J Biomol NMR (2012) 54:97–114
DOI 10.1007/s10858-012-9653-2

ARTICLE

Transforming between discrete and continuous angle distribution models: application to protein χ_1 torsions

Jürgen M. Schmidt

Received: 11 March 2012 / Accepted: 30 June 2012 / Published online: 31 July 2012

Abstract Two commonly employed angular-mobility models for describing amino-acid side-chain χ_1 torsion conformation, the *staggered-rotamer jump* and the *normal probability density*, are discussed and performance differences in applications to scalar-coupling data interpretation highlighted. Both models differ in their distinct statistical concepts, representing *discrete* and *continuous* angle distributions, respectively. Circular statistics, introduced for describing torsion-angle distributions by using a universal circular order parameter central to all models, suggest another distribution of the continuous class, here referred to as the *elliptic* model. Characteristic of the elliptic model is that order parameter and circular variance form complementary moduli. Transformations between the parameter sets that describe the probability density functions underlying the different models are provided. Numerical aspects of parameter optimization are considered. The issues are typified by using a set of χ_1 related ^3J coupling constants available for FK506-binding protein. The discrete staggered-rotamer model is found generally to produce lower order parameters, implying elevated rotatory variability in the amino-acid side chains, whereas continuous models tend to give higher order parameters that suggest comparatively less variation in angle conformations. The differences perceived regarding angular mobility are attributed to conceptually different features inherent to the models.

Keywords Circular statistics · Directional data · Probability density · Gaussian model · Elliptic model · Jinc function · Bessel function · Torsion angle conformation · Rotamer equilibria · Differential probability · ^3J · Vicinal coupling constants · Amino-acid side chain · Protein structure · FKBP

Introduction

When increasing the resolution of a molecular model through measurement, evaluation and interpretation of, for example, NMR data, dynamical phenomena will invariably be encountered at some point in the structure derivation process. Biomolecular NMR structure determination refers, amongst other parameters, to ^3J coupling constants measured as the observable spin-system responses to varying torsion angles (Karplus 1963). The significance of rotatory oscillation about idealised torsion-angle geometries and their implications with respect to the interpretation of dynamically averaged NMR parameters was previously recognised (Jardetzky 1980; Nagayama and Wüthrich 1981; Hoch et al. 1985).

Torsion angles, especially their temporal and spatial value distributions within a given molecular moiety, such as an amino-acid residue in a protein, are obtained from coupling constants only indirectly through the application and testing of suitable models, such that—on principle—all data interpretation is made on the basis of inferences. Without exception, all models for the description of torsion-angle conformation and dynamics represent probability density functions of the angle, $P(\theta)$. These broadly fall into two classes, *discrete* and *continuous* distributions.

Conformational analysis on the basis of ^3J coupling constants usually begins with attempts at obtaining for the

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9653-2) contains supplementary material, which is available to authorized users.

J. M. Schmidt (✉)
School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK
e-mail: j.m.schmidt@kent.ac.uk

sought torsion parameter a single unique value that satisfies all experimental constraints, resulting in a discrete single-point or δ distribution. A popular discrete model that includes torsion-angle dynamics was introduced by Pachler (1963, 1964) and involves jumps between the three energetically favourable ideal staggered-rotamer states. This model proved particularly efficient in the analysis of amino-acid side-chain χ_1 torsions (Hansen et al. 1975).

Continuous probability density functions employed so far include the uniform distribution (Nagayama and Wüthrich 1981), the Gaussian or normal distribution (Hoch et al. 1985; Karimi-Nejad et al. 1994; Brüschweiler and Case 1994; Polshakov et al. 1995), as well as extensions to the normal distribution model in the form of bimodal (Avbelj and Baldwin 2003) or multimodal shapes, such as the CUPID approach (Džakula et al. 1992a, b).

The present investigation is to shed light on the similarities and differences between the two arguably most popular angular-motion models employed in the interpretation of data for the amino-acid side-chain torsion χ_1, the Pachler model and the Gaussian model. The former being a discrete, the latter a continuous distribution function, each can be considered a limiting case in their own right. Transformation between both models will be detailed, so as to permit interconversion between the parameter sets and to help develop a consistent picture and understanding of the underlying molecular motional process. In practice, results from fitting both models to the same experimental data differ for reasons that will be addressed in the present work.

From the above, the need arises for a measure of circular variability that is independent of the chosen model of torsion-angle dynamics. It is also an aim of the present work to develop an alternative and improved model suitable for describing molecular torsion-angle mobility.

Theory

Circular order parameter

One approach to expressing conformational variability is the use of order parameters that, owing to their fractional value range, between 0 and 100 %, facilitate grasp of the extent of dynamical effects. Commonplace in the analysis of spin-relaxation data are dimensionless *angular* order parameters, usually signified by S^2, descriptive of bond-vector correlation in *three*-dimensional spherical space that are taking on limit values of 1 and 0, respectively, for a single static orientation and for isotropic reorientation (Lipari and Szabo 1982a).

In *J*-coupling analysis, dimensionality is reduced insofar as dihedral-angle variation is in a circular rather than spherical context. This involves correlating difference angles as produced by an ensemble of bond vectors projected onto a *two*-dimensional circle, as opposed to vectorial correlation on a three-dimensional sphere. Vector motion in the spherical context (Lipari and Szabo 1982b), if restricted to within a plane intersecting a sphere, would result in a residual baseline of $S^2 = \{P_2(\cos \pi/2)\}^2 = 0.25$, as motional freedom is then deprived of one space dimension.

A *circular order parameter*, R, suitable for summarising dihedral-angle dynamics takes on a value of 1 for a single static rotamer, and a value of 0 for complete circular averaging. Consider an ensemble of torsions represented by polar vectors of unit length (directors), their phases reflecting the directions or dihedral-angle values, θ. Then, the circular order parameter is derived from the vector sum as given by

$$R = \left(C_1^2 + S_1^2\right)^{1/2} \tag{1}$$

with C_m and S_m the mth trigonometric moments, defined as

$$\begin{aligned} C_m &= N^{-1}\Sigma_k \cos m\theta_k \\ S_m &= N^{-1}\Sigma_k \sin m\theta_k \end{aligned} \tag{2}$$

where k runs over all N angle items of the distribution in a particular torsion. Notice that the second sine and cosine moments, S_2 and C_2, differ from the squares of first moments, S_1^2 and C_1^2, employed in Eq. 1.

Circular mean direction and deviation

For any *circular* distribution in *discrete* angle data, estimates for both mean direction and mean angular deviation, θ_o and σ_o, respectively,[1] are simultaneously obtained following the principle of vector addition in the complex number plane, using the complex operator $i = (-1)^{1/2}$,

$$\begin{aligned} R\exp(i\theta_o) &= N^{-1}\Sigma_k \exp(i\theta_k) \\ &= C_1 + iS_1 \end{aligned} \tag{3}$$

The mean *direction* or *location* of the circular distribution is thus encoded in the *phase* of the sum vector, which involves trigonometric first moments only, and is given by[2]

$$\theta_o = \arctan(S_1/C_1). \tag{4}$$

Descriptive of the *concentration* of the angles about their mean direction is the length of the normalized sum vector, R, as defined in Eq. 1, which is dimensionless and bounded

[1] The nought subscript may designate, by virtue of its shape, the circular statistical variables.

[2] Numerical implementations usually benefit from provision in mathematical libraries of function `atan2` which takes sine and cosine terms as separate arguments, allowing to determine the correct quadrant into which the resultant vector falls.

on the interval [0, 1]. The complementary quantity, $(1 - R)$, must then relate to the *dispersion* in the circular data about their mean (Mardia 1972; MacArthur and Thornton 1993). Various approaches have been suggested to extract circular variance measures from R.

Circular variance according to Batschelet

Batschelet (1965, 1981) defines angular variance in directional data as

$$\sigma_{\mathrm{B}}^2 = 2(1 - R). \tag{5}$$

The factor of 2 reflects the small-angle approximation, $\sin\Delta\theta \approx \Delta\theta$, where $\Delta\theta = \theta_k - \theta_{\mathrm{o}}$, following which

$$\begin{aligned}\sigma_{\mathrm{B}}^2 &= 2(1 - N^{-1}\Sigma_k \cos(\theta_k - \theta_{\mathrm{o}})) \\ &= 2(1 - \cos\sigma_{\mathrm{B}})\end{aligned} \tag{6}$$

rationalising the identity

$$R = \cos\sigma \tag{7}$$

suggested by Janin et al. (1978) for the analysis of torsion-angle variation. Batschelet's approach makes no assumption regarding the shape of the circular distribution and, therefore, works with any discrete dataset, such as, in the present case, a distribution of staggered χ_1 torsion-angle rotamers.

Circular variance according to Mardia

Conceptually different, Mardia (1972) assumes the circular distribution to be characterised by a continuous Normal or Gaussian distribution, whose width parameter or standard deviation is that abscissa at which the probability density equals $e^{-1/2}$ of its maximum value assumed to be located at the origin. Then, the corresponding sum-vector length, i.e., the order parameter, identifies with

$$R = \exp(-\sigma_{\mathrm{M}}^2/2), \tag{8}$$

and angular variance in the Gaussian context equates to

$$\sigma_{\mathrm{M}}^2 = -2\ln(R). \tag{9}$$

Notice that circular variance as given by Batschelet remains finite on the interval [0, 2], whereas that given by Mardia diverges to infinity for $R \to 0$.

Circular variance in wrapped normal distributions

Mardia's approach seeks to impose Gaussian statistics for the linear case onto the circle. The circular equivalent of expanding the exponential function into a series of terms involves a series of (symmetrical) cosine functions which automatically take wrapping effects into account. Regarding circular variance measures, the properties of the respective model, known as the wrapped normal distribution (Fisher 1993), are virtually identical to that of the linear Gaussian model.

Truncating the cosine-series expansion in the wrapped normal distribution after the first cosine mode yields the so-called cardioid distribution (Jeffreys 1961). This being essentially a simplified model, its discussion is implicit in that of the parent.

Circular variance according to Hyberts

Hyberts et al. (1992) suggested an entirely empirical relationship between the order parameter and a circular standard deviation, which they define as

$$\sigma_{\mathrm{H}} = 2\arccos(1 + 0.5\ln R). \tag{10}$$

For small values of R (designated S in their original paper), this circular standard deviation trends toward infinity. The authors argue that "the angular order parameter is a better quantity to describe the precision of dihedral angles than a standard deviation. For a completely random angle distribution the standard deviation of the angles is not defined." Indeed, referring to the model of Eq. 10, σ_{H} is undefined if $R = 0$.

The circular uniform distribution

The obvious model of a completely random angle distribution is the circular uniform distribution. Representing a limiting case, the circular uniform distribution has an order parameter of zero and—by definition—its circular standard deviation is infinite. All directions being equally likely, $P = (2\pi)^{-1}$, the impossibility of fitting any direction parameter leaves the circular uniform of no further practical use than as a null model against which to test other distributions (Fisher 1993).

Circular variance from trigonometric moments

It is here suggested to obtain the circular variance, σ_{o}^2, by an approach motivated entirely by the definition of variance on the basis of statistical moments. Trigonometric moments as given in Eq. 2 are the sample analogues of the coefficients in the Fourier series expansion of the circular probability density. Variance—in circular as well as linear statistics—would generally be obtained from the second moment of the distribution, taken away the mean-square magnitude, as follows

$$\begin{aligned}\sigma_{\mathrm{o}}'^2 &= \mu_2' - \mu_1'^2 \\ &= N^{-1}\Sigma_k \exp(i2\theta_k) - N^{-2}\{\Sigma_k \exp(i\theta_k)\}^2 \\ &= Q^2 \exp(i2\theta_{\mathrm{o}}) - \{R\exp(i\theta_{\mathrm{o}})\}^2 \\ &= \left(Q^2 - R^2\right)\exp(i2\theta_{\mathrm{o}})\end{aligned} \tag{11}$$

The factor Q^2 is only temporarily introduced to distinguish sum-vector shortening related to the second moment from that due to the first moment, as both are unlikely identical. Thus, circular variance manifests in a length difference between resultant vectors obtained from averaging over single-angle and double-angle terms, modulated with an angular frequency corresponding to twice the phase of the mean direction.

Following statistical theorems that the mean minimizes the variance, the net effect arising from the average location is yet to be eliminated from Eq. 11. Typically a complex number, any standard deviation according to Eq. 11 therefore remains incomplete until including the 'mirror-image' moments to which circular data naturally give rise. Employing products between complex numbers and their complex conjugates (signified by asterisk), rather than straight squares, ensures real values result for both the second moment and the squared first moment,

$$\sigma_o^2 = \mu_2'^* - \mu_1' \mu_1'^*. \tag{12}$$

Referring to Eq. 3, the mean-square magnitude is obtained as

$$\begin{aligned} \mu_1' \mu_1'^* &= N^{-2} \Sigma_k \exp(i\theta_k) \Sigma_k \exp(-i\theta_k) \\ &= R \exp(i\theta_o) R^* \exp(-i\theta_o) \\ &= R^2 \end{aligned} \tag{13}$$

and the completed second moment is the auto-correlation

$$\begin{aligned} \mu_2'^* &= N^{-1} \Sigma_k \{\exp(i\theta_k) \exp(-i\theta_k)\} \\ &= 1 \end{aligned} \tag{14}$$

Eventually, the circular variance is given (in rad^2) as

$$\sigma_o^2 = 1 - R^2 = (1 + R)(1 - R) \tag{15}$$

and the (squared) circular order parameter as

$$R^2 = 1 - \sigma_o^2 = (1 + \sigma_o)(1 - \sigma_o) \tag{16}$$

Reminiscent of a semicircle function, henceforth referred to as the *elliptic* model, the circular standard deviation (in rad) and the (dimensionless) circular order parameter form a pair of *complementary moduli*. As R^2 is proportional to $\cos^2\sigma_o$ (Eq. 7), so is $(1 - R^2)$ proportional to $\sin^2\sigma_o$.

Notice that Eq. 15 resembles Eq. 5, except the fixed factor of 2 is replaced by the variable term $(1 + R)$, which, departing from 2, dies away with increasing dispersion, effectively removing the small-angle assumption made in Batschelet's approach.

Applying Eq. 8, the Gaussian variance maps onto the circular variance of Eq. 15 according to

$$\sigma_o^2 = 1 - \exp(-\sigma_M^2) \tag{17}$$

and, conversely, the Gaussian variance (Eq. 9) is obtained from the circular variance using

$$\sigma_M^2 = -\ln(1 - \sigma_o^2). \tag{18}$$

Clearly, the Gaussian variance, which is appropriate for the linear case, and the circular variance suggested in the present work differ somewhat and deserve more detailed comparison by example.

Circular variance in the two-state exchange case

Following the concepts just outlined, standard deviations were calculated for a two-site jump between discrete angle states, represented by a pair of directions on the circle deliberately centered at zero location, while incrementing separation half-angle $\Delta\theta$ (Fig. 1).

Averaging in this two-state case essentially amounts to taking the sum of a complex number and its complex conjugate, corresponding to locations $+\Delta\theta$ and $-\Delta\theta$, respectively. The farthest two directions can be placed apart is at π separation. Due to the then vanishing magnitude of R, the mean direction must remain undefined in this special case, which is plausible and intuitively appropriate.

For combinations of two equally populated torsion-angle states separated by $2\Delta\theta = 120°$, for example, averaging any two out of the three possible staggered-rotamer states in an aliphatic hydrocarbon moiety, the resultant mean direction between the two populated directions is the exact antipode of the unoccupied third direction. In these situations, σ_B takes a value of 1 rad or 57.3° as the order parameter $R = 0.5$, whereas the Gaussian deviation, σ_M, would be larger at 67.5°, and the circular deviation of the elliptic model, σ_o, would be smaller at 49.6° (Table 1).

Circular variance in the rotatory averaging case

In the absence of angular variation, $R = 1$ and $\sigma_o = 0$ holds for all models. However, the models differ in the other extreme of complete circular averaging, when $R = 0$. Standard deviations according to Mardia, σ_M, or Hyberts, σ_H, become infinite, whereas Batschelet's deviation, σ_B, reaches its maximum of $\sqrt{2}$ rad (Eq. 5), or just above 81°. Decreasing R also increases the mean circular deviation in the elliptic model (Eq. 15), however, σ_o falls progressively short of σ_B, until, in the limit, at $R = 0$, a variance of 1 rad^2 is obtained, just half the value of Batschelet's variance.

As R approaches zero, singularities in the models by Mardia or Hyberts cause infinite slopes, $d\sigma/dR$. This contrasts Batschelet's and the linear model, both of which share the feature of a defined finite slope at $R = 0$, equaling $-1/\sqrt{2}$ and $-\sqrt{2}$, respectively, suggesting residual propensity for σ to rise as if the angle range were to extend beyond full circle (Fig. 1). The mean deviation in the elliptic model, σ_o, however, meets the point $R = 0$ with zero slope. This seems most plausible, because circular

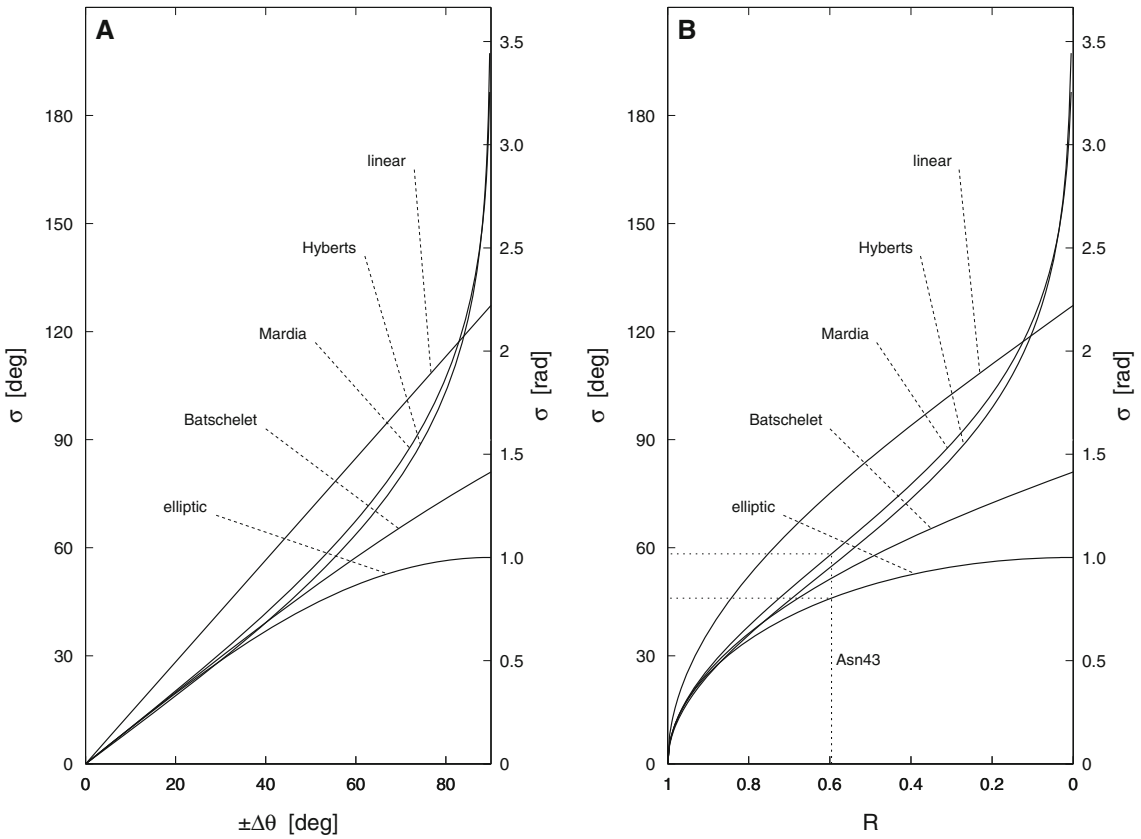

Fig. 1 Comparison of circular standard deviation measures for a two-site conformational exchange process involving dihedral-angle reorientation by $\pm\Delta\theta$ degrees. Notice that $\pm 60°$ corresponds to a jump between two staggered amino-acid side-chain rotamers. Circular standard deviations obtained according to concepts by Batschelet, Mardia, Hyberts, and this work, all connect through a concept-independent, dimensionless order parameter, R, as detailed in the text. The standard deviation corresponding to linear statistics is also shown. **a** Circular standard deviation in degrees or radians versus reorientation half angle. **b** Circular standard deviation in degrees or radians versus circular order parameter. Mapping of an order parameter is exemplified for an R value of 0.596, as fitted for the Asn43 side chain in FKBP, yielding Gaussian and elliptic standard deviations of 58.3° and 46.0°, respectively. The elliptic function obtained according to the approach suggested in the present study demonstrates the complementary-modulus property, $R^2 + \sigma_o^2 = 1$

variance cannot possibly increase any further once all directions are populated, and no propensity for increase ought to remain either.

It may be remarked that outliers do not really exist in circular statistics, a fact that naturally imposes bounds on distribution spreads.

Circular variance in the three-state exchange case

Amino-acid side-chain conformation is often analyzed under the premise of three staggered torsion-angle rotamer states being present at $\chi_1 = -60°, +60°$, and $\pm 180°$ (Pachler 1963, 1964), signified as *gauche*$^+$ (g^+), *gauche*$^-$ (g^-), and *trans* (t) conformations (Janin et al. 1978), respectively.[3] Whether continuous or discrete, two parameters are necessary to model either a Gaussian distribution about a mean torsion angle or, alternatively, a set of probabilities associated with the fixed, staggered rotamers. Thus, both models are of identical complexity and supposedly exploit the available experimental information equally. Both models interconvert as follows.

Discrete → continuous model transformation

Equate both probability density models at the level of the torsion angle distribution and claim that

$$\exp(i\theta_o - \sigma_M^2/2) = P_{g^+}\exp(-i\pi/3) + P_{g^-}\exp(i\pi/3) + P_t\exp(i\pi) \quad (19)$$

The right-hand-side constitutes a weighted sum of fixed directions, the exponential components being given constants, such that, using Eq. 8, the apparent mean direction and mean deviation of an equivalent circular normal distribution are obtained from the probabilities only,

[3] Pachler chose the rotatory sense of substituent placement opposite to the convention later recommended by the IUPAC-IUB (1970), and Janin et al. chose an opposite angle convention, explaining apparent sign inconsistencies.

Table 1 Mean directions from staggered-rotamer equilibria and circular standard deviations within various statistical frameworks

Case	P_{g^+}	P_t	P_{g^-}	R	$\theta_o \pm$	σ_o	σ_B	σ_M	σ_H
Single state									
	1	0	0	1	−60 ±	0	0	0	0
	0	1	0		180 ±				
	0	0	1		+60 ±				
2-State jump, 2:1 occupancies									
	2/3	0	1/3	0.5774[a]	−30 ±	46.8	52.7	60.1	56.5
	2/3	1/3	0		−90 ±				
	1/3	2/3	0		−150 ±				
	0	2/3	1/3		+150 ±				
	0	1/3	2/3		+90 ±				
	1/3	0	2/3		+30 ±				
2-State jump, 1:1 occupancies									
	1/2	1/2	0	0.5	−120 ±	49.6	57.3[b]	67.5	63.7
	1/2	0	1/2		0 ±				
	0	1/2	1/2		+120 ±				
3-State jump, uniform occupancies, rotational average									
	1/3	1/3	1/3	0	n/d ±	57.3[b]	81.0[c]	∞	n/d

Circular mean directions θ_o and deviations σ (in degrees) following transformations described in the text: σ_o (Eq. 15), σ_B (Eq. 5), σ_M (Eq. 9), σ_H (Eq. 10), and R (Eq. 1)

n/d not defined

[a] Equates to $1/\sqrt{3}$

[b] Equates to 1 rad

[c] Equates to $\sqrt{2}$ rad

$$R\exp(i\theta_o) = P_{g^+}(1 - i\sqrt{3})/2 + P_{g^-}(1 + i\sqrt{3})/2 - P_t \tag{20}$$

To eliminate any one of the three interdependent probability parameters, most conveniently and merely for symmetry reasons, we choose the substitution $P_t = 1 - P_{g^-} - P_{g^+}$, and obtain

$$\begin{aligned} \sin\theta_o &= (3^{1/2}/2)(P_{g^-} - P_{g^+}) \\ \cos\theta_o &= (3/2)(P_{g^-} + P_{g^+}) - 1 \end{aligned} \tag{21}$$

The composite probabilities $(P_{g^-} + P_{g^+})$ and $(P_{g^-} - P_{g^+})$, respectively, determine the proportion complementary to the *trans* rotamer and the balance between the two *gauche* rotamers, the latter representing circular anisotropy. The apparent mean direction in the context of the continuous normal distribution then results from (compare Eq. 4)

$$\begin{aligned} \tan\theta_o &= \sin\theta_o/\cos\theta_o \\ &= (P_{g^-} - P_{g^+})\left\{3^{1/2}(P_{g^-} + P_{g^+}) - 0.5\right\}^{-1} \end{aligned} \tag{22}$$

If Eq. 22 represents the phase or direction of a polar, then its modulus or length, i.e., the order parameter, expressed in terms of the probability parameters, is given by

$$R = \left\{1 + 3\left(P_{g^-}{}^2 + P_{g^+}{}^2 - P_{g^-} - P_{g^+} + P_{g^-}P_{g^+}\right)\right\}^{1/2} \tag{23}$$

For any uniquely occupied single-rotamer state, $P_{g^+} = 1$ and $P_{g^-} = 0$, say, $R = 1$, as expected. For equal occupancies, $P_{g^+} = P_{g^-} = 1/3$, i.e., complete 'rotational' averaging, $R = 0$. For any two-site jump between equally populated staggered states, such as $P_{g^+} = P_{g^-} = 1/2$, $R = 0.5$, consistent with the previous section (Table 1). It is noted though that many other distributions would also give rise to an order parameter of 0.5, for example, Gaussian-type models including skewness or kurtosis.

Continuous → discrete model transformation

By reversing Eqs. 22 and 23, θ_o and σ_o transform into apparent staggered-rotamer probabilities once the intermediary circular order parameter R is available (e.g., from Eq. 8 in the context of a normal distribution), according to

$$\begin{aligned} P_{g^+} &= 1/3(1 + R\cos\theta_o - 3^{1/2}R\sin\theta_o) \\ P_{g^-} &= 1/3(1 + R\cos\theta_o + 3^{1/2}R\sin\theta_o) \\ P_t &= 1 - P_{g^+} - P_{g^-} \end{aligned} \tag{24}$$

Application to NMR data

Coupling constant averaging

To delineate conformation averaged J coupling constants, appropriate expressions can be substituted for the various terms in the original Karplus equation (Karplus 1963),

$$^3J(\theta) = C_0 + C_1 \cos\theta + C_2 \cos 2\theta \quad (25)$$

The Karplus coefficients, C_m, identify with the first three Fourier coefficients of the probability density in the torsion angle obtained when transforming Eq. 25 into the reciprocal-angle domain.

Averaging effects on J coupling constants were previously modelled by convolving Karplus curves with the continuous Gaussian probability density function (Hoch et al. 1985; Karimi-Nejad et al. 1994; Polshakov et al. 1995),

$$\langle {}^3J(\theta)\rangle = C_0 + C_1\langle\cos\theta\rangle + C_2\langle\cos 2\theta\rangle \quad (26)$$

Alternatively, variance-dependent scaling was applied to the angle-related Karplus coefficients,

$$\langle C_m\rangle = F_m C_m, \quad (27)$$

such as to yield (Brüschweiler and Case 1994)

$$\langle {}^3J(\theta)\rangle = C_0 + \langle C_1\rangle \cos\theta + \langle C_2\rangle \cos 2\theta \quad (28)$$

At any rate, the correspondingly modified Karplus equation, Eqs. 26 or 28, is subsequently used for interpreting J values.

The scaling factors, F_m in Eq. 27, arise from applying convolution theorems, whereby 'folding' the instantaneous coupling with a probability density involves multiplying the Karplus-curve transform (coefficients C_m) with the probability-density transform (factors F_m) in order to yield expectation values for the conformation averaged coupling constant. Circular variability thus manifests as certain scaling ratios between 'adjacent' Karplus coefficients, and the ratios depend on the chosen model of torsion-angle dynamics. Amenable to manipulation, the scaling factors and their ratios provide a convenient and efficient way of modelling conformation averaged coupling constants.

Scaling Karplus coefficients in the Gaussian model context

An angle distribution of Gaussian shape imposes a Gaussian envelope on the series of Karplus coefficients in Eq. 28 (Fig. 2). The mth Karplus coefficient subjected to angular averaging in the Gaussian context scales with a simple dependence on the angular order parameter,

$$\langle C_m\rangle = R^{(m^2)} C_m \quad (29)$$

At the order-parameter level, averaging the coupling constant then stipulates scaling the Karplus coefficients C_0, C_1, and C_2 by explicit factors of $F_0 = 1$, $F_1 = R$, and $F_2 = R^4$, respectively. At the standard-deviation level (Eq. 8), scaling factors were given by Brüschweiler and Case (1994) generically as $F_m = \exp(-m^2\sigma_\theta^2/2)$, and explicitly as $F_0 = 1$, $F_1 = \exp(-\sigma_\theta^2/2)$, and $F_2 = \exp(-2\sigma_\theta^2)$. Figure 2 shows these relations for various degrees of circular dispersion. Higher modes of the angular distribution associated with scaling factors, F_3, F_4, etc. are not normally discernable by applying the standard Karplus model of J-coupling that comprises the three lowest modes only.

Scaling Karplus coefficients in the elliptic model context

The combination of nested exponential and trigonometric functions in Eq. 10 is reminiscent of Bessel functions. Designed for application to circular statistics, Richard von Mises (1918) suggested probability distributions that employ a Bessel function kernel. The von-Mises distribution later became known as the circular normal distribution (Fisher 1993; Jammalamadaka and SenGupta 2001; Fernández-Durán 2004). Where cosine and sine functions satisfy rectangular boundary conditions defined in Cartesian coordinates, Bessel functions satisfy circular or cylindrical boundary conditions involving polar coordinates. They form the Fourier series for the arc of a circle.

Analogous to the sine-related *sinc* function as the Fourier transform of a rectangular profile (tophat function), the semicircle function (compare Eq. 16),

$$f_l = (1 - (l/\pi\sigma_o)^2)^{1/2}; \quad 0 \text{ if } l > |\pi\sigma_o| \quad (31)$$

has the *jinc* function as its Fourier transform (Bracewell 1986, Figure 2),

$$F_m = jinc(mx) = J_1(\pi mx)/\pi mx \quad (32)$$

where $J_1(x)$ is a Bessel function of the first kind.[4] Bearing resemblance to the power series representation for $\sin(x)$, the series representation of $J_1(x)$ is

$$J_1(x) = x/2 - x^3/(2^2 4) + x^5/(2^2 4^2 6) - \cdots \quad (33)$$

For convenience, the expression of Eq. 33 is usually multiplied by $2/x$ to accomplish $jinc(0) = 1$. Notice that the *jinc* series amplitude decays only as $x^{-1/2}$, contrasting the faster decay of x^{-1} for the *sinc* series. Benefitting speedy numerical optimization, the derivative of Eq. 32 is obtained as (Abramowitz and Stegun 1972; Spiegel and Liu 1999)

[4] Coincidentally, the same symbols, J, are being used for both Bessel functions and NMR coupling constants. The Bessel function is distinguished by a subscript numeral.

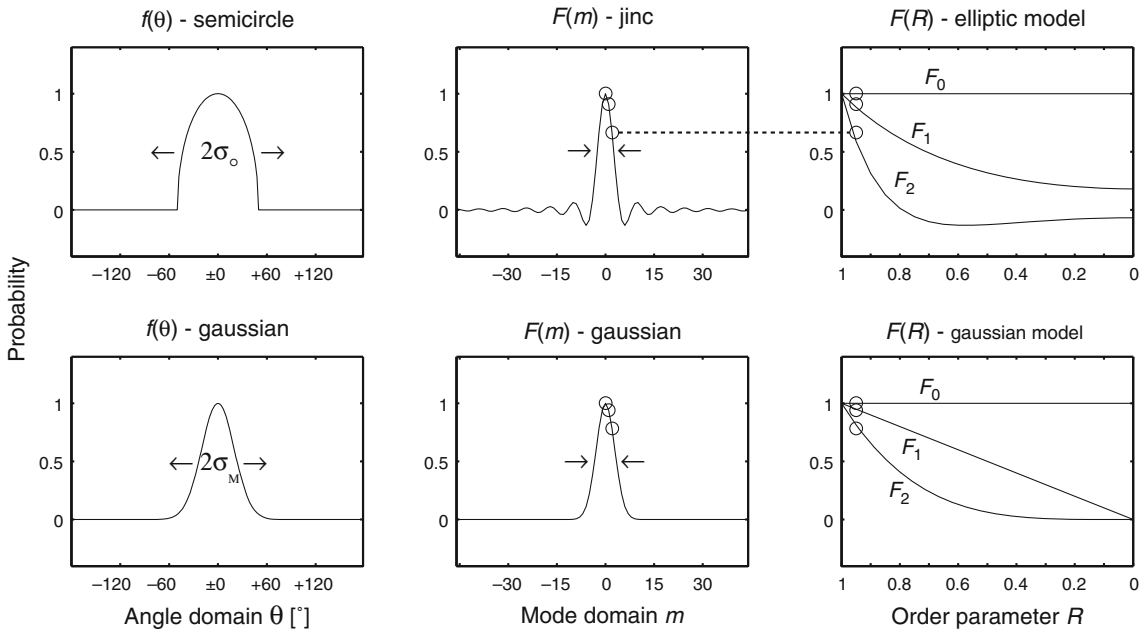

Fig. 2 (*top panels*) Semicircle function, $f(\theta) = (1 - \theta^2)^{-1/2}\ \Pi(\theta/2)$, providing angular probability density in the torsion-angle domain to characterise conformation averaging, and its Fourier transform, $F(m) = J_1(m\pi)/m\pi$, the *jinc* function in the mode domain; both functions scaled to unit amplitude for convenience. (*bottom panels*) Gaussian function, $f(\theta) = (2\sigma^2)^{-1/2}\ \exp(-\theta^2/2\sigma^2)$, providing angular probability density in the framework by Mardia, and its Fourier transform, $F(m) = \exp(-m^2\sigma^2/2)$; both functions scaled to unit amplitude. *Arrows* indicate the reciprocal relationship of (circular or Gaussian) standard deviations with their Fourier transform counterpart. The broader the distribution in the angle, the more concentrated the intensity about the conformation independent mode-0 in the co-domain, and vice versa. *Circles* in central and right-hand-side panels indicate relative function values relevant to the scaling of the first three Karplus modes, C_m, $m = 0, 1, 2$, and their dependencies on variation in the order parameter, the *dotted line* showing the connection. Explicit values of $F(R)$ are provided as Supplementary Material

$$jinc'(x) = -J_2(x)/x \tag{34}$$

Coupling-constant averaging by way of scaling coefficients in contexts other than the Gaussian model is possible in principle, albeit less straightforward. An analytical solution to the Bessel function integral for the purpose of averaging in the context of the elliptic model cannot be obtained in closed form and typically requires an iterative approach.

Scaling factors to the Karplus coefficients that emulate coupling-constant averaging in the framework of the elliptic model were computed and inserted in Fig. 2 also (see Supplementary Material). It is seen that multiplier F_1 remains positive throughout the range of R. With the elliptic model, descent in the value of F_1 is steepest when the order parameter R is highest, only to level off at around 0.18 toward full dispersion. This contrasts the strict proportionality between F_1 and R in the Gaussian model. Interestingly, this lends to an interpretation by which the magnitude difference between primary and secondary maxima of the apparent, modified Karplus curve, which is characterised by coefficient C_1, cannot disappear entirely upon complete rotational averaging, as it does in the Gaussian case.

While Karplus coefficients C_2 in the Gaussian context also scale with positive factors F_2 throughout, scaling as derived for the elliptic model turns negative once R decreases below 0.8. The negative excursion reflects the waviness radiating from the center of the Bessel function and accounts for the discontinuities in the angular probability density where the semicircle meets the baseline (Fig. 2).

Scaling Karplus coefficients in the Pachler model context

Population parameters P_{g^+}, P_t, and P_{g^-} for the three staggered rotamers are normally calculated using linear combinations of *trans* and *gauche* couplings, J_t and $J_{g^\pm}$ (Hansen et al. 1975). The corresponding conformations of torsion angle χ_1 in amino acids possessing a C^β methylene group are signified as t_2g_3, g_2t_3 and g_2g_3, indicating the spatial relationship between the unique H^α atom and both $H^{\beta 2}$ and $H^{\beta 3}$.

The question now arises as to what scaling factors, F_m, would possibly represent the Pachler model. Appealing as the concept may be, imposing conformation averaging on the Karplus coeffients is impossible in the case of the staggered-rotamer model. As the directions in the angle domain are invariant, Fourier transform of an angle triplicity typical of discrete probability density at three fixed equidistant directions commands *static* baseline intensity at

mode $m = 3$ in the other, reciprocal-angle co-domain, regardless of the populations. In fact, the Fourier transform of a discrete equidistant three-point *uniform* probability distribution on the circle is a grating with identical intensity at every third mode ($m = 0, 3, 6, \ldots$), and zero otherwise, which causes degeneracy between lower ($m - 3$) and higher (m) trigonometric modes. All corresponding to the zero-th mode effectively, this is consistent with the notion that a uniform distribution in the staggered rotamers should produce precisely that mean J coupling constant that is signified by the respective Karplus coefficient C_0 for the coupling type in question.

The intricacies just outlined suggest that there are certain limitations to the staggered-rotamer model, and that an iterative approach might necessitate special precautions (see "Appendix").

Results

Transformation between discrete and continuous torsion-angle distributions shall be exemplified by using ^3J coupling data measured by Xu et al. (1992) for the side-chain torsions χ_1 in recombinant human FK506-binding protein (FKBP) complexed with ascomycin. Comparison torsion angles were obtained from crystal coordinates of the FKBP/FK506 complex (PDB accession code 1FKF) resolved at 1.7 Å (Van Duyne et al. 1991).

Pairs of $^3J_{\mathrm{H}\alpha,\mathrm{H}\beta}$ and $^3J_{\mathrm{N}',\mathrm{H}\beta}$ coupling constants available for 31 out of the total 107 amino acids in FKBP permitted solving for the respective χ_1 torsion angles. Entirely independent of reference conformations, such as dihedral angles derived from X-ray data, both torsion angles for all included amino-acid residues as well as Karplus coefficients for all included J-coupling types were iteratively adjusted by applying a self-consistency approach that exploits parameter over-determination (Schmidt et al. 1999; Pérez et al. 2001). Best illustrated by an n-by-m spreadsheet of Karplus functions (Eq. 28), built from n residue entries and m coupling data per residue, the variable torsion-angle parameters and the variable Karplus coefficients would act on *row* and *column* elements, respectively, in order to optimally reproduce the constant measured values of ^3J.

By optimizing all variables at once, potential bias due to setting parameters in advance is avoided. One such factor in the J coupling analysis is the choice of Karplus coefficients, which has previously been demonstrated critically to affect the interpretation of ^3J coupling constants in terms of torsion angles and possible inferences regarding angular mobility (Karimi-Nejad et al. 1994; Blümel et al. 1998). Fixing Karplus coefficients often results in elevated order parameters making torsion-angle ranges appear more concentrated. Only unrestricted simultaneous fitting of all relevant variables allows adequately to apportion the residual fit error between those contributions from Karplus coefficients, those from torsion-angle dynamics, and those due to truly random uncertainty.

As side-chain related coupling constants are affected by variation of the local topology across different amino-acid types, the self-consistent optimization was supplemented by amino-acid specific Karplus curve increments, ΔC_0, in addition to the principal angular dependencies of $^3J_{\mathrm{H}\alpha,\mathrm{H}\beta}$ and $^3J_{\mathrm{N}',\mathrm{H}\beta}$ given by the fundamental coefficients, C_m (Schmidt 2007a).

The 31 side-chain conformations were first subjected to a single-rotamer search, in which three-quarters of the torsions unambiguously converged in the vicinity of an ideal staggered state, in agreement with angle values seen in the X-ray comparison set. Possibly due to ignoring angular dynamics at this stage, the remaining quarter of residues exhibited distorted angles near energetically less plausible eclipsed states. For convenient comparison, results from all models are visualised as dial graphs in Fig. 3, detailed parameters are provided as Supplementary Material.

The modelling of torsion-angle dynamics, however, required additional variables be fit as follows: (i) the *discrete* three-site jump employed two independent probability parameters per residue (see "Appendix"); and (ii) the *continuous* distribution function was generically fitted at the level of mean direction and order parameter, from which circular standard deviations for both Gaussian and elliptic models were ultimately derived by calculation. Thus, each approach involved 2×31 model specific mobility parameters, in addition to the 6 fundamental and 13 incremental Karplus coefficients. All parameters were iterated while continually calculating coupling constants and comparing these with the 124 experimental values of $^3J_{\mathrm{H}\alpha,\mathrm{H}\beta}$ and $^3J_{\mathrm{N}',\mathrm{H}\beta}$. In a complementary manner, parameters obtained from optimization according to one approach were eventually transformed to those corresponding to the other approach.

Approach (i): Mean direction and dispersion parameters from staggered-rotamer probabilities

Probabilities P_{g^+} and P_{g^-} were obtained numerically for the discrete three-point distribution model and, using Eqs. 22 and 23, respectively, were converted to apparent mean directions, θ_o, and circular order parameters, R. Applying Eqs. 9 and 15, the order parameters were subsequently transformed into mean angular deviations in the context of the continuous distribution models introduced. The transformed parameters displayed in Fig. 3 (right-hand-side) suggest extensive side-chain mobility, linked to comparatively low R values between 0.692 and 0.185 (Supplementary Material).

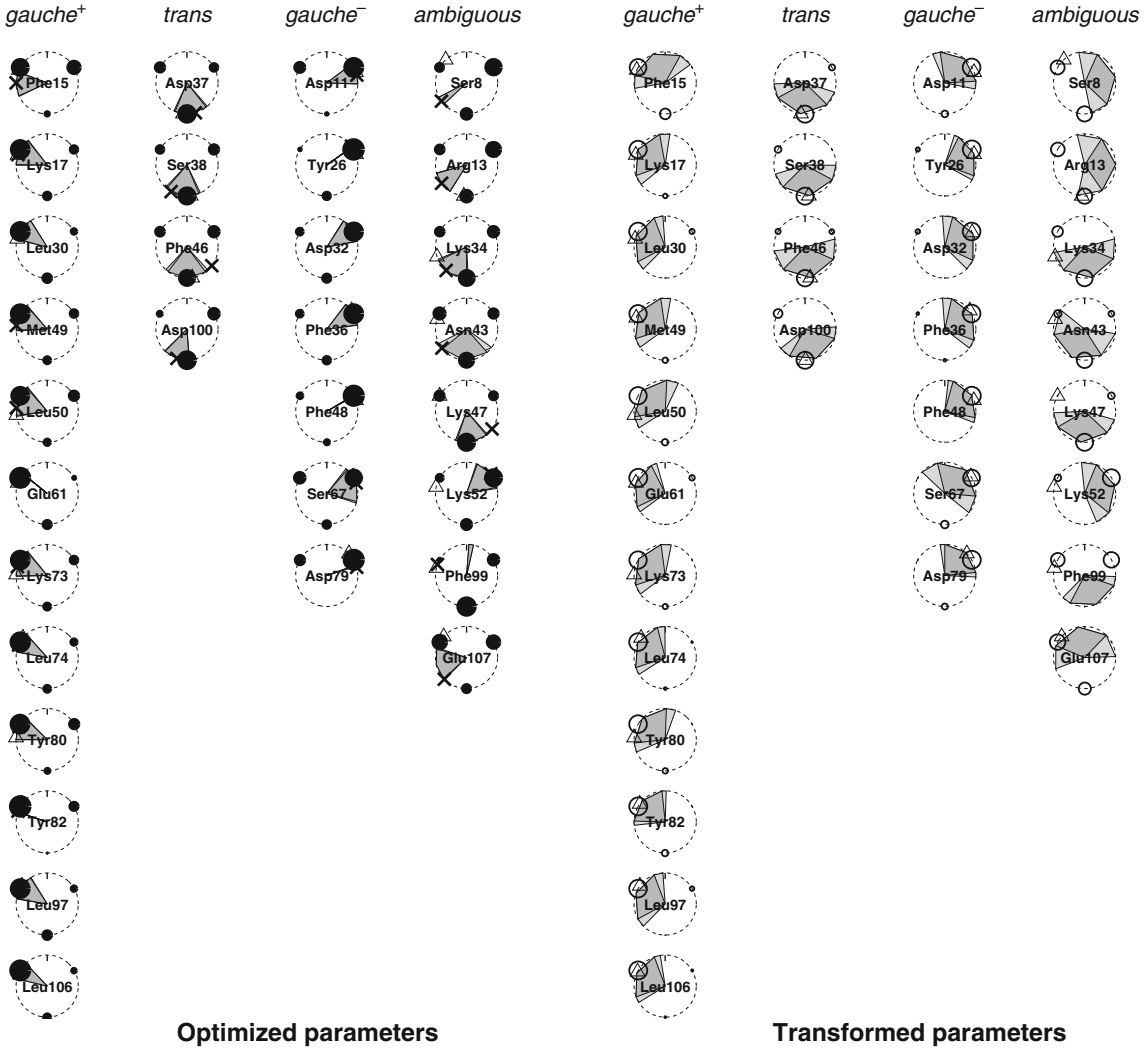

Fig. 3 Dial graphs showing all results from fitting various mobility models to ^3J coupling constants for selected χ_1 torsion angles in FKBP. Residues are grouped according to their predominant staggered rotamer as indicated, or conformations are considered ambiguous in case of divergent results. Tick marks at the top of each dial denote the angle origin, 0°. (*left*) For each amino acid, *crosses* show the iteratively optimized torsion angle assuming a single rotamer state. *Closed circles* represent in area-proportional manner the fitted probabilities P_{g^+}, P_t, and P_{g^-} of the three discrete staggered rotamers. Heavier and lighter *shaded dial segments* indicate, in contexts of elliptic and Gaussian models, respectively, the fitted continuous probability distributions with circular direction and dispersion. *Triangles* indicate comparison torsion angles derived from crystal structure coordinates. (*right*) Model parameters obtained by transformation of the optimized parameters: *Open circles* indicate transformed staggered-rotamer probabilities from continuous-model mean torsions and order parameters. *Shaded segments* show corresponding elliptic/Gaussian continuous model parameters as obtained from transforming fitted staggered-rotamer probabilities. Single-rotamer results cannot be transformed. Most noticeably, the fitted discrete staggered-rotamer probabilities (on the *left*) generally transform into wider apparent circular dispersions (on the *right*) than are being obtained when fitting the continuous model directly. Conversely, the narrower fitted circular dispersions (on the *left*) transform into more pronounced probabilities for the staggered main rotamer (on the *right*)

Large (transformed) standard deviations hint at considerable 'leakage of order' (disorder) in which the two minor rotamer states are both significantly populated. For example, the predominant *trans* state in Phe46 is only half populated, while both *gauche* states are populated roughly equally at around one-quarter each and, as a consequence, σ_M equals 94°. The most mobile side chains in Asn43, Arg13, and the C-terminal residue Glu107, all partaking in the group labelled 'ambiguous' in Fig. 3, delivered values of R so small that the transformed Gaussian dispersion parameters resulted in remarkably high values σ_M of up to 105°. Mobility parameters calculated following Batschelet's transformation (Eq. 5) indicate similar trends, except that values of σ_B (not shown) are generally smaller than Gaussian standard deviations σ_M, yet larger still than the circular standard deviations σ_o promoted in this investigation.

In essence, populating all three rotamer states gives rise to excessive apparent circular dispersion parameters. This does not, however, place constraints on the apparent mean direction. For example, as anticipated from the symmetrical disposition of both minor state populations in Phe46, the resultant transformed mean direction came out near $\theta_o = 180°$, in agreement with the *trans* state seen in the crystal structure.

If, however, precisely one rotamer state remains unpopulated, the mean direction becomes limited in its accessible range and, invariably, distorted torsions emerge that must deviate from ideal staggered states. The effect shall be demonstrated by example. Tyr82 exhibited the highest order parameter and appeared to be the most rigid of those side chains included in the analysis. Its fitted *trans* state is practically absent (2 %), and the second minor state converged at a lowly 20 % *gauche*$^+$ population, such that the predominant 78 % *gauche*$^-$ population would lend to an interpretation that this χ_1 torsion preferentially dwells in the $-60°$ state. Virtually zero population was also fitted for the *trans* rotamer in Asp79, except the permuted roles of *gauche*$^+$ and *gauche*$^-$ make the $+60°$ state pre-eminent in Asp79. Importantly, however, the transformed mean directions for both residues must deviate somewhat from the ideal $-60°$ or $+60°$ angle value, because a minor staggered component would not otherwise arise. Indeed, mean angles calculated from the probabilities as $-47°$ (Tyr82) and $+42°$ (Asp79), respectively, agree well with the slightly twisted torsion angles of $-55°$ and $+45°$ seen in the crystal structure.

Surprising on another note is that even for these residues perceived as rotationally rigid, the calculated dispersion parameters were as high as 40° to 50°. Values exceeding 60° are perceived as extensive rotational averaging, implying significant population of all three staggered rotamers (Schmidt 1997). It is unlikely that all side chains in FKBP should be this flexible. Clearly, although the transformed mean directions look acceptable, the dispersion values emerging from fitting the staggered-rotamer model appear exaggerated, due likely to artificially attenuated order parameters.

Approach (ii): Staggered-rotamer probabilities from mean direction and dispersion parameters

In contrast, dispersion parameters obtained from fitting the continuous model ranged only from virtually 0° (Tyr82) up to 58.3° (Asn43), qualitatively consistent with the line-up found with approach (i), yet considerably smaller in value. With order parameters significantly higher between 1 and 0.596, these dispersion figures appear more plausible than the elevated values of σ derived from the fitted staggered-state model, and mean directions are more frequently in accord with ideal staggered states also (Fig. 3; Supplementary Material).

The transformed rotamer-state populations as calculated, using Eq. 24, from the continuous-model parameters for the most rigid Tyr82 side chain qualitatively agreed with results of approach (i), except for the emergence of spurious negative probabilities (Fig. 4) which are being discussed in connection with skew effects below.

Contrasting approach (i), which implied complete angular revolution for the most flexible Asn43 side chain, fitting the continuous model resulted in a preference for the 180° state, with suitably—instead of excessively—high circular deviation (Fig. 3). Interestingly, the comparison X-ray structure indicates $\chi_1 = -71°$ for conformationally 'ambiguous' residue Asn43.

Similar to approach (i), previously highlighted Phe46 exhibited one predominant rotamer and two equally populated subordinate states also in the continuous-model optimization followed by parameter transformation. This time, however, the distribution shifted significantly in favour of the main conformer (84 %), and the minor state populations decreased to 8 % each. The higher order parameter of 0.76 and smaller circular standard deviations at around 40° with either Gaussian or elliptic model, compared with values from approach (i) of 0.26 and 55°–94°, respectively, look again more plausible than the complete angle revolution implied by the discrete model fitted previously.

Error considerations

Errors affecting the fit results are linked, first, to the reliability of the experimental *J* coupling data and, second, to their propagation into the fitted model parameters.

The precision of the experimental *J* coupling constants was apparently limited by the digital spectral resolution Xu et al. (1992) applied in their procedure of shifting selected E.COSY multiplet traces against one another. Matching the quoted 1.0 and 0.6 Hz per point for $^3J_{H\alpha,H\beta}$ and $^3J_{N',H\beta}$, respectively, a uniform standard error of ± 0.5 Hz was therefore applied to normalize ε_J^2, the *J* residual, in all angle-model optimizations. The actual statistical error in the data is likely smaller, because a correspondingly different coupling constant would have been found, should a trace have been mis-aligned by just over half a data point and subsequently subjected to re-evaluation. This provides reason to assume that the one-sigma interval of 0.5 Hz applied to the *J* data is, in fact, to be considered an upper error limit rather. It is remarked that tightening/relaxing the uniform standard error would merely increase/decrease the value of ε_J^2, that drives the minimization algorithm, but would not affect the best-fit parameter set as such.

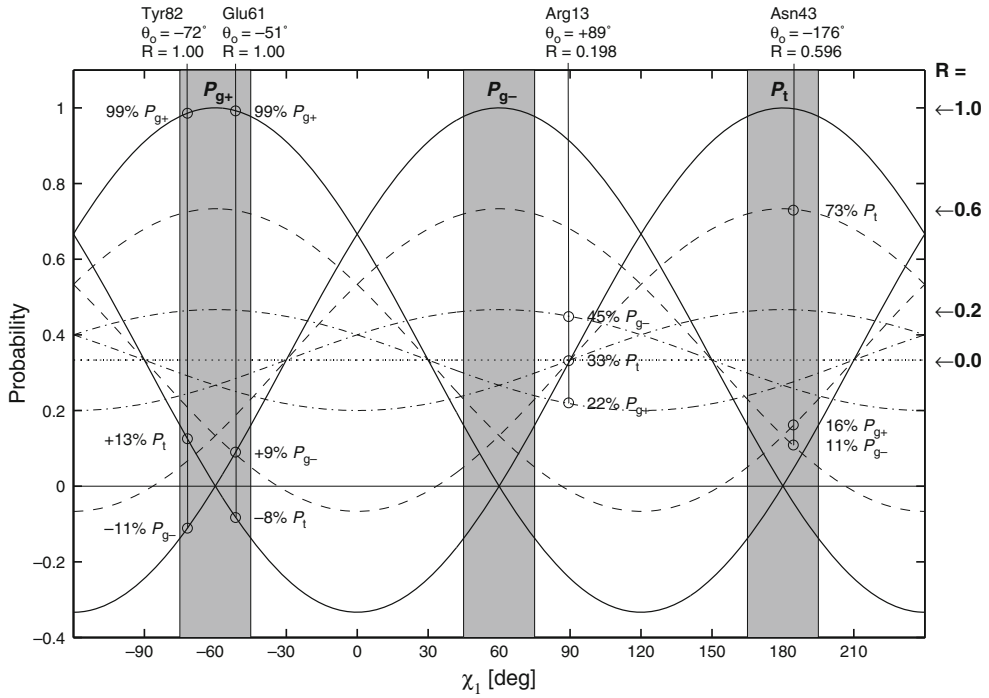

Fig. 4 Populations of the discrete staggered rotamers as calculated for arbitrary mean torsion angles and for selected values of the order parameter using Eq. 24. Mean torsion angles deviating from ideal staggered states produce differential probability in the two minor conformers. For example, for Tyr82, the continuous model fits order parameter 1 and mean χ_1 angle −72°, away from ideal staggered −60°, causing the apparent (transformed) minor populations to deviate from 0 % toward 13 % for P_t, and toward −11 % for P_{g^-}, whilst maintaining the main-rotamer likelihood P_{g^+} at 99 % (also see Supplementary Material). Roles are reversed for the minor populations in Glu61 whose χ_1 deviates toward the other side of −60°. Similar applies with permuted indeces to situations with mean χ_1 in the ±180° or +60° bands. Reduced order parameters equalize the populations, as example Asn43 shows. If $R < 0.5$, negative probabilities no longer emerge, regardless of mean angle. *Arrows* and differing *line styles* indicate the effect of increasing the amplitude of libration about the mean angle, as decreasing order parameters manifest in an associated decrease in the main-rotamer population, while the pivot probability for the two minor populations rises, until all probabilities reach 1/3 eventually in the extreme of complete rotatory averaging. The graph can be used in reversed manner, for example, the three fitted staggered-rotamer populations equate to a low order parameter in highly flexible Arg13 and an apparent (transformed) circular mean direction of +89°

Error propagation affecting the J coupling analysis was previously investigated by employing variance–covariance matrix methods (Schmidt et al. 1999), which are not always suitable for highly non-linear systems. The more robust alternative approach applied here estimated the output fit-parameter variation by repeat optimizations whilst adding to all input J coupling data uncorrelated random Gaussian noise of zero mean and 0.25-Hz standard deviation (Box and Muller 1958). The tighter interval was chosen to contain 95 % of the variation within one nominal J standard error. Results from 31 repetitions were collected and variation around the optimized unmodified (noise-free) set averaged. Averages and error margins for fundamental Karplus coefficients, which scale directly with the input J data, typically reproduced within the one standard-error interval in J, i.e., $C_0 = 6.10 \pm 0.10$ Hz, $C_1 = -0.92 \pm 0.23$ Hz, and $C_2 = 4.57 \pm 0.14$ Hz for $^3J_{H\alpha,H\beta}$, and $C_0 = 2.25 \pm 0.05$ Hz, $C_1 = -1.34 \pm 0.14$ Hz, and $C_2 = 1.01 \pm 0.17$ Hz for $^3J_{N',H\beta}$ couplings. Discarding the occasional record where a torsion angle converged at a different conformational well, torsion-angle parameters of the continuous model fluctuated by less than ±3°. Disregarding their intrinsic non-linearity, order parameters appeared to vary by ±0.06 around the best-fit set. Finally, the probability parameters of the discrete model reproduced within ±2 %.

Characteristic of self-consistent modelling is that discrepancies in the torsion angles ($RMSD_\theta$), which are merely by comparison, do not correlate with violations of J constraints ($RMSD_J$). Excluding those residues showing the largest torsion discrepancies is insignificant for improving the overall fit to the J coupling data and, vice versa, omitting data causing the largest coupling discrepancies is unlikely to give a better match with comparison geometries.

Fitting single-state, discrete staggered-rotamer, and continuous distribution models accomplished normalized ε_J^2 residuals of 215.6, 194.0, and 89.5 Hz^2, respectively, with $RMSD_J$ violations of J coupling constraints of 0.66, 0.63, and 0.42 Hz. Clearly, the continuous distribution

model adapted best to the experimental constraints, keeping the average deviate within the 0.5-Hz error interval. Considering that the single-state model comprises fewer variables, both discrete models fared similarly in terms of normalized error.

Although the model fitting does not reference any X-ray derived torsion angles, comparison with angles in the crystal structure revealed $RMSD_\theta$ discrepancies of 49.8°, 57.0°, and 48.0°, where the value for the staggered rotamers was calculated using transformed apparent mean torsion values. $RMSD_\theta$ numbers always look larger for protein side-chain torsions than for an analogous evaluation of main-chain angles (Schmidt et al. 1999), because of a certain likelihood for side-chain torsions to lock into different conformations under solution and solid state conditions. Larger angle deviations stem from those side chains labelled 'ambiguous' in Fig. 3, at times yielding conflicting torsion-angle conformations. For the three clear-cut staggered-conformation groups, stereospecific H^β chemical-shift assignments adjusted on-the-fly reliably reproduced those given by Xu et al. (1992). Fluctuating permutations were seen in repeat iterations only for the 'ambiguous' group (as marked in the Supplementary Material), explaining some of the angle discrepancy with the comparison crystal structure. The numbers corroborate that the continuous distribution model reproduced the comparison angles better, even though this was not a primary aim of this analysis.

Discussion

The adaptive nature of self-consistent parameter fitting, one may argue, would impede comparison of different models of angular motion as the effect of changing the model may be concealed by the many variables present. However, all models investigated employed the same number of fit variables for modelling torsion-angle mean and dispersion, i.e., two per residue. Other things being equal, any bias or discrepancy perceived must be attributed to differences in the concepts underlying the models.

Model comparison

One fundamental difference between discrete and continuous angular-distribution models concerns their performance and interpretation with respect to conformation averaging.

In the normal distribution, the *average* conformation is—by definition—the *most probable* state. Viewed through 'Gaussian lenses', the mean torsion angle is meant truly to exist as one of many states with finite, non-zero probability—or lifetime, for that matter.

In the discrete model of averages over distinct states, however, the mean conformation typically is *not*—actually *cannot at all* be—represented by any one of the constituent states alone and, therefore, must have vanishing probability or lifetime. It is imperative here to realise that, conversely, if a single constituent state represented the average structure already, then a superposition of multiple states would be utterly obsolete. Consequently: *When averaging discrete states, each constituent must deviate to some degree from the mean*. Seen in the light of the discrete model, the mean torsion angle is 'hidden from view' and meant to be non-existent.

Choosing the discrete over the continuous model, the question then arises as to how many distinct ensemble constituents would be required for satisfactory explanation of the data (Blackledge et al. 1993). There is no limit on this number, however, the simpler the model, the higher its significance.

Multiple conformations

Whenever coupling constants disagree with the most simplistic and straightforward single-conformer model, equilibria between at least two conformers need be considered. Unless exchange rates are sufficiently low to give rise to distinct spectral subsets, ensembles must be treated simply as rapidly interconverting conformations. Accounting for such effects is the purpose of fitting, in addition to the most probable torsion angle, a somewhat abstract order parameter. Its general interpretation is the extent of molecular mobility, albeit without a chance of resolving any individual ensemble members or pinning down the timescale on which conformational exchange processes take place.

Amino-acid side chains often exhibit restricted motion, for example, exchange between only two rather than all three staggered states. Fitting continuous models in these situations often manifests in elevated dispersion parameters, and occasionally in the artefactual inversion of the circular mean direction. This is a consequence of applying a unimodal model, which the continuous models typically are, to a multimodal distribution, such as the exchange between two clearly defined distinct conformations.

Owing to its predefined fixed angle values, the staggered-rotamer model is notoriously inadequate in situations in which the torsion is topologically restricted to non-staggered states. Proline is a pertinent example where staggered states cannot be adequately fit. Yet, this model is satisfactory, especially, in cases for which near-eclipsed mean χ_1 angles of narrow spread emerge with the Gaussian model, even though the ^3J data compellingly point at a staggered conformation (e.g., Ser8 and Phe99 in Fig. 3).

Skew-angle effects

Among those residues labelled 'ambiguous' in Fig. 3, the Lys34 side-chain χ_1 torsion adopts in the crystal structure a near-eclipsed, possibly conformation averaged value of $-108.2°$. Bearing in mind that conformation averaging as recorded in a crystal-structure dataset may differ in origin from that observed in solution, the former likely involves distinct conformations held by distinct molecules, each of which can be of a comparatively static texture, whereas in solution, the differing conformations are being traversed by one and the same molecule. To this extent, conformation averaging over molecules placed at distinct crystal lattice positions denotes 'space averaging', whereas flexible molecules tumbling in solution give rise to 'time averaging'. Whichever way, the resultant observable disorder causes dispersion in the angular distribution. The challenge is to work out the relative proportions of the contributing conformers. The NMR data for said Lys34 (Xu et al. 1992) show relative coupling-constant pairs in qualitative agreement with a single principal rotamer, which can only be the g_2t_3 or *trans* state. And yet, angle values converging away from the ideal $\pm 180°$ staggered conformer, depending on the applied model, hint at some 'skew' torsion conformation in solution.

Studies on rotamericity (Schrauber et al. 1993; Carugo and Argos 1997; MacArthur and Thornton 1999) found that a substantial proportion of amino-acid side chains in protein structures deviate by more than 20° from ideal staggered conformation, and skewness even tends to rise with increased resolution of the crystal structure. Distorted conformations might, in fact, be due to subtle averaging effects over otherwise near-ideal torsion rotamers, rather than long-lived skew-torsion-angle states, and the lowly populated rotamers possibly become observable only with increasing data resolution, as suggested initially.

While the Gaussian model accommodates skew simply by shifting the location of the distribution, the effect cannot however be readily reproduced by a discrete model that does not exhibit density at the angle of the twist-away torsion. Possibilities for adjusting the location of the distribution are greatly reduced with a model limited to fixed angle values. Even though the Pachler model does not preclude the mean location from pointing in an arbitrary direction (Table 1), this comes at the price of high angular disorder (small value of R), or high circular variance (large value of σ), for that matter.

Artificial superposition of staggered states to emulate skew effects in the Pachler model invariably results in *differential probability* for the two minor states, that is, a non-zero positive probability contribution from one of the minor states is being compensated for by an almost, yet not exactly, equal probability of opposite sign for the other minor state (Fig. 4). Differential probability between those minor states depends on the distance the skew torsion angle is away from the nearest staggered state. Compare, for example, Asp79 at $\chi_1 = +71°$ with Phe48 at near-ideal $+60°$, both of which exhibit the same low circular standard deviation and an order parameter near 1 (Fig. 3; Supplemantary Material). The probability difference between the minor states is negligible for Phe48, whereas that for Asp79 amounts to more than ± 10 % at the given 11° deviation in the torsion angle. Notice how the relative populations of the minor states reverse signs if the mean angle deviates from the ideal value in the opposite direction, for example, in the case of Tyr26 with $\chi_1 = +55°$. At any rate, a negative state probability encountered in the Pachler-type analysis of rotamer equilibria is a somewhat artefactual, model-inherent expression of an underlying skew (non-staggered) torsion angle with comparatively little dispersion (Fig. 4).

Pachler model

A distribution mean located away from an ideal staggered angle commands non-zero probability not only for the predominant rotamer but also for one or both of the minor rotameric states. Consequently, that proportion of probability which is scattered over all three rotamers corresponds to a uniform baseline distribution on the circle, causing the circular order parameter R to decrease noticeably. Any spurious negative probability that may arise in the Pachler-type analysis (Fig. 4) contributes as much to increased disorder as does a positive-only set of p_i parameters. If negative probabilities are not acceptable, an approach to fitting range-bound all-positive probability parameters is devised in the "Appendix". However, forcing probabilities to remain positive has the adverse effect of lowering the order parameter, thereby artificially exaggerating the extent of circular mobility. A high value of R can only be maintained in the staggered-rotamer model if all but one probabilities remain near zero, which is often impossible in practice and also would make a three-state fit obsolete. As a consequence, the Pachler model tends to emphasise the second moment or dispersion of the angular distribution, and underestimates its first moment, resulting typically in low order parameters and comparatively higher angular dispersion.

Gaussian model

In the Gaussian distribution model, mean angle and angular dispersion, as the first and second distribution moments, respectively, are orthogonal parameters, permitting—in theory—the perfect separation of effects. Both location and dispersion parameters are somewhat coupled though, namely by the actual set of experimental data constraints, which themselves may not be entirely independent observations.

The possibility of adjusting the mean to the optimal location, at any granularity, primarily ensures angular variance is kept at minimum. The residual fit error is subsequently reduced by adjusting the spread parameter. As a consequence, the Gaussian model tends to emphasise the first moment of the angular distribution, while second-moment effects appear underweight, resulting in high apparent order parameters and comparatively lower angular dispersion.

Elliptic model

The involvement of both first and second moments of the distribution according to the elliptic model of torsion-angle mobility by using the circular statistic suggested in this work (Eq. 12), is a consequence of averaging on a par both sine and cosine components of a torsion angle. Rather than a single value as in linear statistics, two equally ranked descriptors of direction are averaged for each torsion. Thus, measures of direction, θ_o, and concentration, R, are concurrently obtained, with no preference for one or other. Both elliptic as well as Gaussian models yield identical mean directions and differ only in the dispersion part. A reassuring aspect of the elliptic-model approach to obtaining circular variance on the basis of moments is that the dispersion measure, derived as the complementary order parameter $(1 - R^2)$, embodies some *squared* property, which meets with general concepts of variance and co-variance.

Conclusions

New improved approaches to the measurement of molecular structure data by NMR spectroscopy increasingly reveal ambiguities in the analysis of molecular conformation that are not always caused by lack of data or accuracy or precision in the procedures (Schmidt and Löhr 2012). Data interpretation then needs to consider conformational variability that is not necessarily resolved by collecting additional experimental data. Instead, development is encouraged of models, and an understanding of molecular structure, that integrate both dynamic as well as static aspects of conformation. This seems true, especially, for research into challenging subjects, such as partially or intrinsically disordered protein structures (Dyson and Wright 1998; Hennig et al. 1999; Vajpai et al. 2010).

The present work investigated and compared different concepts of characterising torsion-angle mobility as derived from ^3J coupling constants. Given the same number of fitted variables, the Gaussian model frequently performs statistically better than the staggered-rotamer model (Schmidt 1997; Pérez et al. 2001). It was one aim of this work to put those models through rigorous inspection and to give this perception a sound basis. As the actual molecular motion process does not depend on the model used for data analysis, one would expect from applying different but equally complex models to obtain similar and consistent statements about that molecular process. After all, the facts about the molecular process lie entirely and exclusively in the observable experimental data. Even so, application of those models to real data conveys performance differences that must be attributed to the conceptual differences between the models, originating, amongst other reasons, from the highly non-linear character of the Karplus relationship. Differing results from applying different angle-mobility models have been observed and were investigated previously, for example, by Karimi-Nejad et al. (1994) and Schmidt (1997), and therefore were expected to be seen also with the present J coupling data. As often only one model is being applied in common practice, differences would not arise or apparent inconsistencies would go unnoticed.

Mapping between *circular* variance defined on the interval $[0, 2\pi]$ or $[-\pi, +\pi]$ and *linear* variance defined on the interval $[-\infty, +\infty]$ is best accomplished, regardless of the concept chosen, by employing a circular order parameter, R. Resembling the correlation coefficient in linear statistics, the order parameter is oftentimes more accessible to interpretation than the variance. Furthermore, fitting of the model-independent parameter, R, rather than the model-dependent standard deviation, σ, offers the advantage of permitting to associate σ with a particular model later.

When applied to ^3J coupling based analysis of amino-acid side-chain χ_1 torsion angles, common measures of standard deviation in circular data were found to yield differing and often too large values, especially in cases where angle distributions are notionally narrow. This is, in part, a consequence of the limitation in the Karplus model of ^3J to only two circular modes and, thus, due to the ignorance of higher moments that would help describe skewness and kurtosis of a potentially multimodal circular distribution. The primary reason for the different performances lies, however, in the way the various models separate the properties related to aspects of first and second moments of the circular distribution. This does not mean that one or other is less sensible or realistic, as they are only models. In fact, the Gaussian mean angle and the most populated staggered rotamer often agree. However, the need arises for a model of circular statistics that offers a compromise between both aspects and places equal emphasis on both moments. The elliptic model suggested in the present investigation seems best positioned to fulfil these requirements.

Appendix: Fitting range-bound probability parameters

The following devises a method for optimizing sets of interdependent probability parameters bounded on the interval [0, 1], so as to remain meaningful in their application to averaging distinct states.

Let the observed J coupling constant represent an average due to a weighted superposition of three point distributions in circular torsion-angle space, with fixed directions and associated coupling values given as

$$\langle J\rangle = p_1 J_1 + p_2 J_2 + p_3 J_3. \tag{35}$$

The three dihedral-angle states, may—but do not have to—coincide with the staggered-rotamer conformations, in which case the identities $p_1 = P_{g^+}$, $p_2 = P_t$, $p_3 = P_{g^-}$, and likewise for J, apply as follows,

$$\langle J\rangle = P_{g^+} J_{g^+} + P_t J_t + P_{g^-} J_{g^-}. \tag{36}$$

The normalising condition, $\Sigma_k p_k = 1$, constrains the value of the terminal probability such that only two independent probability parameters need be determined.

However, two similarly designed parameters p_1 and p_2 would be mathematically and numerically interchangeable and, importantly, would not adhere to the normalisation condition, resulting at times in spurious negative probabilities for p_3. It may therefore be desirable to contain the effective value range of each probability parameter p in an interval $[0, a]$, which is accomplished by parameter transformation employing the *logistic sigmoid*,

$$p = a(1 + e^{-p'})^{-1}. \tag{37}$$

Here, an external programme control variable p' can conveniently be maintained on the *unbounded* interval $[-\infty, +\infty]$. Change in the *constrained* variable p as a result of change in the *unconstrained* variable p' (notice the prime) is then expressed by the derivative

$$dp/dp' = p(a - p) \tag{38}$$

where the complementary probability $(a - p) = a(1 + e^{-p'})^{-1} e^{-p'}$. Applying the chain rule, the unconstrained probability parameter affects the model function (Eq. 35) according to $dJ/dp' = dJ/dp \times dp/dp'$.

Let $a = 1$. Ideally, the primary parameter p_1 would explore the whole interval [0, 1], while the secondary parameter p_2 would fall in the reduced interval $[0, 1 - p_1]$. Letting $p_2 = p_2(p_1)$ and $p_3 = 1 - p_1 - p_2$, the model of Eqs. 35 and 36 is re-defined to feature the sought auto-normalising property as follows,

$$\begin{aligned}\langle J\rangle &= p_1 J_{g^+} + p_2(1 - p_1) J_t + \{1 - p_1 - p_2(1 - p_1)\} J_{g^-} \\ &= p_1 \left(J_{g^+} - J_{g^-}\right) + p_2(1 - p_1)\left(J_t - J_{g^-}\right) + J_{g^-}\end{aligned} \tag{39}$$

The *actual* probability variables, signified in Eq. 36 by uppercase P_i, that need be established at some stage in the procedure are thus obtained from the constrained probability parameters as

$$\begin{aligned} P_{g^+} &= p_1 \\ P_t &= (1 - p_1) p_2 \\ P_{g^-} &= (1 - p_1)(1 - p_2) \end{aligned} \tag{40}$$

The appearance of product probabilities in Eq. 40 commands the re-definition of all derivatives with respect to each probability, such that the change in function value J with respect to *two unconstrained* probability parameters is expressed entirely in terms of *two constrained* probability variables according to

$$\begin{aligned} dJ/dp'_1 &= p_1(1 - p_1)\left(J_{g^+} - J_{g^-}\right) - p_1 p_2(1 - p_1)\left(J_t - J_{g^-}\right) \\ dJ/dp'_2 &= p_2(1 - p_1)(1 - p_2)\left(J_t - J_{g^-}\right). \end{aligned} \tag{41}$$

Expressed in terms of the *three actual* probabilities the derivatives read

$$\begin{aligned} dJ/dp'_1 &= P_{g^+}\left(1 - P_{g^+}\right)\left(J_{g^+} - J_{g^-}\right) - P_{g^+} P_t \left(J_t - J_{g^-}\right) \\ dJ/dp'_2 &= P_t P_{g^-}\left(1 - P_{g^+}\right)^{-1}\left(J_t - J_{g^-}\right) \end{aligned} \tag{42}$$

where identities of Eq. 40 were used and, resulting from these, $(1 - p_2) = P_{g^-}/(1 - P_{g^+})$. Written as

$$\begin{aligned} P_{g^+} &= \left(1 + e^{-p'_1}\right)^{-1} \\ P_t &= \left(1 - P_{g^+}\right)\left(1 + e^{-p'_2}\right)^{-1} \\ P_{g^-} &= 1 - P_{g^+} - P_t \end{aligned} \tag{43}$$

the re-defined parameter transformations, exempt of all auxiliary variables p_i, exclusively require the unconstrained parameters p'_1 and p'_2 be parsed to the optimizer engine.

It is noteworthy that the first derivative in Eqs. 41 and 42 involves two coupling differences between two pairs of staggered-rotamer states, whereas the second derivative involves only one such difference. This is consistent with the notion that one fewer degree of freedom is left once the first probability has been assigned. Assigning the second probability obviates the need for any further coupling information as the remaining third probability is fully determined.

The difference couplings are given solely by the phase offset associated with the respective coupling type

$$\left(J_{g^+} - J_{g^-}\right) = -(3)^{1/2}(C_1 \sin\Delta\theta + C_2 \sin 2\Delta\theta) \tag{44a}$$

$$\begin{aligned}\left(J_t - J_{g^-}\right) = &-(3/4)^{1/2}(C_1 \sin\Delta\theta + C_2 \sin 2\Delta\theta) \\ &- (3/2)(C_1 \cos\Delta\theta - C_2 \cos 2\Delta\theta)\end{aligned} \tag{44b}$$

where C_1 and C_2 are Karplus coefficients appropriate for the coupled pair of nuclei in the χ_1 torsion-angle topology considered, and $\Delta\theta$ is the phase increment to χ_1 associated with that particular coupled spin pair.

For symmetric Karplus curves, the *gauche* couplings J_{g^+} and J_{g^-} are identical, thereby effectively removing the first term from the derivative dJ/dp'_1. If asymmetric curves are to be considered that include a first-order sine component (Schmidt 2007b), the above Eqs. 44a and 44b are extended by the terms $-(3)^{1/2}\ S_1\ \cos\ \Delta\theta$ and $-(3/4)^{1/2}\ S_1\ \cos\ \Delta\theta + (3/2)\ S_1\ \sin\ \Delta\theta$, respectively.

References

Abramowitz M, Stegun IA (1972) Handbook of mathematical functions with formulas, graphs, and mathematical tables. In: 20th repr. National Bureau of Standards, Applied Mathematics Series 55, Washington

Avbelj F, Baldwin RL (2003) Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi. Proc Natl Acad Sci USA 100:5742–5747

Batschelet E (1965) Statistical methods for the analysis of problems in animal orientation and certain biological rhythms. American Institute of Biological Sciences, Washington

Batschelet E (1981) Circular statistics in biology. Academic Press, London

Blackledge MJ, Brüschweiler R, Griesinger C, Schmidt JM, Xu P, Ernst RR (1993) Conformational backbone dynamics of the cyclic decapeptide antamanide. Application of a new multiconformational search algorithm based on NMR data. Biochemistry 32:10960–10974

Blümel M, Schmidt JM, Löhr F, Rüterjans H (1998) Quantitative ϕ torsion angle analysis in *Desulfovibrio vulgaris* flavodoxin based on six ϕ related ^3J couplings. Eur Biophys J 27:321–334

Box GEP, Muller ME (1958) A note on the generation of random normal deviates. Ann Math Stat 29:610–611

Bracewell RN (1986) The Fourier transform and its applications, 2nd Intl Edn. McGraw-Hill, New York

Brüschweiler R, Case DA (1994) Adding harmonic motion to the Karplus relation for spin–spin coupling. J Am Chem Soc 116: 11199–11200

Carugo O, Argos P (1997) Correlation between side chain mobility and conformation in protein structures. Prot Eng 10:777–787

Dyson HJ, Wright PE (1998) Equilibrium NMR studies of unfolded and partially folded proteins. Nature Structural Biology, Supplement, vol 5, pp 499–503, ISSN 1072-8368

Džakula Ž, Westler WM, Edison A, Markley JL (1992a) The CUPID method for calculating the continuous probability distribution of rotamers from NMR data. J Am Chem Soc 114:6195–6199

Džakula Ž, Edison A, Westler WM, Markley JL (1992b) Analysis of χ_1 rotamer populations from NMR data by the CUPID method. J Am Chem Soc 114:6200–6207

Fernández-Durán JJ (2004) Circular distributions based on nonnegative trigonometric sums. Biometrics 60:499–503

Fisher NI (1993) Statistical analysis of circular data. Cambridge University Press, Cambridge

Hansen PE, Feeney J, Roberts GCK (1975) Long range ^{13}C-^{1}H spin–spin coupling constants in amino acids. Conformational applications. J Magn Reson 17:249–261

Hennig M, Bermel W, Spencer A, Dobson CM, Smith LJ, Schwalbe H (1999) Side-chain conformations in an unfolded protein: χ_1 distributions in denatured hen lysozyme determined by heteronuclear ^{13}C, ^{15}N NMR spectroscopy. J Mol Biol 288:705–723

Hoch JC, Dobson CM, Karplus M (1985) Vicinal coupling constants and protein dynamics. Biochemistry 24:3831–3841

Hyberts SG, Goldberg MS, Havel TF, Wagner G (1992) The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. Protein Sci 1:736–751

IUPAC-IUB Commision on Biochemical Nomenclature (1970) J Mol Biol 52:1–17

Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics. Series on Multivariate Analysis vol 5. World Scientific, Singapore

Janin J, Wodak S, Levitt M, Maigret B (1978) Conformation of amino acid side-chains in proteins. J Mol Biol 125:357–386

Jardetzky O (1980) On the nature of molecular conformations inferred from high-resolution NMR. Biochem Biophys Acta 621:227–232

Jeffreys H (1961) Theory of probability, 3rd edn. Oxford University Press, Oxford

Karimi-Nejad Y, Schmidt JM, Rüterjans H, Schwalbe H, Griesinger C (1994) Conformation of valine side chains in ribonuclease T1 determined by NMR studies of homonuclear and heteronuclear ^3J coupling constants. Biochemistry 33:5481–5492

Karplus M (1963) Vicinal proton coupling in nuclear magnetic resonance. J Am Chem Soc 85:2870–2871

Lipari G, Szabo A (1982a) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104:4546–4559

Lipari G, Szabo A (1982b) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J Am Chem Soc 104:4559–4570

MacArthur MW, Thornton JM (1993) Conformational analysis of protein structures derived from NMR data. Prot Struct Funct Genet 17:232–251

MacArthur MW, Thornton JM (1999) Protein side-chain conformation: a systematic variation of χ_1 mean values with resolution—a consequence of multiple rotameric states. Acta Crystallogr D55:994–1004

Mardia KV (1972) Statistics of directional data. Academic Press, London

Nagayama K, Wüthrich K (1981) Systematic application of two-dimensional ^{1}H nuclear-magnetic-resonance techniques for studies of proteins. Eur J Biochem 115:653–657

Pachler KGR (1963) Nuclear magnetic resonance study of some α-amino acids—I. Coupling constants in alkaline and acidic medium. Spectrochim Acta 19:2085–2092

Pachler KGR (1964) Nuclear magnetic resonance study of some α-amino acids—II. Rotational isomerism. Spectrochim Acta 20:581–587

Pérez C, Löhr F, Rüterjans H, Schmidt JM (2001) Self-consistent Karplus parametrization of ^3J couplings depending on the polypeptide side-chain torsion χ_1. J Am Chem Soc 123:7081–7093

Polshakov VI, Frenkiel TA, Birdsall B, Soteriou A, Feeney J (1995) Determination of stereospecific assignments, torsion-angle constraints, and rotamer populations in proteins using the program AngleSearch. J Magn Reson B108:31–43

Schmidt JM (1997) Conformational equilibria in polypeptides. II. Dihedral-angle distribution in antamanide based on three-bond coupling information. J Magn Reson 124:310–322

Schmidt JM (2007a) A versatile component-coupling model to account for substituent effects. Application to polypeptide ϕ and χ_1 torsion related ^3J data. J Magn Reson 186:34–50

Schmidt JM (2007b) Asymmetric Karplus curves for the protein side-chain ^3J couplings. J Biomol NMR 37:287–301

Schmidt JM, Löhr F (2012) Refinement of protein tertiary structure by using spin–spin coupling constants from nuclear magnetic resonance measurements. In: Faraggi E (ed) Protein structure. Intech, Rijeka, ISBN 979-953-307-576-0. Available from http://www.intechopen.com/books/protein-structure

Schmidt JM, Blümel M, Löhr F, Rüterjans H (1999) Self-consistent ^3J coupling analysis for the joint calibration of Karplus coefficients and ϕ-torsion angles. J Biomol NMR 14:1–12

Schrauber H, Eisenhaber F, Argos P (1993) Rotamers: to be or not to be? An analysis of amino-acid side-chain conformations in globular proteins. J Mol Biol 230:592–612

Spiegel MR, Liu J (1999) Schaum's outline series: Mathematical Handbook of Formulas and Tables, 2nd edn. McGraw Hill, New York

Vajpai N, Gentner M, Huang J, Blackledge M, Grzesiek S (2010) Side-chain χ_1 conformations in urea-denatured ubiquitin and protein G from ^3J coupling constants and residual dipolar couplings. J Am Chem Soc 132:3196–3203

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1991) Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. Science 252:839–842

Von Mises R (1918) Über die Ganzzahligkeit der Atomgewichte und verwandte Fragen. Physikal Z 19:490–500

Xu RX, Olejniczak ET, Fesik SW (1992) Stereospecific assignments and χ_1 rotamers for FKBP when bound to ascomycin from $^3J_{H\alpha,H\beta}$ and $^3J_{N,H\beta}$ coupling constants. FEBS Lett 305:137–143