J Biomol NMR (2007) 37:287–301
DOI 10.1007/s10858-006-9140-8

ARTICLE

Asymmetric Karplus curves for the protein side-chain ^3J couplings

Jürgen M. Schmidt

Received: 8 September 2006 / Revised: 27 November 2006 / Accepted: 4 December 2006 / Published online: 27 February 2007

Abstract The standard Karplus equation for calculating ^3J coupling constants from any given dihedral angle requires three empirical coefficients be determined that relate to the magnitudes of three modes of the angle dependency of ^3J. Considering cosine modes only (bimodal, unimodal and baseline component), Karplus curves are generally symmetric with respect to the sign of the angle argument. Typically, their primary and secondary maxima differ in amplitude, whereas the two minima are of equal depth. However, chiral molecular topologies, such as those surrounding the-main-chain and side-chain torsions in amino-acid residues, preclude, as regards substituent positioning, exact mirror-image conformations from being formed—for any given torsion-angle value. It is therefore unlikely that ^3J couplings assume identical values for the corresponding positive and negative dihedral angles. This suggests that a better empirical fit of the torsion-angle dependency of ^3J could be obtained when removing the constraint of symmetrically identical coupling constants. A sine term added to the Karplus equation allows independent modelling of both curve minima typically located near dihedral-angle values of +90° and –90°. Revisiting an extensive ^3J coupling dataset previously recorded to determine the side-chain torsions χ_1 in the protein flavodoxin, the asymmetric Karplus model accomplishes a more accurate fit to the experimental data. Asymmetries revealed in the angle dependencies exceed the experimental precision in determining ^3J. Accounting for these effects helps improve molecular models.

Keywords Data redundancy · Flavodoxin · Optimisation · Self consistency · Torsion angle · Vicinal coupling constants

Electronic Supplementary Material The online version of this article (doi:10.1007/s10858-006-9140-8) contains supplementary material, which is available to authorized users.

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

Introduction

Accurate parameterisation of the dihedral-angle dependencies of ^3J coupling constants is critical to the NMR analysis of molecular geometry. Regarding protein-structure determination, investigations into the side-chain torsion χ_1 in amino-acid residues (Pérez et al. 2001) concluded that amino-acid specific coefficients inserted into the Karplus equations for calculating ^3J from any given torsion angle value (Karplus 1963) more accurately reflect the impact from the variety of substituent patterns present in different residue types than the conventional one-size-fits-all approach.

The commonly used three cosine-related coefficients (C_0, C_1 and C_2) bestow Karplus curves with a primary and a secondary maximum, and with two minima of equal depth. From the periodicity and symmetry, degenerate ^3J values ensue for dihedral-angle pairs symmetric with respect to both *trans* and *cis* orientations of the four-atom fragment in question, given by ±180° and 0°, respectively. However, as amino-acid topologies are typically chiral in nature, the mirror-image symmetry between ^3J values for positive and negative angle arguments in the Karplus equation may not always be justified in protein studies. It is likely,

rather, that chiral torsion topologies give rise to asymmetric minima of ^3J for dihedral angles around +90° and –90° (Fig. 1).

Previous attempts at optimising coupling-angle relationships indicated that the symmetry of the simple Karplus equation may have to be relinquished in the presence of substituent effects (Pachler 1972; Donders et al. 1989). Homonuclear vicinal J_{HH} coupling constants in substituted ethanes were found to vary asymmetrically with the torsion angle, depending on the relative position of the substituent (Pachler 1970, 1971). Even though asymmetries were barely 0.2 Hz in monofluoroethane and came out only slightly larger at 0.35 Hz in 1,2-difluoroethane, the substantially more complex situation in amino acids may give rise to significant effects. Indeed, investigations into substituent effects on ^3J in biomolecules initially focused on cyclic topologies like prolyl and ribosyl moieties. Empirical coefficients were established that depend on the electronegativity of the substituent as well as on their phase angle relative to the torsion in question, also known as the Haasnoot–Altona equations (Haasnoot et al. 1980, 1981a, b).

These early studies were limited to scant proton–proton coupling data only, to the effect that experimental information needed in the angle determination was quickly exhausted. We are nowadays in a position also to exploit the abundance of heteronuclear J coupling data brought about by isotopic enrichment. Unlike studies on proton–proton couplings, heteronuclear couplings involve substituents not only attached to the central atoms (2 and 3) but also to the terminal positions (1 and 4).

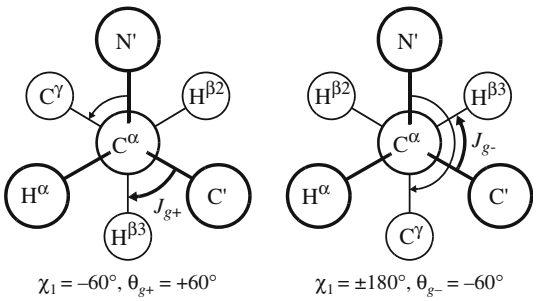

Fig. 1 Amino-acid side-chain χ_1 fragment viewed down the C^{α}–C^{β} bond demonstrating that mirror-image symmetry of ^3J coupling constants is not normally maintained due to asymmetric substituent patterns. The coupling $^3J(C',H^{\beta})$ is found to exhibit the largest asymmetry, exemplified here by *gauche* conformations g^+ (+60°) and g^- (–60°) for the fragment C'–C^{α}–C^{β}–$H^{\beta 3}$ (heavy arrows), encountered for nominal χ_1 angles of –60° and ±180°, respectively (light arrows). With respect to the carbonyl atom, the influence of the C^{γ} substituent on that coupling is from a *trans* position (left) and from a *gauche* position (right), respectively. Clearly, the two conformations are not mirror images of one another. Similar considerations apply to other substituent placements

Data redundancy arising from at least four coupling values per torsion angle can be tapped by using an unbiased self-consistency approach to parameter estimation (Schmidt et al. 1999), permitting the study of effects other than the basic angle-dependency of ^3J in addition. In order to account for a possible presence of asymmetric torsion-angle dependencies of polypeptide ^3J couplings, the standard Karplus model is extended here by a sine-related coefficient (S_1), and all coefficients are being re-optimised for the six χ_1 related ^3J coupling types encountered in polypeptides (Fig. 2).

Methods

Karplus (1963) suggested to collate a series of trigonometric modes, m, each individually scaled by

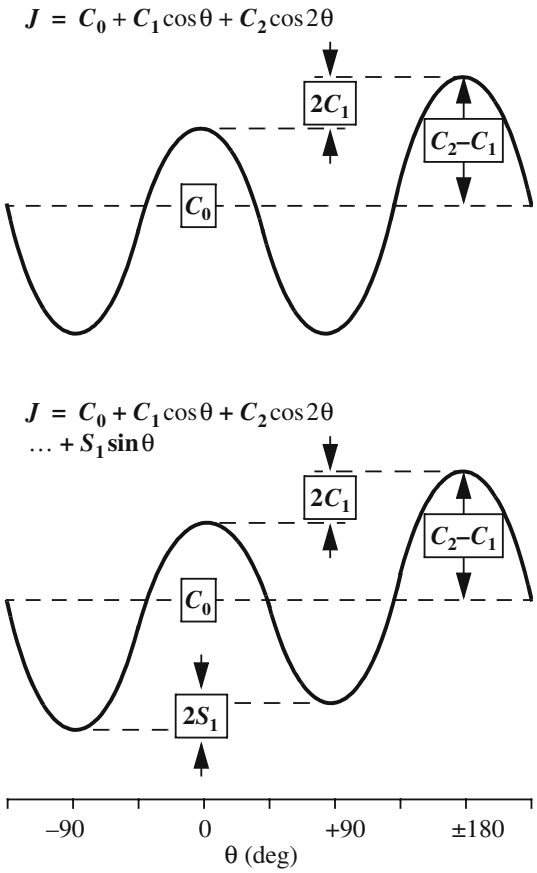

Fig. 2 Comparison of a regular Karplus curve (top) and one including an asymmetric sine mode (bottom). Interpretation of the coefficients identifies C_0 as the mean J value obtained upon complete torsion-angle revolution, (C_2–C_1) as the largest deflection in J from the mean, and *primary* and *secondary* maxima differ by $2C_1$. The fourth coefficient S_1 established in the present work denotes the difference between the minima

empirical coefficients, C_m, that eventually became known as Karplus parameters, their values being given in Hertz. The model resembles, though not originates from (Karplus 1959; Barfield and Karplus 1969), the real portion of a Fourier expansion in the dihedral angle θ

$$J(\theta) = \Sigma_m C_m \cos m\theta. \quad (1)$$

In ^3J analysis, m typically runs from 0 to 2, resulting in the superposition of only the first three symmetric modes, as depicted in Fig. 2A, the explicit equation being

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

Signs alternating between even and odd mode coefficients are typically obtained, due solely to the choice of the dihedral-angle origin (IUPAC-IUB 1970), stipulating that 0° be assigned to θ when the four-atom molecular fragment in question is in a *cis*-coplanar conformation. A negative coefficient C_1 ensures that the primary maximum of the Karplus curve related to the *trans*-coplanar conformation exceeds the value of the secondary maximum. Furthermore, values of $J(\theta)$ form mirror-image pairs between both semi-circles, $\theta = 0° \ldots +180°$ and $0° \ldots -180°$.

Asymmetry in the angle dependencies of ^3J must arise from at least two effects that would suggest different ways of modelling. These effects result from the presence of a particular substituent type and the actual positioning of that substituent on a given coupling path J(X,Y). Focusing on the amino-acid side-chain fragment X–C$^\alpha$–C$^\beta$–Y, and leaving aside glycine, the presence of three different substituents (N′, H$^\alpha$, C′) bonded in the same order to each *L*-amino-acid α-carbon constitutes a recurring pattern of asymmetry that could be modelled globally by adding a fourth term to the respective fundamental Karplus equations (Fig. 2B), such that

$$J(\theta) = C_0 + C_1 \cos\theta + C_2 \cos 2\theta + S_1 \sin\theta. \quad (3)$$

The positioning of substituents becomes an issue if the above sense of substituent ordering were selectively inverted, consider, for example, a sequence of *L*- and *R*-amino acids. Even though substituent types and numbers were identical, the pattern would no longer be recurring and asymmetry would need to be considered on a per-residue basis. To account for 'positive' and 'negative' substituents on a H–C–C–H coupling path, Haasnoot et al. (1980) introduced a sign convention, subsequently applied to sine-term coefficients by Donders et al. (1989). As proteins ordinarily consist of *L*-amino acids only, a component of global asymmetry according to Eq. 3 is implicated in the analysis of coupling constants ^3J(H$^\alpha$,H$^\beta$), ^3J(N′,H$^\beta$), ^3J(C′,H$^\beta$), ^3J(H$^\alpha$,C$^\gamma$), ^3J(N′,C$^\gamma$) and ^3J(C′,C$^\gamma$) related to the amino-acid side-chain torsion χ_1 by the mere presence of the chiral and, thus, asymmetric C$^\alpha$ substituent pattern on the coupling path, regardless of the C$^\beta$ substituent pattern.

The additional variability of C$^\beta$ substituents in amino acids results in a non-uniform distribution of substituents around the same torsion type in different amino acids. Asymmetry arising from different substituent types and numbers would need to be implemented in a residue-specific manner, or, in practice, be associated with substituent atoms. For example, as their C$^\beta$-bound substituents differ, ^3J(H$^\alpha$,C$^\gamma$) couplings in threonine and asparagine likely require distinct parameterisations. An extended Karplus model capable of accounting for specific substituents has been introduced by Pérez et al. (2001) as given by

$$J_{XY}(\theta) = C_0 + C_1 \cos\theta + C_2 \cos 2\theta + \Sigma\Delta J^R_{XY} \quad (4)$$

The sought asymmetry could then be modelled in an extended cumulative coupling increment due to all relevant substituents,

$$\Sigma\Delta J^R_{XY} = \gamma_{XY}\Sigma f_R(\Delta C^R_0 + \Delta S^R_1 \sin\theta), \quad (5)$$

where f_R accounts for the frequencies of like-substituent atoms, R, and γ_{XY} is the product gyromagnetic ratio of the interacting nuclei in the fragment X–C$^\alpha$–C$^\beta$–Y relative to hydrogen according to

$$\gamma_{XY} = (\gamma_X\gamma_Y)^{1/2}(\gamma_H)^{-1}. \quad (6)$$

γ_{XY} equals 1.0000, 0.5014, 0.3183, 0.2515 and 0.1596 for ^{1}H–^{1}H, ^{1}H–^{13}C, ^{1}H–^{15}N, ^{13}C–^{13}C and ^{13}C–^{15}N couplings, respectively. For the purpose of numerical simplicity, the negative signs of both γ_N and the nitrogen-related coupling constants are disregarded.

As Eq. 3 cannot account for residue-specific substituent patterns and Eq. 4–5 do not account for residue-independent global asymmetry effects, a combined approach will be necessary, resulting in

$$J_{XY}(\theta) = C_0 + C_1 \cos\theta + C_2 \cos 2\theta + S_1 \sin\theta + \Sigma\Delta J^R_{XY}. \quad (7)$$

A generalised polar Karplus model

In a generalised model, each cosine mode would be supplemented by a corresponding sine component, such that

$$J(\theta) = \Sigma_m (C_m \cos m\theta + S_m \sin m\theta). \quad (8)$$

Pairs of coefficients C_m and S_m can be identified with *Cartesian* components of the m-th coupling mode. The model can be recast in *polar* coordinates comprising component-coupling magnitudes J_m given in Hertz, associated with phases φ_m given in radians. The following transformation identities hold:

$$C_m = J_m \cos \varphi_m; \quad S_m = J_m \sin \varphi_m \quad (9)$$

and

$$J_m = (C_m^2 + S_m^2)^{1/2}; \quad \varphi_m = \text{atan}(S_m/C_m) \quad (10)$$

The *polar* representation of Eq. 8 is then given by

$$J(\theta) = \Sigma_m J_m \cos m(\theta + \varphi_m). \quad (11)$$

Trivially, the sin(0) term in Eq. 8 vanishes. Setting also the coefficient $S_0 = 0$ defines the torsion-angle phase origin in Eq. 11 as $\varphi_0 = 0$, and the mean coupling magnitude as $J_0 = C_0$. Limiting the analysis to a total of three modes results in

$$J(\theta) = J_0 + J_1 \cos(\theta + \varphi_1) + J_2 \cos 2(\theta + \varphi_2), \quad (12)$$

from which the original Karplus equation of Eq. 2 is recovered by setting all phases φ_m at 0, whereas equation Eq. 3 is obtained by setting only φ_2 at 0.

Phases φ_1 and φ_2 in Eq. 12 must not to be confused with J-coupling type related phases stemming from $\theta = \chi_1 + \Delta\chi_1$, where the increment $\Delta\chi_1$ is a fixed multiple of $2\pi/3$, relating the actual dihedral angle θ(X–C$^{\alpha}$–C$^{\beta}$–Y) to the nominal amino-acid side-chain torsion angle χ_1 defined by θ(N′–C$^{\alpha}$–C$^{\beta}$–C$^{\gamma(1)}$) (IUPAC-IUB 1970). As the increment also defines the diastereotopic position of atom Y, which equates to C$^{\gamma}$, H$^{\beta 2}$ or H$^{\beta 3}$ in the majority of amino-acid types, $\Delta\chi_1$ takes values of +120°, –120° and ±0° in ^3J(H$^{\alpha}$,C$^{\gamma(1)}$), ^3J(H$^{\alpha}$,H$^{\beta 2}$) and ^3J(H$^{\alpha}$,H$^{\beta 3}$), respectively. Likewise, $\Delta\chi_1$ takes ±0°, +120° and –120° in ^3J(N′,C$^{\gamma(1)}$), ^3J(N′,H$^{\beta 2}$) and ^3J(N′,H$^{\beta 3}$), and is –120°, ±0° and +120° in ^3J(C′,C$^{\gamma(1)}$), ^3J(C′,H$^{\beta 2}$) and ^3J(C′,H$^{\beta 3}$). In branched amino-acid side chains, C$^{\gamma 2}$ would replace H$^{\beta 2}$, with respective increments referring to those $\Delta\chi_1$ values. However, this is true only in the case of valine. Substituent rankings according to the IUPAC-IUB naming conventions result in switched priorities in isoleucine and threonine, i.e., assigning clockwise C$^{\gamma 2}$, C$^{\gamma 1}$, H$^{\beta(3)}$ and C$^{\gamma 2}$, O$^{\gamma 1}$, H$^{\beta(3)}$, respectively. In computational practice, it is convenient temporarily to assume χ_1 were taken from θ(N′–C$^{\alpha}$–C$^{\beta}$–C$^{\gamma 2}$) in Ile and Thr, such that the above $\Delta\chi_1$ mappings can be applied uniformly (Pérez et al. 2001).

Expressing the generalised Karplus equation in polar terms helps rationalise as to how additive sine components shape the graphical torsion–coupling relationship. The effect of the unimodal sine component S_1 in Eq. 3 is to shift the first mode away from θ by an increment φ_1. While the maxima of that mode are little changed, the displaced cosine wave is most felt at angles where its value change is steepest. Given a constant magnitude J_1, the opposing slopes of the cosine function at corresponding angles around +90° and –90° will increase the coupling constant on one side, and decrease it on the other side. The resulting asymmetry is then seen in different curve minima. Similarly, a non-zero bimodal component S_2 would chiefly affect couplings taken at angles of ±45° and ±135°. However, a term $S_2 \sin 2\theta$ was not included in Eq. 3 as the discrimination of global asymmetry effects attributed to the four different quadrants of the angle range was not required.

The polar representation of the extended Karplus model offers numerical advantages as magnitudes J_m and phases φ_m form an orthogonal parameter set, affecting 'vertical' and 'horizontal' curve features, respectively, contrasting the correlated parameters C_m and S_m both of which relate equally to amplitude, and only indirectly to phase. Their periodic and thus naturally confined range makes unconstrained optimisation of phase parameters φ_m also more robust and efficient than searching unbounded value ranges for parameters S_m. Search properties of J_m equal those of C_m.

Angular motion

Taking angular mobility effects on the NMR parameters into account has been critical to accurate amino-acid side-chain analysis (Karimi-Nejad et al. 1994; Schmidt 1997; Pérez et al. 2001). Harmonic motion about mean torsion angles is often adequate to reflect dynamic averaging of the model ^3J coupling constants. To approximate convolution of the angle with a unimodal Gaussian probability distribution (Karimi-Nejad et al. 1994; Brüschweiler and Case 1994), trigonometric modes were multiplied by exponential terms, $\exp(-0.5m^2\sigma_\theta^2)$. To help convergence and to prevent Gaussian-width parameters σ_θ and also Karplus coefficients C_m (or J_m) from venturing into proscribed regions, both these variable types were fitted in the logarithm (Bates and Watts 1988), thereby constraining them to positive values, necessitating the sign of the odd-mode coefficient C_1 (or J_1) be inverted after reversing the variable transformation.

Conceptually different, staggered-rotamer models were also fitted to the experimental data. Probabilities

p_1, p_2 and p_3, obtained for superimposing χ_1 conformers of –60°, ±180° and +60°, respectively, complement the results. Both Gaussian and staggered-rotamer models consume the same number of degrees of freedom, however, the staggered-rotamer model often proves statistically less significant (Schmidt 1997; Pérez et al. 2001). The logistic sigmoid (Bates and Watts 1988) was applied to contain the free variables p_1 and p_2 within the [0,1] interval. The dependent parameter p_3 occasionally results in spurious small negative values as a consequence of restricting the angles in the model to only three distinct values.

Results

Referring solely to a redundant set of ^3J coupling constants, self-consistent parameterisation of ^3J couplings related to the protein torsion χ_1 was carried out taking possible asymmetry in the coupling-angle relationship into account. Self-consistent multi-parameter optimisation has been described in detail for ^3J-coupling-based analyses of polypeptide ϕ and χ_1 torsions (Schmidt et al. 1999; Pérez et al. 2001). Suffice it that least-squares regression of an over-determined dataset iteratively minimises the residual between calculated and experimental coupling constants by *simultaneously varying all model variables*, i.e., torsion angles, Karplus coefficients and any additional parameters. Crucially, the method does not require reference geometries, such as torsion angles derived from crystallographic coordinates. The only link between experimental data and fitted parameters is the (instantaneous) model of the torsion-angle dependency of ^3J.

Flavodoxin ^3J coupling constraints

Revisiting the dataset previously recorded for flavodoxin (Pérez et al. 2001), Karplus curves were re-iterated including a fourth parameter in each fundamental coefficient set to accomplish the asymmetric shapes shown in Fig. 3. The new asymmetric Karplus parameterisations are given in Table 1. As before, incremental component couplings accounted for any dependency of ^3J on amino-acid topology (Pérez et al. 2001), except these were extended here to include substituent effects up to the second-sphere, where substituents of type '1' are bonded directly to any of the atoms in the χ_1 coupling path X–C^α–C^β–Y, and those of type '2' via two bonds. Type '*inner*' encompasses substituents bonded to central atoms C^α and C^β, and '*outer*' encompasses those attached to terminal sites X and Y (Table 2).

Calculations were carried out using the extended Karplus model according to Eq. 7. Asymmetry effects broken down into *substituent*-atom type related sine components according to Eq. 5 did not perform better than a global *coupling*-type related sine term according to Eq. 3. The limited number of amino acids in the dataset leaves very roughly only 6–7 experimental constraints per J coupling type per amino-acid type and, as a result of χ_1 torsions tending to cluster around residue-type dependent preference angles in addition (Janin et al. 1978; Ponder and Richards 1987), may have been too few to discriminate 'positive' and 'negative' substituents (Haasnoot et al. 1980). Final results were obtained with ΔS_1^R kept at zero, using instead J-coupling type related uniform asymmetry coefficients S_1. In contrast, substituent-type related angle-independent increments ΔC_0^R proved highly significant and were added to the fundamental coupling magnitudes C_0.

Values of $^3J(H^\alpha,H^\beta)$ and $^3J(C',H^\beta)$, among those couplings conveniently measured in χ_1 angle determination, were found to differ between opposing +90° and –90° conformations by approximately 1 Hz. This may also explain why χ_1 analysis was found most sensitive to uncertainties in $^3J(C',H^\beta)$ coupling constants (Dzakula et al. 1996). Asymmetries clearly exceed the estimated experimental precision of 0.5 Hz, highlighting the effect this may have on determining accurate torsion angles from coupling constants.

Out of 713 experimental ^3J coupling constants (disregarding duplicate entries for Ala), 587 (82%) were reproduced by the fit within one standard-error interval, $\sigma_J = 0.5$ Hz, and another 113 (16%) matched within 0.5–1.0 Hz. This is an improvement over the reference calculation by Pérez et al. (2001) in no less than 25 constraints. Violated by 1.0–1.5 Hz were 12 constraints, and only a single one exceeded the $3\sigma_J$ interval slightly.

Out of the 95 non-alanine side chains, the top 10 J-coupling violations, possibly due to problematic experimental data such as weak signal intensity or signal overlap, accounted for around 30% of the normalised residual of 435.6 units and include Phe91 (5.2%), Trp60 (4.6%), Asp143 (2.7%), Ser97 (2.6%), Tyr17 (2.3%), Leu55 (2.2%), Ser10 (2.2%), Ser58 (2.1%), Glu109 (2.1%) and Gln68 (2.0%). A noticeable proportion of asparagine couplings, especially, improved with the new model.

Statistical significance

The present investigation into self-consistent asymmetric Karplus curves refers to the previous experimental

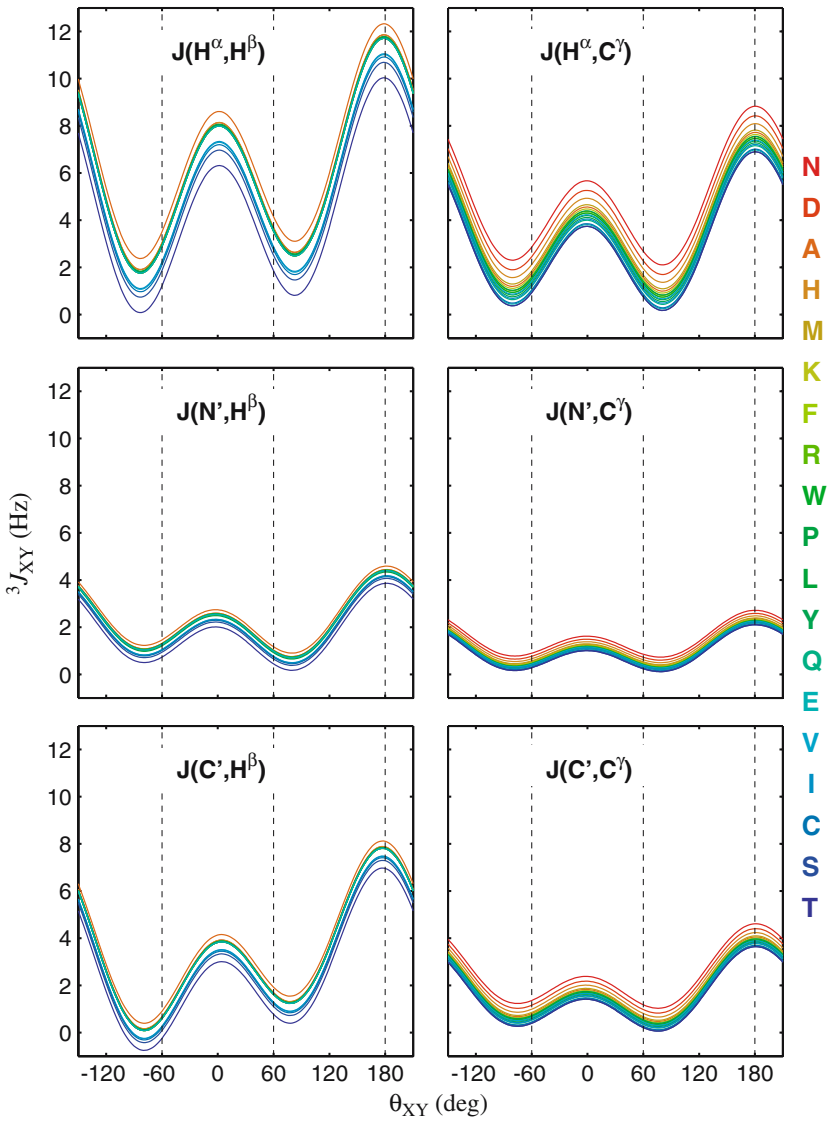

Fig. 3 Self-consistently optimised asymmetric dependencies of protein ^{3}J couplings on the side-chain torsion angle χ_1. Vertical dashed lines indicate *trans* and *gauche* orientations of the dihedral angle subtended between X–C$^{\alpha}$–C$^{\beta}$–Y, signposting ^{3}J values referred to when determining staggered-rotamer distributions. Amino-acid type specific Karplus curves differ by angle-independent offsets reflecting topological variation. One-letter codes on the right rank the amino-acid types by their average coupling magnitude (matching the rainbow-coloured curves) as encountered in the flavodoxin dataset. It is seen that the depths of the minima, for example, in the ^{3}J(C′,H$^{\beta}$) curve, can differ by as much as the spread between the curves. Both variation of the side-chain topology and asymmetries due to chiral substituent environment are equally relevant to *J*-coupling-based side-chain analysis

J-coupling database related to the χ_1 torsions in flavodoxin (Pérez et al. 2001) comprising $n = 763$ experimental ^{3}J values (including 50 duplicate alanine couplings). The number of simultaneously adjusted variables, $p = 260$, in the new asymmetric model included 112 torsion-angle parameters (auxiliary alanine as well as non-alanine residues), 112 angular-mobility parameters, 12 incremental couplings to account for first- and second-sphere substituent effects, 6×3 Karplus coefficients related to the fundamental cosine modes and, finally, 6 coefficients related to the additional sine modes. Void of sine components, the reference set employed only 249 fit parameters. This included an incremental coefficient for first-sphere nitrogen substituents, which turned out to be irrelevant, given the amino-acid topologies considered here, and was therefore omitted in the current χ_1 analysis.

Everything else left unchanged, the new asymmetric model improved the normalised fit residual ε_J^2 to 435.6, from 500.2 in the respective reference calculation (Pérez et al. 2001). The decrease substantially exceeds the number of additional fit parameters (11), indicating that the present results are statistically of higher significance. Indeed, the two residuals related to 503 and 514 degrees of freedom, respectively, yield an *F*-ratio of 1.148, equivalent to a *relative* probability of just under 6% that the better of both fits is a mere chance incidence.

The average violation between calculated and experimental ^{3}J coupling constants (RMSD$_J$) improved to 0.38 Hz from 0.40 Hz in the reference set. Assuming a uniform standard deviation of 0.50 Hz for the individual *J*-coupling measurement, the *absolute* significance in the context of χ^2 statistics rose to 98.6%, from 66.0% reported previously.

Table 1 Self-consistent amino-acid specific asymmetric Karplus coefficients related to the χ_1 torsion in proteins[a]

J(X,Y) type	χ_1 substituent pattern	C_0	C_1	C_2	S_1	J_t	J_{g^+}	J_{g^-}
$^3J(H^{\alpha},H^{\beta})$	Consensus	5.86	–1.86	3.81	–0.37	11.52	3.34	2.71
	Ala	6.66				12.32	4.14	3.51
	His	6.20				11.86	3.68	3.04
	Asn	6.19				11.85	3.67	3.04
	Phe,Tyr,Trp	6.11				11.78	3.59	2.96
	Asp	6.10				11.76	3.58	2.95
	Pro	6.07				11.74	3.56	2.92
	Arg,Glx,Leu,Lys,Met	6.04				11.71	3.52	2.89
	Ile	5.38				11.05	2.87	2.23
	Val	5.35				11.02	2.83	2.20
	Cys	5.25				10.91	2.73	2.10
	Ser	5.02				10.69	2.51	1.87
	Thr	4.37				10.03	1.85	1.21
$^3J(N',H^{\beta})$	Consensus	2.15	–0.93	1.26	0.17	4.33	0.91	1.20
	Ala	2.41				4.59	1.17	1.46
	Asn,His	2.26				4.44	1.02	1.31
	Asp,Phe,Tyr,Trp	2.23				4.41	0.99	1.28
	Arg,Glx,Leu,Lys	2.21				4.39	0.97	1.26
	Met	2.20				4.38	0.96	1.25
	Pro	2.16				4.34	0.92	1.21
	Ile	2.00				4.18	0.76	1.05
	Val	1.99				4.17	0.75	1.04
	Cys	1.96				4.14	0.72	1.01
	Ser	1.88				4.07	0.65	0.94
	Thr	1.68				3.86	0.44	0.73
$^3J(C',H^{\beta})$	Consensus	3.24	–1.99	2.48	–0.59	7.71	1.52	0.50
	Ala	3.64				8.11	1.92	0.90
	Asn,His	3.41				7.87	1.69	0.67
	Phe,Tyr,Trp	3.37				7.83	1.65	0.63
	Asp	3.36				7.83	1.64	0.62
	Pro	3.35				7.81	1.63	0.61
	Arg,Glx,Leu,Lys,Met	3.33				7.80	1.61	0.59
	Ile	3.01				7.47	1.28	0.27
	Val	2.99				7.45	1.26	0.25
	Cys	2.94				7.40	1.21	0.20
	Ser	2.82				7.29	1.10	0.08
	Thr	2.49				6.96	0.77	–0.25
$^3J(H^{\alpha},C^{\gamma})$	Consensus	3.41	–1.58	2.46	0.10	7.45	1.30	1.48
	Asn	4.79				8.83	2.68	2.86
	Asp	4.38				8.42	2.27	2.45
	His	4.05				8.09	1.94	2.12
	Met	3.78				7.82	1.67	1.84
	Ala	3.68				7.72	1.57	1.75
	Lys	3.59				7.64	1.49	1.66
	Pro	3.51				7.56	1.41	1.58
	Arg	3.49				7.54	1.39	1.56
	Phe	3.43				7.47	1.32	1.49
	Trp	3.42				7.46	1.31	1.49
	Val	3.35				7.39	1.24	1.42
	Leu	3.31				7.36	1.21	1.38
	Tyr	3.24				7.28	1.13	1.30
	Ile	3.17				7.21	1.06	1.24
	Gln	3.12				7.16	1.01	1.18
	Cys	2.97				7.01	0.86	1.04
	Glu	2.93				6.97	0.82	1.00
	Ser,Thr	2.86				6.90	0.75	0.93

Table 1 Continued

J(X,Y) type	χ_1 substituent pattern	C_0	C_1	C_2	S_1	J_t	J_{g^+}	J_{g^-}
$^3J(N',C^\gamma)$	Consensus	1.05	–0.55	0.68	0.02	2.27	0.41	0.46
	Asn	1.49				2.71	0.85	0.89
	Asp	1.36				2.58	0.72	0.77
	His	1.25				2.48	0.62	0.66
	Met	1.16				2.39	0.53	0.57
	Ala	1.13				2.36	0.50	0.54
	Lys	1.11				2.33	0.47	0.51
	Arg	1.07				2.30	0.44	0.48
	Phe,Trp	1.05				2.28	0.42	0.46
	Val	1.03				2.26	0.39	0.44
	Leu,Pro	1.02				2.24	0.38	0.43
	Tyr	0.99				2.22	0.36	0.40
	Ile	0.97				2.20	0.34	0.38
	Gln	0.95				2.18	0.32	0.36
	Cys	0.91				2.14	0.27	0.32
	Glu	0.90				2.12	0.26	0.30
	Ser,Thr	0.87				2.10	0.24	0.28
$^3J(C',C^\gamma)$	Consensus	1.69	–1.11	1.11	0.10	3.92	0.49	0.67
	Asn	2.38				4.61	1.18	1.36
	Asp	2.18				4.40	0.97	1.16
	His	2.01				4.24	0.81	0.99
	Met	1.87				4.10	0.67	0.85
	Ala	1.82				4.05	0.62	0.80
	Lys	1.78				4.01	0.58	0.76
	Pro	1.74				3.97	0.54	0.72
	Arg	1.73				3.96	0.53	0.71
	Phe,Trp	1.70				3.92	0.49	0.67
	Val	1.66				3.89	0.46	0.64
	Leu	1.64				3.87	0.44	0.62
	Tyr	1.60				3.83	0.40	0.58
	Ile	1.57				3.80	0.36	0.55
	Gln	1.54				3.77	0.34	0.52
	Cys	1.47				3.70	0.27	0.45
	Glu	1.45				3.68	0.25	0.43
	Ser,Thr	1.41				3.64	0.21	0.39

[a] Coefficients (in Hz) for use with the extended Karplus equation, $^3J(\theta) = C_0 + C_1 \cos\theta + C_2 \cos 2\theta + S_1 \sin\theta$, where $\theta = \chi_1 + \Delta\chi_1$, with $\Delta\chi_1$ depending on the actual diastereospecific positions of the coupled nuclei. Differential effects ΔJ^R_{XY} compiled from substituent related increments (Table 2) weighted by frequencies f_R depending on topology (Pérez et al. 2001) are already included in the amino-acid specific coefficients C_0 given. Consensus C_0 coefficients are averages over all 112 residues included in the fit, weighted by their fractional type occurrence. Coefficients C_1, C_2 and S_1 use the amino-acid independent consensus values. Results derive from the model including second-sphere substituent effects and Gaussian-random fluctuation as described in the text. Useful to staggered-rotamer analysis, values given as J_t, J_{g^+}, J_{g^-} refer to θ angles of ±180°, +60° and –60°, respectively

Analogous comparison of the staggered-rotamer model applied in both studies also showed an improved ε_J^2 of 517.0, compared with 546.4 previously, a decrease in $RMSD_J$ to 0.41 Hz from 0.42 Hz, and a significance increase to 32% from 16% previously.

Flavodoxin side-chain χ_1 geometries

It is not the aim of self-consistent modelling to reproduce a reference structure, yet, it is instructive to compare the results obtained with independent data available (Fig. 4). Within ±30° tolerance, 75 out of 95 rotatable non-alanine side-chain torsions included in the fit agree with consensus crystallographic data computed from eight protein data bank (PDB) coordinate sets: 2FX2, 3FX2, 4FX2, 5FX2 (Watt et al. 1991), two chains A and B in 1BU5 (Walsh et al. 1998), 1J8Q (Artali et al. 2002), and one referred to as 'Walsh' (Walsh 1994).

Asymmetry between positive and negative *dihedral* angles, as signified by θ_{XY}, does not necessarily translate into asymmetry between positive and negative χ_1 *torsion* angles, as both angle types differ by *J*-coupling type related phase increments as detailed above. In

Table 2 Substituent related increments ΔC_0^R in Hz[a]

Substitution	Sphere	Inner	Outer
H → C	1	–0.65	–0.36
H → N	1	n/a	+1.28
H → O	1	–1.64	+0.47
H → S	1	–1.41	n/a
H → C	2	+0.04	+0.20
H → N	2	+0.12	±0.00
H → O	2	+0.03	–0.38

[a] Values self-consistently obtained from the optimisation carried out in this work (n/a, not applicable to the present flavodoxin dataset)

fact, opposing signs to the various asymmetry coefficients S_1 may lead overall to part-cancellation of the effects stemming from the individual coupling types. Thus, different densities seen around χ_1 values of + 60° and –60° in the scattergram of Fig. 4 by no means reflect Karplus curves asymmetric with respect to θ_{XY} but reflect primarily the χ_1 torsion-angle preferences in proteins rather (Janin et al. 1978; Ponder and Richards 1987).

The somewhat tilted correlation seen around $\chi_1 = -60°$ means that angle values for that staggered rotamer are more narrowly distributed in the crystal structure than in the NMR analysis, possibly due to the use of energy potentials in the crystallographic refinement, whilst no such terms are applied in the self-consistency protocol. Similarly, ±180° and +60° rotamers tend to be closer to ideal angles in the crystal structure, whereas the NMR analysis indicates χ_1 conformations skewed toward smaller values. Such differences between solution and solid state may indeed be genuine as both methods are subject to distinct intrinsic effects regarding, for example, conformational averaging.

Referencing the averaged X-ray derived torsion angles, the majority of the side chains shown in Table 3 assume unique χ_1 conformations in solution and agree with those seen in crystal structures. Although the overall violation measures improved with the new model, some NMR-derived torsion angles deviate from consensus X-ray-derived angles by more than ±30°, those marked in Table 3 and as seen in Fig. 4. Torsion-angle dynamics and localisation uncertainties, effects which are experienced and interpreted differently in NMR and X-ray structure determination, help explain some of the discrepancies. In the X-ray datasets, as many as 14 side chains appear disordered (standard deviations above 40°), with variability often linked to different oxidation states of the FMN co-factor bound, of which eight residues at positions 40, 42, 62, 63, 64, 69, 118 and 144 coincide with the disagreements found between NMR and X-ray analysis. For example, coupling constants in Ser64 point to an unambiguous +60° staggered-rotamer conformation in solution, whereas this side chain looks disordered across X-ray datasets. Notably, electron density in the dataset 1BU5 was ill-defined in the FMN-binding region around residues 62–64 and at the C-terminus 147–148 (Walsh et al. 1998), helping explain the χ_1 angle discrepancy seen between NMR and X-ray structures of Asp62, Asp63, Ser64, and Ile148 (Knauf et al. 1996).

Genuinely different orientations were found in three instances. For the Ser10 side chain, which hydrogen bonds with the FMN-phosphate group (Walsh et al. 1998), as well as for Glu110 and surface-exposed Asp34, both NMR and X-ray analysis find χ_1 angles twisted away from the ideal staggered ±180° conformation, albeit in opposite directions in solution and solid state.

Compared with the previous ^3J analysis using symmetric Karplus curves, improvement was seen in the overall torsion-angle discrepancy, as $RMSD_\theta$ decreased to 39.2° from 46.3°. The largest contributions came from residues that appear conformationally averaged in either the Gaussian or the staggred-rotamer model. Noteworthy is Ser97 which flipped from previously –113° into a new conformation of +31°, considerably closer to the χ_1 angles around +80° seen in X-ray structures. Similar was observed for Asp63 whose χ_1 angle effectively changed sign, and Glu20 now matches the X-ray conformation within 10°, compared with 40° discrepancy previously.

Discussion

Redundancy in the Karplus equation requires at least *four* coupling constants be available to solve for *one* angle value. This applies in favourable cases only, as redundancy between various pairs of Karplus curves for χ_1 complicates the analysis in addition. The remaining coupling constants in excess of the minimum four allow angular dynamics to be studied. Two parameters are necessary to model either a Gaussian distribution for each torsion or, alternatively, a staggered-rotamer set. Thus, both models are of identical complexity and exploit the available experimental information equally. Both models, however, have their advantages and drawbacks also. The Gaussian mean angle and the most populated staggered rotamer agree usually, yet, the Gaussian model often performs statistically better than the staggered-rotamer model (Schmidt 1997; Pérez et al. 2001). However, the latter sometimes appears to give a more plausible interpretation. In the present study, both Ser40 and Glu99 are

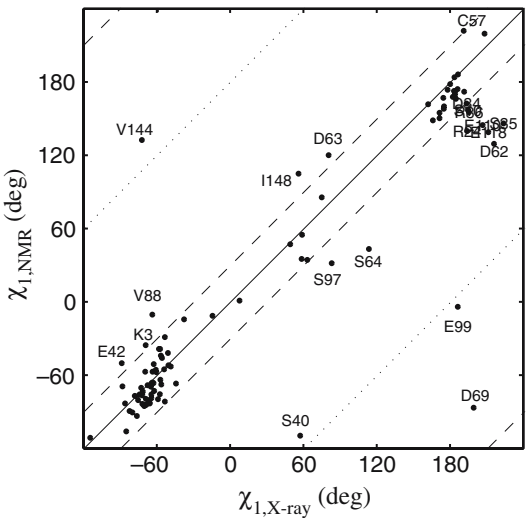

Fig. 4 Side-chain torsions χ_1 in *D. vulgaris* flavodoxin from self-consistent *J* coupling analysis using an asymmetric Karplus model compared with consensus crystallographic data. Residues falling outside the ±30° tolerance margin (dashed bounds) are labeled. Discrepancies of about 180° (dotted lines) are due likely to numerical ambiguity. Shifts by 120° likely result from staggered-rotamer transitions. Properly to represent χ_1 rotamer preferences (Janin et al. 1978; Ponder and Richards 1987), all angles reported follow the IUPAC-IUB definitions (1970) with phase shifts temporarily applied to Ile and Thr during optimisations removed. See text for further details

such cases, exhibiting less likely near eclipsed Gaussian mean χ_1 angles with little spread when the preferred staggered angles are +60° and ±180°, respectively, in agreement with the distributions seen in crystal structures. Such results are likely to be attributed to ambiguity arising from the limited experimental data as just five coupling constants were available for each of these residues (Table 3).

The staggered-rotamer model appears inadequate though in situations when the torsion's range of motion is topologically restricted to non-staggered states. Proline χ_1 angles ordinarily vary only between –30° and +30°, making the Gaussian model the only plausible choice as only this model can adequately cover the range around $\chi_1 = 0°$. By applying the improved asymmetric model of Karplus curves, in connection with Gaussian motion, the numerical ambiguity held responsible in the previous analysis for the insensible torsion-angle value of $\chi_1 = +136°$ adopted by Pro73 (Pérez et al. 2001) was resolved. In agreement with X-ray data, χ_1 converged at –12°, consistent with a constrained polypeptide proline five-ring geometry and a dynamically averaged C^{β}-*endo*/C^{γ}-*exo* ring pucker (Schmidt et al. 1993). Also, χ_1 in Pro2 shifted toward a more positive value and thus closer to consensus X-ray data. The near-zero χ_1 angle obtained for this N-terminal proline is consistent with ultra-rapid pucker interconversion in a pyrrolidine ring not constrained by any peptide link (Sarkar et al. 1986).

In other situations, lack of experimental data cannot be the reason for torsion discrepancies. Examples are Arg24 and Val144 both of which contribute all nine possible coupling constants. While the latter side chain is more compatible with a 3-site jump between all staggered rotamers, equivalent to full rotational averaging, the former appears to be subjected to 2-site jumps between –60° and ±180° rotamers. In both cases, the Gaussian model found conformations skewed from ideal staggered states.

Amino-acid side-chain torsions are often analysed under the premise of staggered-rotamer conformations of $\chi_1 = -60°$, ±180° and +60° being present (Pachler 1963, 1964), solving for their respective populations, p_1, p_2 and p_3, on the basis of so-called *trans* and *gauche* couplings. The presence of a S_1-sine component barely affects the unique *trans* coupling J_t related to $\theta_{XY} = \pm 180°$. However, the difference between the pair of *gauche* couplings J_{g^+} and J_{g^-} obtained for positive and negative 60° angles from asymmetric Karplus curves is given by

$$\Delta J_{\mathrm{gauche}} = J_{g^+} - J_{g^-} = 2S_1 \sin(\pi/3) = 1.732\, S_1. \qquad (13)$$

Depending on the S_1 coefficient, significant variation may accrue to the population ratios.

A formula was previously devised for assessing the average effect two alternative sets of Karplus coefficients may have on a calculated ^3J coupling constant (Blümel et al. 1998; Pérez et al. 2001). The root-mean-square (rms) measure of the anticipated change in the coupling magnitude thus derives from integrating the (squared) difference Karplus equation within bounds $-\pi$ and $+\pi$, in essence, finding the area enclosed by the two respective curves which is entirely defined by the difference between the two sets of Karplus coefficients. The inclusion of a sine mode does not affect the simplicity of the comparison, given here in the general form as

$$(\Delta J_{\mathrm{rms}})^2 = (\Delta C_0)^2 + 0.5\Sigma_{m=1}^{M}\{(\Delta C_m)^2 + (\Delta S_m)^2\}. \qquad (14)$$

As a consequence, torsion angles interpreted on the basis of different Karplus coefficients may critically depend on the set of coefficients used if ΔJ_{rms} exceeds the experimental standard deviation of *J*. Average differential effects between the new asymmetric and the previous symmetric Karplus parameterisations for the χ_1-related couplings $^3J(H^{\alpha},H^{\beta})$, $^3J(N',H^{\beta})$, $^3J(C',H^{\beta})$, $^3J(H^{\alpha},C^{\gamma})$, $^3J(N',C^{\gamma})$ and $^3J(C',C^{\gamma})$ were 0.46,

Table 3 Side'-chain torsion angles χ_1 and angular motion in *Desulfovibrio vulgaris* flavodoxin as inferred from various self-consistent *J* coupling models

Residue[a]	No. *J*	Comments[b]	NMR[c]				X-ray[d]
			Gaussian mean ± std	$p_1(-60°)$	$p_2(\pm180°)$	$p_3(+60°)$	8 datasets mean ± std
Pro-2	5	u/a (proline pucker)	0.5 ± 0.5	*26*	*51*	*23*	6.9 ± 11.3
Lys-3	9	N and X wide	*–35.7 ± 38.6	61	26	13	–70.0 ± 28.4
Leu-5	7	u/a	177.8 ± 27.4	8	80	12	179.1 ± 9.1
Ile-6	7	u/a	–177.4 ± 22.5	7	83	10	175.7 ± 4.4
Val-7	7	u/a	–176.7 ± 30.3	9	76	15	–177.5 ± 6.8
Ser-10	5	N and X opposite signs	*156.2 ± 23.5	21	70	9	–166.5 ± 8.9
Thr-12	5	u/a	–34.8 ± 26.4	65	30	6	–46.0 ± 10.5
Asn-14	7	u/a	–76.1 ± 26.5	75	11	14	–75.0 ± 5.7
Thr-15	4	u/a	–149.3 ± 26.0	14	62	23	–174.2 ± 6.4
Glu-16	8	u/a	169.8 ± 31.9	15	73	12	–177.0 ± 15.0
Tyr-17	8	u/a	148.0 ± 14.6	31	66	3	164.9 ± 6.0
Thr-18	5	u/a	–164.3 ± 0.1	0	86	14	–177.5 ± 9.1
Glu-20	9	N averaged, X single	–85.6 ± 53.1	52	32	16	–71.1 ± 14.8
Thr-21	5	u/a	–158.9 ± 24.1	0	77	23	–178.9 ± 2.1
Ile-22	7	u/a	170.2 ± 20.9	13	82	5	174.2 ± 9.4
Arg-24	9	N skewed, X wide	*139.4 ± 10.2	39	53	8	–167.1 ± 24.5
Glu-25	8	N wide, X single	–93.7 ± 35.8	57	23	21	–77.1 ± 15.2
Leu-26	7	u/a	–79.9 ± 14.1	81	2	16	–69.7 ± 11.7
Asp-28	8	u/a	–90.8 ± 21.9	66	10	25	–80.7 ± 9.5
Tyr-31	9	u/a	–89.7 ± 17.1	69	6	24	–83.1 ± 7.3
Glu-32	5	N skewed, X disordered	–14.7 ± 1.4	52	38	10	–38.6 ± 50.8
Val-33	9	u/a	172.0 ± 32.0	15	72	13	–177.9 ± 4.6
Asp-34	9	N and X opposite signs	*161.7 ± 33.1	18	68	14	–167.5 ± 13.1
Ser-35	6	N skewed, X disordered	*145.5 ± 18.8*	32	59	9	–137.5 ± 48.4
Arg-36	7	u/a	–42.2 ± 0.2	85	22	–7	–51.6 ± 8.4
Ser-40	5	N skewed, X disordered	*–109.7 ± 0.1	15	39	46	56.5 ± 43.7
Val-41	9	N single, X disordered	–52.2 ± 32.3	71	21	9	–51.4 ± 52.6
Glu-42	8	N wide, X disordered	*–50.5 ± 37.5	68	22	11	–89.4 ± 45.4
Leu-46	7	u/a	173.0 ± 32.8	14	73	13	177.1 ± 8.3
Phe-47	9	u/a	–55.7 ± 21.7	83	12	4	–54.8 ± 7.4
Phe-50	7	u/a	–83.9 ± 0.6	79	1	20	–72.7 ± 6.0
Asp-51	7	u/a	–46.2 ± 0.0	89	17	–7	–56.5 ± 3.1
Val-53	8	u/a	166.3 ± 25.5	14	77	9	173.6 ± 4.3
Leu-54	6	u/a	–63.9 ± 28.1	76	13	11	–57.8 ± 6.5
Leu-55	7	u/a	–84.7 ± 0.3	79	1	20	–69.5 ± 15.3
Cys-57	6	N skewed, X single	*–138.8 ± 22.3	9	59	32	–170.0 ± 4.9
Ser-58	6	u/a	46.6 ± 35.6	14	18	68	48.3 ± 6.8
Thr-59	5	u/a	172.7 ± 22.4	16	79	5	–165.2 ± 8.8
Trp-60	9	u/a	–76.2 ± 31.7	68	12	20	–65.1 ± 6.3
Asp-62	7	N averaged, X disordered	*128.8 ± 0.1*	52	34	14	–145.0 ± 44.1
Asp-63	8	N averaged, X disordered	*119.6 ± 24.1*	29	42	28	79.8 ± 57.8
Ser-64	6	N wide, X disordered	*42.8 ± 34.0	15	18	67	112.6 ± 55.7
Ile-65	6	N wide, X disordered	162.6 ± 35.1	23	64	13	161.1 ± 52.6
Glu-66	7	u/a	–82.1 ± 27.0	71	13	17	–54.2 ± 14.7
Leu-67	7	u/a	–69.7 ± 32.6	72	16	12	–88.8 ± 20.5
Gln-68	7	u/a	171.5 ± 38.5	13	68	19	–169.6 ± 5.5
Asp-69	8	N single, X disordered	*–87.0 ± 27.8	66	13	21	–161.8 ± 51.8
Asp-70	7	u/a	–51.4 ± 0.0	91	13	–3	–63.4 ± 6.7
Phe-71	9	u/a	–141.0 ± 0.0	0	64	36	–152.9 ± 4.0
Ile-72	5	u/a	171.4 ± 16.1	10	87	3	171.5 ± 5.3
Pro-73	5	u/a (proline pucker)	–11.8 ± 15.9	*46*	*45*	*9*	–15.3 ± 9.2
Leu-74	8	u/a	154.2 ± 13.8	24	74	3	170.3 ± 4.4
Phe-75	9	u/a	167.3 ± 24.3	14	79	8	–178.9 ± 2.4
Ser-77	4	N and X wide	–68.0 ± 42.9	61	20	19	–57.1 ± 29.6
Leu-78	9	N and X wide	–111.6 ± 30.1	40	30	30	–115.3 ± 24.7
Glu-80	7	N wide, X disordered	–75.9 ± 24.5	77	10	13	–58.3 ± 40.7
Thr-81	5	u/a	–85.4 ± 23.1	72	6	22	–62.4 ± 5.3
Gln-84	9	u/a	149.8 ± 31.5	26	59	15	170.3 ± 18.4
Arg-86	8	N averaged, X single	*154.6 ± 52.6	22	51	27	–165.6 ± 5.6
Lys-87	7	u/a	–73.7 ± 17.7	84	6	10	–72.1 ± 10.1
Val-88	5	N skewed, X wide	*–10.8 ± 25.7	40	43	18	–64.5 ± 36.3
Phe-91	7	u/a	33.9 ± 18.4	6	22	72	62.3 ± 5.0
Cys-93	4	u/a	–57.6 ± 19.4	85	13	2	–70.2 ± 4.7
Ser-96	6	u/a	54.4 ± 30.9	15	10	74	57.9 ± 7.0
Ser-97	6	N averaged, X single	*31.2 ± 58.5	34	15	50	82.2 ± 13.6

Table 3 continued

Residue[a]	No. *J*	Comments[b]	NMR[c]				X-ray[d]
			Gaussian mean ± std	p_1(–60°)	p_2(±180°)	p_3(+60°)	8 datasets mean ± std
Tyr-98	6	u/a	–79.3 ± 22.6	74	7	19	–65.3 ± 9.1
Glu-99	5	N averaged, X single	*–4.3 ± 11.2	32	52	16	–174.7 ± 17.2
Tyr-100	6	u/a	–53.2 ± 33.3	71	21	8	–49.4 ± 12.0
Phe-101	8	u/a	157.4 ± 0.1	25	82	–7	173.8 ± 4.4
Cys-102	5	u/a	–80.8 ± 17.0	76	2	22	–76.1 ± 4.2
Val-105	8	u/a	167.9 ± 20.7	12	82	6	–177.4 ± 5.1
Ile-108	7	u/a	173.6 ± 22.7	12	82	6	175.3 ± 4.5
Glu-109	9	u/a	–83.0 ± 20.1	75	7	18	–66.7 ± 13.5
Glu-110	5	N and X opposite signs	*144.2 ± 26.5	33	57	10	–154.5 ± 13.7
Lys-111	8	u/a	165.6 ± 37.1	15	67	19	–176.4 ± 6.7
Leu-112	6	u/a	–83.4 ± 0.3	81	0	19	–86.8 ± 16.9
Asn-114	9	u/a	–68.9 ± 36.3	69	18	13	–65.9 ± 6.9
Leu-115	7	u/a	–77.0 ± 30.5	71	14	16	–72.7 ± 14.3
Glu-118	6	N skewed, X disordered	*138.4 ± 8.6*	43	51	6	–149.7 ± 47.3
Ile-119	8	u/a	–176.1 ± 18.1	5	86	9	178.4 ± 4.5
Val-120	9	N single, X disordered	–38.9 ± 31.3	65	27	8	–58.1 ± 53.1
Asp-122	8	N and X wide	–66.8 ± 32.4	73	15	12	–63.5 ± 35.6
Leu-124	7	u/a	159.3 ± 17.8	20	77	3	174.0 ± 4.2
Arg-125	4	u/a	–83.3 ± 23.9	72	6	22	–66.6 ± 3.5
Ile-126	8	u/a	172.3 ± 18.6	14	83	3	174.6 ± 5.4
Asp-127	8	u/a	173.6 ± 23.6	8	83	10	–175.0 ± 6.7
Arg-131	8	u/a	–73.2 ± 21.6	81	7	12	–63.0 ± 7.6
Arg-134	8	u/a	–174.4 ± 37.8	13	70	17	–174.5 ± 11.5
Asp-136	5	u/a	–70.9 ± 22.6	81	9	11	–73.6 ± 14.5
Ile-137	7	u/a	–178.0 ± 0.0	4	92	4	178.8 ± 4.4
Val-138	7	u/a	161.3 ± 23.9	17	75	8	161.0 ± 9.1
Asp-143	7	u/a	–80.0 ± 28.3	72	12	16	–59.9 ± 5.6
Val-144	9	N averaged, X disordered	*131.9 ± 20.4*	35	44	21	–73.0 ± 58.1
Arg-145	8	u/a	–106.5 ± 0.2	48	13	39	–85.7 ± 14.0
Ile-148	7	C-term. flexible	*–15.7 ± 27.9	45	40	15	–65.0 ± 17.5
			Statistics[e]				
$RMSD_J$ [Hz]			0.38				0.41
$RMSD_\theta$ [deg]			39.2				n/a
Observables (n)			763				763
Adjustables (p)			260				260
Deg. of freedom (n–p)			503				503
Normalized error ε_J^2			435.6				517.0
Abs. significance Q [%]			98.6				32.3

[a] Includes residues providing a minimum four coupling constants per residue, the actual number of experimental *J* data is put against the residue name. Degenerate H^β proton resonances prevented measuring 4 coupling constants in Tyr8, Asp37, Glu48, Leu52, Asp76, Glu79, Cys90, Asp95, Asp106, Lys113, Gln121, Pro130, Asp135, Trp140, and His142. Torsion angles reported for isoleucine and threonine are taken from $N–C^\alpha–C^\beta–C^{\gamma 2}$ to unify phase shifts (Pérez et al. 2001)

[b] Short-hand to interpretation: (X) X-ray sets; (N) NMR data, this work; (u/a) unambiguous rotamer in both X-ray and NMR datasets; (single) one predominant narrowly distributed staggered rotamer, typical standard deviation below 20° in averaged X-ray data, or typical Gaussian spread below approximately 30° in NMR analysis; (wide) predominant conformer with wider distribution, standard deviation of 20–40° in averaged X-ray data or Gaussian spread around 40° in NMR analysis; (disordered) no discernable rotamer preference and standard deviation of above 40° in X-ray data; (averaged) rotationally averaged, no discernable rotamer preference from either Gaussian or staggered-rotamer model in NMR analysis; (skewed) eclipsed conformation or severly distorted from ideal staggered state possibly due to two-site jumps in NMR analysis

[c] Gaussian model mean torsion angles and spread parameters (in degrees) and staggered-rotamer populations in percent. Asterisks mark torsions deviating by more than 30° from consensus X-ray coordinates. Italics indicate a less plausible model, giving preference to the other model

[d] Torsion angle averages and standard deviations (in degrees) from eight coordinate sets (see text and Supplementary Material for details)

[e] Normalization used a uniform experimental standard deviation of $\sigma_J =$ 0.50 Hz. $RMSD_J$ is the average violation per *J* constraint, and $RMSD_\theta$ the difference from consensus torsion angles in the X-ray structures, applicable only to the Gaussian model. Notice that $RMSD_\theta$ is *not* a measure of quality of the NMR analysis. Significance measures $Q = 1 - P$, with P the probability that the fitted parameter set is a chance event, where $P \approx 0.5$ if the *normalized* residual approaches the number of degrees of freedom in the respective model, i.e., $\varepsilon_J^2 \approx n - p$

0.20, 0.61, 0.47, 0.06 and 0.19 Hz, respectively. Accordingly, the new parameterisation of $^3J(C',H^\beta)$ couplings would be most decisive in the interpretation of the torsion angles, followed by both H^α related couplings.

Given the above coupling-type order, transformation of the Cartesian pair coefficients C_1 and S_1 according to Eq. 10 yields polar magnitudes, J_1, for the first-mode sinusoids of 1.90, 0.94, 2.07, 1.58, 0.55 and 1.12 Hz, respectively, consistently larger than the (negated) coefficients C_1 of 1.37, 0.75, 1.58, 0.96, 0.49 and 0.87 Hz obtained in previous symmetric modelling. Asymmetry from the presence of first-mode sinusoids transforms into phase increments, given here as complementary ($\pi-\phi_1$) values of –11.2°, +10.3°, –16.5°, +3.7°, +2.5° and +5.4°, respectively. Different signs relate to the different slants seen in the various Karplus curves of Fig. 3. Again, parameterisation of $^3J(C',H^\beta)$ couplings would seem most decisive in the torsion-angle interpretation. Interestingly, absolute asymmetries in all H^β related Karplus curves are found to be about three times larger than in those related to C^γ couplings.

As the mean coupling magnitudes were to be maintained, the C_0 coefficients do not vary significantly between the previous symmetric and the improved asymmetric model. Similarly, the new coefficients C_2 agree with previous values within measurement uncertainty, with a tendency towards larger values though. The fact that J_{trans} values, dominated by the superposition of C_0 and C_2 coefficients, came out consistently larger with the new model indicates a cleaner separation of the intrinsic factors (angle, mobility, substituents) influencing ^3J.

Parameterisations of Karplus curves including phase shifts have previously been performed on selected amino-acid types (Chou et al. 2003). These parameterisations for the branched amino-acid side chains Val, Ile and Thr in a number of proteins differ from the present work in two respects. Firstly, optimisations carried out referred to reference structure data from residual dipolar coupling measurements while the present work parameterises the Karplus curves self-consistently, i.e., solely on the basis of over-determined information contained in the J coupling constants, without making any assumptions on the structure. Secondly, Chou et al. implemented an overall phase shift in the Karplus curve by inserting the same angle argument into both angle-dependent terms of the Karplus equation, to the effect that the extrema of the Karplus curve were all simultaneously displaced from the usual locations of –90°, ±0°, +90° and ±180°. In the present study, only the first-mode component is being displaced, affecting the minima at –90° and +90° only. This differs from shifting the curve as a whole and instead affects the shape of the curve, resulting in asymmetric amplitudes of the minima. Although of similar value range, phase increments from both studies cannot be compared directly for the reasons stated, and also because different underlying phenomena are at issue. The approach by Chou et al. basically tests distortions from ideal tetrahedral geometry at the C^β site by narrowing or widening the 120° phase increments between the different dihedral angles. The present approach tests stereochemical influences, i.e., directional differences upon angle rotation while keeping the inter-dihedral phase relations constant at multiples of 120°. Distortion of tetrahedral geometry at the C^α site was previously found to be negligible (Schmidt et al. 1999). What can be concluded though, is that quantitative ^3J-coupling analysis in proteins appears to have reached a level of detail and accuracy at which a change of a few degrees in a torsion angle, on the order of thermal librational amplitudes, make a noticeable difference, allowing, for example, genuine differences between NMR-based solution and X-ray-based crystal structures to be detected.

Concluding remarks

Key to a successful determination of angular-geometry constraints is the availability of accurately parameterised torsion-angle dependencies of all ^3J couplings involved in the analysis. Taking a self-consistency approach to molecular geometry refinement, this study was aimed at investigating possible shortcomings in the commonly employed J-coupling model, and to help explain any fit imperfections resulting from these. Exploiting redundant structure information inherent in large sets of experimental data, self-consistent analysis allows to consider other influences on the ^3J coupling constant, in addition to the basic angular dependency, and was demonstrated here to reveal asymmetries in the various Karplus curves related to the torsion χ_1 in amino acids.

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

The present study focused on the analysis of amino-acid side-chain torsions. Protein main-chain folds do

not normally sample sufficient a range of ϕ torsion values to enable reliable detection of asymmetries in $^3J(\phi)$ in a single protein. For example, only two residues in flavodoxin assume positive values of ϕ. More data would need to be acquired from proteins exhibiting a larger proportion of positive torsion angles.

Protein side-chain couplings pose significant challenges to their interpretation that need to be addressed when analysing molecular geometry. Contrasting less problematic ϕ torsion analyses, side-chain related coupling constants reflect not only angular dynamics but also the variety of amino-acid topologies and the substituent positioning within these. This was dealt with in the present study by extending the *J*-coupling model to include an asymmetric mode for the torsion-angle dependency. The asymmetric Karplus parameterisations obtained for ^3J related to protein side chains were demonstrated to accomplish improved interpretation of molecular geometries, both static and dynamic, and are deemed applicable to other polypeptides.

Acknowledgements Frank Löhr is thanked for stimulating discussion and his continual interest in this *J*-coupling study.

References

Artali, R., Bombieri, G., Meneghetti. F., Gilardi. G., Sadeghi, S.J., Cavazzini, D., Rossi, G.L.: Comparison of the refined crystal structure of wild-type (1.34 Å) flavodoxin from *Desulfovibrio vulgaris* and the S35C mutant (1.44 Å) at 100 K. Acta. Crystallogr. Sect. D **58**, 1787–1792 (2002)

Barfield, M., Karplus, M.: Valence-bond bond-order formulation for contact nuclear spin-spin coupling. J. Am. Chem. Soc. **85**, 2870–2871 (1969)

Bates, D.M., Watts, D.G.: Nonlinear Regression Analysis and its Applications. Wiley, New York (1988)

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

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

Chou, J.J., Case, D.A., Bax, A.: Insights into the mobility of methyl-bearing side chains in proteins from $^3J_{CC}$ and $^3J_{CN}$ couplings. J. Am. Chem. Soc. **125**, 8959–8966 (2003)

Donders, L.A., de Leeuw, F.A.A.M., Altona, C.: Relationship between proton–proton NMR coupling constants and substituent electronegativities. Magn. Res. Chem. **27**, 556–563 (1989)

Dzakula, Z., Westler, W.M., Markley, J.L.: Continuous probability distribution (CUPID) analysis of potentials for internal rotations. J. Magn. Reson. B **111**, 109–126 (1996)

Haasnoot, C.A.G., de Leeuw, F.A.A.M., Altona, C.: The relationship between proton–proton NMR coupling constants and substituent electronegativities—I. An empirical generalization of the Karplus equation. Tetrahedron **36**, 2783–2792 (1980)

Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.P.M., Altona, C.: The relationship between proton–proton NMR coupling constants and substituent electronegativities. II. Conformational analysis of the sugar ring in nucleosides and nucleotides in solution using a generalized Karplus equation. Org. Magn. Reson. **15**, 43–52 (1981a)

Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.P.M., Altona, C.: Relationship between proton–proton NMR coupling constants and substituent electronegativities. III. Conformational analysis of proline rings in solution using a generalized Karplus equation. Biopolymers **20**, 1211–1245 (1981b)

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

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

Karimi-Nejad, Y., Schmidt, J.M., Rüterjans, H., Schwalbe, H., Griesinger, C.: Conformation of valine side chains in ribonuclease T_1 determined by NMR studies of homonuclear and heteronuclear ^3J coupling constants. Biochemistry **33**, 5481–5492 (1994)

Karplus, M.: Contact electron-spin coupling of nuclear magnetic resonance. J. Chem. Phys. **30**, 11–15 (1959)

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

Knauf, M.A., Löhr, F., Blümel, M., Mayhews, S.G., Rüterjans, H.: NMR Investigation of the solution conformation of oxidized flavodoxin from *Desulfovibrio vulgaris*: determination of the tertiary structure and detection of protein-bound water molecules. Eur. J. Biochem. **238**, 423–434 (1996)

Pachler, K.G.R.: Nuclear magnetic resonance study of some a-amino acids—I. Coupling constants in alkaline and acidic medium. Spectrochim. Acta **19**, 2085–2092 (1963)

Pachler, K.G.R.: Nuclear magnetic resonance study of some a-amino acids—II. Rotational isomerism. Spectrochim. Acta **20**, 581–587 (1964)

Pachler, K.G.R.: Extended Hückel theory MO calculations of proton–proton coupling constants. The substituent effect in fluoroethane. Tetrahedron Lett. **22**, 1955–1958 (1970)

Pachler, K.G.R.: Extended Hückel theory MO calculations of proton–proton coupling constants—II. The effect of substituents on vicinal couplings in monosubstituted ethanes. Tetrahedron **27**, 187–199 (1971)

Pachler, K.G.R.: The dependence of vicinal proton–proton coupling constants on dihedral angle and substituents. J. Chem. Soc. Perkin Trans. **II**, 1936–1940 (1972)

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

Ponder, J.W., Richards, F.M.: Tertiary templates for proteins—Use of packing criteria in the emmeration of allowed sequences for different structural chesses. J. Mol. Biol. **193**, 775–791 (1987)

Sarkar, S.K., Young, P.E., Torchia, D.A.: Ring dynamics of D,L-proline and D,L-proline hydrochloride in the solid state: a ^{2}H nuclear magnetic resonance study. J. Am. Chem. Soc. **108**, 6459–6464 (1986)

Schmidt, J.M., Brüschweiler, R., Ernst, R.R., Dunbrack, R.L. Jr, Joseph, D., Karplus, M.: Molecular dynamics simulation of the proline conformational equilibrium and dynamics in antamanide using the CHARMM force field. J. Am. Chem. Soc. **115**, 8747–8756 (1993)

Schmidt, J.M.: Conformational equilibria in polypeptides. II. Dihedral-angle distribution in antamanide based on

three-bond coupling information. J. Magn. Reson. **124**, 310–322 (1997)

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

Walsh, M.A.: Structural studies on a genetically engineered flavodoxin. Dissertation, National University of Ireland (1994)

Walsh, M.A., McCarthy, A., O'Farrell, P.A., McArdle, P., Cunningham, P.D., Mayhew, S.G., Higgins, T.M.: X-ray crystal structure of the *Desulfovibrio vulgaris* (Hildenborough) apoflavodoxin–riboflavin complex. Eur. J. Biochem. **258**, 362–371 (1998)

Watt, W., Tulinsky, A., Swenson, R.P., Watenpaugh, K.D.: Comparison of the crystal structures of a flavodoxin in its three oxidation states at cryogenic temperatures. J. Mol. Biol. **218**, 195–208 (1991)