13C ppm

Showcasing research from Professor Hansen's laboratory, Institute of Structural and Molecular Biology, Division of Biosciences, University College London, UK.

Determining isoleucine side-chain rotamer-sampling in proteins from ^{13}C chemical shift

Chemical shifts are often the only nuclear magnetic resonance parameter that can be obtained for challenging macromolecular systems such as low-populated states. We developed a framework to derive the conformational sampling of isoleucine side-chains in proteins from ^{13}C chemical shifts.
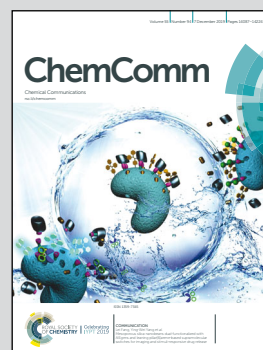
ROYAL SOCIETY OF CHEMISTRY | Celebrating IYPT 2019

rsc.li/chemcomm

Registered charity number: 207890

Check for updates

Determining isoleucine side-chain rotamer-sampling in proteins from ^{13}C chemical shift†

Lucas Siemons, [a] Boran Uluca-Yazgi,[bc] Ruth B. Pritchard, [a] Stephen McCarthy,[d] Henrike Heise[bc] and D. Flemming Hansen ⋆[a]

Chemical shifts are often the only nuclear magnetic resonance parameter that can be obtained for challenging macromolecular systems. Here we present a framework to derive the conformational sampling of isoleucine side chains from ^{13}C chemical shifts and demonstrate that side-chain conformations in a low-populated folding intermediate can be determined.

Protein side chains have the ability to sample several conformations, which is important for many biological processes. Side-chain motions are often de-correlated from the backbone,[1–3] making it important to be able to specifically characterise them. The structure and conformational sampling of side chains are often derived from a combination of several NMR measurements, such as nuclear Overhauser effects (NOEs), three-bond scalar couplings,[4,5] residual dipolar couplings and spin-relaxation measurements.[6] Although these measurements are feasible for most proteins smaller than 20 kDa, an accurate description of side-chain behaviour of larger systems often becomes challenging. Moreover, characterising the conformational sampling of side chains in low-populated states[2,7] is still more difficult, because neither NOEs nor scalar couplings can currently be obtained. Relating chemical shifts, the most easily assessible NMR parameter, to structure and motions provides an attractive alternative to characterise side chains in many challenging systems.[8–10]

Below we show a framework for relating ^{13}C chemical shifts to the conformational sampling of side chains, with focus on the isoleucine side chain. This side chain is composed of sp^3 hybridised carbon atoms allowing both side-chain dihedral angles, χ_1 (N–C$^\alpha$–C$^\beta$–C$^{\gamma 1}$) and χ_2 (C$^\alpha$–C$^\beta$–C$^{\gamma 1}$–C$^{\delta 1}$), to sample the

three canonical states $\{\sim 60°, \sim 180°, \sim 300°\}$, referred to as *gauche+* (g_p); *trans* (t); and *gauche−* (g_m), respectively, Fig. 1a. Using a combination of density functional theory (DFT) calculations and a comprehensive set of experimental long-range scalar coupling constants for model proteins, chemical shift profiles are created for the most populated side-chain rotameric states. These chemical shift profiles, in turn, are used to provide a near complete description of the side-chain rotamer distribution, *i.e.* determine the populations of the $\{\chi_1, \chi_2\}$ rotameric states from experimental chemical shifts alone. The readily available nature of chemical shifts greatly extends the systems where side-chain rotamer distributions can be obtained. To demonstrate this, the framework is applied to characterise isoleucine side chains in an 'invisible', low-populated protein folding intermediate.[11]
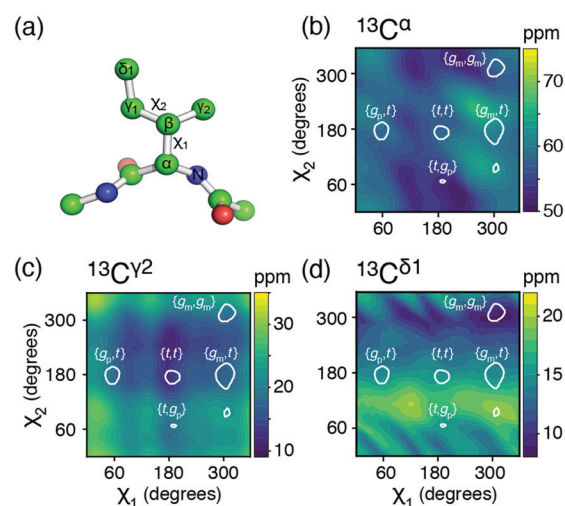

Fig. 1 (a) The isoleucine side-chain is shown with the dihedral angles and side-chain carbons labelled. (b)−(d) Theoretical chemical shift surfaces for ^{13}C$^\alpha$, ^{13}C$^{\gamma 2}$ and ^{13}C$^{\delta 1}$ in a β-sheet backbone conformation, respectively. The shielding constants were calculated using the GIAO method with a B3LYP functional and the EPR-III basis set and referenced as described in ESI.† The white contours indicate the regions, which together comprise 90% of the total populations observed in high-resolution crystal structures.[12]

^{a} *Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London, UK WC1E 6BT. E-mail: d.hansen@ucl.ac.uk*

^{b} *Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany*

^{c} *Institute of Complex Systems, ICS-6: Structural Biochemistry and JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich, Jülich, Germany*

^{d} *Department of Chemistry, University College London, 20, Gordon Street, London, WC1H 0AJ, UK*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cc06496f

The potential energy surfaces derived from the DFT calculations show that the rotameric states $\{g_p, g_m\}$, $\{t, g_m\}$, and $\{g_p, g_p\}$ are all populated to less than 1% and $\{g_m, g_p\}$ is only populated around 3%, which is in agreement with an analysis of a large set of high-resolution crystal structures,[12,13] Table S1 and Fig. S1 (ESI†). The top five states, in decreasing order of overall population: $\{g_m, t\}$, $\{g_m, g_m\}$, $\{g_p, t\}$, $\{t, t\}$, $\{t, g_p\}$, account for >97%, while the top four states account for >95%. Below, mainly the top four states will be considered.

Chemical shift (δ) surfaces *vs.* $\{\chi_1, \chi_2\}$, were obtained from DFT calculations (ESI†) on a model peptide representing the isoleucine side chain, Fig. 1a. These chemical shift surfaces, Fig. 1b–d and Fig. S2 (ESI†), show a strong dependence of the ^{13}C chemical shift on the side-chain conformation and that these shifts frequently vary by up to 5 ppm between each of the allowed four rotameric states. For ^{13}C$^{\delta1}$ and ^{13}C$^{\gamma2}$ the change in chemical shift between states approximately follows a γ-*gauche* effect,[9,14] whereas the surfaces for ^{13}C$^\alpha$, ^{13}C$^\beta$ and ^{13}C$^{\gamma1}$ are more complex. For example, the surface of $\delta(^{13}C^\beta)$ *vs.* $\{\chi_1, \chi_2\}$ shows a 4 ppm change between the rotameric states even though the dihedral angles between ^{13}C$^\beta$ and its γ-substituents do not change. Frequently ^{13}C$^\alpha$ and ^{13}C$^\beta$ chemical shifts are used for backbone secondary structure determination.[15,16] It is interesting to note that the ^{13}C$^\alpha$ surface shows a variation of about 5 ppm between the commonly populated states, $\{g_m, t\}$ and $\{g_m, g_m\}$, similar to the differences in chemical shift between α-helical and β-sheet conformations.

The five aliphatic ^{13}C chemical shifts for isoleucine, $\boldsymbol{\delta} = \{\delta_{C^\alpha}, \delta_{C^\beta}, \delta_{C^{\gamma1}}, \delta_{C^{\gamma2}}, \delta_{C^{\delta1}}\}$, can be calculated if the populations, $\mathbf{p}$, of the most populated rotameric states are known. For four states, $\mathbf{p} = \{p_{mt}, p_{mm}, p_{pt}, p_{tt}\}$,

$$\boldsymbol{\delta}_{\text{calc}}(\mathbf{p}) = (p_\alpha \mathbf{D}^\alpha + (1 - p_\alpha)\mathbf{D}^\beta)\mathbf{p} \qquad (1)$$

where p_α is the probability of an α-helix backbone conformation and $\mathbf{D}^\alpha$ and $\mathbf{D}^\beta$ are matrices representing the five ^{13}C chemical shifts in each rotameric state for α-helix and β-sheet conformations respectively. Eqn (1) holds here because the exchanges between the rotameric states are in the fast-exchange regime[9,10] and observed chemical shifts represent a population-weighted average. Equally, the populations of the rotameric states, $\mathbf{p}$, can be determined from experimentally observed ^{13}C chemical shifts when the secondary structure is known, since (1) the number of significantly populated states is less than the number of ^{13}C chemical shifts available and (2) the chemical shifts of the rotameric states are linearly independent meaning that the square matrices $(\mathbf{D}^\alpha)^T\mathbf{D}^\alpha$ and $(\mathbf{D}^\beta)^T\mathbf{D}^\beta$ are non-singular. Thus, the populations, $\mathbf{p}$, can be determined by minimising the target function,

$$\chi^2(\mathbf{p}) = \|(\boldsymbol{\delta}_{\text{calc}} - \boldsymbol{\delta}_{\text{obs}}) \circ \mathbf{W}\|_2^2 =$$
$$\|(\mathbf{D}_{\text{calc}}\mathbf{p} - \boldsymbol{\delta}_{\text{obs}}) \circ \mathbf{W}\|_2^2 =$$
$$\|(-(\boldsymbol{\sigma}_{\text{calc}} - \boldsymbol{\sigma}_{\text{ref}})\mathbf{p} - \boldsymbol{\delta}_{\text{obs}}) \circ \mathbf{W}\|_2^2$$
$$\text{subject to } \|\mathbf{p}\|_1^1 = 1 \text{ and } 0 \leq p_i \leq 1 \qquad (2)$$

where $\mathbf{D}_{\text{calc}} = p_\alpha \mathbf{D}^\alpha + (1 - p_\alpha)\mathbf{D}^\beta$; $\boldsymbol{\sigma}_{\text{calc}} = -\mathbf{D}_{\text{calc}} + \boldsymbol{\sigma}_{\text{ref}} = p_\alpha \boldsymbol{\sigma}^\alpha + (1 - p_\alpha)\boldsymbol{\sigma}^\beta$ is a matrix with the isotropic shielding constants for the five ^{13}C chemical shifts in the rotameric states and $\boldsymbol{\sigma}_{\text{ref}}$ is the reference shielding constant. Finally, $\mathbf{W}$ is a vector with the weights of the

five ^{13}C chemical shifts, {1/2.7 ppm, 1/2.0 ppm, 1/1.7 ppm, 1/1.3 ppm, 1/1.7 ppm} determined from the standard deviation of previously assigned ^{13}C chemical shifts in the BMRB database. The DFT calculations, Fig. 1, provide the shielding constants, $\boldsymbol{\sigma}_{\text{calc}}$, for the aliphatic ^{13}C in each rotameric state and in the two backbone conformations, α-helix and β-sheet (ESI†).

Calculated chemical shifts, δ_{calc}, are related to the shielding constants by $\delta_{\text{calc}} = -(\sigma_{\text{calc}} - \sigma_{\text{ref}})$, and an accurate determination of the reference shielding constants is therefore essential. The reference shielding constant were initially estimated by requiring that random coil populations derived from a large set of crystal structures[12] yield random-coil chemical shifts. In a subsequent optimisation, nuclei-specific reference shielding constants were obtained using a comprehensive set of experimentally derived long-range scalar coupling constants, Tables S2–S4 (ESI†). Specifically, long-range $^3J_{C^{\gamma1}-N}$, $^3J_{C^{\gamma1}-CO}$, $^3J_{C^{\gamma2}-N}$, $^3J_{C^{\gamma2}-CO}$, $^3J_{C^{\gamma2}-C^{\delta1}}$ and $^3J_{C^\alpha-C^{\delta1}}$ couplings were measured experimentally for 17 isoleucine residues in two model proteins, T4 lysozyme (T4L) and ubiquitin. The nuclei-specific reference shielding constants, $\sigma_{\text{ref},i}$, $i = \{C^\alpha, C^\beta, C^{\gamma1}, C^{\gamma2}, C^{\delta1}\}$, were optimised by minimising the RMSD between experimentally measured scalar couplings and scalar couplings back-calculated from chemical-shift-derived populations, eqn (2), using standard Karplus curves,[17] Fig. S3a (ESI†). In these optimisations the backbone conformations from crystal structures were used. The correlation between experimental ^3J-coupling constants and ^3J-coupling constants calculated using the optimised constants, Fig. 2a, strongly indicate that rotamer
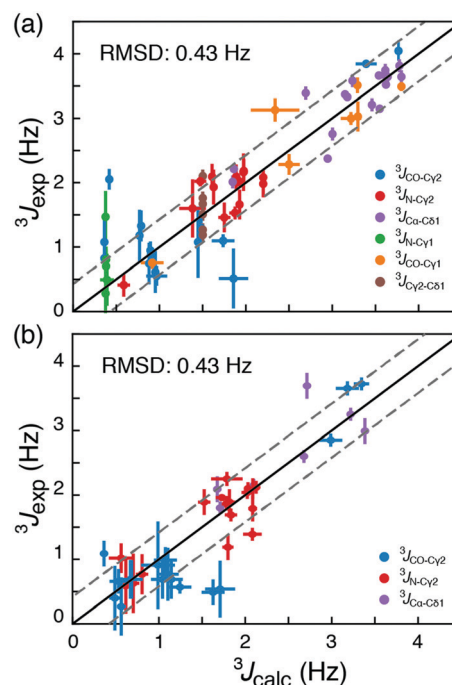

Fig. 2 Comparison of long-range experimentally measured ^3J scalar coupling constants with those derived from ^{13}C chemical shifts. (a) Couplings from the two proteins T4L and ubiquitin used in the optimisation of $\sigma_{\text{ref},i}$. (b) Couplings from the four other proteins: GB3, C-SH2 PLC-γ, HIV protease and protein L published previously. The black line is $y = x$ and the grey lines are $y = x \pm$ RMSD.

distributions, Tables S5 and S6 (ESI†), can be determined from ^{13}C chemical shift. The obtained rotamer distributions also agree well with those present in crystal structures, Tables S5 and S6 (ESI†), although a substantially broader distribution of populated states is observed in solution from ^{13}C chemical shifts and ^3J-couplings.

To assess the accuracy of the method, rotamer populations were determined from ^{13}C chemical shifts for four additional proteins, HIV protease,[18] GB3,[18] C-SH2 PLC-γ1,[19] and protein L,[20] where long-range scalar couplings have been previously obtained. An RMSD of 0.43 Hz was obtained for this cross-validation set, when the back-calculated ^3J-couplings were compared to published experimental ^3J-couplings; Tables S7–S10 (ESI†). This corresponds to an RMSD in populations of the rotameric states between 0.16 and 0.19, Fig. S4 and S5 (ESI†). In a cross-validation akin to Fig. 2, but including the $\{t,g_p\}$ state and using five rotameric states, gives an RMSD between measured and back-calculated ^3J-couplings of 0.43 Hz. From our data it is therefore not statistically justified (p-value = 0.55) to include the additional $\{t,g_p\}$ state in the analysis. While the RMSDs of the obtained populations provide an upper bound for the error in determining rotamer populations from chemical shifts, it is expected that the true standard error is smaller due to uncertainties associated with the Karplus parametrisations (RMSD ≈ 0.04 Hz; Fig. S6, ESI†) and the assumption that ^3J couplings depend only on the intervening dihedral.[16,17]

Until this point only folded proteins were considered and the ϕ, ψ angles of available structures were used to select the set of DFT calculations, $\mathbf{D}^\alpha$ or $\mathbf{D}^\beta$, used for calculating the rotamer populations from ^{13}C chemical shifts. The probability of α-helix, p_α, can be predicted from backbone chemical shifts using TALOS-N.[16] Doing so gives similar RMSD values to those in Fig. 2, 0.42 Hz (Tables S5–S10, ESI†). This means that if an accurate backbone structure is not available, but a backbone chemical shift assignment is, then the backbone conformations can be derived from TALOS-N and still provide a robust description of the side-chain conformations in ca. 95% of cases.

To explore the intrinsic conformational preference of isoleucine side chains, the rotamer distributions were calculated form experimental chemical shifts and scalar couplings for two peptides, Ace-Ile-NMe (AIN) and Gly-Ile-Gly (GIG). For these 'random coil representations' the backbone was assumed to be 50% β-sheet and 50% α-helix; changing this by up to 10% affected the derived populations, $\mathbf{p}$, by less than 0.02. As shown in Fig. 3a the chemical-shift-derived ^3J scalar couplings agree very well with experimental scalar couplings (RMSD = 0.22 Hz). Both the AIN and the GIG peptide show very similar rotamer distributions and a broad similarity between the distributions determined from scalar couplings and from chemical shift is also seen, Table S11 (ESI†). In these random coil models $\{g_p,t\}$, $\{g_m,g_m\}$, and $\{g_m,t\}$ are the three major populated states in agreement with deposited crystal structures. However, the distributions are substantially different from the statistical potential derived from the high resolution crystal structures.[12] The populations derived here are $p_{pt} \approx$ 42%, $p_{mm} \approx$ 35%, and $p_{mt} \approx$ 22%, whereas $p_{pt} \approx$ 14%, $p_{mm} \approx$ 17%, and $p_{mt} \approx$ 57% are obtained from the statistical potential.
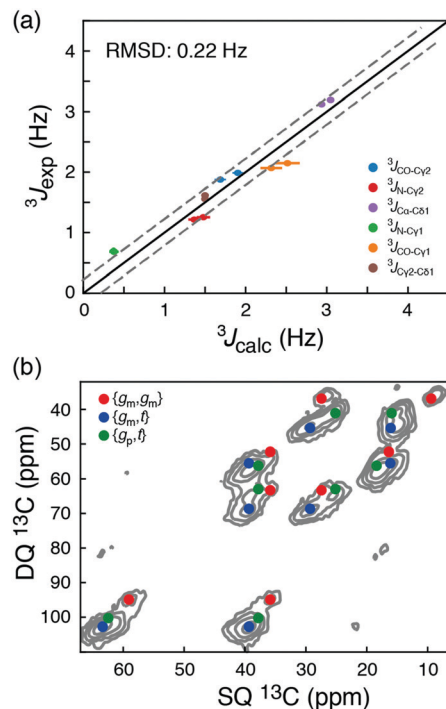

Fig. 3 (a) Comparison of long-range ^3J scalar coupling constants derived from ^{13}C chemical shifts and experimentally measured couplings for the AIN and GIG peptide. (b) A DNP-enhanced solid-state double-quantum-single-quantum (DQSQ) NMR-spectrum of the AIN peptide in frozen solution (100 K). See also Fig. S3 (ESI†).

To explore these random coil models further, a dynamic nuclear polarization (DNP) enhanced solid-state NMR ^{13}C–^{13}C DQSQ spectrum was recorded on a frozen sample of AIN at 100 K. Under these conditions, exchanges between the side-chain rotameric states are so slow that separate NMR signals are observed for each of the major states, Fig. 3b and Fig. S7 (ESI†).[21] Calculating the peak positions using the chemical shift profiles also readily identifies $\{g_m,t\}$, $\{g_m,g_m\}$ and $\{g_p,t\}$ as the major states. Moreover, the populations of the three major states determined from peak volumes in the DQSQ spectrum agree well with those determined in solution from ^{13}C chemical shifts, in particular a substantial population of $\{g_p,t\}$. Importantly, as demonstrated both with solution-state measurements at room temperature and solid-state NMR measurements at 100 K, the random coil sampling of AIN and GIG is significantly different from that obtained from the average of all isoleucine side chains in a large set of high-resolution crystal structures.[12] In particular, the population of $\{g_p,t\}$ found here is substantially larger than what is predicted from the statistical potential derived from crystal structures.

Backbone chemical shifts frequently play an important role in structure determination, providing information on the backbone dihedral angles, yet side-chain chemical shifts are rarely used for structural characterisations. The excellent agreement between back-calculated long-range scalar couplings and those measured experimentally, Fig. 2 and 3, suggests that side-chain chemical shift can also be readily used in protein structure determination protocols. Structural characterisations of low-populated and excited states have emerged over the last decade[2,22]
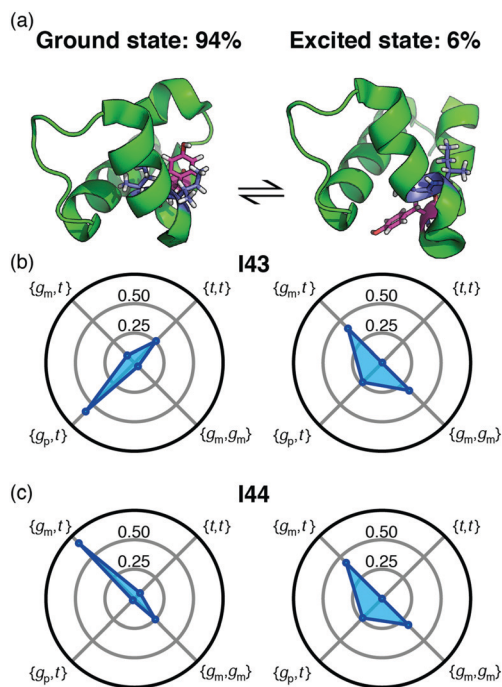
Fig. 4 (a) Exchange between the ground state of the L24A FF domain and a folding intermediate. Tyr49, which change substantially between the two structures is shown in magenta and the isoleucine residues are shown in blue. (b) and (c) The rotamer distributions for Ile43 and Ile44 in each of the two states.

and these characterisations largely hinge on chemical-shifts derived constraints. Recently, side-chain ^{13}C chemical shifts[11] have become available for low-populated states, which allow one to obtain rotamer distributions for side chains in low-populated states using the approach described above. One example is the L24A FF domain, which exchanges between a folded ground state and a protein folding intermediate with an exchange rate[11] of 130 s^{-1}. Side-chain chemical shifts were recently obtained for the folding intermediate using CEST NMR experiments.[11] Chemical-shift-derived side-chain rotamer populations for the isoleucine residues were obtained using backbone conformations derived from TALOS-N, Fig. 4. In the ground state I43 predominantly samples the $\{g_p,t\}$ with $60 \pm 6\%$ and $\{t,t\}$ with $26 \pm 3\%$, while I44 predominantly samples $\{g_m,t\}$ with $68 \pm 10\%$ and $\{g_m,g_m\}$ with $23 \pm 2\%$. This agrees with TALOS-N, where the I43 χ_1 is predicted to be g_p and I44 χ_1 is predicted to be in a g_m conformation. In the folding intermediate both I43 and I44 adopt a much broader distribution with $p_{pt} = 24 \pm 6\%$, $p_{mm} = 33 \pm 1\%$, and $p_{mt} = 42 \pm 6\%$, similar to the random coil distribution.

In conclusion, a close relationship between aliphatic ^{13}C chemical shifts and the conformational sampling of the isoleucine side chain was shown. Using this relationship, side-chain rotamer distributions can be determined directly from the side-chain ^{13}C chemical shifts. The method presented here allows for a determination of full rotameric states represented by both χ angles as opposed to projections along each χ angle independently, as is the case for previous methods[4,5] based on ^3J long-range scalar couplings. Since chemical shifts can be obtained from a wide variety of sources, such as relaxation dispersion experiments, chemical exchange saturation transfer and solid state NMR, this method greatly increases the situations where side-chain conformational samplings can be obtained. Although the method for determining side-chain conformations from chemical shifts here is shown for isoleucine, it is anticipated that similar approaches will allow for characterisations of other side chains in proteins.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 K. M. Anderson, A. Esadze, M. Manoharan, R. Brüschweiler, D. G. Gorenstein and J. Iwahara, *J. Am. Chem. Soc.*, 2013, **135**, 3613–3619.
2 G. Bouvignies, P. Vallurupalli, D. F. Hansen, B. E. Correia, O. Lange, A. Bah, R. M. Vernon, F. W. Dahlquist, D. Baker and L. E. Kay, *Nature*, 2011, **477**, 111–117.
3 C. Zeymer, N. D. Werbeck, S. Zimmermann, J. Reinstein and D. F. Hansen, *Angew. Chem., Int. Ed.*, 2016, **55**, 11533–11537.
4 A. Bax, D. Max and D. Zax, *J. Am. Chem. Soc.*, 1992, **114**, 6923–6925.
5 S. Grzesiek, G. W. Vuister and A. Bax, *J. Biomol. NMR*, 1993, **3**, 487–493.
6 S. F. Cousin, P. Kadeřávek, N. Bolik-Coulon, Y. Gu, C. Charlier, L. Carlier, L. Bruschweiler-Li, T. Marquardsen, J.-M. Tyburn, R. Brüschweiler and F. Ferrage, *J. Am. Chem. Soc.*, 2018, **140**, 13456–13465.
7 D. F. Hansen, P. Vallurupalli and L. E. Kay, *J. Am. Chem. Soc.*, 2009, **131**, 12745–12754.
8 R. E. London, B. D. Wingad and G. A. Mueller, *J. Am. Chem. Soc.*, 2008, **130**, 11097–11105.
9 D. F. Hansen, P. Neudecker and L. E. Kay, *J. Am. Chem. Soc.*, 2010, **132**, 7589–7591.
10 D. F. Hansen and L. E. Kay, *J. Am. Chem. Soc.*, 2011, **133**, 8272–8281.
11 G. Bouvignies, P. Vallurupalli and L. E. Kay, *J. Mol. Biol.*, 2014, **426**, 763–774.
12 V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 12–21.
13 S. C. Lovell, J. M. Word, J. S. Richardson and D. C. Richardson, *Proteins*, 2000, **40**, 389–408.
14 A. E. Tonelli and F. C. Schilling, *Macromolecules*, 1981, **14**, 74–76.
15 D. S. Wishart and B. D. Sykes, *J. Biomol. NMR*, 1994, **4**, 171–180.
16 Y. Shen and A. Bax, *J. Biomol. NMR*, 2013, **56**, 227–241.
17 J. M. Schmidt, *J. Biomol. NMR*, 2007, **37**, 287–301.
18 J. J. Chou, D. A. Case and A. Bax, *J. Am. Chem. Soc.*, 2003, **125**, 8959–8966.
19 L. E. Kay, D. R. Muhandiram, N. A. Farrow, Y. Aubin and J. D. Forman-Kay, *Biochemistry*, 1996, **35**, 361–368.
20 O. Millet, A. Mittermaier, D. Baker and L. E. Kay, *J. Mol. Biol.*, 2003, **329**, 551–563.
21 A. König, D. Schölzel, B. Uluca, T. Viennet, Ü. Akbey and H. Heise, *Solid State Nucl. Magn. Reson.*, 2019, **98**, 1–11.
22 A. J. Baldwin and L. E. Kay, *Nat. Chem. Biol.*, 2009, **5**, 808–814.