Journal of Biomolecular NMR, **24:** 149–154, 2002.
KLUWER/ESCOM

A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics

Mario Schubert[a,b,c], Dirk Labudde[a,*], Hartmut Oschkinat[a,b] & Peter Schmieder[a,*]
[a]*Forschungsinstitut für Molekulare Pharmakologie, Robert-Rössle-Str. 10, D-13125 Berlin, Germany;* [b]*Freie Universität Berlin, Takustr. 3, D-14195 Berlin, Germany;* [c]*Present address: Department of Biochemistry & Molecular Biology, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada*

Received 4 July 2002; Accepted 3 September 2002

Key words: ^{13}C chemical shifts, cis peptide bond, peptide bond conformation, proline, protein structure

Abstract

The chemical shift difference ($\delta[^{13}C^{\beta}] - \delta[^{13}C^{\gamma}]$) is a reference-independent indicator of the Xaa-Pro peptide bond conformation. Based on a statistical analysis of the ^{13}C chemical shifts of 1033 prolines from 304 proteins deposited in the BioMagRes database, a software tool was created to predict the probabilities for cis or trans conformations of Xaa-Pro peptide bonds. Using this approach, the conformation at a given Xaa-Pro bond can be identified in a simple NOE-independent way immediately after obtaining its NMR resonance assignments. This will allow subsequent structure calculations to be initiated using the correct polypeptide chain conformation.

In peptides and proteins, the planar peptide bond occurs predominantly in the trans conformation (Ramachandran and Sasisekharan, 1968). In general the cis form is energetically less favorable due to the steric repulsion of the C^{α}/H^{α} atoms of the two sequential amino acids. However, in peptide bonds preceding prolines (Xaa-Pro), the C^{δ}/H^{δ} in the pyrolidine ring and the C^{α}/H^{α} atoms of the preceding residue experience a comparable repulsion and the energy difference between the cis and the trans conformation is reduced. Therefore an appreciable fraction of the Xaa-Pro peptide bonds occur in the cis form. A recent survey of a non-redundant database of 571 high resolution protein structures found 5.2% of all Xaa-Pro peptide bonds occur in the cis conformation, as compared to only 0.03% of all Xaa-nonPro peptide bonds (Weiss et al., 1998; Jabs et al., 1999). The latter numbers may even be an under-estimate because in case of low resolution data in X-ray crystallography or in the absence of a characteristic NOE pattern for a cis peptide bond in NMR spectroscopy the common programs used for protein structure calculation by default assume a trans peptide bond conformation. Information about the presence of a cis peptide bond thus has to be provided as input for structure determination programs like, X-PLOR (Brünger, 1992; Badger et al., 1999), DYANA (Güntert et al., 1997) or ARIA (Nilges et al., 1997; Linge et al., 2001). Accordingly, a reliable identification of Xaa-Pro peptide bonds in the cis conformation is highly desirable.

The spectroscopic identification of a cis peptide bond with NMR traditionally has relied on the observation of a strong H^{α}-H^{α} NOE between the two sequential residues (Wüthrich, 1986). Chemical shift degeneracies and artifacts originating from insufficient water suppression, however, frequently hinder the identification of such NOE signals and thus a cis peptide bond remains unnoticed. Depending on quality and amount of the remaining restraints, structures with a wrong conformation around the Xaa-Pro peptide bond may result.

Alternatively, it has been noted that an indication of a Xaa-Pro cis conformation in proteins can be obtained from ^{13}C chemical shift data (Stanczyk et al., 1989; Torchia et al., 1989). However, the limited amount of ^{13}C chemical shift data available at that time prevented an extensive statistical analysis

*To whom correspondence should be addressed. E-mails: schmieder@fmp-berlin.de, labudde@fmp-berlin.de

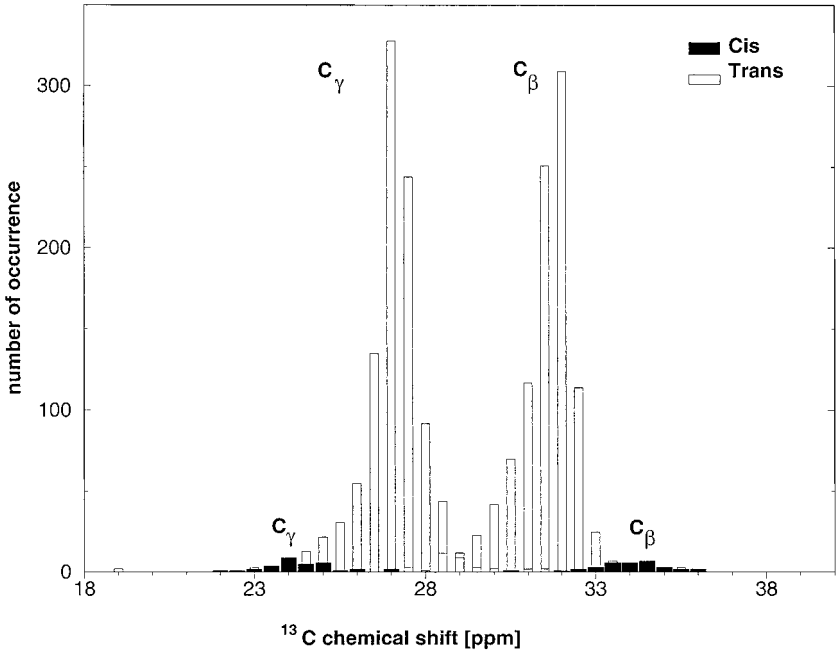

Figure 1. Number of occurrence of Xaa-Pro peptide bond conformations as a function of the proline $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts, plotted in 0.5 ppm intervals. The statistical values are listed in Table 1.

of this correlation. In contrast, ^{13}C chemical shifts have long been used to differentiate between cis and trans Xaa-Pro peptide bonds in small peptides (Dorman and Bovey, 1973). Signature features of the cis conformation include an upfield change in the $^{13}C^{\gamma}$ chemical shift and a downfield change in the $^{13}C^{\beta}$ chemical shift. Thus, the chemical shift difference $\Delta_{\beta\gamma}$ ($\Delta_{\beta\gamma} = \delta[^{13}C^{\beta}] - \delta[^{13}C^{\gamma}]$) was proposed as an indicator for cis or trans conformation (Siemion et al., 1975).

The chemical shift differences can be explained by different puckering of the pyrolidine ring, which is flexible and dynamic on the $\sim 10^{-11} - 10^{-12}$ s time scale (London, 1978; Schmidt et al., 1993), interconverting mainly between two pucker conformations: the UP (C^{γ}-exo) conformation and the DOWN (C^{γ}-endo) conformation (Haasnoot et al., 1981). The UP/DOWN equilibrium is influenced by the peptide bond conformation. Prolines in cis peptide bonds prefer the DOWN pucker due to a small energy difference between UP and DOWN of estimated 1.1 kcal/mol (Haasnoot et al., 1981). In contrast, the two states are close to isoenergetic in a trans peptide bond. Ab initio molecular orbital calculations on Ac-cis-Pro-$NHCH_3$ (Kang, 1996) yielded a population of the DOWN conformation of approx. 85–90% in cis-Xaa-prolines and approx. 50% in trans-Xaa-prolines, in close agreement with the populations estimated by analyzing X-ray structures of proteins (Milner-White et al., 1992) and NMR experiments on Ac-cis-Pro-NH_2 in solution (Haasnoot et al., 1981). The influence of the proline pucker and the angle ψ on the ^{13}C chemical shifts and thus on $\Delta_{\beta\gamma}$ was analyzed by ab initio quantum mechanical calculations (Giessner-Prettre et al., 1987). In agreement with the experimental findings by Siemion et al., the DOWN conformation which is favored by cis-Xaa-prolines shows a larger $\Delta_{\beta\gamma}$ compared to the UP conformation .

Chemical shift values of proteins are much more dispersed in comparison to those of small peptides due to their secondary and tertiary structures. Therefore, a statistical analysis of the ^{13}C chemical shift data of prolines in proteins was carried out in order to verify the applicability of using this information to determine the conformations of Xaa-Pro peptide bonds. A Tcl/Tk-script was written to extract proline $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts, along with the protein identification and the residue number, from the BioMagResBank (BMRB) database (http://www.bmrb.wisc.edu, Seavey et al. 1991). 446 protein entries in the BMRB, representing an overall number of 2051 prolines, were included in the analysis. Paramagnetic proteins were excluded. For only 1033 out of the 2051 prolines from 304 proteins both the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts

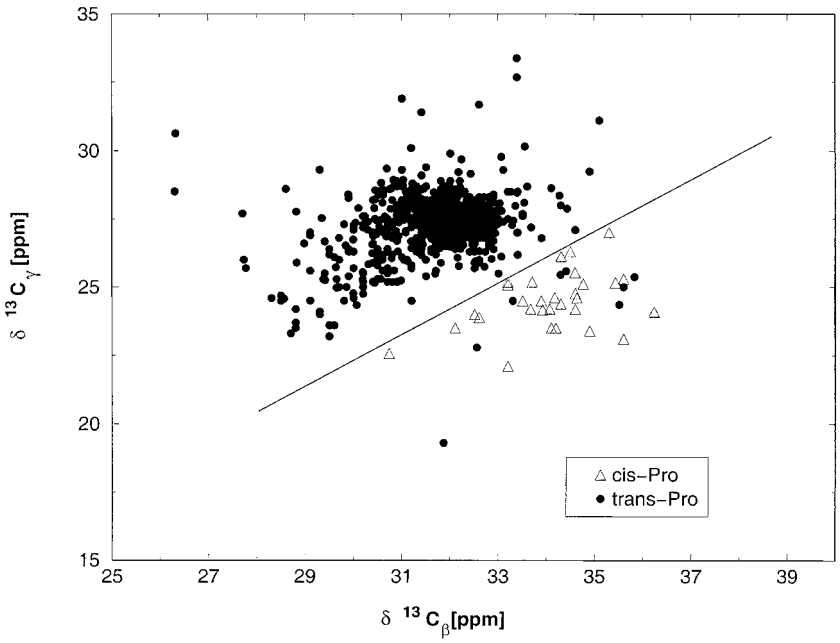

Figure 2. Two-dimensional plot of $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts as a function of the Xaa-Pro peptide bond conformation. The solid line shows approximately the border between the two clusters of cis and trans proline conformations. Eight data points resulting from trans Xaa-Pro peptide bonds occur among the region of cis Xaa-Pro peptide bonds. It is possible that these reflect incorrectly assigned peptide bond conformations.

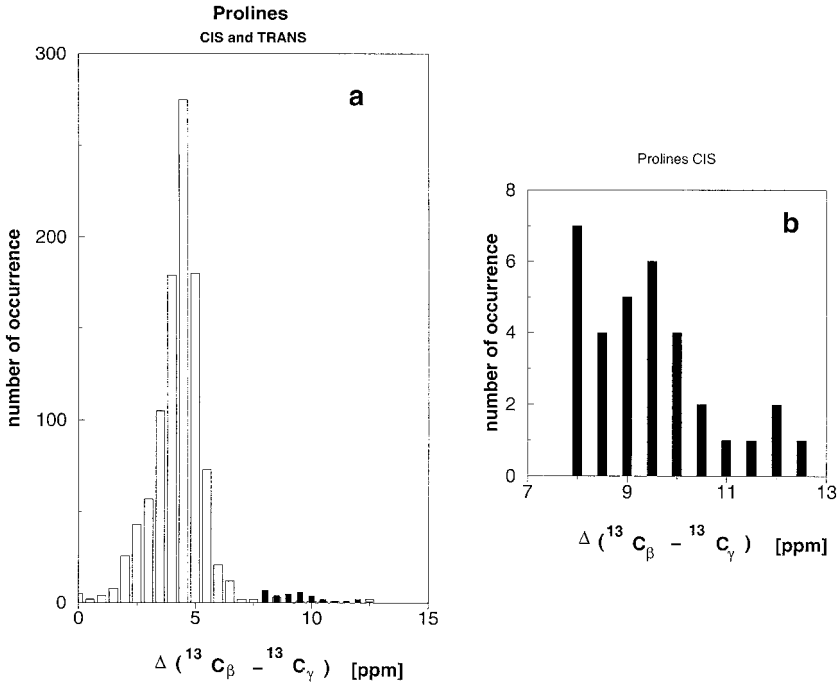

Figure 3. Number of occurrence of the Xaa-Pro peptide bond conformation as a function of the chemical shift difference $\Delta_{\beta\gamma}$ ($\Delta_{\beta\gamma} = \delta[^{13}C^{\beta}] - \delta[^{13}C^{\gamma}]$), plotted in 0.5 ppm intervals. The statistical values are listed in Table 2. (a) both cis and trans conformation are shown; (b) only the cis conformation is shown on a larger scale.

Table 1. Summary of the statistical values of the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts for 1033 prolines in cis or trans conformations

	$^{13}C^{\beta}$ Pro trans	$^{13}C^{\beta}$ Pro cis	$^{13}C^{\gamma}$ Pro trans	$^{13}C^{\gamma}$ Pro cis
Average value $\bar{\delta}$ (ppm)	31.75	34.16	27.26	24.52
Min value δ_{min} (ppm)	26.30	30.74	19.31	22.10
Max value δ_{max} (ppm)	35.83	36.23	33.39	27.01
Standard deviation σ_{δ} (ppm)	0.98	1.15	1.05	1.09

and a crystal or NMR structure were available. Visualization of the corresponding crystal or NMR structures from the PBD database (http://www.pdb.org, Bernstein et al., 1977) revealed 1000 trans and 33 cis Xaa-Pro peptide bonds. The smaller number of cis Xaa-Pro bonds (3.18%) compared to the 5.2% in the literature (Weiss et al., 1998; Jabs et al., 1999) is mainly due to missing $^{13}C^{\gamma}$ chemical shift data. The chemical shifts with the information on the conformation were grouped in 0.5 ppm intervals for statistical analysis as shown in Figure 1. Using the program Grace (http://plasma-gate.weizmann.ac.it/Grace) all statistical standard parameters were calculated (for the resulting histograms/based on the frequency of occurrence), e.g. average values $\bar{\delta}$, standard deviation σ_{δ} and extreme values δ_{min}, δ_{max}, and are given in Table 1. Visual inspection of the histograms in Figure 1 clearly indicates that, on average, the $^{13}C^{\beta}$ chemical shifts of cis-Xaa-Pro are displaced upfield while $^{13}C^{\gamma}$ chemical shifts occur downfield compared to data for trans-Xaa-Pro. There is, however, considerable overlap between the chemical shifts of both conformations due to additional structure-dependent effects in proteins. Thus, consideration of only one chemical shift ($\delta[^{13}C^{\gamma}]$ or $\delta[^{13}C^{\beta}]$) is not a reliable indicator of the peptide bond conformation in proteins.

Figure 2 shows a two-dimensional landscape of $^{13}C^{\gamma}$ versus $^{13}C^{\beta}$ chemical shifts. Because cis-Xaa-prolines tend to have higher $^{13}C^{\beta}$ and lower $^{13}C^{\gamma}$ chemical shift values compared to trans-Xaa-prolines, data for the two Xaa-Pro conformations form two distinct clusters, separated by the approximately diagonal line drawn in Figure 2. Based on this clustering the simplest method to monitor this separation is the analysis of the difference $\Delta_{\beta\gamma}$ of both chemical shifts. In addition, such an approach is a reference independent indicator of Xaa-Pro peptide bond conformations.

The complete histogram showing the distribution of $\Delta_{\beta\gamma}$ is presented in Figure 3. $\Delta_{\beta\gamma}$ ranges from 8.0 to 12.5 ppm and from 0.0 ppm to 12.5 ppm for Xaa-Pro bonds in the cis and trans conformation, respectively. However, the numbers of occurrence for these conformations differ distinctly in the overlap region. The histograms lead to average values $\bar{\Delta}_i$ and standard derivations σ^i_{Δ} ($i = cis,trans$) of 4.51 ppm $\pm$ 1.17 ppm and 9.64 ppm $\pm$ 1.27 ppm for trans- and cis-Xaa-prolines, respectively (Table 2). The parameters obtained from both histograms can be used to derive a Gaussian distribution.

Table 2. Statistical values of the chemical shift difference $\Delta_{\beta\gamma}$ ($\Delta_{\beta\gamma} = \delta\,[^{13}C^{\beta}] - \delta[^{13}C^{\gamma}]$) as a function of the Xaa-Pro peptide bond conformation

	$\Delta_{\beta\gamma}$ Pro trans	$\Delta_{\beta\gamma}$ Pro cis
Average value $\bar{\Delta}$ (ppm)	4.51	9.64
Variance σ^2_{Δ} (ppm^2)	1.37	1.62
Standard deviation σ_{Δ} (ppm)	1.17	1.27

The density $p(t)$ and the error integral $F(x)$ of a Gaussian distribution are defined as:

$$p(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\{(t-\mu)^2/2\sigma^2\}} \tag{1}$$

$$F(x) = \int_{-\infty}^{x} p(t)dt. \tag{2}$$

Using the error integral it is possible to derive the probability of a value x to be found in an interval $(x_1, x_2]$ (Bronstein and Semendjajew, 1985).

A Gaussian distribution with the average value μ and the standard deviation σ can be transformed into a Normalized Gaussian distribution using the relationship

$$\lambda = {}^{(x-\mu)}/_{\sigma}. \tag{3}$$

It is then possible to use tabulated values for the error integral

$$\Phi(\lambda) = \int_0^{\lambda} \phi(t)dt = 1/\sqrt{2\pi} \int_0^{\lambda} e^{-\{t^2/2\}}\, dt \qquad (4)$$

in the calculations. The error integrals can in turn be used to calculate the single probabilities of the cis or trans conformer (P_S^{cis} (Δ) and P_S^{trans} (Δ)) from a given $\Delta_{\beta\gamma}$. The resulting equations are

$$P_S^i(\Delta) = \Phi(\lambda_2^i) - \Phi(\lambda_1^i);\ i(cis, trans). \qquad (5)$$

In our application we chose an integration interval of 0.6 ppm, which leads to

$$\lambda_{1,2}^i = \frac{(\Delta \pm 0.3) - \bar{\Delta}_i}{\sigma_\Delta^i};\ i(cis, trans). \qquad (6)$$

After the standardization of both single probabilities, values in the ranges from 0.0% to 100.0% (P_{cis} (Δ) and P_{trans} (Δ)) can be derived from a given $\Delta_{\beta\gamma}$.

In the range from 0.0 ppm to 4.8 ppm for $\Delta_{\beta\gamma}$ the peptide bond conformation is predicted to be 100% trans, whereas from 9.15 ppm to 14.4 ppm it is 100% cis. In the range from 4.8 ppm to 9.15 ppm, the prediction is ambiguous and only probabilities can be given for both conformers. If the data are ambiguous, the conformation should be confirmed using the conventional NOE-based method: a cis peptide bond Xaa-Pro is indicated by a strong H^{α}-H^{α} NOE in NOESY spectra whereas trans peptide bond by strong H^{α}-H^{δ} NOE (Wüthrich, 1986).

To ease the utilization of $\Delta_{\beta\gamma}$ for the prediction of the Xaa-proline peptide bond conformation the program **POP** (**p**rediction **o**f **p**roline conformations) was created (The source code for POP is available under http://www.fmp-berlin.de/~labudde). $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts serve as input for the prediction. The program calculates the chemical shift difference $\Delta_{\beta\gamma}$ and the single probabilities P_S^{cis} (Δ) and P_S^{trans} (Δ). After standardization of both values the probabilities of the cis and trans conformer, P_{cis} (Δ) and P_{trans} (Δ), are reported.

In conclusion, we have shown that Xaa-Pro peptide bond conformations in proteins can be derived from $^{13}C^{\gamma}$ and $^{13}C^{\beta}$ chemical shifts. With the help of a standard statistical analysis, the densities and the error integral of the distributions of cis and trans conformations were obtained. On the basis of the error integral, the trans/cis conformation for an Xaa-Pro peptide bond is predicted. In most cases the conformation will be identified unambiguously. Because the only input to the program POP are chemical shift data, the method can be easily implemented into high-throughput protein structure determination procedures. The method provides an important parameter for automated or manual protein structure calculation. The a priori knowledge of the presence of a cis Xaa-Pro peptide bond is crucial for automated structure calculation programs, e.g., ARIA (Nilges et al., 1999; Linge et al., 2001) or CANDID/DYANA (Herrmann et al., 2002). Furthermore, in many cases, cis Xaa-Pro peptide bonds occur in exposed loop regions, which may be difficult to analyze structurally due to limited NOE information and/or dynamic effects. Thus we anticipate that use of chemical shift information will aid in the identification and distinction of characteristic turns like type VIa and type VIb.

Acknowledgements

Support from the Forschungsinstitut für Molekulare Pharmakologie is gratefully acknowledged. The work was supported by a grant of the BMBF (01 GG 9812, Leitprojekt 'Proteinstrukturfabrik'). Mario Schubert was supported by the DFG Graduiertenkolleg GRK 80 'Modellstudien' and acknowledges a Feodor-Lynen Fellowship from the Alexander-von-Humboldt Foundation. The authors thank D. Leitner and L. McIntosh for carefully reading the manuscript.

References

Badger, J., Kumar, R.A., Yip, P. and Szalma, S. (1999) *Proteins*, **35**, 25–33.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) *Nucl. Acids Res.*, **28**, 235–242.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O. Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Bronstein, I.N. and Semendjajew, K.A. (1985) *Taschenbuch der Mathematik*, Teubner Verlag, Leipzig.

Brünger, A.T. (1992) *X-Plor, Version 3.1: A system for X-Ray Crystallography and NMR*, Yale University Press, New Haven.

Dorman, D.E. and Bovey, F.A. (1973) *J. Org. Chem.*, **38**, 2379–2383 and references cited therein.

Giessner-Prettre, C., Cung, M.T. and Marraud, M. (1987) *Eur. J. Biochem.*, **163**, 79–87.

Haasnoot, C.A.G., de Leeuw, F.A.A.M., de Leeuw, H.P.M. and Altona, C. (1981) *Biopolymers*, **20**, 1211–1245.

Herrmann, T., Güntert, P. and Wüthrich, K. (2002) *J. Mol. Biol.*, **319**, 209–227.

Jabs, A., Weiss, M.S. and Hilgenfeld, R. (1999) *J. Mol. Biol.*, **286**, 291–304.

Kang, Y.K. (1996) *J. Phys. Chem.*, **100**, 11589–11595.

Linge, J.P., O'Donoghue, S.I. and Nilges, M. (2001) *Meth. Enzymol.*, **339**, 71–90.
London, R.E. (1978) *J. Am. Chem. Soc.*, **100**, 2678–2685.
Milner-White, E.J., Bell, L.H. and Maccallum, P.H. (1992) *J. Mol. Biol.*, **228**, 725–734.
Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.
Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.
Ramachandran, G. N. and Sasisekharan, V. (1968) *Adv. Prot. Chem.*, **23**, 283–437.
Schmidt, J.M., Brüschweiler, R., Ernst, R.R., Dunbrack, R.L., Joseph, D. and Karplus, M. (1993) *J. Am. Chem. Soc.*, **115**, 8747–8756.
Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J.L. (1991) *J. Biomol. NMR*, **1**, 217–236.
Siemion, I.Z., Wieland, T. and Pook, K.-H. (1975) *Angew. Chem.*, **87**, 712–714.
Stanczyk, S.M., Bolton, P.H., Dell'Acqua, M. and Gerlt, J.A. (1989) *J. Am. Chem. Soc.*, **111**, 8317–8318.
Torchia, D.A., Sparks, S.W., Young, P.E. and Bax, A. (1989) *J. Am. Chem. Soc.*, **111**, 8315–8317.
Weiss, M.S., Jabs, A. and Hilgenfeld, R. (1998) *Nat. Struct. Biol.*, **5**, 676.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.