The Impact of Direct Refinement against Three-Bond HN–C"H Coupling Constants on Protein Structure Determination by NMR

DANIEL S. GARRETT, JOHN KUSZEWSKI, TIMOTHY J. HANCOCK,* PATRICA J. LODI, GEERTEN W. VUISTER, ANGELA M. GRONENBORN,† AND G. MARIUS CLORE‡

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received January 11, 1994

Over the past few years considerable improvements have been made in the precision and accuracy of protein structures determined by NMR such that it is possible to obtain structures of approximately the same quality as 2.0–2.5 Å resolution crystal structure (1, 2). These high-resolution structures have been based on a large number of approximate interproton distance restraints (an average of 15 or more per residue), supplemented by loose torsion angle restraints derived from coupling constant measurements and systematic conformational grid searches. A few years ago, Kim and Prestegard (3) proposed that refinement against $J_{HN,Ha}$ coupling constants, which are directly related to the $\phi$ backbone torsion angles, be also included in the structure determination. They were not able, however, to assess the real impact of this procedure because methods for either measuring accurate $J_{HN,Ha}$ couplings or determining very-high-resolution protein NMR structures were not available at the time.

Recently a new 3D NMR experiment, known as HNHA, has been developed which allows $J_{HN,Ha}$ coupling constants of larger (> 10 kDa) $^{15}$N-enriched proteins to be obtained at an accuracy of 0.5 Hz or better by measuring the diagonal-peak to cross-peak intensity ratio in a 3D $^{15}$N-separated quantitative J-correlation spectrum (4). To assess the impact of incorporating $J_{HN,Ha}$ coupling constants directly into the target function employed in the structure calculation, we have measured $J_{HN,Ha}$ coupling constants using the HNHA experiment for three proteins, namely the IgG binding domain of protein G (56 residues), interleukin-4 (133 residues), and interleukin-1β (153 residues), for which there exist both high-resolution NMR (5–7) and crystal structures (8–12). We show that the experimental coupling constants obtained using the HNHA experiment are in better agreement with the X-ray structures than the original NMR structures. However, the latter are easily refined against the coupling constants to yield excellent agreement between calculated and experimental values with minimal shifts in atomic coordinates and no impairment in covalent geometry, nonbonded contacts, or agreement with the interproton distance and torsion angle restraints.

The NMR structures for the three proteins chosen for study were all determined on the basis of an average of more than 15 experimental restraints per residue, and the backbone atomic root mean square (rms) difference between the ensemble of calculated structures and their respective mean coordinate positions was less than 0.5 Å (5–7). In addition, the structures exhibited no violations in interproton distance or torsion angle restraints greater than 0.3 Å or 2°, respectively, displayed very small deviations from idealized covalent geometry, and exhibited good nonbonded contacts (as exemplified by large negative values for the calculated 6-12 Lennard-Jones van der Waals energy) (5–7). The X-ray structures of protein G, interleukin-4 (IL-4), and interleukin-1β (IL-1β) were solved at 1.67 Å (8), 2.25 Å (9), and 2.0 Å (10–12) resolution, respectively, and all were refined to R factors less than 20% with good stereochemistry and covalent geometry.

All three protein samples were uniformly (>95%) $^{15}$N labeled, and their expression, purification, and spectral assignment have been described previously (5–7, 13). The NMR experiments were recorded on ~1.5 mM samples dissolved in 90% H$_2$O/10% D$_2$O (at 26°C for protein G, and 36°C for IL-4 and IL-1β) using a Bruker AMX600 spectrometer equipped with a Bruker self-shielded z-gradient triple-resonance probe. The 3D HNHA experiment was carried out using sine-bell pulse field gradients (10 G/cm at the center of the sine bell) exactly as described by Vuister and Bax (4).

All the spectra were recorded with an acquisition time of 10.56 ms and 48 complex points in $t_1$ (1H), and an acquisition time of 51.22 ms and 512 complex points in $t_2$ (1H). The spectra of IL-4 and IL-1β were recorded with an acquisition time of 40.8 ms and 50 complex points in $t_2$ (15N), while that for protein G was recorded with an acquisition time of 34.272 ms and 42 complex points in $t_2$ (15N).

* Present address: Center for Molecular Bioscience and Biotechnology, Lehigh University, 111 Research Drive, Bethlehem, Pennsylvania 18015.
† To whom correspondence should be addressed.

Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.
data were processed using the in-house processing software nmrPipe (F. Delaglio, unpublished). The intense water resonance was digitally filtered along $t_1$ (14) prior to apodization with a 63° shifted sine bell, zero-filling to 1024 complex points, Fourier transformation, and phasing. The data were then Fourier transformed along $t_1$, prior to mirror-image linear prediction (16) along $t_2$ to predict 42 complex points, prior to zero-filling to 128 complex points, Fourier transformation, and phasing. Finally, the data in $t_1$ were inverse Fourier transformed, followed by forward-backward linear prediction (15) along $t_1$ to predict 24 complex points prior to apodization with a 90° shifted square sine bell, zero-filling to 256 complex points, Fourier transformation, phasing, and shifting the data by 45 complex points.

Peak assignments, positions, and intensities were determined using the in-house programs PIPP (17) and STAPP (D.S.G., unpublished). The absolute intensities for the diagonal ($I_{\text{diag}}$) and cross ($I_{\text{cross}}$) peaks were then employed to determine the experimental $^3J_{\text{HN,Ha}}$ coupling constant values using the equation

$$^3J_{\text{HN,Ha}} = C_{\text{relax}} \tan^{-1} \left( \frac{I_{\text{cross}}}{I_{\text{diag}}} \right)^{1/2} \left( \frac{2\pi T}{1\text{ms}} \right),$$

where $C_{\text{relax}}$ is a correction term for the faster relaxation of the 2$\text{I}_2$ $^3J_{\text{HN,Ha}}$ antiphase magnetization compared to that of the in-phase $^3J_{\text{HN,Ha}}$ magnetization, and 2$\text{I}_2$ is the time (26.1 ms) during which dephasing occurs as a result of the $^3J_{\text{HN,Ha}}$ coupling. A value of 1.0555 (corresponding to a $T_{\text{relax}}$ of 200 ms) was employed for $C_{\text{relax}}$ in the case of protein G, while a value of 1.11 for $C_{\text{relax}}$ (corresponding to a $T_{\text{relax}}$ of 100 ms) was used for the larger IL-4 and IL-1β proteins (4).

It is important to note that $I_{\text{cross}}$ and $I_{\text{diag}}$ represent the true intensities observed directly from the NMR experiments and were not the result of spectral deconvolution. Accurate $^3J_{\text{HN,Ha}}$ coupling constants were obtained for 52 (of 56) residues of protein G, 99 (of 133) residues of IL-4, and 113 (of 153) residues of IL-1β. For the remaining residues, diagonal-peak overlap precluded the measurement of accurate $^3J_{\text{HN,Ha}}$ coupling constants.

Table 1 provides a summary of the agreement between the measured and calculated $^3J_{\text{HN,Ha}}$ coupling constants for the NMR and crystal structures of protein G, IL-4, and IL-1β. The calculated $^3J_{\text{HN,Ha}}$ coupling constants were obtained from the Karplus equation

$$J(\phi) = A \cos^2(\phi - 60) + B \cos(\phi - 60) + C$$

with values of 6.51, -1.76, and 1.60 for $A$, $B$, and $C$, respectively (4). These parameters are the result of nonlinear optimization to yield the best fit between the measured $^3J_{\text{HN,Ha}}$ values for Staphylococcal nuclease in solution and the known backbone $\psi$ torsion angles of its highly refined 1.65 Å resolution crystal structure (4). For Staphylococcal nuclease the rms difference between the observed $^3J_{\text{HN,Ha}}$ couplings and calculated $^3J_{\text{HN,Ha}}$ values were 1.82 Å for the crystal structure. The agreement with the calculated values is excellent, indicating that the Karplus equation is a good fit for these data.
plings and the \( ^3J_{\text{HNHe}} \) couplings calculated from the crystal structure is 0.73 Hz (4). A comparison between the Karplus curve generated by Eq. [2] and the calculated values of \( ^3J_{\text{HNHe}} \) as a function of \( \phi \) for the X-ray and NMR structures of protein G, IL-4, and IL-1\( \beta \) is shown in Figs. 1A and 1B, respectively.

From the data presented in Table 1 and Figs. 1A and 1B, it is clear that the agreement between observed and calculated \( ^3J_{\text{HNHe}} \) values is consistently better for the X-ray structures than the corresponding NMR structures. (Note that this was not altered by any attempts to reparameterize Eq. [2] on the basis of either the NMR or the X-ray structures of protein G, IL-4, or IL-1\( \beta \).) This is evident not only from the larger rms differences but also by the larger number of deviations greater than 2 Hz for the NMR structures relative to the corresponding X-ray structures. Interestingly, in both the NMR and the X-ray structures the location of residues with deviations > 2 Hz is distributed approximately equally between secondary structure elements (particularly at their beginning and end) and connecting (i.e., loops and turns) segments. When \( ^3J_{\text{HNHe}} \) couplings corresponding to deviations of more than 2 Hz between observed and calculated values are excluded, the rms difference between the observed and calculated values is significantly improved for the NMR structures although it is still 10–20% worse than that for the corresponding X-ray structures.

FIG. 1. Comparison of the calculated values of the \( ^3J_{\text{HNHe}} \) coupling constants as a function of the backbone torsion angle \( \phi \) for the (A) X-ray structures, (B) original NMR structures, and (C) \( J \)-refined NMR structures of protein G (\( \bigcirc \)), IL-4 (\( \bigtriangleup \)), and IL-1\( \beta \) (\( \bigcirc \)) with the theoretical Karplus curve obtained using Eq. [2] with values of 6.51, -1.76, and 1.60 for the \( A \), \( B \), and \( C \) parameters, respectively (4). The dashed lines represent the theoretical Karplus curves calculated with values of \( A \), \( B \), and \( C \) set to \( \pm \) their standard deviations reported in Ref. (4).
TABLE 2  
Effect of J Refinement on the Precision of the Backbone Coordinates and on the Atomic rms Shifts in the Mean Backbone Coordinate Positions

<table>
<thead>
<tr>
<th>Backbone atomic rms differences (Å)</th>
<th>Protein G' (residues 1–56)</th>
<th>IL-4' (residues 8–129)</th>
<th>IL-1β' (residues 3–151)</th>
</tr>
</thead>
<tbody>
<tr>
<td>⟨NMR_{original}⟩ vs NMR_{original}</td>
<td>0.27 ± 0.03</td>
<td>0.44 ± 0.03</td>
<td>0.41 ± 0.04</td>
</tr>
<tr>
<td>⟨NMR_{refined}⟩ vs NMR_{refined}</td>
<td>0.26 ± 0.04</td>
<td>0.49 ± 0.04</td>
<td>0.46 ± 0.05</td>
</tr>
<tr>
<td>NMR_{refined} vs NMR_{original}</td>
<td>0.27</td>
<td>0.41</td>
<td>0.46</td>
</tr>
<tr>
<td>NMR_{original} vs X-ray</td>
<td>1.14</td>
<td>1.35</td>
<td>0.87 ± 0.02</td>
</tr>
<tr>
<td>NMR_{refined} vs X-ray</td>
<td>1.08</td>
<td>1.30</td>
<td>0.81 ± 0.02</td>
</tr>
</tbody>
</table>

a ⟨NMR_{original}⟩ and ⟨NMR_{refined}⟩ are the ensemble of simulated annealing structures obtained before (5–7) and after J refinement (the number of structures in each ensemble is given in footnotes b, d, e, and f of Table 1). NMR_{original} and NMR_{refined} are the mean coordinates obtained by averaging the coordinate positions of the individual NMR_{original} and NMR_{refined} structures, respectively best-fitted to each other. The residues used in the best-fitting and the computation of the rms differences are 1–56, 8–128, and 3–151 for protein G, IL-4, and IL-1β, respectively. The N- and C-termini of IL-4 and IL-1β are disordered in solution (6, 7).

b The backbone atomic coordinates comprise the N, Cα, C, and O atoms.

c The original protein G (5), IL-4 (6) and IL-1β (7) NMR structures are based on a total of 1058 (914 distance and 54 φ, 51 ψ and 39 χ1 torsion angle), 2974 (2617 distance and 130 φ, 119 ψ, 74 χ1, 32 χ2, and 2 χ3) and 3146 (2780 distance and 152 φ, 115 ψ, and 99 χ1 torsion angle) experimental restraints, respectively. This corresponds to an average of 18.9, 22.4, and 20.2 experimental restraints per residue, respectively. The reason that the NMR structure of protein G is determined to much higher precision than the other two structures is its very high (~95%) secondary structure content. The total number of restraints used in the calculation of the J-refined structures is the same as that used for the original structures, with J_{HN,Hα} coupling constant restraints, where available, replacing the corresponding φ torsion angle restraints. The number of J_{HN,Hα} restraints employed is 52 for protein G, 99 for IL-4, and 113 for IL-1β.

d The coordinate precision is defined as the average atomic rms difference between the ensemble of calculated structures and their mean coordinate positions (25).

e The three X-ray structures of IL-1β (10–12) were employed.

It is also worth noting that only the 1.67 Å resolution X-ray structure of protein G [after eliminating the ψ angles for Gly-9, located at the beginning of the turn between the first and second β-strands, and Gly-14, located at the beginning of the second β-strand)] comes close to that of the 1.65 Å resolution X-ray structure of Staphylococcal nuclease in terms of the agreement between calculated and observed J_{HN,Hα} values. Indeed, there is a clear correlation between the extent of agreement between observed and calculated J_{HN,Hα} couplings and the resolution and degree of refinement of the X-ray structures. Thus, the rank order is Staphylococcal nuclease, protein G, interleukin-1β, and interleukin-4, which were solved at 1.65, 1.67, 2.0, and 2.25 Å resolution, respectively. [Note that although the X-ray structures of Staphylococcal nuclease and protein G were solved at similar resolutions, the former (18) is more highly refined than the latter (8).] Such a correlation has previously been observed for the X-ray structure of turkey egg-white lysozyme solved at varying degrees of resolution and refinement (19).

The observation that the agreement between calculated and observed J_{HN,Hα} values is clearly better for the X-ray structures than the corresponding high-resolution NMR structures, despite the fact that the latter represent the most precise structures determined by NMR to date (1), suggests that the incorporation of J_{HN,Hα} coupling constants directly into the NMR structure determination procedure may be a worthwhile undertaking. We therefore added a term given by (3, 20)

$$E_J = k_J (J_{obs} - J_{calc})^2$$  \[3\]

into the target function employed in the simulated annealing calculations (where k_J is a force constant, J_{obs} the observed value of J_{HN,Hα} and J_{calc} the calculated value of J_{HN,Hα} using Eq. [2]), and proceeded to further refine the NMR structures. Thus, the target function minimized by simulated annealing (21) comprised terms for the interproton distance restraints, the torsion angle restraints (excluding φ angles for which J_{HN,Hα} couplings were available), the J_{HN,Hα} coupling constant restraints, covalent geometry (i.e., bonds, angles, chirality, and planarity), and a repulsive quartic van der Waals term for the nonbonded contacts. The results of this refinement are provided in Tables 1 and 2 and Fig. 1C. The rms difference between calculated and observed J_{HN,Hα} values is reduced to approximately the level of the experimental error in the measurement of J_{HN,Hα}. Thus, the rms differences for the J-refined NMR structures of protein G, IL-4, and IL-1β are 0.62, 0.37, and 0.65 Hz, respectively. In addition, the agreement with the experimental interproton distance and torsion angle restraints, the deviations from idealized geometry, and the quality of the nonbonded contacts remain unaltered and have approximately the same values.
as those reported for the original NMR structure determinations (5–7). Finally, the backbone atomic rms difference between the ensemble of calculated structures and their corresponding mean coordinate positions is only minimally affected and the atomic rms shift in the mean coordinate positions before and after \( J \) refinement lies within the positional errors of the coordinates (Table 2). This is readily explained as only small changes in \( \phi \) are required to fit the Karplus equation (Eq. [2]) and correlated compensatory changes in the backbone torsion angle \( \psi \) ensure that the coordinate positions are only minimally altered.

In this paper we have shown that with the availability of accurate \( ^3J_{\text{HN,Ha}} \) coupling constants, such as those measured by 3D heteronuclear \( J \) quantitative \( J \)-correlation experiments, large improvements in the agreement between calculated and observed \( ^3J_{\text{HN,Ha}} \) coupling constants can be readily obtained for protein structures derived from NMR data by directly incorporating \( ^3J_{\text{HN,Ha}} \) coupling restraints into the refinement procedure. Moreover, in the case of high-resolution NMR structures, such as the three examples used in the present paper, this is achieved without any impairment in the agreement with the other restraints in the target function (i.e., experimental interproton distance and torsion angle restraints, covalent geometry, and nonbonded contacts), and results in only minimal atomic rms shifts with no increase in precision at the expense of accuracy (cf. Table 2). Indeed, these observations provide a measure of the high quality of the three structures prior to \( J \) refinement as only minimal perturbations are required to satisfy the \( ^3J_{\text{HN,Ha}} \) restraints within experimental error.

The inclusion of direct refinement against \( ^3J_{\text{HN,Ha}} \) coupling constants offers a number of clear-cut benefits. First, as the agreement between observed and calculated values of \( ^3J_{\text{HN,Ha}} \) coupling constants is better for X-ray structures than for the corresponding high-resolution NMR structures obtained without \( J \) refinement, a clear case can be made that any improvement in agreement between observed and calculated coupling constants obtained upon \( J \) refinement should result in a corresponding improvement in the quality of the resulting NMR structure. Second, \( J \) restraints impose almost no cost in computational efficiency since the introduction of \( J \) restraints results in a less than 1% increase in the CPU time required for simulated annealing refinement. Third, motional effects simply average \( J \) values linearly with occupancy, and are therefore easy to interpret. This is in stark contrast to direct refinement against NOE intensities (22, 23) as the latter not only are subject to \( \langle r^{-6} \rangle^{-1/6} \) averaging but can also be attenuated by variations in the effective correlation times of the interproton distance vectors. In conclusion, it is our view that the direct incorporation of \( ^3J_{\text{HN,Ha}} \) coupling constant restraints into NMR structure determinations offers a simple and reliable means of improving the accuracy of protein NMR structures.

**ACKNOWLEDGMENTS**

This work was supported by the AIDS Directed Anti-Viral Program of the Office of the Director of the National Institutes of Health (G.M.C. and A.M.G.).

**REFERENCES**