+\phi_m)+\Omega_j] + b_j \cos[\chi_{max}\times \\ &\quad \cos(P+\phi_m)+\Omega_j] + c_j\} \frac{\exp[-(\chi_{max}-\chi^0_{max})^2/(2D^2)]}{D\sqrt{2\pi}}\, d\chi_{max} \\ &= \frac{a_j}{2}\{1+\exp[-2D^2\cos^2(P+\phi_m)]\cos[2\chi^0_{max}\cos(P+\phi_m)+ \\ &\quad 2\Omega_j]\} + b_j \exp\left[-\frac{D^2}{2}\cos^2(P+\phi_m)\right] \\ &\quad \cos[\chi^0_{max}\cos(P+\phi_m)+\Omega_j] + c_j \end{aligned} \quad (18)$$

In the limiting case of a single value for χ_{max} (*D* infinitely sharp), eq 18 for the projection of Karplus equation reduces to eq 8 for the section at $\chi_{max} = \chi^0_{max}$. Fourier coefficients of the "projected" Karplus eq 18 are obtained straightforwardly using eqs 10−12.

CUPID-5 combines eqs 13−18 with the CUPID linear regression procedure[9,11] to yield a truncated Fourier expansion of the pseudorotational probability distribution when applied to experimentally measured, pseudorotationally averaged scalar couplings from a five-membered ring. The actual distribution can be obtained from the truncated Fourier series by two methods. Preferably, it should be treated as a Boltzmann distribution whose exponent (the pseudorotational potential) is derived from the truncated Fourier expansion of the probability distribution.[11] Alternatively, it can be approximated by a superposition of Gaussian probability peaks centered at different pseudorotation values (not to be confused with the Gaussian distribution of χ_{max} from eq 18) and having the same Fourier coefficients as those obtained from experimental measurements.[10]

3. Illustrative Examples

We applied the expressions derived above to 42 experimental proton−proton coupling data sets from molecules containing L-prolines, D-prolines, 4-hydroxy-L-prolines, and 4-fluoro-L-prolines. The data (Supporting Information) were taken from refs 7 and 8. Karplus parameters (Table 1) were calculated from eqs 2−6. The phase factors Ω_{1j}, Ω_{2j}, and ϕ_m for all proton pairs in the proline ring are listed in Table 1. The rms values for couplings improved significantly when Ω_{1j} and Ω_{2j} were corrected for deviations from ideal tetrahedral geometry in sp^3 carbons.[8] The factors ϕ_m were determined from eq 7 with $M = 2$.[7,8]

Figure 1 shows representative probability distributions of pseudorotations in six proline rings calculated by the projection method. The results for an additional 36 data sets are detailed in the Supporting Information. The most likely values for the pucker amplitude, $\chi^0_{max} = 40°$, and the width of its Gaussian distribution, $D = 15°$, are based on force field calculations[6] which indicate that the strain energy of the proline ring increases sharply when χ_{max} is outside the interval (25°,50°). This range may be even narrower (35−45°) in sugars.[2] The probability distributions in Figure 1 were reconstructed from the truncated Fourier series using the pseudorotational potential and the Boltzmann relationship.[11] The same experimental data, Karplus parameters, and phase factors (Table 1S, Supporting Information) were used to obtain the truncated Fourier series of the probability distributions for pseudorotations with the section method (eqs 13−17) on varying χ_{max} from 30 to 55°. Both the projection and the section approaches result in similar probability distributions, regardless of whether the Gaussians or the pseudorotational potential were used (Figure 2; Table 2S, Supporting Information). Furthermore, the Gaussian probability peak positions (Figure 2A) and areas (Figure 2B) obtained by CUPID-5 in most cases agreed with those determined from energy calculations[7] or from the discrete model[7,8] when the same

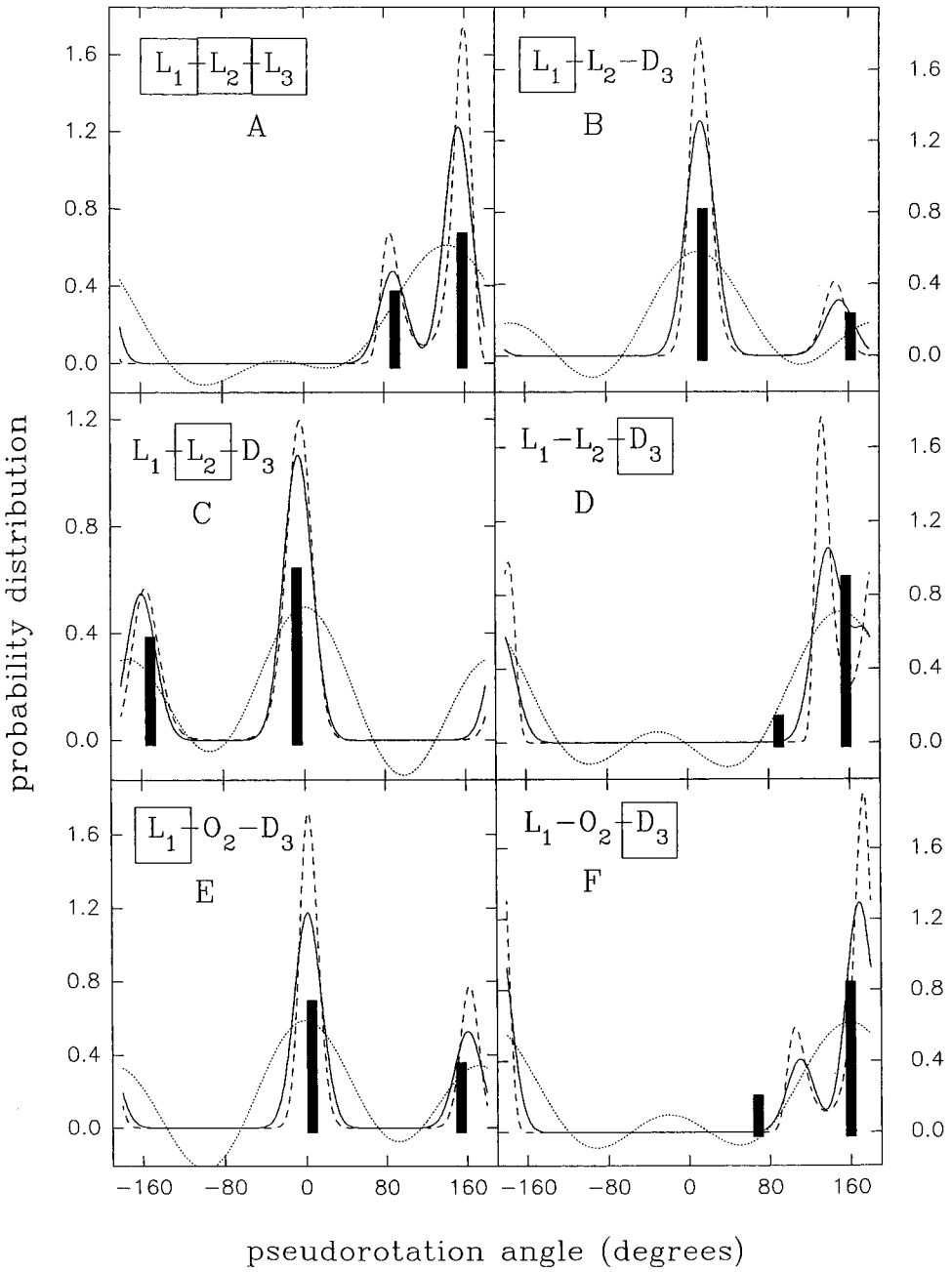

Figure 1. Probability distributions of pseudorotations in (A) L-Pro of *cyclo*(L-Pro$_1$-L-Pro$_2$-L-Pro$_3$) ($L_1L_2L_3$), (B) L-Pro$_1$ of *cyclo*(L-Pro$_1$-L-Pro$_2$-D-Pro$_3$) ($L_1L_2D_3$), (C) L-Pro$_2$ of $L_1L_2D_3$, (D) D-Pro$_3$ of $L_1L_2D_3$, (E) L-Pro$_1$ of *cyclo*(L-Pro$_1$-BzlPro$_2$-D-Pro$_3$) ($L_1O_2D_3$), and (F) D-Pro$_3$ of $L_1O_2D_3$ obtained with the projection method. The Gaussian distribution for χ_{max} was centered at 40°, with standard deviation equal to 15°. Truncated Fourier series (dotted lines), Gaussian probability peaks (solid lines), and Boltzmann factors (dashes) were computed according to the procedures described in refs 9−11. Thick vertical bars represent pseudorotation values and probabilities reported by ref 7. Pseudorotations in D-Pro residues were shifted by 180° to enable comparison with the conformations of L-Pro rings.

experimental input data and Karplus parameters were used. The coupling data (Figure 3), however, were modeled significantly better by CUPID-5 than by the conventional, discrete model[7,8] (Table 2).

Unlike analysis by the discrete model, CUPID-5 provides error ranges for the pseudorotation angle values (Figure 2A) and populations (Figure 2B). In most cases, the positions of small peaks (probability less than 0.25, *cf.* Supporting Information) are extremely imprecise; they are indicated by open symbols in Figure 2A, and their computed error ranges are not shown. The pseudorotation distributions in *cyclo*(Pro-Aib), *cyclo*(Pro-D-Phe), *cis*-NAcPro-NH_2, and all substituted proline rings except *cyclo*(Pro-Pro-Hyp), *cis*-NAcHyp, and *cis*-Gly-Hyp are represented adequately by a single Gaussian. Large error ranges for probability peak positions and areas also were obtained whenever the calculated pseudorotation values were less than $\pi/2$ away from one another, because the resolution of the method is limited by the truncation of the Fourier series for the probability distribution.[11] This situation exists in *cyclo*(Pro-Hyp), *cyclo*(D-Ala-Hyp), and D-Pro$_3$ in *cyclo*(Pro-BzlGly-D-Pro). Whereas application of the conventional method[8] to data from these molecules left the authors of ref 8 unaware of the problem, large error ranges were reported in the CUPID-5 analysis. In the cases of *cyclo*(Pro-D-Pro), *cyclo*(Pro-Pro-Pro), D-Pro$_3$ in *cyclo*(Pro-Pro-D-Pro), and Hyp in *cyclo*(Pro-Pro-Hyp), on the other hand, reduction of the distribution from two to one Gaussian resulted in an unacceptable increase in the fitting penalty. Two Gaussians were therefore retained for these molecules in spite of the large error ranges for the calculated parameters (*cf.* Supporting Information).

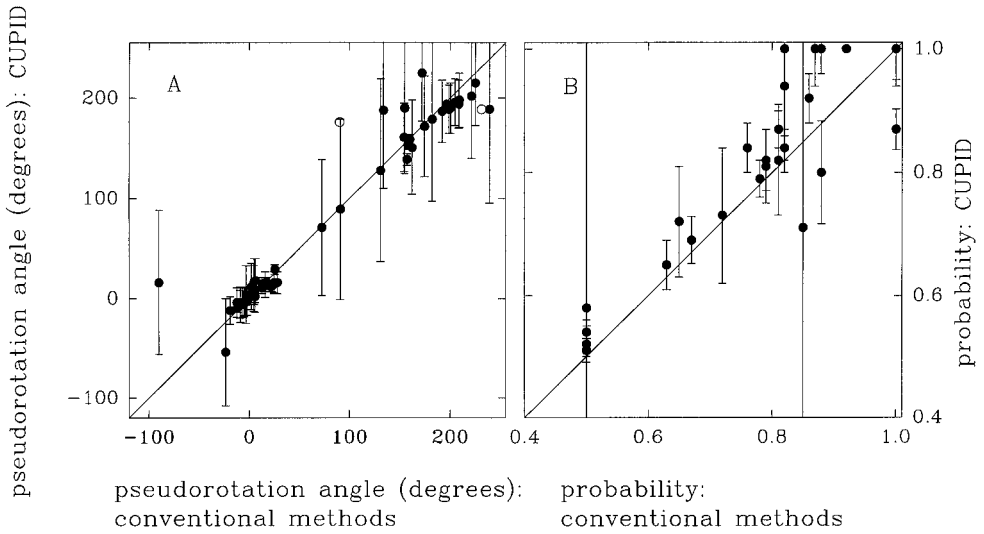

Figure 2. Pseudorotations and their populations in proline rings from 42 data sets. The molecules and their pseudorotation values are listed in the Supporting Information. (A) Correlation plot of pseudorotation angles derived with CUPID-5 and conventional[7,8] methods. (B) Correlation plot of conformational probabilities derived from CUPID-5 and conventional[7,8] methods. The same experimental data and Karplus parameters were used for both calculations. Error ranges are shown only for CUPID results; errors were not reported in the previous analyses.[7,8] Open circles in (A) indicate conformers with error ranges that approximate the full pseudorotation cycle (error bars not shown; *cf.* Supporting Information). Typically, these are conformers with small populations or multiple conformers whose χ_{max} values differ by less than $\pi/2$.

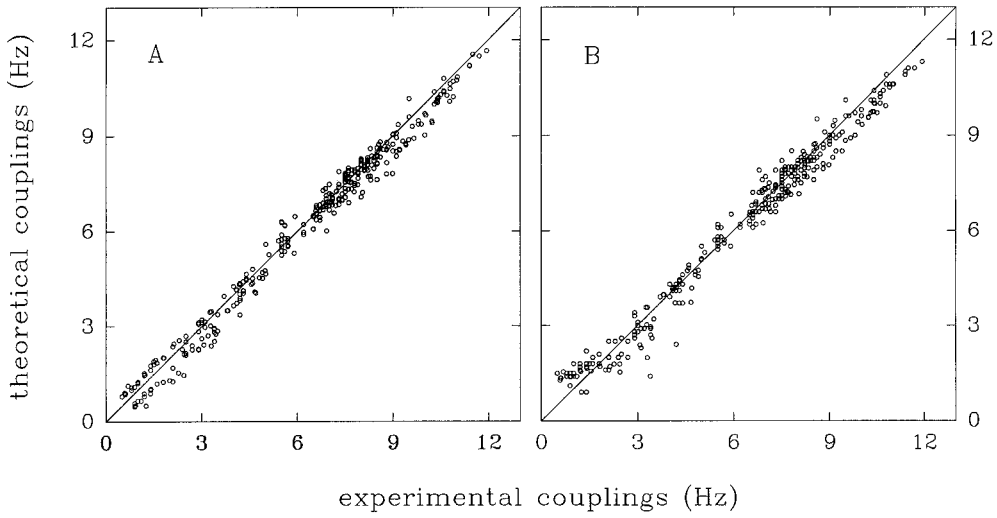

Figure 3. Correlation plots of experimental spin−spin couplings in 42 data sets with those calculated by the (A) CUPID-5 and (B) discrete[7,8] methods.

Table 2. Average Root-Mean-Square Values (Hz) for Proton−Proton Couplings Obtained with CUPID and with the Conventional, Discrete Method[a]

ring type	no. of data sets	CUPID (projection)	CUPID (section)	discrete model[7,8]
proline	23	0.39	0.40	0.42
hydroxyproline	15	0.25	0.26	0.32
fluoroproline	4	0.27	0.28	0.45

[a] The rms values for individual data sets are listed with other CUPID results in the Supporting Information. In the CUPID analysis, all couplings were considered. In the analysis of these data by the discrete method,[8] all couplings of *cyclo*(L-Pro-L-Hyp) and of *cyclo*(L-Pro-D-Pro) were considered, but only eight couplings were fitted.

Special attention has been paid to *cyclo*(L-Pro-L-Pro-L-Pro) (LLL) and its analog *cyclo*(L-Pro-L-Pro-Hyp) (LLH). CUPID successfully reproduced two independently measured data sets for LLL[7,17] and another data set for LLH.[8] All three calculations resulted in similar bimodal distributions ($P \approx 90°$ and 160°), in agreement with refs 7 (LLL) and 8 (LLH). Because of the unusual conformations calculated for LLL (close to one another and to the energy barrier), the authors of ref 7 doubted the feasibility of the two-state model in this case and interpreted the results in terms of a single shallow potential well with unusual broadness, centered at the weighted average of the two peaks. However, the consistent pattern emerging from three different data sets and the agreement between the pseudorotational potentials derived from CUPID and from force field calculations (Figure 4) suggest that the bimodal distribution represents a valid model for the system despite the large error ranges.

(17) Deber, C. M.; Torchia, D. A.; Blout, E. R. *J. Am. Chem. Soc.* **1971**, *93*, 4893.

4. Discussion

Good agreement was found between pseudorotations and populations estimated by CUPID-5 and by the discrete model (Figure 2) generally whenever the results of the discrete model were judged to be trustworthy.[7,8] However, it is important to note that CUPID-5 did not simply give the same results obtainable using the conventional method. The CUPID-5 approach consistently yielded a better fit to the input data (Table 2) and led to some additional improvements: (i) The experi-

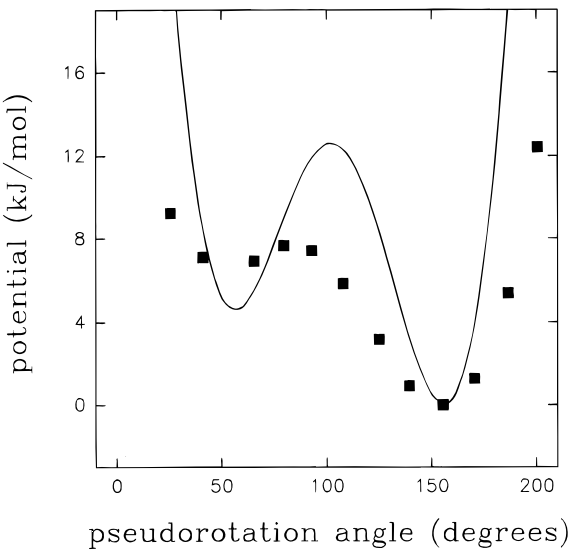

Figure 4. Comparison of the pseudorotational potential for *cyclo*(L-Pro$_1$-L-Pro$_2$-L-Pro$_3$) (LLL) calculated with CUPID-5 (solid line) with that derived from force field calculations (data points).[7] The CUPID-derived potential was obtained as described in ref 11 from the Fourier coefficients of the distribution derived from eq 18 ($\chi^0_{max} = 40°$, $D = 15°$).

mental couplings for LLL from ref 17 were reproduced successfully with a relatively small rms value (0.37 Hz) by CUPID-5, whereas the discrete model failed to obtain a meaningful result[8] from the same set of coupling data. The probability distribution resulting from CUPID-5 analysis of that data set was almost the same as that obtained by conventional or CUPID-5 analysis of a different set of spin−spin couplings, measured for the *same* molecule in *another* laboratory.[7] This shows that the discrete model can either be successful[7] or unsuccessful,[8] depending on minor variations in input data, whereas CUPID-5 is more robust. (ii) Difficulties encountered on fitting the couplings in Hyp$_3$ from LLH by the discrete model (rms 0.49 Hz) were attributed to contaminations of experimental data with errors.[8] However, the same data were analyzed by CUPID-5 with significantly smaller rms values (0.36 Hz). Remarkably, the distribution of pseudorotations in Hyp$_3$ obtained with CUPID-5 is similar to that based on the discrete model,[8] and both are similar to the distribution in the related peptide LLL, which was treated by the discrete approach with variable success[7,8] (*vide supra*). (iii) The amplitude χ_{max} obtained with the discrete model (Table IV in ref 8) for the *S* form of *trans*-NAc-Hyp (25−30°) suggests an extremely flattened ring. The flatness is even more emphasized by the predicted twist conformation,[8] ${}^{\beta}_{\alpha}T$. The amplitude for the *S* form (25°) is almost two times smaller than χ_{max} for the *N* form[8] (45−50°; Table IV in ref 8). These two values represent extremes within the energetically favorable range[6] for χ_{max} (25−50°). CUPID-5, however, reproduced the same experimental data with a better rms value (0.21 Hz vs. 0.26 Hz obtained with the discrete model[8]) without resorting to the strain from extreme flattening or puckering of the ring ($\chi_{max} = 35−40°$). (iv) Data for L-Pro$_1$ in *cyclo*(Pro-Hyp) and in *cyclo*(Pro-D-Pro) could be fitted by the discrete model only when the couplings from $H^{\gamma 2}$ to the protons on C^{δ} were omitted.[8] With CUPID-5, the complete data set for each molecule could be utilized (rms 0.47 and 0.54 Hz, respectively).

The key factor in the success of the CUPID approach is its use of linear regression, which ensures *the best possible* fit. The Fourier coefficients of the distribution are unique, linearly independent, effortlessly computed from the measured data, and easily transformed into either a sum of Gaussians or a Boltzmann exponential factor. Therefore, bypassing the determination of Fourier coefficients and employing a direct nonlinear fitting of the discrete pseudorotation angles and their populations to experimental data is inferior to the CUPID-5 procedure. In addition, the discrete model fits a larger number of parameters (five parameters) than CUPID-5 (four parameters, since χ_{max} is either fixed or sampled at a small number of discrete values). Even worse, parameters fitted by the discrete model can be highly correlated.[8] These differences are particularly weighty in substituted proline rings, which yield fewer input data. The combined effect of all these factors probably explains the significantly lower rms values found for substituted prolines analyzed by CUPID-5 than by the discrete model (Table 2). The same reasoning applies to direct fitting of Gaussians centered at the most likely pseudorotation angles. A grid search over the free parameters should suffice to construct the multidimensional probability, but the computational effort would be much greater than in the case of CUPID-5, and most calculations would be wasted on unlikely combinations of the parameters.

Another strength of the CUPID-5 approach is its straightforward, built-in error analysis,[11] which is not available with conventional methods. Error estimates can warn the investigator against overinterpreting the calculated conformations and their populations. We presume that the pseudorotations and populations calculated with conventional approaches also have large error ranges when conformers are close to one another or when one conformer has a low population. A straightforward error analysis could, in principle, be performed on the nonlinear fit of discrete conformers. However, the discrete method as implemented in ref 8 analyzes the error propagation by fitting the data repetitively with a series of restrictions on a selected parameter instead of analytically estimating its error range. This is why uncertainties are more difficult to detect in results from the discrete approach[8] than in those from the CUPID-5 method.

Errors in the input couplings, NOEs, and Karplus parameters, as well as the unknown χ_{max}, are the factors that limit the accurate prediction of the probability distribution of conformations. These errors are larger than those introduced by rejecting higher-order Bessel functions from eqs 13−17 (indicated by the ellipsis) and justify our use of approximate expressions for C_{1j}, S_{1j}, C_{2j}, and S_{2j}, (eqs 13−17), which consist of only the first several orders of Bessel functions. The higher-order Fourier coefficients (C_{3j}, S_{3j}, C_{4j}, S_{4j}, ...) may safely be neglected for the same reason. To probe the validity of these approximations, we investigated the dependences of two representative vicinal proton−proton couplings on the pseudorotation angle. The exact pseudorotation profiles (a projection, eq 18, and three sections with three different χ^0_{max} values, eq 8) of spin−spin couplings are compared in Figure 5 with their truncated trigonometric expansions that were actually used in CUPID-5 (eq 9 with $N_j = 2$ and eqs 13−17). The results indicate that errors from unknown χ^0_{max} are larger than those from truncation, and that both are small in comparison to those introduced by typical experimental errors. In a few individual cases, however, the rms value derived from CUPID-5 analysis was no better (or even worse) than that from the discrete model (*cf.* Supporting Information); in these exceptional cases, the approximations made in CUPID-5, in particular the omission of $\{C_{3j}, S_{3j}, C_{4j}, S_{4j}, ...\}$ from eq 9, may be the reason. In the vast majority of cases, CUPID-5 analysis led to smaller rms

(18) Other proton−proton couplings in proline rings exhibited pseudorotational dependences similar to these but shifted in phase (Supporting Information). Karplus parameters used to calculate pseudorotational profiles of couplings are listed in Table 1.

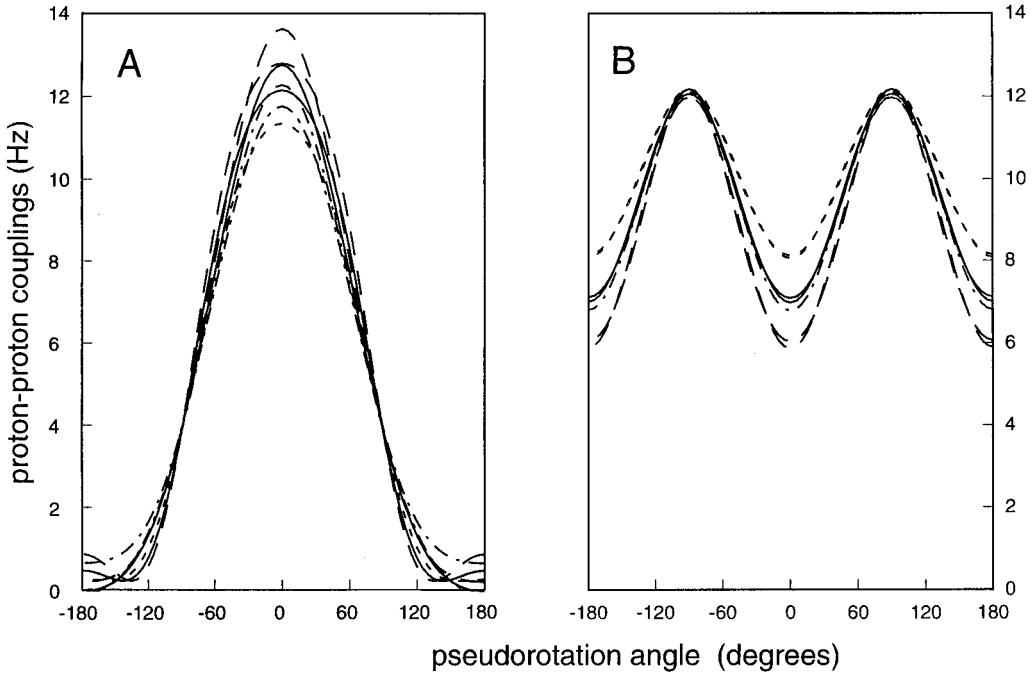

Figure 5. Pseudorotation profiles of two typical spin−spin couplings. (A) $H^{\beta 2}-H^{\gamma 3}$ coupling (an example of a "pseudorotational Karplus curve" with one maximum). (B) $H^{\beta 3}-H^{\gamma 3}$ (representative of a "pseudorotational Karplus curve" with two maxima). Both panels display the results obtained with the projection ($\chi^0_{max} = 40°$, $D = 15°$; dash-dot-dash) and with the section method ($\chi_{max} = 35°$, short dashes; $\chi_{max} = 40°$, solid line; $\chi_{max} = 45°$, long dashes). Both the exact curves (eqs 8 and 18 for the section and projection, respectively) and their truncated Fourier expansions (eqs 9 and 13−17) are shown for each choice of χ_{max}.[18]

differences (Table 2). With better quality data, it may be possible to include higher order Bessel functions and Fourier coefficients in the CUPID-5 calculations, provided that a reasonably precise estimate of χ^0_{max} is available.

An additional reason why the fit achieved with CUPID-5 is superior may be that the discrete model ignores fluctuations around pseudorotational equilibrium states,[7] whereas CUPID-5 improves the fit by taking the contributions of these fluctuations into account. The extents of oscillatory motions around the minima of the pseudorotational potential are measured by the widths of the Gaussian probability peaks or, equivalently, by the second derivatives of the potential curves around the minima, which are proportional to the potential barrier heights. In general, the fitted Gaussian peak widths are unreliable (Supporting Information) and cannot be used to characterize the angular fluctuations.[19] However, the equilibrium conformations (the positions of potential minima or Gaussian peak locations) and their populations (energy differences between them or areas of Gaussians) are reliable, as attested by the agreement between the potential calculated for LLL with that derived from force field calculations[7] (Figure 4).

Cis couplings ($H^{\alpha}-H^{\beta 3}$, $H^{\beta 2}-H^{\gamma 2}$, $H^{\beta 3}-H^{\gamma 3}$, $H^{\gamma 2}-H^{\delta 2}$, $H^{\gamma 3}-H^{\delta 3}$) have relatively large second-order Fourier coefficients C_{2j} and S_{2j} compared to experimental errors (Figure 5). The pseudorotational profile of the *trans* couplings ($H^{\alpha}-H^{\beta 2}$, $H^{\beta 3}-H^{\gamma 2}$, $H^{\beta 2}-H^{\gamma 3}$, $H^{\gamma 2}-H^{\delta 3}$, $H^{\gamma 3}-H^{\delta 2}$), on the contrary, are dominated by C_{1j} and S_{1j}. Large C_{2j} and S_{2j} values improve the reliability of the derived probability distribution when multiple conformers coexist. One can conclude, therefore, that, contrary to previously stated opinions,[20] *cis* couplings *are* helpful in conformational analyses. However, the pseudorotational profiles of *cis* couplings $H^{\beta 2}-H^{\gamma 2}$ and $H^{\beta 3}-H^{\gamma 3}$ are expected to be identical (within ±0.5 Hz) unless the Barfield transmission effect is included.[7] The inclusion of both couplings is still beneficial, since it reduces the effects of experimental errors. The same is true for the couplings $H^{\gamma 2}-H^{\delta 2}$ and $H^{\gamma 3}-H^{\delta 3}$. In cases where the second-order Fourier coefficients of the available couplings are not large enough with respect to experimental error, additional NMR input should be used, such as NOEs and heteronuclear, geminal,[21−23] and one-bond couplings.[22,24,25]

Bond lengths must be considered when torsion angles are reconstructed from pseudorotation data.[26] In proline, where the $N-C^{\alpha}$ bond is considerably shorter than the other bonds within the ring, a problem occurs when the pseudorotational wheel starts at different atoms. One solution has been to average the

(19) In most cases, in order to allow jumps between conformers,[11] the lower limit on Gaussian peak widths must be set at 20° (Supporting Information). For the same reason, the upper limit on energy barriers was set at 100 kJ/mol when the pseudorotational potential was fitted (Supporting Information). This upper limit easily accommodates the expected energy barrier between potential wells of around 13 kJ/mol.[2,6,7] However, the fitted Fourier coefficients of the potential are sensitive to changes in χ_{max} and have large error ranges. Thus, neither the Gaussian widths nor the CUPID-derived potential barriers between conformers provide useful information about the extent of fluctuations or the energetics of transition states on the pseudorotation wheel. That is to be expected, since the experimental inputs for CUPID-5 represent equilibrium averages and do not contain any information on the kinetics of transitions.[11] In five-membered rings, the characterization of transition states is further impeded by the fact that CUPID-5 uses two geometric descriptors, P and χ_{max}, which are enough only for low-energy conformations but may have to be replaced with four descriptors in high-energy states.[6] Similar doubts about the usefulness of calculated potential barriers were expressed in ref 7.

(20) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1973**, *95*, 2333.
(21) Marino, J. P.; Reif, B.; Zimmer, D. P.; Schwalbe, H.; Crothers, D. M.; Griesinger, C. *ENC: Book of Abstracts;* 37th Experimental Nuclear Magnetic Resonance Conference, Asilomar, CA, 1996, MP25, p 79.
(22) Podlasek, C. A.; Stripe, W. A.; Carmichael, I.; Shang, M.; Basu, B.; Serianni, A. S. *J. Am. Chem. Soc.* **1996**, *118*, 1413.
(23) Church, T.; Carmichael, I.; Serianni, A. S. *J. Am. Chem. Soc.* **1996**, *280*, 177.
(24) Carmichael, I.; Chipman, D. M.; Podlasek, C. A.; Serianni, A. S. *J. Am. Chem. Soc.* **1993**, *115*, 10863.
(25) Serianni, A. S.; Wu, J.; Carmichael, I. *J. Am. Chem. Soc.* **1995**, *117*, 8645.
(26) Juranić, N.; Niketić, S. R.; Juranić, I. *J. Mol. Struct.* **1992**, *271*, 209.

results starting at all five atoms.[2] The error in torsion angles introduced by not doing so in CUPID-5 is below the experimental error. The *direct* effects that bond lengths and bond angles have on scalar couplings[22,25] also has been ignored in CUPID-5, because these effects, too, are below the uncertainties in measured couplings. Also, the introduction of the Barfield transmission effect[7] into CUPID-5 was deemed unjustifiable in light of the simplifications and approximations that have already been included.

The projection method (eq 18) gave a slightly better rms value (Table 2) and almost the same Fourier coefficients of spin−spin couplings (within the experimental error range) as the method in which χ_{max} was assumed to be constant (eqs 8, 9, and 13−17). Two extreme interpretations of the projection are possible. In the first, which is physically unrealistic, χ_{max} is assumed to adopt a fixed, discrete value that is not well defined by the data. In the second, the ensemble of molecules is characterized by a distribution of χ_{max} values, and each molecule acquires different χ_{max} values in the course of time comparable with the duration of the measurement. In either extreme, because the width D of the probability distribution for χ_{max} is relatively large while sections at different χ_{max} values are similar (Supporting Information), it is not necessary to identify the exact value of χ^0_{max} with a precision better than $\pm 2°$.[12] Molecular mechanics calculations suggest that χ_{max} in sugars varies with P;[27,28] thus, CUPID's assumption of constant χ_{max} must be viewed as an approximation.

5. Conclusion

The CUPID-5 method offers several advantages over conventional approaches to the analysis of conformations of five-membered rings from NMR data: (i) CUPID-5 uses linear regression and, therefore, always yields *the best possible* fit; (ii) the number of parameters fitted in a CUPID-5 analysis (Fourier coefficients of the probability distribution) is smaller (four parameters) than that in a conventional analysis (five parameters); (iii) the Fourier coefficients used in CUPID-5 analyses are independent (whereas the parameters of the discrete model can be correlated); (iv) CUPID-5 provides analytic expressions for error propagation, which are not available with the discrete model; (v) CUPID-5 uses the same formalism to treat different types of experimental input data; and (vi) CUPID-5 analysis is easy to implement with a variety of input data types (*e.g.*, vicinal, geminal, and one-bond couplings, NOEs, chemical shifts, and cross-correlations) and has modest hardware requirements (a personal computer suffices for all CUPID calculations). It is our hope that CUPID-5 will make NMR-based conformational analyses of five-membered rings in biomolecules easier and more accurate.[13]

Acknowledgment. This work was supported by NIH Grants RR 02301 and GM 35976. The authors thank Dr. Arthur S. Edison, Dr. George Maalouf, Dr. Nenad Juranić, and Dr. Slobodan Macura for stimulating discussions, and the reviewers for useful suggestions.

Supporting Information Available: Listings of experimental values of proton−proton couplings in proline rings, Fourier coefficients and Gaussian parameters of probability distributions for the pseudorotation, Fourier coefficients of potential barriers for the pseudorotation in prolines, notation for atoms and dihedral angles in L-proline and D-proline, and Karplus curves for all the proton pairs in L-proline (31 pages). See any current masthead page for ordering and Internet access instructions.

JA962160+

(27) French, A. D.; Tran, V. *Biopolymers* **1990**, *29*, 1599.
(28) French, A. D.; Dowd, M. K. *J. Comput. Chem.* **1994**, *15*, 561.