

Communication: Accurate determination of side-chain torsion angle χ_1 in proteins: Phenylalanine residues

R. Suardíaz,¹ R. Crespo-Otero,² C. Pérez,¹ J. San Fabián,³ and

J. M. García de la Vega^{3,a)}

¹Departamento de Química Física, Facultad de Química, Universidad de la Habana, La Habana 10400, Cuba

²Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

³Departamento de Química Física Aplicada, Universidad Autónoma de Madrid, Madrid 28049, Spain

(Received 5 November 2010; accepted 19 January 2011; published online 10 February 2011)

Quantitative side-chain torsion angle χ_1 determinations of phenylalanine residues in *Desulfovibrio vulgaris* flavodoxin are carried out using exclusively the correlation between the experimental vicinal coupling constants and theoretically determined Karplus equations. Karplus coefficients for nine vicinal coupling related with the torsion angle χ_1 were calculated using the B3LYP functional and basis sets of different size. Optimized χ_1 angles are in outstanding agreement with those previously reported by employing x ray and NMR measurements. © 2011 American Institute of Physics. [doi:10.1063/1.3553204]

Knowledge on the three-dimensional structure of biological macromolecules is a key factor for understanding molecular processes occurring in living systems.¹ NMR spectroscopy is the only method that allows the determination of 3D-structures of proteins in solution. Measurement of nuclear Overhauser effects complemented with scalar coupling constants and chemical shifts provides the constraints needed for the determination of 3D-structures. Additionally, residual dipolar couplings make it possible to estimate the relative orientation of a number of bond vectors. Three-bond scalar coupling constants (3J) benefit NMR structure refinements through the well known Karplus equation² and have been demonstrated to deliver quantitative constraints on backbone and side-chain torsional angles of proteins.^{3,4}

The use of the vicinal coupling constants $^3J_{XY}$ for obtaining torsional angles relies on the accurate knowledge of the angular dependence on those couplings. A truncated Fourier series is used to describe this relationship, shown here for the couplings related to the side-chain dihedral angle χ_1 ,

$$^3J_{XY}^{cal}(\theta) = C_0 + C_1 \cos(\theta) + C_2 \cos(2\theta) + C_3 \cos(3\theta) + S_1 \sin(\theta) + S_2 \sin(2\theta), \quad (1)$$

where θ is the dihedral angle between the planes $X - C^\alpha - C^\beta$ and $C^\alpha - C^\beta - Y$, which can be related to the χ_1 angle assuming perfect tetrahedral geometries at the C^α and C^β atoms, $\theta = \chi_1 + \Delta\theta$ (see Fig. 1). Phase shifts $\Delta\theta$ for each coupling type are shown in Table I. Equation (1) reduces to the usual Karplus equation² when the coefficients C_3 , S_1 , and S_2 are neglected.

Here, we report the determination of side-chain torsion angle χ_1 of a well known protein, *Desulfovibrio vulgaris* flavodoxin, using experimental vicinal coupling constants and theoretical Karplus-like equations. The Fourier coefficients are obtained theoretically. Therefore, none of the

coefficients of Eq. (1) are neglected. We follow a seven steps procedure: (1) An amino acid residue phenylalanine (PHE) from *Desulfovibrio vulgaris* flavodoxin³ is chosen. Six PHE residues are found in that protein. Between seven and nine coupling constants $^3J_{XY}$ related to χ_1 angle are available for each of those residues. A total of 49 experimental $^3J_{XY}$ values have been reported.³ (2) The amino acid geometry was fully optimized at the B3LYP/6-31G** level of theory^{5,6} using the GAUSSIAN program.⁷ (3) The dihedral angle χ_1 ($N' - C^\alpha - C^\beta - C^\gamma$) is scanned from 0° to 300° in 60° steps with all the remaining degrees of freedom optimized at the same level of theory as described in previous work.⁸ A set of six geometries with different torsion angle χ_1 is obtained. (4) For each of these geometries with various values of torsion, nine coupling constants $^3J_{XY}(\chi_1)$ related to the side-chain torsion χ_1 are calculated at B3LYP level of theory and the following five basis sets: TZVP,⁹ 6-31G(d,p),¹⁰ EPR-III,¹¹ aug-cc-pVTZ-J,¹² and pcJ-2 (Ref. 13) using standard procedure.¹⁴⁻¹⁸ Although these basis sets are of different size and hence require different computational effort, the results are similar (see Tables II and III). Therefore, the inexpensive TZVP basis set can be an acceptable choice for future calculations. B3LYP/TZVP level of theory has proved to provide similar results to those of *ab initio* SOPPA and SOPPA(CCSD) methods with larger basis sets¹⁹ and has been used successfully in the calculations of J and hyperfine couplings.²⁰⁻²⁴ (5) The nine sets of coupling constants $^3J_{XY}$ and the angles from geometry optimizations were used to calculate the six corresponding Fourier coefficients in Eq. (1), solving 6×6 system of nonhomogeneous linear equations. The Fourier coefficients for the B3LYP/TZVP level and for other levels of theory are shown in Table I and in the supplementary information,²⁵ respectively. (6) For each amino acid residue, single χ_1 sweep is used to look for the minimum of the root mean square deviation (rmsd) between the calculated couplings $^3J_{XY}^{cal}$, using the Karplus equation (1) for the angle χ_1 , and the experimental ones $^3J_{XY}^{exp}$.

^{a)}Electronic mail: garcia.delavega@uam.es.

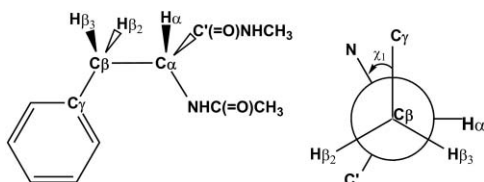


FIG. 1. Molecular structure and Newman projection of phenylalanine.

These rmsd have been represented in Fig. 2 for B3LYP/TZVP results. (7) A three-site jump model is also considered as an additional and more complete approach that consider the existence of not only one position of the side-chain. The three canonical rotamers were assumed with χ_1 angles of -60° , 180° , and 60° . Populations p_i were calculated minimizing the functional $\sigma = \sum_i (p_1 J_i^{-60} + p_2 J_i^{180} + p_3 J_i^{60} - J_i^{\text{exp}})^2$ with the condition of $p_1 + p_2 + p_3 = 1$. In some cases, the obtained populations have small negative values due to the restriction of considered angles and errors on experimental and theoretical models (see Table II). Estimated populations are similar to those reported using empirical asymmetric Karplus-like equations.²⁶

The method outlined uses sets of experimental homo- and heteronuclear 3J coupling constants which, in spite of their different magnitude, have been considered into the rmsd minimization with the same weight. Largest couplings, as $^3J_{\text{HH}}$, dominate unweighted minimization process. This fact can be justified, in part, because those coupling constants with larger values are, generally, determined with less uncertainty. On the other hand, they present larger variation with the torsional angles allowing a more accurate determination of those angles. An alternative to the unweighted minimization is to introduce some weights for each type of coupling considering, for instance, weights proportional to the coupling magnitude or experimental uncertainties. The magnitude of the coupling constants depends on gyromagnetic ratios (γ) of the coupled nuclei. Therefore, the use of reduced coupling constants K_{AB} ($J_{\text{AB}} = h\gamma_A\gamma_B K_{\text{AB}}/4\pi^2$) will make more uniform the different type of couplings. In order to simplify the large numbers derived for the reduced couplings, we will use reduced units (r.u.) (Ref. 27) defined as $10^{19} \text{ T}^2 \text{ J}^{-1}$. The results obtained with the TZVP basis set using reduced coupling constants are summarized in Fig. 2(b) and in Table II. Results from the remaining basis set are presented in the supplementary information.²⁵ In general, results obtained from the reduced couplings are qualitatively similar to those obtained from frequency ones. The use of weighted minimizations with differ-

ent kind of weights should be considered with more detail over larger datasets.

The uncertainties of χ_1 are estimated from the shape of the rmsd minima. From Fig. 2, we can determine the range of angles comprises in a specific contour around the rmsd minima, (angles within $\text{rmsd}_{\text{min}} \pm \text{rmsd}_{\text{interval}}$). Those uncertainties were calculated using selected intervals (0.2 Hz and 0.1 a.u.) and they are depicted in Table II. The angles uncertainties for the amino acids with nine coupling constants are between 6° and 10° for the selected rmsd intervals. Some amino acid residues show an asymmetric uncertainty (see Table II) which corresponds to the asymmetric shape of the minima (see Fig. 2). Larger uncertainties are obtained for amino acid residues with only seven experimental couplings.

It is interesting to note that the intrinsic degeneracy of the Karplus equation which gives rise to multivalued solutions²⁸ is not fully resolved simply using simultaneous equations for different coupling constants that related to the same torsion angle. In fact, in our study, where seven or more couplings are used, it is observed that at least two minima which are separated approximately in 180° are obtained for each amino acid (see Fig. 2).

This behavior can be explained as following, when we use an extended Karplus equation, Eq. (1), the difference

$$^3J(\theta) - ^3J(\theta + 180) = 2C_1 \cos \theta + 2C_3 \cos 3\theta + 2S_1 \sin \theta$$

is small because C_1 , C_3 , and S_1 are usually the smallest Fourier coefficients. In fact, these coefficients and the larger S_2 are neglected in the empirical parametrization of Karplus equations. Thus, this ambiguity will be even more significant with the usual three coefficients (C_0 , C_1 and C_2) Karplus equation.²

In addition to Fig. 2, the ambiguities indicated above can be detected using the so-called dihedral angle ambiguity map,²⁸ which is the sum of n two-dimensional self-correlation diagram of a Karplus curve, i.e., the representation of rmsd_J vs χ_1 and χ'_1 defined as

$$\text{rmsd}_J(\chi_1, \chi'_1) = \sqrt{\frac{\sum^n [^3J_{X,Y}^{\text{cal}}(\chi_1) - ^3J_{X,Y}^{\text{cal}}(\chi'_1)]^2}{n-1}}, \quad (2)$$

where n is the number of different coupling constants types or Karplus equations considered. The individual two-dimensional self-correlation diagrams for the nine types of coupling constants studied in this work are presented in the supplementary information.²⁵ This kind of maps shows the

TABLE I. Phase shifts (deg) and Fourier coefficients (Hz) for each of the nine vicinal couplings^a in phenylalanine calculated at the B3LYP/TZVP.

	$^3J_{H_\alpha, H_{\beta 2}}$	$^3J_{H_\alpha, H_{\beta 3}}$	$^3J_{N', H_{\beta 2}}$	$^3J_{N', H_{\beta 3}}$	$^3J_{C', H_{\beta 2}}$	$^3J_{C', H_{\beta 3}}$	$^3J_{H_\alpha, C_{\gamma 1}}$	$^3J_{N', C_{\gamma 1}}$	$^3J_{C', C_{\gamma 1}}$
$\Delta\theta$	-120.0	0.0	120.0	-120.0	0.0	120.0	120.0	120.0	-120.0
C_0	6.338	6.086	1.814	1.899	3.782	3.748	3.910	0.852	1.940
C_1	0.862	0.878	-0.307	0.273	-0.717	-0.750	-0.649	-0.227	-0.807
C_2	4.963	5.467	1.420	2.054	3.572	3.174	3.810	0.933	1.583
C_3	-0.247	-0.335	-0.201	-0.019	-0.089	-0.422	-0.162	-0.018	-0.109
S_1	0.024	0.156	-0.288	0.030	0.133	0.253	-0.001	0.006	0.246
S_2	1.100	0.707	0.350	-0.070	-1.141	-1.469	0.748	0.006	-0.763

^aThe isotopes for the indicated couplings correspond to ^1H , ^{13}C , and ^{15}N .

TABLE II. Optimized side-chain torsion angles χ_1 (deg) and obtained staggered-rotamer populations (in percent) for phenylalanine residues of *Desulfovibrio vulgaris* flavodoxin (Ref. 3). Results obtained with TZVP basis set using frequency (A) or reduce coupling (B). Uncertainties are given in parentheses.

Residue	X ray ^a	Schmidt ^b	Pérez ^c	TZVP (this work)		P_1 (-60°)	P_2 (180°)	P_3 (60°)	n^d
				(A)	(B)				
PHE47	-54.8±7.4	-55.7±21.7	-52.5	-56.5(-6.6/+6.4)	-55.4(-6.8/+6.1)	95	6	-1	9
PHE50	-72.7±6.0	-83.9±0.6	-78.0	-68.4(-7.0/+7.5)	-91.5(-12/+42) ^e	94	-9	15	7
PHE71	-152.9±4.0	-141.0±0.0	-147.9	-157.7(-8.9/+8.6)	-147.9(-9.0/+7.7)	-5	69	36	9
PHE75	-178.9±2.4	167.5±24.3	172.2	-178.6(-7.4/+7.5)	174.6(-8.1/+9.8)	7	88	5	9
PHE91	62.3±5.0	33.9±18.4	46.9	62.9(-30/+11) ^a	53.5(-8.3/+8.7) ^e	1	24	75	7
PHE101	-173.8±4.4	157.4±0.1	162.2	176.7(-7.0/+7.1)	173.0(-9.2/+11)	13	90	-3	8

^aConsensus crystallographic data computed from eight protein data bank coordinate sets (Ref. 26).^bUsing an asymmetric Karplus equation in a self-consistency approach in which the Karplus coefficients and the torsional angles are fitted simultaneously (Ref. 26).^cUsing a symmetric Karplus equation in a self-consistency approach (Ref. 3)^dNumber of experimental couplings for the indicated residue.^eCorresponds to a second minimum (see Fig. 2).

regions where dihedral angle χ_1 approaches those of another angle χ'_1 , i.e., regions with low rmsd_J and therefore, with ambiguities in the determination of the torsional angle. In Fig. 3(a), we present a dihedral-angle ambiguity map calculated with $n = 9$ and the Fourier coefficients of Table I. The dark (blue) areas, with low rmsd_J , correspond to ambiguity values. In Fig. 3(a), an example of ambiguity in $\chi_1 = 60^\circ$ and 240° is shown.

It should also be noted that the residue PHE91, which has only seven experimental couplings instead of nine, presents an additional ambiguity close to the chosen minimum located at 64° . This ambiguity disappears when couplings in reduced units are used, however, in this case a similar ambiguity appears for PHE50 [see Fig. 2(b)]. Figure 3(b) represents the dihedral angle ambiguity map where only seven couplings have been considered [$n = 7$ in Eq. (2)]. Chosen coupling constants are those of the PHE91 residue. In this figure it is observed that for χ'_1 close to 240° there is a wide blue area (low rmsd_J) around χ_1 angle between 30° and 60° . In addition, the rmsd representation (blue curve in Fig. 2) shows a flat area between 30° and 60° . When Fourier coefficients calculated with the aug-cc-pVTZ-J and pcJ-2 basis sets are used to find the torsional angle χ_1 , three minima instead of two are found at 26° , 64° , and 240° for the PHE91 residue (see rmsd representation for these basis sets in the supplementary information²⁵).

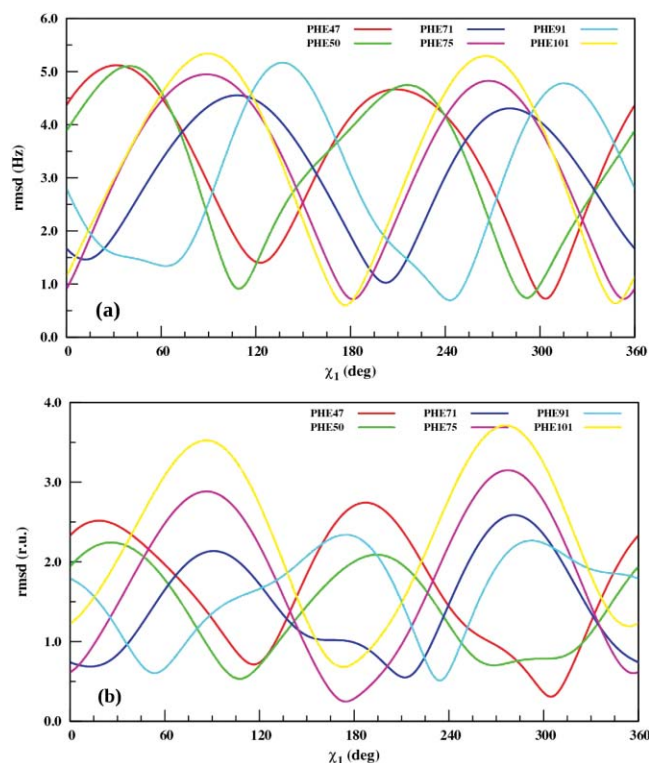
The ambiguities previously reported show, at least, two minima separated 180° . In this work, results are compared with x ray and NMR measurements and, therefore, there is

no doubt about which one has to be chosen. Furthermore, the study of side-chain rotamer libraries^{29,30} indicates that the choice of relaxed and noneclipsed side-chain conformations is justified even in protein interiors³⁰ not only for PHE but for majority of amino acid residues.

Taking into account the complexity of the problem, using theoretical Karplus equations can present several advantages: (1) Theoretical results allow to use Karplus equations which include all of the needed coefficients. Empirical parametrization of Karplus equations required the use of large data sets of well known geometries and coupling constants. These requirements limit the parametrizations to three coefficient Karplus equations. (2) The geometries and therefore the torsional angles cover the whole conformational space, includ-

TABLE III. Side-chain torsion angles χ_1 (deg) obtained at the B3LYP level and the indicated basis sets using couplings in frequency units.

Residue	6-31G(d,p)	EPR-III	aug-cc-pVTZ-J	pcJ-2
PHE47	-55.3±0.8	-58.0	-58.6	-58.5
PHE50	-70.3±0.8	-73.8	-75.3	-75.3
PHE71	-157.5±0.8	-151.5	-149.6	-149.6
PHE75	-177.7±0.5	-174.9 ^a	-173.6 ^a	-173.8 ^a
PHE91	63.3±0.8	63.5 ^a	63.6 ^b	64.2 ^b
PHE101	177.0±0.5	179.2 ^a	-179.5 ^a	-180.0 ^a

^aCorresponds to a second minimum.^bCorresponds to a third minimum. For the PHE91 residue, the plot of the rmsd vs χ_1 present a flat area with two minima around 60° (see Fig. 2).FIG. 2. Root mean square deviation (in Hz and r.u.) as function of the dihedral angle χ_1 (deg) for the six phenylalanine residues. Calculated from coupling constants in frequency units (a) and in reduced units (b).

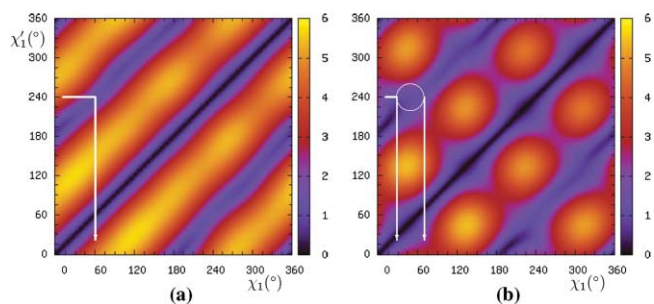


FIG. 3. Dihedral angle ambiguity maps ($\text{rmsd}_J(\chi_1, \chi'_1)$) corresponding to the sum of the nine (a) or seven (b) involved couplings (see text).

ing regions which are difficult to find in experimental data sets and avoiding that the Fourier coefficients can be biased toward the gauche/anti conformation as addressed in previous work.⁸ In this way, obtained equations provide a possible prediction of conformations away from the traditional gauche/anti arrangements. (3) Fourier coefficients obtained theoretically include implicitly several of the coupling constants effects which make difficult the empirical parametrization. Two main observed effects are those due to the substituents (their electronegativity, position and conformation) and changes into the local geometry (bond angles and bond lengths). In general, empirical Karplus parametrization^{3,31} is applied to the six couplings which are significantly different ($^3J_{H_\alpha, H_\beta}$, $^3J_{N', H_\beta}$, $^3J_{C', H_\beta}$, $^3J_{H_\alpha, C_{\gamma 1}}$, $^3J_{N', C_{\gamma 1}}$, and $^3J_{C', C_{\gamma 1}}$), while theoretical parametrization can be done for the nine types of couplings. This means that effects due to different substituent orientation or local geometry are included, for instance, in $^3J_{H_\alpha, H_{\beta 2}}$ and in $^3J_{H_\alpha, H_{\beta 3}}$.

As expected for this kind of molecular systems positive C_1 coefficient is obtained. This fact has been recently interpreted as a consequence of hyperconjugative interactions between orbitals of the coupling pathway and those of carbonyl group.^{8,23,24} It is also noteworthy that the value of coefficient S_2 is larger than those of S_1 and C_3 (extra coefficient regarding classic Karplus equation) suggesting a more important role in the fitting of truncated Fourier series.

Torsional angles χ_1 obtained with this method are compared with those previously obtained by x ray and NMR (Table II). Outstanding results are obtained using theoretical Karplus-like equation. Torsion angles calculated using a set of extended basis, defined above, are summarized in Table III. A general inspection of these data indicates that torsion angles χ_1 are predicted in very good agreement with the experimental values regardless the size of basis sets.

In conclusion, we describe a simple and effective method to obtain protein side-chain torsional angles. To the best of our knowledge, this is the first report where theoretically obtained Karplus equation has been used to determine protein torsional angles. The inexpensive B3LYP/TZPV level yields highly accurate values of χ_1 . Moreover, larger basis sets pro-

duce similar results. Further developments for all remaining amino acids are currently being carried out.

Financial support from the MICINN of Spain (Project No. CTQ2007-66547 and CTQ2010-19232), the Comunidad de Madrid (Project Nos. S2009/ENE-1743) and AECID (Project No. D/023653/09) is gratefully acknowledged. Computational facilities have been provided by CCC-UAM.

- ¹W. Gronwald and H. R. Kalbitzer, *Prog. Nucl. Magn. Reson. Spectrosc.* **44**, 33 (2004).
- ²M. Karplus, *J. Am. Chem. Soc.* **85**, 2870 (1963).
- ³C. Pérez, F. Löhr, H. Rüterjans, and J. M. Schmidt, *J. Am. Chem. Soc.* **123**, 7081 (2001).
- ⁴J. M. Schmidt, M. Blümel, F. Löhr, and H. Rüterjans, *J. Biomol. NMR* **14**, 1 (1999).
- ⁵A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- ⁶C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ⁷M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, revision D.01, Gaussian, Inc., Wallingford CT, 2004.
- ⁸R. Suardíaz, C. Pérez, J. M. García de la Vega, J. San Fabián, and R. H. Contreras, *Chem. Phys. Lett.* **442**, 119 (2007).
- ⁹N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, *Can. J. Chem.* **70**, 560 (1992).
- ¹⁰P. C. Hariharan and J. A. Pople, *Theor. Chem. Acc.* **28**, 213 (1973).
- ¹¹V. Barone, *J. Chem. Phys.* **101**, 6834 (1994).
- ¹²P. F. Provasi, G. A. Aucar, and S. P. A. Sauer, *J. Chem. Phys.* **115**, 1324 (2001).
- ¹³F. Jensen, *J. Chem. Theory Comput.* **2**, 1360 (2006).
- ¹⁴T. Helgaker, M. Watson, and N. C. Handy, *J. Chem. Phys.* **113**, 9402 (2000).
- ¹⁵V. Sychrovský, J. Gräfenstein, and D. Cremer, *J. Chem. Phys.* **113**, 3530 (2000).
- ¹⁶V. Barone, J. E. Peralta, R. H. Contreras, and J. P. Snyder, *J. Phys. Chem. A* **106**, 5607 (2002).
- ¹⁷J. E. Peralta, G. E. Scuseria, J. R. Cheeseman, and M. J. Frisch, *Chem. Phys. Lett.* **375**, 452 (2003).
- ¹⁸W. Deng, J. R. Cheeseman, and M. J. Frisch, *J. Chem. Theory Comput.* **2**, 1028 (2006).
- ¹⁹R. H. Contreras, R. Suardíaz, C. Pérez, R. Crespo-Otero, J. San Fabián, and J. M. García de la Vega, *J. Chem. Theory Comput.* **4**, 1494 (2008).
- ²⁰L. Hermosilla, P. Calle, J. M. García de la Vega, and C. Sieiro, *J. Phys. Chem. A* **109**, 1114 (2005).
- ²¹A. B. Sahakyan, A. A. Shahkhatuni, A. G. Shahkhatuni, and H. A. Panosyan, *Magn. Reson. Chem.* **46**, 63 (2008).
- ²²A. B. Sahakyan, A. G. Shahkhatuni, A. A. Shahkhatuni, and H. A. Panosyan, *J. Phys. Chem. A* **112**, 3576 (2008).
- ²³R. Suardíaz, C. Pérez, R. Crespo-Otero, J. M. García de la Vega, and J. San Fabián, *J. Chem. Theory Comput.* **4**, 448 (2008).
- ²⁴R. H. Contreras, R. Suardíaz, C. Pérez, R. Crespo-Otero, J. San Fabián, and J. M. García de la Vega, *Int. J. Quantum Chem.* **110**, 532 (2010).
- ²⁵See supplementary material at <http://dx.doi.org/10.1063/1.3553204> for additional Fourier coefficients, rmsd curves, side-chain torsional angles, and dihedral-angle ambiguity maps.
- ²⁶J. M. Schmidt, Y. Hua, and F. Löhr, *J. Biomol. NMR* **37**, 287 (2007).
- ²⁷J. S. Fabián, E. Díez, J. M. G. de la Vega, and R. Suardíaz, *J. Chem. Phys.* **128**, 084108 (2008).
- ²⁸M. Blümel, J. M. Schmidt, F. Löhr, and H. Rüterjans, *Eur. Biophys. J.* **27**, 321 (1998).
- ²⁹S. C. Lovell, J. M. Word, J. S. Richardson, and D. C. Richardson, *Proteins* **40**, 389 (2000).
- ³⁰S. C. Lovell, I. W. Davis, W. B. A. III, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, and D. C. Richardson, *Proteins* **50**, 437 (2003).
- ³¹J. M. Schmidt, *J. Magn. Reson.* **186**, 34 (2007).