Supporting information

An Efficient 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.*

[‡]*The Procter & Gamble Company, 8611 Beckett Road, West Chester, Ohio 45069, U.S.A.*

Table of Contents

[*] To whom correspondence should be addressed. E-mail: pingzhou@fudan.edu.cn.

Tel. +86-21-55664038. Fax: +86-21-65640293.

1. Basic theory of FT2D-NMR spectroscopy

Pertinent components of signals of free induction decay (FID) obtained in a 2D time domain of t_d and t_e can be, in principle, expressed as $x(t_d, t_e)$, a function of convoluted frequency modulation, giving the form:

$$x(t_d, t_e) = \sum_{i=1}^{m}\sum_{j=1}^{n} \cos(\omega_d^i t_d) \exp(-\lambda_d^i t_d) \cos(\omega_e^j t_e) \exp(-\lambda_e^j t_e) Z^{i,j} \quad \text{(S1)}$$

where $Z^{i,j}$ is the amplitude of the correlation, which becomes significant only if the resonance frequency ω_d^i correlates with ω_e^j.

When the first Fourier transformation is applied along the t_d axis, the 2D time domain signals $x(t_d, t_e)$ are converted to a series of spectral traces $y(\omega_d, t_e)$ in the frequency domain ω_d, and then $y(\omega_d, t_e)$ is phased to yield the absorption lineshapes:

$$y(\omega_d, t_e) = \sum_{i=1}^{m} \mathrm{Abs}(\omega_d^i) \sum_{j=1}^{n} \cos(\omega_e^j t_e) \exp(-\lambda_e^j t_e) Z^{i,j} \quad \text{(S2)}$$

where *Abs* represents an absorption lineshape function.

A series of traces $y(\omega_d, t_e)$ are then converted to a 2D-NMR spectrum, $z(\omega_d, \omega_e)$, in the 2D frequency domain of ω_d and ω_e (corresponding to ω_2 and ω_1 in common FT2D-NMR terminology) by the second Fourier transform along the t_e axis. The hypercomplex Fourier transform or real Fourier transform with time proportional phase increment (TPPI) is often applied to the t_e domain to obtain the absorption lineshapes in the real part of FT2D-NMR:

$$\mathrm{Re}[z(\omega_d, \omega_e)] = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Abs}(\omega_d^i)\, \mathrm{Abs}(\omega_e^j) Z^{i,j} \quad \text{(S3)}$$

where *Re* presents the real part of spectrum.

2. The theory of generalized 2D correlation NMR spectroscopy

Given a set of traces $y(\omega_d, t_e)$ in Eq. (S2), a pair of two arbitrarily chosen resonances are

considered at ω_d^1 and ω_d^2 in the frequency domain ω_d, namely $y(\omega_d^1, t_e)$ and $y(\omega_d^2, t_e)$. The generalized 2D correlation NMR (GEN2D-NMR) spectrum is obtained by Eq. (S4):

$$\Phi(\omega_d^1, \omega_d^2) + i\,\Psi(\omega_d^1, \omega_d^2) = \frac{1}{\pi(T_{\max} - T_{\min})} \int_0^{\infty} z(\omega_d^1, \omega_e) \cdot z^*(\omega_d^2, \omega_e)\, \mathrm{d}\omega_e \qquad \text{(S4)}$$

where $\Phi(\omega_d^1, \omega_d^2)$ is referred to as synchronous GEN2D-NMR spectrum and $\Psi(\omega_d^1, \omega_d^2)$ as asynchronous GEN2D-NMR spectrum. $z(\omega_d^1, \omega_e)$ is the Fourier transform of $y(\omega_d^1, t_e)$ along the t_e axis, and $z^*(\omega_d^2, \omega_e)$ is the conjugate Fourier transform of $y(\omega_d^2, t_e)$. $T_{\min}$ and $T_{\max}$ are, respectively, the starting and the ending points of time in the evolution period.

We will demonstrate in following that the correlation information in the synchronous GEN2D-NMR spectrum is essentially equivalent to that in the real part of homo-nuclear FT2D-NMR spectrum, and the asynchronous spectrum corresponds to the imaginary part of the FT2D-NMR spectrum. The significant aspects of the generalized 2D correlation analysis is that the GEN2D spectrum formally defined in Eq. (S4) can be actually calculated directly from a series of spectrum slices along t_e domain in Eq. (S2) after the time-domain data in Eq. (S1) is Fourier transformed, without using the second Fourier transform. Thus, the burdensome need to densely and broadly collect data points along the t_e is eliminated.

Based on the theory described in literature (S1), $\Phi(\omega_d^1, \omega_d^2)$ and $\Psi(\omega_d^1, \omega_d^2)$ are computed by Eq. (S5) and Eq. (S6):

$$\Phi(\omega_d^1, \omega_d^2) = \frac{1}{T_{\max} - T_{\min}} \int_{T_{\min}}^{T_{\max}} y(\omega_d^1, t_e) \cdot y(\omega_d^2, t_e)\mathrm{d}t_e \qquad \text{(S5)}$$

$$\Psi(\omega_d^1, \omega_d^2) = \frac{1}{T_{\max} - T_{\min}} \int_{T_{\min}}^{T_{\max}} y(\omega_d^1, t_e) \cdot h(\omega_d^2, t_e)\, \mathrm{d}t_e \qquad \text{(S6)}$$

where $h(\omega_d^2, t_e)$ is Hilbert transformation of $y(\omega_d^2, t_e)$:

$$h(\omega_d^2, t_e) = \frac{1}{\pi}_{pv} \int_{-\infty}^{+\infty} \frac{y(\omega_d^2, t_e')}{t_e' - t_e} \mathrm{d}t_e \qquad \text{(S7)}$$

The integration symbol ${}_{pv}\int$ denotes that the *Cauchy principal value* is taken, so that the singularity at the point where $t_e^{'} = t_e$ is excluded from the integration.

Lemma for GEN2D-NMR. Provided a given function $\cos(\omega_1 t)\exp(-\lambda_1 t)$, the function $\mathrm{Y}(\omega) = \int_0^\infty \cos(\omega_1 t)\exp(-\lambda_1 t)\cdot\cos(\omega t + \phi)\exp(-\lambda t)\,\mathrm{d}t$ has a absorption lineshape varied with phase ϕ shown in Figure S1a (ϕ = 0°), S1b (ϕ = 30°) and S1c (ϕ = -30°) when $\omega = \omega_1 \pm \Delta\omega$. On the other hand, the function $\mathrm{Y}^{'}(\omega) = \int_0^\infty \cos(\omega_1 t)\exp(-\lambda_1 t)\cdot\sin(\omega t + \phi)\exp(-\lambda t)\,\mathrm{d}t$ has a dispersed lineshape varied with phase ϕ shown in Figure S1d (ϕ = 0°), S1e (ϕ = 30°) and S1f (ϕ = -30°) when $\omega = \omega_1 \pm \Delta\omega$. $\Delta\omega$ is λ_1 and λ dependent. The integral will be zero when ω is far away from ω_1.

For the convenience of discussion, only center frequency $\omega = \omega_1$ is considered, namely, the term in the previous functions:

$$\int_0^\infty \cos(\omega_1 t)\exp(-\lambda_1 t)\cdot\cos(\omega_1 t + \phi)\exp(-\lambda t)\,\mathrm{d}t$$

or $$\int_0^\infty \cos(\omega_1 t)\exp(-\lambda_1 t)\cdot\sin(\omega_1 t + \phi)\exp(-\lambda t)\,\mathrm{d}t$$

is retained in the expressions. This simplification would not influence the final result.

Synchronous GEN2D-NMR spectrum. For synchronous GEN2D-NMR spectrum, based on Eq. (S2), the individual expression of $y(\omega_d^1, t_e)$ and $y(\omega_d^2, t_e)$ gives

$$y(\omega_d^1, t_e) = \mathrm{Abs}(\omega_d^1)\sum_{i=1}^{n}\cos(\omega_e^i t_e)\exp(-\lambda_e^i t_e)Z^{1,i} \qquad \text{(S8)}$$

$$y(\omega_d^2, t_e) = \mathrm{Abs}(\omega_d^2)\sum_{j=1}^{n}\cos(\omega_e^j t_e)\exp(-\lambda_e^j t_e)Z^{2,j} \qquad \text{(S9)}$$

When ω_d^1 correlates with ω_e^1, ω_e^2, ω_e^3, … ω_e^{s-1}, ω_e^s $(s < n)$, and ω_d^2 correlates with ω_e^1, ω_e^2, ω_e^3, … ω_e^{v-1}, ω_e^v $(v \le s < n)$, we have

$$y(\omega_d^1, t_e) = \mathrm{Abs}(\omega_d^1)\sum_{i=1}^{s}\cos(\omega_e^i t_e)\exp(-\lambda_e^i t_e)Z^{1,i} \qquad \text{(S10)}$$

$$y(\omega_d^2, t_e) = \mathrm{Abs}(\omega_d^2)\sum_{j=1}^{v}\cos(\omega_e^j t_e)\exp(-\lambda_e^j t_e)Z^{2,j} \tag{S11}$$

When Eq. (S10) and (S11) are substituted into Eq. (S5), we have

$$\begin{aligned}
&\Phi(\omega_d^1, \omega_d^2) * (T_{\max} - T_{\min}) \\
&= \int_{T_{\min}}^{T_{\max}} \{\mathrm{Abs}(\omega_d^1)\sum_{i=1}^{s}\cos(\omega_e^i t_e)\exp(-\lambda_e^i t_e)Z^{1,i}\} \\
&\{\mathrm{Abs}(\omega_d^2)\sum_{j=1}^{v}\cos(\omega_e^j t_e)\exp(-\lambda_e^j t_e)Z^{2,j}\}\mathrm{d}t_e \\
&= \mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_d^2) \\
&\{\sum_{i=1}^{v}\sum_{i=j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\cos(\omega_e^j t_e)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e \\
&+ \sum_{i=1}^{v}\sum_{i\neq j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\cos(\omega_e^j t_e)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e\}
\end{aligned}$$

According to the previous lemma, if only center frequency is concerned, i.e., the term

$A = \sum_{i=1}^{v}\sum_{i=j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\cos(\omega_e^j t_e)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e$ is retained, the result gives rise to

$$\Phi(\omega_d^1, \omega_d^2) = \frac{1}{T_{\max} - T_{\min}}\mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_d^2)A \tag{S12}$$

or

$$\Phi(\omega_d^1, \omega_d^2) = p^{1,2}\cdot\mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_d^2) \tag{S13}$$

where $p^{1,2} = \frac{1}{T_{\max} - T_{\min}}\sum_{i=1}^{v}\sum_{i=j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\cos(\omega_e^j t_e)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e$.

For homo-nuclear FT2D-NMR spectrum, $\mathrm{Abs}(\omega_d^j)$ equals to $\mathrm{Abs}(\omega_e^j)$, therefore, equivalently, there is Eq. (S14) for FT2D-NMR based on Eq. (S13):

$$\Phi(\omega_d^1, \omega_d^2) = p^{1,2}\cdot\mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_e^2) \tag{S14}$$

Eq. (S14) demonstrates that the correlation information in synchronous GEN2D-NMR is equivalent to that in FT2D-NMR as Eq. (S3), but their amplitudes are modulated by a function $p^{1,2}$. The term $p^{1,2}$ represents the overall similarity of the frequency ω_e^i in the elution domain between two traces $y(\omega_d^1, t_e)$ and $y(\omega_d^2, t_e)$, as the value of t_e is changed.

Asynchronous GEN2D-NMR spectrum. $h(\omega_d^2, t_e)$ in Eq. (S7) can be expressed as Eq. (S15) based on reference (S2):

$$h(\omega_d^2, t_e) = \mathrm{Abs}(\omega_d^2)\sum_{j=1}^{n}\sin(\omega_e^j t_e)\exp(-\lambda_e^j t_e) \tag{S15}$$

Similarly, we have

$$\begin{aligned}
&\Psi(\omega_d^1, \omega_d^2) * (T_{\max} - T_{\min}) \\
&= \int_{T_{\min}}^{T_{\max}} y(\omega_d^1, t_e) \cdot h(\omega_d^2, t_e)\mathrm{d}t_e \\
&= \int_{T_{\min}}^{T_{\max}} \{\mathrm{Abs}(\omega_d^1)\sum_{i=1}^{s}\cos(\omega_e^i t_e)\exp(-\lambda_e^i t_e)Z^{1,i}\}\{\mathrm{Abs}(\omega_d^2)\sum_{j=1}^{v}\sin(\omega_e^j t_e)\exp(-\lambda_e^j t_e)Z^{2,j}\}\mathrm{d}t_e \\
&= \mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_d^2) \\
&\{\sum_{i=1}^{v}\sum_{i=j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\sin(\omega_e^j t_e)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e \\
&+ \sum_{i=1}^{v}\sum_{i\neq j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\sin(\omega_e^j t_e)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e\}
\end{aligned}$$

Also according to the previous lemma, if only center frequency is concerned, i.e., the term $A' = \sum_{i=1}^{v}\sum_{i=j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\sin(\omega_e^j t_e)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e$ is retained, we have

$$\Psi(\omega_d^1, \omega_d^2) = \frac{1}{T_{\max} - T_{\min}}\mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_d^2)A' \tag{S16}$$

If there is no phase difference between two resonances ω_d^1 and ω_d^2 in Eq. (S16), the correlation cross-peaks (ω_d^1, ω_d^2) in asynchronous GEN2D-NMR spectrum would show dispersion lineshapes as in Figure S1d.

The relationship between GEN2D-NMR spectrum and FT2D-NMR spectrum. If the two resonances are out of phase by ϕ ($-90^\circ < \phi < 90^\circ$) generally acquired by TPPI (time proportional phase increment) method, the synchronous GEN2D-NMR spectrum will be

$$\Phi(\omega_d^1, \omega_d^2) = p(\cos\phi)^{1,2} \cdot \mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_d^2) \tag{S17}$$

where $p(\cos\phi)^{1,2} = \frac{1}{T_{\max} - T_{\min}}\sum_{i=1}^{v}\sum_{i=j}\sum_{j=1}^{s}\int_{T_{\min}}^{T_{\max}}\cos(\omega_e^i t_e)\cos(\omega_e^j t_e + \phi)\exp[-(\lambda_e^i + \lambda_e^j)t_e]Z^{1,i}Z^{2,j}\mathrm{d}t_e$;

and asynchronous GEN2D-NMR spectrum will be

$$\Psi(\omega_d^1, \omega_d^2) = q(\sin\phi)^{1,2}\ \mathrm{Abs}(\omega_d^1)\,\mathrm{Abs}(\omega_d^2) \tag{S18}$$

where $q(\sin\phi)^{1,2} = \dfrac{1}{T_{\max} - T_{\min}} \sum_{i=1}^{v} \sum_{j=1}^{s} \int_{T_{\min}}^{T_{\max}} \cos(\omega_e^i t_e) \sin(\omega_e^j t_e + \phi) \exp[-(\lambda_e^i + \lambda_e^j) t_e] Z^{1,i} Z^{2,j} \mathrm{d}t_e$

Eq. (S17) shows that the peak lineshape in the synchronous GEN2D-NMR spectrum is like Figure S1b when $\phi > 0$ and Figure S1c when $\phi < 0$, respectively. Eq. (S18) shows that the peak lineshape in the asynchronous GEN2D-NMR spectrum is like Figure S1e (dominantly positive) when $\phi > 0$ and Figure S1f (dominantly negative) when $\phi < 0$, respectively. It is found that the change in lineshape induced by phase function $p(\cos\phi)^{1,2}$ in synchronous GEN2D-NMR spectrum (Figure S1b and S1c) is more limited than that induced by phase function $q(\sin\phi)^{1,2}$ in asynchronous GEN2D-NMR spectrum (Figure S1e and S1f). Thus, the dependence of intensity and its sign on the phase is more sensitive in asynchronous GEN2D-NMR spectrum, capable of conveniently identifying the overlapped peaks among absorption lineshapes in the FT 1D spectrum.

In fact, the information that asynchronous GEN2D-NMR spectrum contains is involved in the imaginary part of FT2D-NMR spectrum, revealed in Eq. (S19). The phase value in imaginary part of FT2D-NMR is referenced to an external function $\sin(\omega_e t_e)$, while that in the asynchronous GEN2D-NMR spectrum is referenced to an internal function $\sin(\omega_e t_e + \phi)$. Interestingly, the property of phase-dependent FT2D-NMR has not yet been well explored up to now. At last we should point out that just because the phase in the synchronous GEN2D-NMR spectrum is referenced to an internal function $\cos(\omega_e t_e + \phi)$, it is not necessary to phase in the synchronous GEN2D-NMR spectrum.

$$\begin{aligned} z(\omega_d, \omega_e) &= \int_{-\infty}^{+\infty} y(\omega_d, t_e) e^{-i\omega_e t_e}\, \mathrm{d}t_e \\ &= \int_{-\infty}^{+\infty} y(\omega_d, t_e) \cos(\omega_e t_e)\, \mathrm{d}t_e + i \int_{-\infty}^{+\infty} y(\omega_d, t_e) \sin(\omega_e t_e)\, \mathrm{d}t_e \end{aligned} \tag{S19}$$

3. Application of GEN2D-NMR to liquid-state NMR

Figure S2 shows another application of GEN2D-NMR (Figure S2b), comparing this time with homo-nuclear ^{1}H-^{1}H NOESY FT2D-NMR spectrum (Figure S2a) in the liquid state. The experiment was performed on Bruker DRX-500 at 500.13 MHz for ^{1}H. Like the example in text, both spectra were recorded with the same pulse sequence and parameters for acquisition and process, except using a different number of acquisition points in the t_e domain, i.e., 256 points in FT2D-NMR as compared to only 96 points in GEN2D-NMR. These two spectra are displayed with the same thresholds for the impartial comparison. It can be seen that most of the correlation information are retained although their cross-peak intensities in two spectra are incomparable because the cross-peak amplitudes in the synchronous GEN2D-NMR spectrum is modulated by a phase function. The axial peaks appearing at 0.8 and 3.5 ppm in both spectra (Figure S2a and S2b) could be eliminated if a higher baseline threshold were used. Detailed comparison of resolution and sensitivity among the selected numbers of acquisition points in GEN2D-NMR spectra is shown in Table S1. It is found that the considerable reduce in 2D experimental time because of the decrease of acquired points along t_e domain merely lead to the decrease in the sensitivity in small extent.

References

(S1) Noda, I. *Appl. Spectrosc.* **2000**, *54*, 994-999.

(S2) Bracewell, R. N. *The Fourier transform and its application*; McGraw-Hill: New York, 1986.

Table S1. Comparisons of resolution and sensitivity in liquid-state GEN2D-NMR spectra with different acquisition points[a]

Techniques	Acquisition Points	Resolution (Hz)[b]	Sensitivity[c]
synchronous GEN2D-NMR	256	12.1	187.3
synchronous GEN2D-NMR	96	13.6	147.4

[a] The data were measured at the cross-peak (2.41, 3.32) ppm. [b] Resolution is measured as the half width at half height. [c] Sensitivity is measured by ratio of signal-to-noise (S/N).

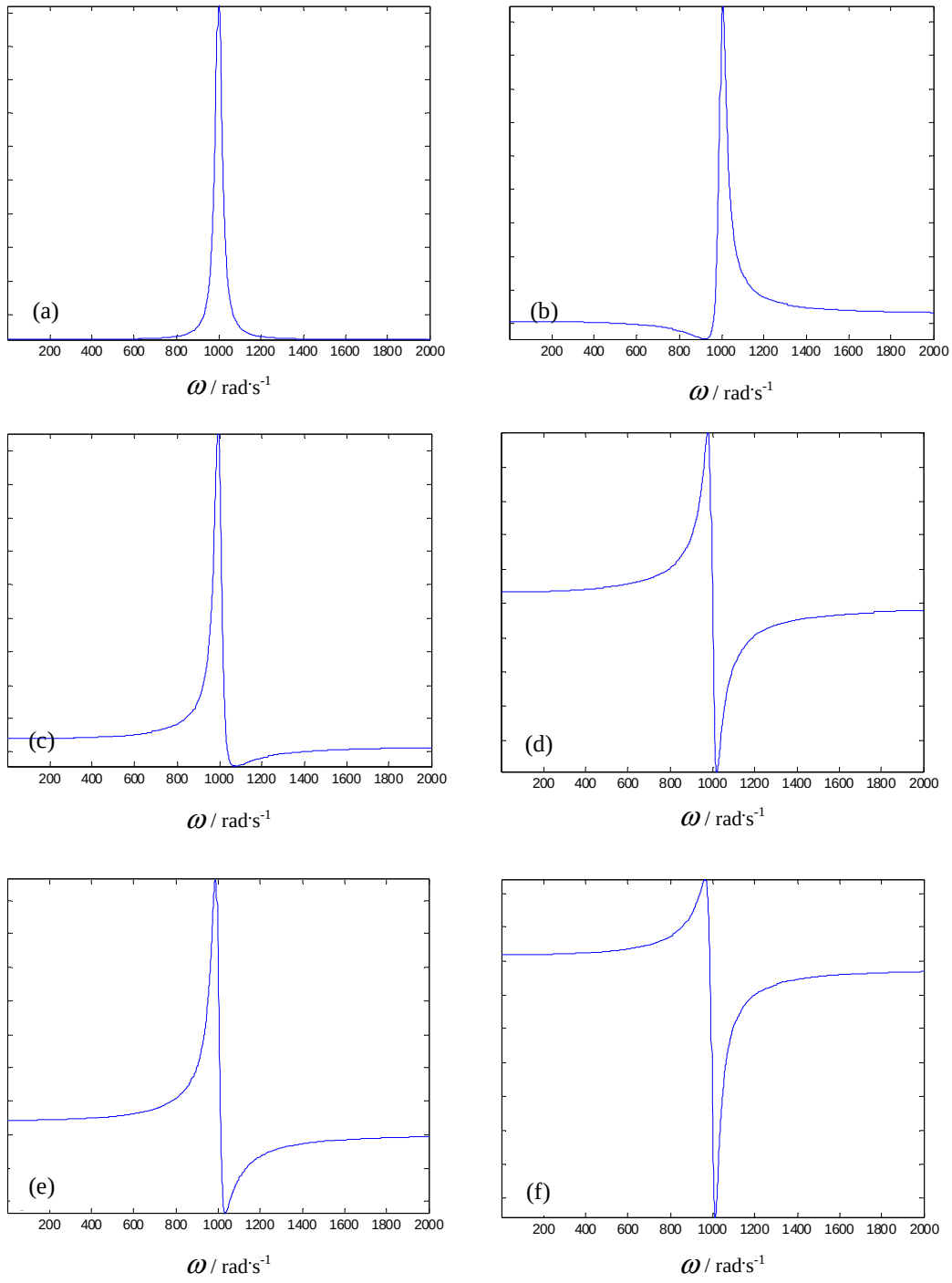

Figure S1. Function $Y(\omega) = \int_0^\infty \cos(\omega_1 t)\exp(-\lambda_1 t)\cdot\cos(\omega t)\exp(-\lambda t)\,dt$ with ϕ = 0° (a), 30° (b) and –30° (c) and Function $Y'(\omega) = \int_0^\infty \cos(\omega_1 t)\exp(-\lambda_1 t)\cdot\sin(\omega t+\phi)\exp(-\lambda t)\,dt$ with ϕ = 0° (d), 30° (e) and –30° (f). ω_1 = 1000 rad/s and $\lambda_1 = \lambda = 20$ rad·s^{-1} are supposed.

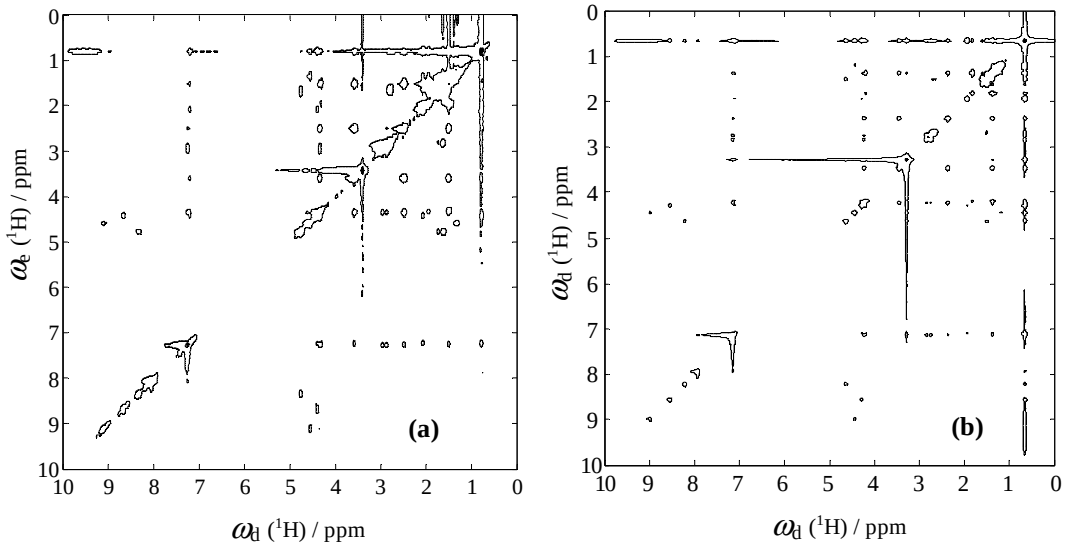

Figure S2. Two-dimensional ^{1}H-^{1}H NOESY NMR spectra in the liquid state. (a) is FT2D-NMR with 256 points along the t_e domain; (b) is synchronous GEN2D-NMR with 96 points along the t_e domain. Two spectra have the same threshold.