An NMR Approach Applicable to Biomolecular Structure Characterization

Bing-Wen Hu, Ping Zhou, Isao Noda, and Guang-Zhi Zhao

The Key Laboratory of Molecular Engineering of Polymers, Ministry of Education, Department of Macromolecular Science, Fudan University, Shanghai 200433, P. R. China, and The Procter & Gamble Company, 8611 Beckett Road, West Chester, Ohio 45069

A generalized 2D correlation NMR (GEN2D-NMR) scheme capable of substantially reducing the experimental time for two-dimensional correlation NMR experiments is described. The experimental time used in GEN2D-NMR is shortened to less than one-tenth of that required in traditional double Fourier transform 2D-NMR (FT2D-NMR) for a $^{13}$C–$^{13}$C spin diffusion experiment for Nephila edulis spider threads. Experimentally, one needs to acquire many fewer points in GEN2D-NMR than that in FT2D-NMR in the evolution time domain. By keeping other acquisition pulse sequence and parameters the same in both methods, the GEN2D-NMR technique can produce a 2D-NMR spectrum using fewer points along the evolution time domain equivalent to that produced by the FT2D-NMR technique using more points. GEN2D-NMR would provide a significant advantage for those molecules that are difficult to highly isotropically label, such as proteins, polypeptides, and polymers, or those which become unstable in a prolonged measurement time.

Classical Fourier transformation 2D-NMR (FT2D-NMR) experiments usually require a relatively large number of acquisition points in the nuclear evolution time ($t_e$) domain. Together with the need for phase cycling, it often becomes rather cumbersome and time-consuming to carry out homo- or heteronuclear correlation 2D-NMR experiments. This is especially true for dilute abundant nuclei correlations, for example, $^{13}$C or $^{15}$N. Therefore, our ability to characterize large biomolecules that are difficult to extensively isotropically label or those samples that are not so stable and cannot tolerate a long measurement period often becomes limited. Herein, a surprisingly efficient alternative approach, generalized 2D correlation NMR (GEN2D-NMR) spectroscopy, is proposed and successfully demonstrated for the structure characterization of Nephila edulis spider threads. Using this method, only a few acquisition points along the $t_e$ domain would be necessary to construct a 2D-NMR spectrum. For example, a $^{13}$C–$^{13}$C coherence obtained with FT2D-NMR is obtained with remarkably reduced experimental time.

Although traditional FT2D-NMR techniques have been successfully applied to numerous structural determinations, researchers have constantly endeavored to develop various methods to reduce the experimental time. For example, single-scan 2D-NMR and radon transform 2D-NMR spectroscopy were recently proposed. We have come to the realization that the fundamental limitation in traditional FT2D-NMR techniques may lie in the well-established experimental scheme based on the double Fourier transformation. GEN2D-NMR may provide an alternative and sometimes more efficient way to treat NMR signals.

The development of GEN2D-NMR spectroscopy was inspired by the success of the so-called generalized 2D correlation spectroscopy scheme. This technique has gained much popularity in recent years in the optical spectroscopy field, such as infrared, Raman, near-infrared and ultraviolet spectroscopy. Eads et al. had applied this method to the analysis of NMR data from a diffusion experiment, but their data were not encoded as typical frequency modulation signals encountered in most FT2D-NMR experiments. The GEN2D-NMR scheme described in this article, addressing the signal correlation encoded as convoluted frequency modulation, circumvents some of the disadvantages associated with the traditional double Fourier transform approach. With this technique, one only needs to acquire many fewer points in the $t_e$ domain than those used in conventional FT2D-NMR. For example, $^{13}$C–$^{13}$C coherence for a partially labeled sample could be observed economically with many fewer acquisition points in the $t_e$ domain. GEN2D-NMR spectrum thus obtained is able to reveal most of the pertinent correlation information of interest. Additionally, this scheme does not even require the complex phase correction applied to the $t_e$ domain. This is a significant advantage because proper phasing for the two-dimensional pure absorption line shapes is often difficult in FT2D-NMR for certain samples, such as solids and polymers, with very broad resonance peaks.


* To whom correspondence should be addressed. Phone: +86-21-55664038.
Fax: +86-21-65640929. E-mail: pingzhou@fudan.edu.cn.
† Fudan University.
‡ The Procter & Gamble Company.
THEORY

Two-dimensional time domain coherence signals in a 2D time domain can be expressed as $x(t_1, t_2)$, a function of convoluted frequency modulation,

$$x(t_1, t_2) = \sum_{r=1}^{m} \sum_{i=1}^{n} \cos(\omega_{id} t_1 + \phi_i) \exp(-\lambda_i t_2) \cos(\omega_{et} t_2 + \phi_e) \exp(-\lambda_e t_2) Z^j$$

(1)

where $t_1$ and $t_2$ are often referred to as $t_1$ and $t_2$ in conventional FT2D-NMR terminology. The amplitude of the correlation $Z^j$ becomes significant only if the resonance frequency, $\omega_i$, in the detection domain, correlates with $\omega_e$ in the evolution domain. The decay constant $\lambda$ specifies the exponential function. After the first Fourier transformation and phase adjustment are applied to a series of FIDs along the $t_1$ axis, the 2D time domain signals $x(t_1, t_2)$ are converted to a series of pure absorption spectral traces $y(\omega_{d}, t_1)$ in the frequency domain $\omega_{d}$. If there is a correlation between $\omega_{d}$ and $\omega_{e}$, the generalized 2D correlation NMR spectrum is constructed by

$$\Phi(\omega_{d}, \omega_{e}) + i \Psi(\omega_{d}, \omega_{e}) = \frac{1}{\pi(T_{\text{max}} - T_{\text{min}})} \int_{T_{\text{min}}}^{T_{\text{max}}} z(\omega_{d}, \omega_{e}) \times (\omega_{d}, \omega_{e}) \, d\omega_{e}$$

(2)

where $z(\omega_{d}, \omega_{e})$ is the Fourier transform of $y(\omega_{d}, t_1)$ along the $t_1$ axis, and $\times(\omega_{d}, \omega_{e})$ is the conjugate Fourier transform of $y(\omega_{d}, t_1)$. $T_{\text{min}}$ and $T_{\text{max}}$ are, respectively, the starting and the ending points of the time in the evolution $t_1$ period. The real component of the correlation function $\Phi(\omega_{d}, \omega_{e})$ is referred to as the synchronous GEN2D-NMR spectrum, which can be expressed as

$$\Phi(\omega_{d}, \omega_{e}) = \frac{1}{T_{\text{max}} - T_{\text{min}}} \int_{T_{\text{min}}}^{T_{\text{max}}} y(\omega_{d}, t_1) y(\omega_{e}, t_2) \, dt_e$$

(3)

The imaginary component of the correlation function $\Psi(\omega_{d}, \omega_{e})$ is referred to as the asynchronous GEN2D-NMR spectrum, which can be expressed as

$$\Psi(\omega_{d}, \omega_{e}) = \frac{1}{T_{\text{max}} - T_{\text{min}}} \int_{T_{\text{min}}}^{T_{\text{max}}} y(\omega_{d}, t_1) h(\omega_{e}, t_2) \, dt_e$$

(4)

where $h(\omega_{e}, t_2)$ is the Hilbert transform of $y(\omega_{e}, t_2)$.

There are certain fundamental relationships between GEN2D-NMR and FT2D-NMR deduced as

$$\Phi(\omega_{d}, \omega_{e}) = p(\cos \phi)^{1.2} Abs(\omega_{d}) Abs(\omega_{e})$$

(5)

$$\Psi(\omega_{d}, \omega_{e}) = q(\sin \phi)^{1.2} Abs(\omega_{d}) Abs(\omega_{e})$$

(6)

where

$$p(\cos \phi)^{1.2} = \frac{1}{T_{\text{max}} - T_{\text{min}}} \sum_{j=1}^{n} \sum_{j=1}^{m} \int_{T_{\text{min}}}^{T_{\text{max}}} \cos(\omega_{i} t_1 + \phi_i) \cos(\omega_{e} t_1 + \phi_e) \exp[-(\lambda_i^1 + \lambda_i^2) t_1] Z^j Z^{2j} \, dt_e$$

and

$$q(\sin \phi)^{1.2} = \frac{1}{T_{\text{max}} - T_{\text{min}}} \sum_{j=1}^{n} \sum_{j=1}^{m} \int_{T_{\text{min}}}^{T_{\text{max}}} \cos(\omega_{i} t_1 + \phi_i) \exp[-(\lambda_i^1 + \lambda_i^2) t_1] Z^j Z^{2j} \, dt_e$$

Here, $\phi$ denotes the phase difference between two nuclei, $p$ and $q$ denote the phase-dependent functions with $-90^\circ < \phi < 90^\circ$, and Abs represents an absorption line shape function. A detailed deduction is shown in the Supporting Information.

Thus, the correlation information $\Phi(\omega_{d}, \omega_{e})$ between $\omega_{d}$ and $\omega_{e}$ in the synchronous GEN2D-NMR spectrum is essentially equivalent to that in the real part of the homonuclear FT2D-NMR spectrum. The correlation amplitude is modulated by a phase-dependent function. The term $p(\cos \phi)^{1.2}$ represents the overall similarity of the frequency $\omega_{d}$ in the elution time domain between two traces, $y(\omega_{d}, t_1)$ and $y(\omega_{e}, t_2)$, as the value of $t_2$ is changed. Therefore, there are clearly some different significances of correlation amplitudes between GEN2D-NMR and FT2D-NMR spectroscopy.

The asynchronous spectrum, on the other hand, corresponds to the imaginary part of FT2D-NMR spectrum, which reveals the explicit phase relationship between two resonances. For example, if there is predominantly a positive asynchronous cross peak $(\omega_{d}^1, \omega_{d}^2)$, there is a phase shift $\phi$ in $\omega_{d}^2$ ahead of that in $\omega_{d}^1$, and vice versa. Thus, the phase discriminating relationship results in different line shapes at different peaks, capable of identifying the overlapped peaks in the FT1D-NMR spectrum.

The significant aspects of the generalized 2D correlation analysis is that the GEN2D spectrum, formally defined in eq 2, can be actually calculated directly from a series of spectrum slices along the $t_2$ axis without the second Fourier transformation once the time-domain data in eq 1 is Fourier-transformed along the $t_1$ axis. Thus, the burdensome need to densely and broadly collect data points along the $t_2$ domain is eliminated.

In this paper, the structural characterization of *Nephila edulis* spider threads is chosen as an illustrative example to demonstrate the potential utility of the generalized 2D correlation analysis. Another successful application of GEN2D-NMR into homonuclear $^1$H–$^1$H NOESY FT2D-NMR spectrum in the liquid state is also demonstrated in Supporting Information.

EXPERIMENTAL SECTION

The sample of *Nephila edulis* spider threads is $^{13}$C isotopically enriched (roughly 10% of all amino acids) on the basis of the method described in the literature.[10]

The NMR experiment was performed on Bruker DMX 400 Advance spectrometer at 400.13 MHz for $^1$H resonance. A 2.5-mm double-resonance magic-angle spinning (MAS) probehead was

used with field strengths of 62.5 kHz for both proton and carbon nuclei. The number of transient scans was 90. A 1H-decoupled strength of 100 kHz, contact time of 1 ms, pulse repeat time of 3 s, acquisition points of 1024 in the $t_d$ domain, spinning rate of 10 kHz, and multiple transient scans of 256 were used. All the chemical shifts were referred to the external reference of methane carbon of 38.5 ppm in adamantane.

The original free induction decays (FIDs) from NMR experiment are Fourier-transformed and phased in the $t_d$ dimension by matNMR software, and then the resultant series of 1D traces along the $t_e$ dimension are used to form a generalized 2D correlation data based on eq 2 using a MatLab-based program.11 The 2D data are displayed by matNMR software.

RESULTS AND DISCUSSION

The sequence of this spider silk is rich in the amino acids glycine (42%) and alanine (28%). Therefore, the resulting solid-state $^{13}$C CPMAS 1D-NMR spectrum (Figure 1) is dominated by the peaks arising from the contributions from these constituents. The chemical shift assignments of the spectrum are tabulated in Table 1.

Figure 2 is the comparison of a conventional spin diffusion $^{13}$C–$^{13}$C correlation solid-state FT2D-NMR spectrum13 (Figure 2a) with a corresponding GEN2D-NMR spectrum (Figure 2b). Both
spectra were recorded with the same pulse sequence and experimental parameters for acquisition and procession, except that a different number of acquisition points were taken in the \( t_e \) domain. It is apparent that although the intensity of each peak in the FT2D-NMR spectrum and that in the GEN2D-NMR spectrum are not completely identical, the correlation information in the FT2D-NMR spectrum (Figure 2a) is essentially the same as that in the GEN2D-NMR spectrum (Figure 2b). Importantly, the former uses as many as 256 acquisition points along the \( t_e \) domain, whereas the latter requires only 24 points instead. That means the nominal experimental running time used in this GEN2D-NMR experiment is only one-tenth of that in FT2D-NMR!

Figure 2c is an asynchronous GEN2D-NMR spectrum. The black and magenta colors indicate positive and negative intensities, respectively. One of the slices at 172.6 ppm (dashed line under the diagonal) in Figure 2c is displayed in Figure 2d. There are clearly different intensities for the peaks in Figure 2d. The peak at 172.6 ppm on the diagonal (the so-called autopeak) of Figure 2c shows a symmetric dispersive line shape. Peaks at 51.0, 49.3, 21.3, and 17.5 ppm in Figure 2d show dominantly positive amplitudes, and the peak at 43.3 ppm, a negative amplitude. These results indicate that there are different phases among these nuclei. The different phases arise from the different motional behavior or relaxation of each nucleus. The nuclei resonating at 51.0, 49.3, 21.3, and 17.5 ppm have phase shifts ahead of that at 172.6 ppm, whereas the nucleus resonating at 43.3 ppm has a phase shift later. In fact, the phase sensitivities for these resonances additionally support the assignments in Table 1, such that the peaks at 51.0, 49.3, 21.3, and 17.5 ppm arise from alanine, and the peak at 43.5 ppm, from glycine. Compared with the FT1D- (Figure 1), FT2D- (Figure 2a), and synchronous GEN2D-NMR spectra (Figure 2b), the asynchronous GEN2D-NMR spectrum is uniquely capable of separating the overlapped peaks in the FT1D-NMR spectrum because of the different phase sensitivities of the nuclei. For example, it clearly separates two groups of overlapped peaks, one at 51.0 and 49.3 ppm, and the other at 21.3 and 17.5 ppm. In fact, the chemical shifts at 51.0 and 17.5 ppm represent the different protein conformations from that at 49.3 and 21.3 ppm. Therefore, the asynchronous GEN2D-NMR spectrum could potentially help researchers to identify the correct chemical shifts in the NMR spectrum.

The actual number of traces required for constructing a GEN2D-NMR may be surprisingly small for some samples with broad resonance peaks in 1D-NMR spectra, that is, with short transverse relaxation time, \( T_2 \). Theoretically, it is possible to construct a crude GEN2D-NMR spectrum using only the first three acquisition points in the \( t_e \) domain. This fact has been demonstrated successfully in Figure 3 for the same sample and technique used in Figure 2. Figure 3, constructed with only three acquisitions points, still shows a remarkably similar feature in the correlation information as compared to that of Figure 2b. Table 2 compares the resolution and sensitivity among the selected numbers of acquisition points in the GEN2D-NMR spectra. It is found that even though the acquisition points in the synchronous GEN2D-NMR spectrum are decreased by up to 100 times, the sensitivity (ratio of signal-to-noise) or resolution is only reduced by 30%. The small price is paid because of the decrease in the acquisition points along the \( t_e \) domain. Experimentally, one could gradually increase the acquisition points in the \( t_e \) domain if the used points were not enough to produce a satisfactory GEN2D-NMR result.

---

**Table 1.** \(^{13}\)C Chemical Shift *a* (ppm) Assignments of *Nephila edulis* Spider Thread

<table>
<thead>
<tr>
<th>amino acid</th>
<th>( C\alpha )</th>
<th>( C\beta )</th>
<th>( C\gamma )</th>
<th>( C\delta )</th>
<th>( C=O )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>49.3, 51.0</td>
<td>17.5, 21.3</td>
<td></td>
<td></td>
<td>172.6*</td>
</tr>
<tr>
<td>Gly</td>
<td>43.3</td>
<td></td>
<td></td>
<td></td>
<td>170.2*</td>
</tr>
<tr>
<td>Leu(^b)</td>
<td>54.3</td>
<td></td>
<td></td>
<td></td>
<td>c</td>
</tr>
<tr>
<td>Ser(^b)</td>
<td>62.2</td>
<td></td>
<td></td>
<td></td>
<td>c</td>
</tr>
<tr>
<td>Tyr(^b)</td>
<td>56.6</td>
<td></td>
<td></td>
<td></td>
<td>c</td>
</tr>
<tr>
<td>Gln</td>
<td>31.2</td>
<td>31.2</td>
<td>178.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Chemical shifts are referenced to adamantine. *b* Values are from ref 10. *c* Could not be determined unambiguously.

---

**Table 2. Comparison of Resolution and Sensitivity in Solid-State GEN2D-NMR Spectra with Different Acquisition Points**

<table>
<thead>
<tr>
<th>technique</th>
<th>acquisition points</th>
<th>resolution (Hz)(^b)</th>
<th>sensitivity(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>synchronous GEN2D-NMR</td>
<td>256</td>
<td>185</td>
<td>158.7</td>
</tr>
<tr>
<td>synchronous GEN2D-NMR</td>
<td>24</td>
<td>223</td>
<td>125.2</td>
</tr>
<tr>
<td>synchronous GEN2D-NMR</td>
<td>3</td>
<td>251</td>
<td>110.2</td>
</tr>
</tbody>
</table>

* The data were measured at cross peak (43.3, 172.9) ppm. *b* Resolution is measured as the half width at half-height. *a* Sensitivity is measured by the signal-to-noise (S/N) ratio.
NMR spectrum. The resulting FIDs, separately acquired at different \( t_e \), can be summed together for the GEN2D-NMR analysis. In addition, there may be some very weak cross peaks that result from the correlation transfer in the GEN2D-NMR spectrum, which is not present in FT2D-NMR. For example, a weak cross peak between \( \text{C}_a \) and \( \text{C}_b \) might be present if \( \text{C}_c \) correlated with \( \text{C}_a \) as well as \( \text{C}_b \), but the intensity of the correlation-transferred cross peak is theoretically very weak, and the correlation indicates that \( \text{C}_a \) and \( \text{C}_b \) are spatially close to each other.

**CONCLUSIONS**

GEN2D-NMR spectroscopy can be successfully applied to the homonuclear correlation 2D-NMR for structure assignment. The GEN2D-NMR method considerably reduces the 2D-NMR experimental time without losing any correlation information. There is no phase adjustment required on the second frequency domain, which is often difficult in the traditional FT2D-NMR method. Furthermore, the asynchronous GEN2D-NMR spectrum shows better identification of peaks than does a traditional 1D-NMR spectrum, which could facilitate the assignment of a 1D-NMR spectrum. Although the cross peak intensity in GEN2D-NMR, which is modulated by a phase function, cannot reveal the information of torsion angles and spatial distances between two nuclei, GEN2D-NMR would still provide a significant advantage for the structural assignment of those molecules that are difficult to label highly isotropically, such as proteins, polypeptides, and polymers, or those which become unstable in a prolonged measurement time.

**ACKNOWLEDGMENT**

Authors thank Prof. Beat H. Meier in ETH Zurich who supported all the facilities to prepare the spider sample and perform solid-state NMR experiments. This work was supported by the Natural Science Foundation of China (Nos. 20274009 and 10475017).

**SUPPORTING INFORMATION AVAILABLE**

The theory of GEN2D-NMR and an additional example of a liquid \(^1\text{H}\)-\(^1\text{H} \) NOESY 2D-NMR experiment are available as Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review June 15, 2005. Accepted September 20, 2005.

AC051061O