Journal of Biomolecular NMR, **21:** 275–280, 2001.
KLUWER/ESCOM

An easy way to include weak alignment constraints into NMR structure calculations

Hans-Jürgen Sass[a], Giovanna Musco[b], Stephen J. Stahl[c], Paul T. Wingfield[c] & Stephan Grzesiek[a]

[a]*Department of Structural Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland;* [b]*Department of Biological and Technological Research, c/o S.Raffaele Hospital Neuroscience Department, 20132 Milano, Italy;* [c]*National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892-2350, U.S.A.*

Received 22 August 2001; Accepted 4 September 2001

Key words: orientation, protein, pseudo-inverse, residual tensorial coupling, structure calculation

Abstract

We have recently shown that an energy penalty for the incorporation of residual tensorial constraints into molecular structure calculations can be formulated without the explicit knowledge of the Saupe orientation tensor (Moltke and Grzesiek, *J. Biomol. NMR*, 1999, **15**, 77–82). Here we report the implementation of such an algorithm into the program X-PLOR. The new algorithm is easy to use and has good convergence properties. The algorithm is used for the structure refinement of the HIV-1 Nef protein using 252 dipolar coupling restraints. The approach is compared to the conventional penalty function with explicit knowledge of the orientation tensor's amplitude and rhombicity. No significant differences are found with respect to speed, Ramachandran core quality or coordinate precision.

The weak alignment of biomacromolecules and the concomitant possibility to observe residual tensorial couplings of magnetic nuclei has led to a wealth of new structural information in high resolution NMR (Tolman et al., 1995; Tjandra and Bax, 1997). Structural restraints from such couplings are usually incorporated into molecular structure calculation programs by an energy penalty function which depends on the Saupe orientation tensor (Saupe, 1964). In principle for an unknown molecular structure, the largest principal value of this orientation tensor, its rhombicity, but not its Euler angles can be determined from a histogram of measured tensorial couplings (Clore et al., 1998a) provided that the number of measured couplings is sufficiently large and that the distribution of individual chemical group orientations is sufficiently isotropic.

*To whom correspondence should be addressed. E-mail: Stephan.Grzesiek@unibas.ch

In a recent simulation we have shown that a weak alignment energy penalty function can be derived which does not require the explicit knowledge of an orientation tensor (Moltke and Grzesiek, 1999). This formulation is possible because the parameters describing the orientation tensor's amplitudes and orientation enter in linear form into the formula for an expected tensorial coupling (Losonczi et al., 1999; Moltke and Grzesiek, 1999). In contrast, parameters describing the internal degrees of freedom of the macromolecule such as the torsion angles are nonlinear. A typical least squares energy penalty for residual couplings is of the following form:

$$\chi^2(\boldsymbol{\alpha}, \mathbf{S}) = \sum_{\lambda} \left[D_\lambda - \sum_{m=-2}^{2} S_m^* T_\lambda^{2m}(\boldsymbol{\alpha}) \right]^2, \quad (1)$$

where D_λ represents a measured residual dipolar or other tensorial coupling, S_m is the orientation tensor in five-dimensional, irreducible notation, T_λ^{2m} ($\boldsymbol{\alpha}$ describes the theoretical tensorial coupling as a non-

linear function of the internal coordinates α, and the summation λ extends over all measured couplings. The formulation of the modified energy penalty function makes use of a general observation by Golub and Pereyra (Golub and Pereyra, 1973) that linear parameters of a least-squares penalty function can always be eliminated by substituting the analytical solution for the linear part of the minimization problem into the penalty function itself. By the same mechanism, it is also possible to calculate in closed form the derivative of the modified energy penalty with respect to the non-linear parameters (Golub and Pereyra, 1973).

Here we report the implementation of this algorithm into the program X-PLOR (Brünger, 1992). The algorithm is used to calculate a refined structure of the HIV-1 Nef protein based on 252 residual dipolar couplings and 1250 conventional NMR restraints. Incorporation of the dipolar coupling information into the structure calculation increases the Ramachandran core quality from 73 to 81% and decreases the backbone rmsd from 0.81 Å to 0.63 Å. The algorithm is easy to use, has good convergence properties, and compares favorably to the conventional approach using explicit knowledge of the orientation tensor's amplitude and rhombicity. Because the algorithm is able to provide alignment constraints by an implicit Saupe tensor we refer to it as ISAC (Implicit Saupe tensor Alignment Constraint).

The mathematical details of the algorithm have been described in detail before (Golub and Pereyra, 1973; Moltke and Grzesiek, 1999) and will not be repeated here. The energy penalty function for the dipolar (and other residual) coupling restraints and its derivatives were included into X-PLOR version 3.851 in a similar way as the conventional procedure (SANI) added by Tjandra and coworkers (Tjandra et al., 1997). The necessary matrix operations were realized using respective routines from the public LAPACK v. 3.0 and LINPACK libraries without changes (Anderson et al., 1999).

One-bond residual dipolar couplings of amide ^{1}H-^{15}N and alpha $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ groups were determined for samples of uniformly ^{15}N or $^{15}N/^{13}C$ labeled HIV-1 Nef protein [construct Δ2–39, Δ159–173, C206A (Grzesiek et al., 1997)] once weakly aligned in a longitudinally compressed 7% polyacryl amide gel (Sass et al., 2000; Tycko et al., 2000) and once by the interaction with 16 mg/ml Pf1 filamentous bacteriophages (Hansen et al., 1998). Nef protein concentrations ranged between 0.3 and 0.6 mM in 95% H_2O/5% D_2O in 5 mm Shigemi tubes, all other conditions were as described previously (Grzesiek et al., 1997). NMR experiments were carried out at 308 K. Amide ^{1}H-^{15}N dipolar couplings were measured on a 600 MHz BRUKER DRX instrument using the DSSE-experiment (Cordier et al., 1999) with data matrices of 100^{*} (t_N) $\times$ 512^{*} (t_H) data points (where n* refers to complex points) and acquisition times of 60 ms (t_N) and 55 ms (t_H). $^{1}H^{\alpha}$–$^{13}C^{\alpha}$ dipolar couplings data were determined on an 800 MHz BRUKER DRX instrument using a constant-time (27 ms) ^{1}H-^{13}C HSQC experiment with J_{CH} modulation of resonance intensities in a third dimension (Tjandra and Bax, 1997). Water suppression was accomplished by gradient selection. Data matrices consisted of 18 (t_J) $\times$ 152^{*} (t_C) $\times$ 512^{*} (t_H) data points with acquisition times of 26 ms (t_J), 27 ms (t_C), and 53 ms (t_H). Data processing and analysis were carried out using the NMRPipe package (Delaglio et al., 1995) and the program PIPP (Garrett et al., 1991).

Estimates for the amplitude and rhombicity of the orientation tensor were derived from a powder pattern analysis of the measured dipolar couplings as well as from the 3.0 Å X-ray structure of HIV-1 Nef (PDB code 1AVV) and the NMR structure 2NEF (Table 1). Clearly, the powder pattern estimate for the gel orientation deviates substantially in its rhombicity from all other estimates using either X-ray or NMR structures or the final results of the simulated annealing (SA) run. This is caused by a poor sampling on the directions of the unit sphere. Likewise, as compared to the other data, the amplitude of the orientation tensor derived from the NMR structure for the phage data seems severely underestimated (Table 1). In addition, the NMR quality factors Q of the 2NEF show larger deviations between measured and calculated dipolar couplings than the 1AVV structure despite the fact that the rmsd of heavy atom backbone positions between the two structures is smaller than 1.5 Å. A detailed analysis of the deviations revealed that these are mainly caused by small local rearrangements of the H^{N}-N and H^{α}-C^{α} vectors and not as preliminary assumed by a slight arrangement of one alpha helix (Sass et al., 2000). For these reasons, the amplitudes and rhombicities derived from the X-ray structure were considered as the best initial estimates for the SANI energy function.

In order to validate the approach for structure calculations without knowledge of the orientation tensor, the structure of HIV-1 Nef was calculated with the same NOE and torsion angle restraints from three-bond couplings as in the original work (Grzesiek et al.,

Table 1. Amplitudes A_{zz} and rhombicities η of the orientation tensors of HIV-1 Nef determined by different methods

		Powder analysis[a]	1AVV[b]	2NEF	ISAC	SANI
Pf1 phage	A_{zz}[c]	17.2	16.1	11.2 ± 0.7	17.4 ± 0.4	16.6 ± 0.1
	η	0.80	0.84	0.71 ± 0.1	0.97 ± 0.03	0.88 ± 0.02
	Q[d]	–	0.44	0.73 ± 0.03	0.17 ± 0.01	0.21 ± 0.02
Gel	A_{zz}	−9	−7.24	−5.19 ± 0.40	−8.98 ± 0.34	−8.22 ± 0.13
	η	0.66	0.14	0.24 ± 0.09	0.09 ± 0.05	0.14 ± 0.04
	Q	–	0.56	0.77 ± 0.03	0.34 ± 0.02	0.41 ± 0.03

[a]The amplitude and rhombicity of the orientation tensor were determined from a powder pattern analysis (Clore et al., 1998a) of the combined and scaled ^{1}H-$^{13}C^\alpha$ and ^{1}H-^{15}N dipolar couplings. For the phage data, the maximum likelihood method (Warren and Moore, 2001) was used to calculate the tensor parameters. For the gel data, this method did not converge and the amplitude and rhombicity were estimated from the extrema of the powder profile.

[b]Amplitude and rhombicity were determined from the X-ray structure 1AVV, the NMR structure 2NEF, or the final 40 structures generated by the ISAC or SANI methods, respectively.

[c]In Hz. The dipolar coupling D is calculated from A_{zz} and η as $D = A_{zz}(3\cos^2\theta - 1 + \eta\sin^2\theta\cos 2\phi)/2$ where θ and ϕ are the polar angles of the internuclear distance vector in the principal axis system of the orientation tensor.

[d]The NMR quality factor Q is defined as the ratio of the rmsd between observed and calculated couplings and the rms of the observed couplings (Cornilescu et al., 1998).

1997) including once the dipolar couplings and energy function by the ISAC method and once with an energy function (SANI) as proposed by Tjandra and coworkers (Tjandra et al., 1997) containing explicit knowledge of orientation tensor amplitude and rhombicity from the 1AVV structure. Clearly, such an approach biases the SANI calculation towards the X-ray data and will yield more favorable results for the SANI method than could be expected for a completely unknown structure. A total of 160 SA structures was calculated by either the ISAC or the SANI method. Subsets of the 40 lowest total energy structures were then selected from each of these sets and used for further numerical comparison.

For the two lowest energy structures resulting from these two approaches, Figure 1 shows a comparison between the total energies and between the dipolar energy contributions as a function of the SA steps during the structure calculations. Clearly, the total energy is almost indistinguishable between both methods with the ISAC method yielding a slightly lower (1189 kcal/mol) final energy than the SANI routine (1251 kcal/mol). Similarly, the dipolar parts of the energy for the ISAC and SANI routines are very close, with the ISAC method yielding slightly smaller contributions for most of the time points in the SA run.

This behavior of the dipolar energy reveals some of the underlying mathematical properties of the ISAC optimization. At the beginning (0–10 ps) of the SA run, the dipolar contribution to the total energy is scaled by a factor of approximately 10^{-4} which increases exponentially to 1 towards the end of the run (45 ps, Figure 2).

Therefore during the initial SA steps, forces resulting from the dipolar energy derivatives are very small and have very little influence on the movement of the molecular coordinates. Nevertheless at this stage, the ISAC method yields about 20% less dipolar energy than the SANI method. This is the result of the implicit linear least squares optimization which always adjusts the orientation, the amplitude, and rhombicity of the orientation tensor to the global minimum position for a given set of internal coordinates $\boldsymbol{\alpha}$. This linear optimization does not require any movement of the molecule itself. In contrast, the SANI method has no flexibility for adjusting the amplitude and rhombicity and needs to reorient the molecule relative to a second virtual molecule which represents the alignment frame. In the SANI method, this reorientation only happens at about 30–35 ps when the dipolar alignment forces acting on the molecule become noticeable due to the increase in the dipolar energy scaling factor. At this point, the dipolar energies for both methods become almost indistinguishable. At the final stage of the SA run (45 ps), the dipolar energy for the ISAC method is again slightly (∼10%) smaller than for SANI. This fact is attributed to the fine tuning of the amplitude and rhombicity in the ISAC method.

Amplitudes and rhombicities of the final orientation tensors derived from the 40 best structures of the

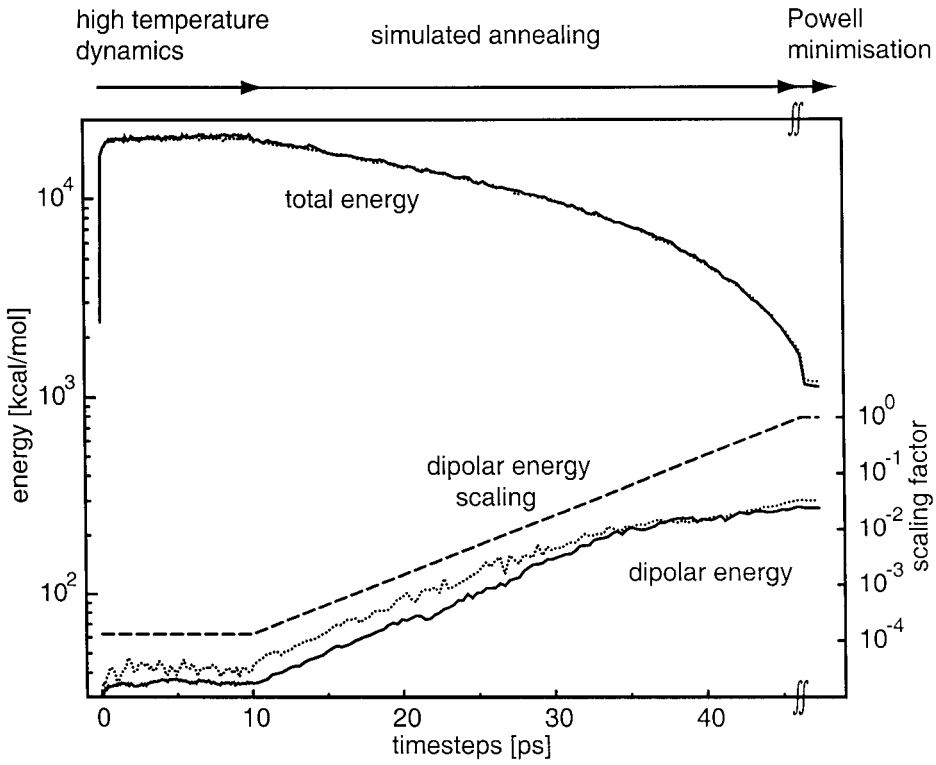

Figure 1. Comparison of simulated annealing runs of HIV -1 using SANI and ISAC dipolar coupling constraints. Data correspond to both lowest energy structures (of 160 total structures) for ISAC (continuous lines) and SANI (dotted lines) procedures. Energies are indicated as the total energy (upper traces) and dipolar energy contribution (lower traces). Individual data points correspond to every 103rd of a series of simulated annealing steps of 3 fs duration. The SA runs consist of a heating period at 3000 K for 10 ps followed by a slow cooling process from 10 to 45 ps during which the temperature is lowered to 100 K and the energy contribution of the experimental constraints is continuously increased by scaling factors. The scaling factor for the dipolar energy contribution is indicated by a broken line. After 45 ps, the molecular dynamics run is followed by a 250 step Powell minimisation.

ISAC and SANI algorithms are also listed in Table 1. Variations between the tensors amount to maximally 9% in the amplitudes (A_{zz}) and to 0.09 in the rhombicities (η). These tensor parameters are also very close to the parameters derived from the 1AVV structure, but deviate somewhat more with respect to the 2NEF structure. A comparison to the powder pattern tensor reveals a reasonable agreement for the amplitudes in both alignment media, but shows a significant deviation for the rhombicities of the gel data (Table 1).We therefore conclude that the determination of the orientation tensor by the ISAC method is more robust than the powder pattern method for cases of poor sampling.

As expected, the inclusion of the 252 dipolar coupling restraints considerably improves the quality of the Nef structure (Table 2). No significant differences are found between using either the ISAC or the SANI routines. For the best 40 structures the precision increases from an RMSD of 0.81 $\pm$ 0.19 to 0.63 $\pm$ 0.13 (0.62 $\pm$ 0.12) Å for the non-mobile backbone heavy atoms and from 1.3 $\pm$ 0.2 to 1.1 $\pm$ 0.2 (1.1 $\pm$ 0.1) Å for all heavy atoms of non-mobile residues by the ISAC (SANI) method. A similar improvement is obtained for the quality of the structures as judged from the percentage of residues in the most favorable region of the Ramachandran plane. This population increases from 72.5 to 80.8% for the selection of non-mobile residues in the Nef core region. Further improvements are obtained for the regularized average structure where the Ramachandran core population increases to 83.3% and other terms such as overall, dipolar, and van der Waals energy, NOE, dihedral violations also become better than the average values for the subset of the 40 best structures (Table 2).

Two additional mathematical properties of the ISAC method should be discussed. (1) In cases where individual estimates ΔD_λ are available for errors of residual coupling measurements, these errors are eas-

Table 2. Structural statistics of the refined HIV-1 Nef structure[a]

RMSDs from experimental distance constrains (Å)			
all (818)	0.079	± 0.005	(0.074)[b]
intraresidual NOEs (70)	0.048	± 0.009	(0.041)
sequential NOEs ($\|i - j\| = 1$) (338)	0.078	± 0.007	(0.071)
short range NOEs ($1 < \|i - j\| <= 5$) (101)	0.094	± 0.005	(0.087)
long range NOEs ($\|i - j\| > 5$) (245)	0.081	± 0.007	(0.081)
H-bonds (64)[c]	0.100	± 0.008	(0.110)
RMSDs (Hz) from dipolar coupling constraints			
oriented in polyacryl amide gels (118)[d]	1.18	± 0.08	(1.09)
oriented by Pf1 phages (134)[e]	1.31	± 0.09	(1.20)
RMSDs from experimental dihedral constraints (°) (157)[f]	0.76	± 0.13	(0.67)
RMSDs from $^3J_{HNHA}$ coupling constant (Hz) (91)	0.98	± 0.10	(0.98)
RMSDs from experimental secondary shifts (ppm)			
$^{13}C^{\alpha}$ (93)	1.32	± 0.05	(1.27)
$^{13}C^{\beta}$ (91)	1.31	± 0.07	(1.27)
Deviation from the idealized covalent geometry			
bonds (Å)	0.0063	± 0.0002	(0.0057)
angles (°)	0.85	± 0.04	(0.87)
improper[g] (°)	0.65	± 0.04	(0.69)
ELJ (kcal/mol)[h]	−370	± 10	(−527)
Coordinate precision (Å)[i]			
backbone non-hydrogen atom	0.63	± 0.13	
all non-hydrogen atoms	1.10	± 0.16	
Percentage of non-gly, non-pro residues in Ramachandran regions[j]			
core	80.8 (83.3)		
allowed	19.0 (16.7)		
generous	0.2 (0.0)		
disallowed	0.0 (0.0)		

[a]The statistics were obtained from a subset of the 40 best energy structures following the simulated annealing protocol with dipolar restraints incorporated by the ISAC routines as described in the text. Individual simulated annealing structures are fitted to each other using residues 76 to 94, 97 to 102, 106 to 147, 181 to 191, and 194 to 199. The number of the various constraints is given in parentheses.
[b]Values in parentheses correspond to data obtained from an average of the 40 structures regularized by a short simulated annealing run (300 K to 100 K in 20 steps of 30 fs) followed by a 1000 step Powell minimization.
[c]For each backbone hydrogen bond constraint, there are two distance restraints: r_{NH-O}, 1.7–2.5 Å, r_{N-O}, 2.3–3.5 Å.
[d]Consisting of 60 ^1H-^{15}N and 58 ^1H-$^{13}C^{\alpha}$ dipolar one-bond couplings.
[e]Consisting of 72 ^1H-^{15}N and 62 ^1H-$^{13}C^{\alpha}$ dipolar one-bond couplings.
[f]The dihedral angle constraints comprise 78 ϕ, 10 ψ, 55 χ^1, and 14 aromatic χ^2 angles.
[g]The improprer torsion restraints serve to maintain planarity and chirality.
[h]ELJ is the Lennard-Jones van der Waals energy calculated with the CHARMM empirical energy function (Brooks et al., 1983) and is not included in the target function for simulated annealing and restrained minimization.
[i]The coordinate precision is defined as the average rms difference between the individual simulated annealing structures and the mean coordinates. Values are reported for residues 58, 72–95, 97–102, 106–148, 181–191, 194–202, i.e., for residues which do not exhibit large amplitude internal motions on the nanosecond time scale (Grzesiek et al., 1997).
[j]These values are calculated with the programm PROCHECK-NMR (Laskowski et al., 1996). Values are reported for the non-mobile residues[i].

ily incorporated into the least squares penalty in a statistically correct way by scaling of the differences:

$$\chi^2(\alpha, S) = \sum_{\lambda} \left[\frac{D_\lambda - \sum_{m=-2}^{2} S_m^* T_\lambda^{2m}(\alpha)}{\Delta D_\lambda} \right]^2 . \quad (2)$$

It is obvious that Equation 2 becomes identical to Equation 1 when the individual couplings D_λ and the functions T_λ^{2m} (α) are scaled by ΔD_λ. Therefore the ISAC method is easily adapted to this case and our current implementation contains this scaling. (2) For cases where residual couplings are determined in the presence of local mobility, Bax and coworkers proposed to use half-open potentials to implement lower limits (Ottiger et al., 1998). Clearly such half-open potentials cannot be expressed in form of the harmonic potentials in Equations 1 or 2 and the Golub–Pereyra algorithm cannot be used. In such cases, it is possible to split the measured couplings into subsets with harmonic potentials and subsets with half-open (or any other form) potentials. At every step of the molecular dynamics run, the optimal orientation tensor can then be taken from the output of the ISAC routine and used as input for the non-harmonic potential. If the number of non-harmonic couplings is small compared to the harmonic couplings, the introduced error will be small.

In summary we have shown that the inclusion of the ISAC energy penalty into a simulated annealing protocol is numerically stable and has good convergence properties. The method is especially user-friendly since only the measured residual couplings are used as input and no estimate for the orientation tensor's amplitude and rhombicity is needed. As compared to the conventional approach, the increase in computational time was found to be very moderate (8.7 % on a SUN SPARC 20 workstation).

The method should yield the best estimates for the orientation tensor in cases where the powder pattern determination (Clore et al., 1998a; Warren and Moore, 2001) of amplitude and rhombicity fails due to poor sampling of the directions on the three-dimensional unit sphere. In the present work only dipolar couplings were used. However the algorithm is easily adaptable to other rank-two interactions such as residual CSA shifts. The method is clearly limited to cases where the five parameters of the orientation tensor are overdetermined by a considerably larger number of measured residual coupling constraints. In practice, this condition is usually met.

Acknowledgements

We are grateful to Nico Tjandra for the X-PLOR code of the original SANI procedure and helpful discussions. The necessary subroutines for the ISAC algorithm implemented in X-PLOR or CNS are available from the authors on request. This work was supported by SNF grant 31-61757.00 (S.G.) and EMBO fellowship ALTF621-1999 (G.M.)

References

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. and Sorenson, D. (1999) *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia.

Brünger, A. (1992) *X-PLOR Version 3.1: A system for Crystallography and NMR*, Yale University New Haven, CT, U.S.A.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1988a) *J. Magn. Reson.*, **133**, 216–221.

Cordier, F., Dingley, A.J. and Grzesiek, S. (1999) *J. Biomol. NMR*, **13**, 175–180.

Cornilescu, G., Marquart, J.L., Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Garrett, D.S., Powers, S., Gronenborn, A.M. and Clore, G.M. (1991) *J. Magn. Reson.*, **95**, 214–220.

Golub, G.H. and Pereyra, V. (1973) *SIAM J. Nummer. Anal.*, **10**, 413–432.

Grzesiek, S., Bax, A., Hu, J.S., Kaufman, J., Palmer, I., Stahl, S.J., Tjandra, N. and Wingfield, P.T. (1997) *Protein Sci.*, **6**, 1248–1263.

Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–74.

Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–86.

Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.

Moltke, S. and Grzesiek, S. (1999) *J. Biomol. NMR*, **15**, 77–82.

Ottiger, M., Delaglio, F., Marquardt, J.L., Tjandra, N. and Bax, A. (1998) *J. Magn. Reson.*, **134**, 365–369.

Sass, H.-J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2000) *J. Biomol. NMR*, **18**, 305–311.

Saupe, A. (1964) *Z. Naturforsch.*, **19a**, 161–171.

Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.

Tjandra, N. and Bax, A. (1997) *J. Magn. Reson.*, **124**, 512–515.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad Sci. USA*, **92**, 9279–9283.

Tycko, R., Blanco, F.J. and Ishii, Y. (2000) *J. Am. Chem. Soc.*, **122**, 9340–9341.

Warren, J.J. and Moore, P.B. (2001) *J. Magn. Reson.*, **149**, 271–275.