Supporting Information

A New Approach to Recognizing the Correct Pattern of Cross-peaks from a Noisy 2D Asynchronous Spectrum by Detecting Intrinsic Symmetry via the Kolmogorov-Smirnov Test

Linchen Xie [a, b], Ran Guo [b, c], Limin Yang [d*], Yukihiro Ozaki [b, e],

Isao Noda [b, f], Yizhuang Xu [b*], Kun Huang [a*]

[a] School of Metallurgical and Ecological Engineering, University of Science and Technology Beijing, Beijing 100083, China

[b] Beijing National Laboratory for Molecular Sciences, State Key Laboratory for Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

[c] Beijing CKC, PerkinElmer Inc., Beijing 100015, PR China

[d] State Key Laboratory of Nuclear Physics and Technology, Institute of Heavy Ion Physics, School of Physics, Peking University, Beijing 100871, China

[e] School of Biological and Environmental Sciences, Kwansei Gakuin University, Sanda, Hyogo 669 - 1330, Japan

[f] Department of Materials Science and Engineering, University of Delaware, Newark, DE 19716, United States

*Corresponding authors: Kun Huang (huangkun@ustb.edu.cn); Limin

Yang (yanglm@pku.edu.cn); Yizhuang Xu (xyz@pku.edu.cn)

Table of contents

Part 1 A brief description of the DAOSD approach

DAOSD (<u>d</u>ouble <u>a</u>synchronous <u>o</u>rthogonal <u>s</u>ample <u>d</u>esign scheme) is used to reveal subtle spectral variation caused by intermolecular interaction. The description of the DAOSD approach in detail can be found in our previous paper [S1-1]. Here we provide a brief description of the DAOSD approach.

We use the following model system to show how the DAOSD approach works. The model system is: two compounds (denoted as P and Q) are dissolved in the same solutions. Under intermolecular interactions between P and Q, part of P undergoes a subtle structural variation and converts into U. Similarly, part of Q converts into V. The interaction could

be described by an equilibrium reaction shown in Eq. S1-1, where K is the equilibrium constant.

$$P + Q \overset{K}{\Leftrightarrow} U + V \qquad (S1\text{-}1)$$

In a spectrum of sample solution, the solvent has no spectral contribution in the spectral region investigated. P has an absorption peak at X_P, and Q has an absorption peak at X_Q. U has an absorption peak at X_U, and V has an absorption peak at X_V. The spectral functions of characteristic peaks of P, Q, U, and V are described by Gaussian functions shown in Eq. S1-2.

$$f_j(x) = \varepsilon_j e^{-(\ln 2)\left[\frac{(x-X_j)^2}{W_j^2}\right]} = \varepsilon_j g_j(x) \qquad (S1\text{-}2)$$

where j stands for the four chemical species P, Q, U, V; x is wavelength; W_j, X_j, and ε_j are the bandwidth, peak position, and intensity of the characteristic peak of the jth chemical species; $g_j(x)$ is the peak-shape function of the jth chemical species.

Since the structural variation caused by the intermolecular interaction is quite weak, the characteristic peak of U is quite close to that of P. In a similar manner, the characteristic peak of V is quite close to that of Q. In other words, the characteristic peaks of P and U are severely overlapped. Moreover, the characteristic peaks of Q and V are severely overlapped. On the other hand, only a very small fraction of P and Q are converted into U and V since the intermolecular interaction between P and Q is quite weak, Thus, the equilibrium concentrations of P and Q are overwhelmingly larger

than those of U and V. Consequently, the characteristic peaks of U and V are buried by the peaks of P and Q. In other words, it is almost impossible to observe any spectral feature of U and V from original 1D spectra of a sample solution containing P and Q.

To characterize the intermolecular interaction between P and Q, a 2D asynchronous spectrum is generated using the DAOSD approach. In the DAOSD approach, two groups of sample solutions were prepared. Each group contains four sample solutions. In the first group of sample solutions, the initial concentrations of P are constants, while the initial concentrations of Q are selected arbitrarily. In the second group of sample solutions, the initial concentrations of Q are constants, while the initial concentrations of P are selected arbitrarily. The initial concentrations of P and Q are listed in **Table S1-1**.

Table S1-1 Initial concentrations of P and Q in the model system, which meet the requirement of the DAOSD approach

Index of the solutions	Initial concentrations of P	Initial concentrations of Q
Group 1		
1	10	0
2	10	4
3	10	6
4	10	10
Group 2		
1	0	10
2	4	10
3	6	10
4	10	10

Then, the 1D spectra of the two groups of sample solutions are

simulated, and we use the two groups of spectra to construct a pair of 2D asynchronous spectra via Eq. S1-3.

$$\Psi = A^T N A \tag{S1-3}$$

where N is the Hilbert-Noda transformation matrix and superscript ^{T} stands for transpose.

The 2D asynchronous spectrum generated via the first group of 1D spectra, where the initial concentrations of P are invariant is denoted as Ψ_P. The 2D asynchronous spectrum generated via the second group of 1D spectra, where the initial concentrations of Q are invariant is denoted as Ψ_Q.

Results from both the mathematical analysis and computer simulation have proved that:

1) If no intermolecular interaction occurs between P and Q (this can be achieved by setting the value of K in Eq. S1-1 as zero), no cross-peak is produced in either Ψ_P or Ψ_Q.

2) If intermolecular interaction indeed occurs between P and Q (this can be accomplished by setting K in Eq. S1-1 as a non-zero value. Herein, the value of K is set as 0.01 L/mol), cross-peaks may be produced in Ψ_P and Ψ_Q.
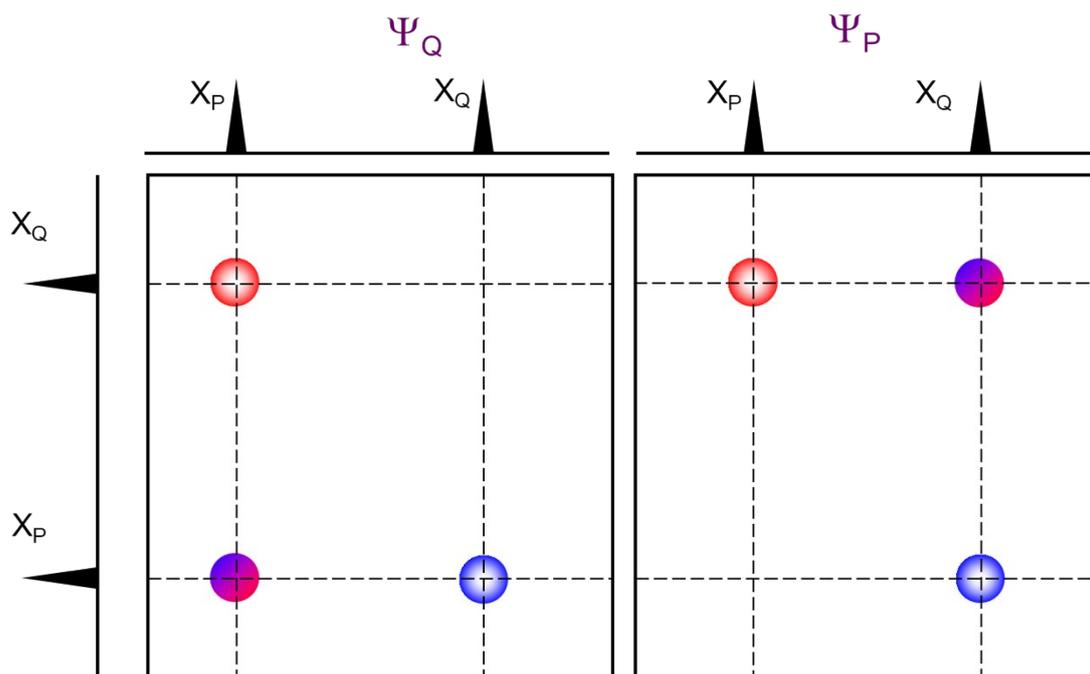
Scheme S1-1 Typical Ψ_P and Ψ_Q.

Scheme S1-1 shows typical Ψ_P and Ψ_Q. In Ψ_P, three groups of cross-peaks appear around (X_P, X_P), (X_P, X_Q), and (X_Q, X_P). In Ψ_Q, three groups of cross-peaks appear around (X_Q, X_Q), (X_P, X_Q), and (X_Q, X_P).

According to the mathematical analysis shown in our previous paper [S1-1], the cross-peak group around (X_Q, X_Q) in Ψ_P reflects the changes of the characteristic peak of Q caused by intermolecular interaction. Furthermore, the spectral function of the cross-peak group can be expressed as Eq. S1-4a. In a similar manner, the cross-peak group around (X_P, X_P) in Ψ_Q reflects the changes of the characteristic peak of P caused by intermolecular interaction. Furthermore, the spectral function of the cross-peak group can be expressed as Eq. S1-4b.

$$\Psi_P(x, y) = \varepsilon_Q \varepsilon_V \left[g_Q(x) g_V(y) - g_Q(y) g_V(x) \right] \left(\vec{\mathbf{C}}_Q^{init} \right)^T \mathbf{N} \vec{\mathbf{C}}_V^{eq} \qquad \text{(S1-4a)}$$

$$\Psi_Q(x,y) = \varepsilon_P \varepsilon_U \left[g_P(x) g_U(y) - g_P(y) g_U(x) \right] \left(\vec{\tilde{C}}_P^{\text{init}} \right)^T N \vec{\tilde{C}}_U^{\text{eq}} \qquad \text{(S1-4b)}$$

where $\vec{\tilde{C}}_P^{\text{init}}$, $\vec{\tilde{C}}_U^{\text{eq}}$ are respectively, the initial concentrations of P and the equilibrium concentrations of U in the second group of solutions shown in **Table S1-1**; $\vec{\tilde{C}}_Q^{\text{init}}$, $\vec{\tilde{C}}_V^{\text{eq}}$ are respectively.

Eq. S1-4a demonstrates that the pattern of cross-peaks around (X_Q, X_Q) in Ψ_P reflects the difference between $g_Q(x)$ and $g_V(x)$. Since $g_Q(x)$ and $g_V(x)$ are relevant to X_Q, X_V, W_Q, and W_V, the pattern of cross-peaks around (X_Q, X_Q) in Ψ_P reflects the changes in peak position and peak width of Q caused by intermolecular interaction.

In a similar manner, Eq. S1-4b demonstrates that the pattern of cross-peaks around (X_P, X_P) in Ψ_Q reflects the difference between $g_P(x)$ and $g_U(x)$. Since $g_P(x)$ and $g_U(x)$ are relevant to X_P, X_U, W_P, and W_U. the pattern of cross-peaks around (X_Q, X_Q) in Ψ_P reflects the changes in peak-position and peak-width of Q caused by intermolecular interaction.

We use the cross-peaks around (X_P, X_P) in Ψ_Q as an example to show the relationship between the patterns of cross-peaks and the variations in peak-position and peak-width. In this case, X_U and W_U are set as variables and we define $\Delta X = X_U - X_P$, $\Delta W = W_U - W_P$. As shown in **Scheme 1**, there is a one-to-one correspondence between the pattern of cross-peaks and the combination of ΔX, ΔW.

From a qualitative point of view, the variation of the peak-position and peak-width of U with respect to those of P can be classified into nine

situations (**Table 1**). The peak parameters of the absorption peak of P are provided in **Table S1-2**. There is a one-to-one correspondence between the patterns of cross-peaks around (X_P, X_P) and the aforementioned nine situations (**Scheme 1**).

Table S1-2 Peak parameters of P and U in the model system

	Situation	Peak position	Peak width	Absorbance
	1	99	19	1.0
	2	100	19	1.0
	3	101	19	1.0
	4	101	20	1.0
U	5	100	20	1.0
	6	99	20	1.0
	7	101	21	1.0
	8	100	21	1.0
	9	99	21	1.0
P		100	20	1.0

From the characteristic pattern of the cross-peak, we can deduce whether the characteristic peak of P undergoes a redshift, or blueshift or remains unchanged under the influence of intermolecular interaction. Moreover, we can also judge whether the peak-width of the peak of P increase, decrease, or remain invariant under the intermolecular interaction.

Similar results can also be obtained from the cross-peaks around (X_Q, X_Q) in Ψ_P.

On the other hand, Eq. S1-4a demonstrates that the difference between

ε_Q and ε_V can only be reflected by the changes in the intensities of the cross-peak (X_Q, X_Q) in Ψ_P. Similarly, Eq. S1-4b demonstrates that the difference between ε_P and ε_U can only be reflected by the changes in the intensities of the cross-peak (X_P, X_P) in Ψ_Q. In a 2D asynchronous spectrum, the intensities of cross-peaks are affected by multiple factors. Thus, neither the cross-peaks around (X_P, X_P) in Ψ_Q, nor those around (X_P, X_P) in Ψ_Q is suitable to reflect the intensity changes of the characteristic peaks caused by intermolecular interaction.

Then, we come to the group of cross-peaks around (X_P, X_Q) and those around (X_P, X_Q) in Ψ_P. Since the two groups of cross-peaks are antisymmetric with respect to the diagonal, we just discuss the cross-peaks around (X_P, X_Q) in Ψ_P. According to the mathematical analysis in our previous paper [S1-1], the cross-peaks reflect changes in the characteristic peak of Q at X_Q only. Moreover, the spectral function of the cross-peaks can be expressed as Eq. S1-5.

$$\Psi_P(x,y) = \varepsilon_Q \varepsilon_U \left[g_U(x) - g_P(x) \right] \left(\vec{\tilde{C}}_U^{\,eq} \right)^T N \vec{\tilde{C}}_Q^{\,init}$$
$$+ \varepsilon_Q (\varepsilon_U - \varepsilon_P) g_P(x) g_Q(y) \left(\vec{\tilde{C}}_U^{\,eq} \right)^T N \vec{\tilde{C}}_Q^{\,init}$$

(S1-5)

From Eq. S1-5, the cross-peaks in this region are composed of two parts: the first part ($\varepsilon_Q \varepsilon_U \left[g_U(x) - g_P(x) \right] \left(\vec{\tilde{C}}_U^{\,eq} \right)^T N \vec{\tilde{C}}_Q^{\,init}$) reflects the variations of bandwidth and peak position of the characteristic peak of Q; The night basic patterns of cross-peaks from the first part (

) are illustrated in **Scheme S1-2**.

$\Delta X<0, \Delta W<0$	$\Delta X<0, \Delta W=0$	$\Delta X<0, \Delta W>0$
$\Delta X=0, \Delta W<0$	$\Delta X=0, \Delta W=0$	$\Delta X=0, \Delta W>0$
$\Delta X>0, \Delta W<0$	$\Delta X>0, \Delta W=0$	$\Delta X>0, \Delta W>0$
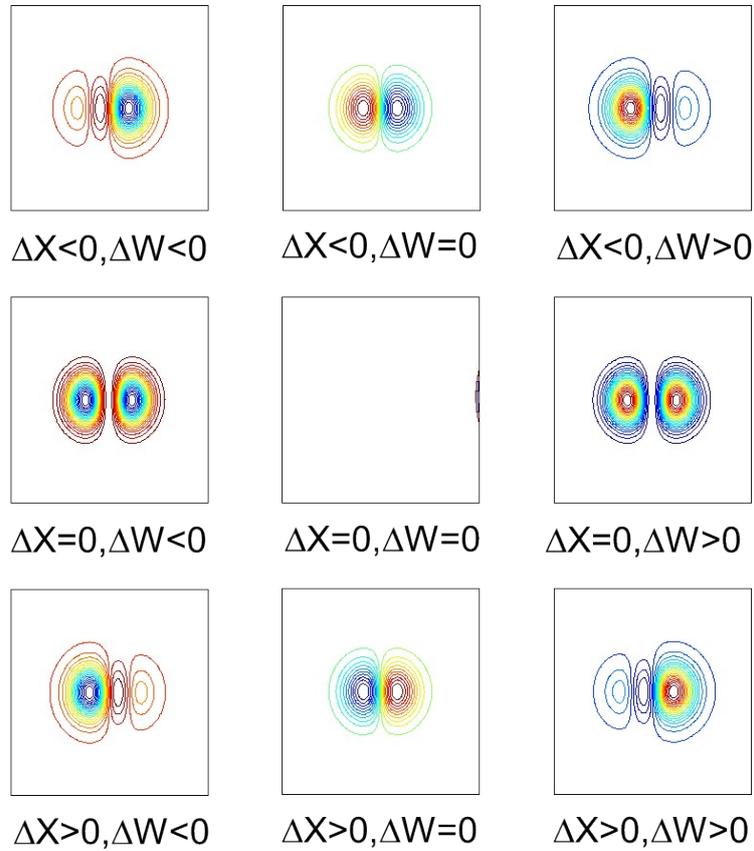
Scheme S1-2 The nine basic patterns of cross-peak around (X_P, X_Q)

derived from _____ versus the combinations of ΔX, ΔW.

The second part (_____) reveals the variation of

intensity of the characteristic peak of P. Herein, we define $\Delta\varepsilon=\varepsilon_U-\varepsilon_P$. Their

basic pattern of $\varepsilon_Q(\varepsilon_U-\varepsilon_P)g_P(x)g_Q(y)\left(\vec{\tilde{C}}_U^{eq}\right)^T N\vec{\tilde{C}}_Q^{init}$ is illustrated in **Scheme S1-3**.
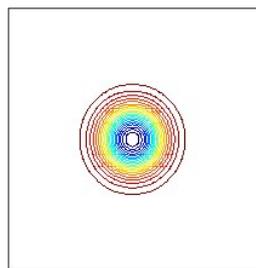
$$\Delta\varepsilon<0$$
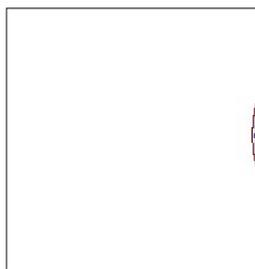
$$\Delta\varepsilon=0$$
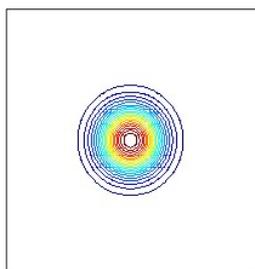
$$\Delta\varepsilon>0$$

Scheme S1-3 The three basic patterns of cross-peaks derived from ▨▨▨▨▨▨▨▨▨▨▨ versus $\Delta\varepsilon$.

The cross-peaks around (X_Q, X_P) in Ψ_Q can be used to reveal the subtle changes of the characteristic peak of Q under the intermolecular interaction via a similar fashion.

We show how to reveal the variation of the characteristic peaks caused by intermolecular interaction via the patterns of cross-peaks from the two 2D asynchronous spectra generated using the DAOSD approach.

The peak parameters for the characteristic peak of P and Q are listed in **Table S1-3**.

Table S1-3 Peak parameters of P and Q in the model system

j	X_j (nm)	W_j (nm)	ε_j
P	100	20	1.0
Q	300	20	1.0

Since the peaks of U and V overlap with those of P and Q severely. It is impossible to obtain the U, V from the original 1D spectra of the P, Q mixture.

When the DAOSD approach is adopted, Ψ_P and Ψ_Q are obtained and shown in **Fig. S1-1A** and **Fig. S1-1B**.

Ψ_P

A

Ψ_Q

B
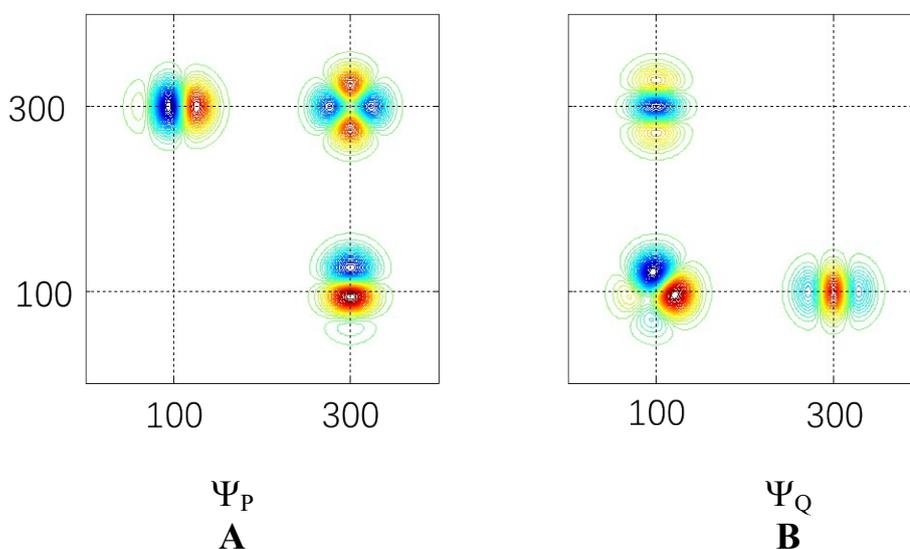
Figure S1-1 A Ψ_P; **B** Ψ_Q.

Comparison between the cross-peaks around (300, 300) in Ψ_P and the basic patterns of cross-peak shown in **Scheme 1**, the peak position of the characteristic peak of Q remains unchanged, but the peak width decrease under the intermolecular interaction ($\Delta X_Q = 0$, $\Delta W_Q < 0$). Under this

situation, the part of cross-peaks around (300, 100) in Ψ_Q, which is relevant to the changes of peak position and peak width is shown in **Fig. S1-2A**. The difference between the cross-peak around (300,100) in Ψ_Q and **Fig. S1-2A** is a positive cross-peak at (300,100) (**Fig. S1-2B**). This result indicates that the intensity of the characteristic peak of Q increases under the intermolecular interaction ($\Delta\varepsilon_Q > 0$).

A   B
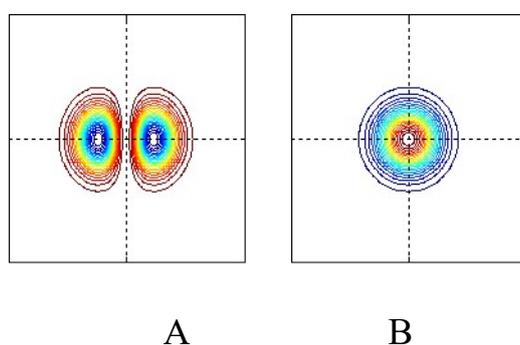
Figure S1-2 A The part of cross-peaks around (300, 100) in Ψ_Q, which is relevant to the changes of peak position and peak width; **B** The difference between the cross-peak around (300,100) in Ψ_Q and **Fig. S1-2A**

Comparison between the cross-peaks around (100, 100) in Ψ_Q and the basic patterns of cross-peak shown in **Scheme S1-2**, both the peak position and the peak width increase under the intermolecular interaction ($\Delta X_P > 0$, $\Delta W_P > 0$).

Under this situation, the part of cross-peaks around (100, 300) in Ψ_P, which is relevant to the changes of peak position and peak width is shown in **Fig. S1-3A**. The difference between the cross-peak around (100,300) in Ψ_P and **Fig. S1-3A** is a negative cross-peak at (100,300) (**Fig. S1-3B**). This result indicates that the intensity of the characteristic peak of P increases

under the intermolecular interaction ($\Delta\varepsilon_P < 0$).

A   B
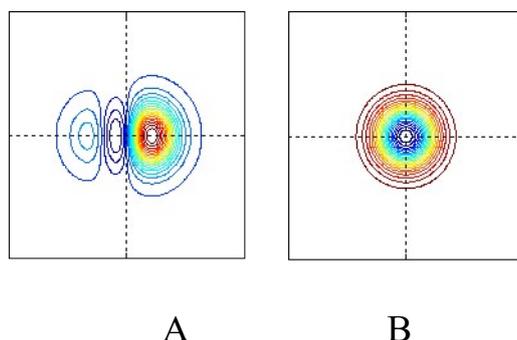
Figure S1-3 A The part of cross-peaks around (100, 300) in Ψ_P, which is relevant to the changes of peak position and peak width; **B** The difference between the cross-peak around (100, 300) in Ψ_P and **Fig. S1-3A**

The preset peak parameters of U and V are listed in **Table S1-4.**

Table S1-4 The preset peak parameters of U and V in the model system

j	X_j (nm)	W_j (nm)	ε_j
U	101	21	0.95
V	300	19	1.05

We confirm subtle changes in peak position, peak width, and intensity can be correctly obtained via the characteristic patterns of cross-peaks of the 2D asynchronous spectra generated via the DAOSD approach.

Then, we consider a chemical system containing P, Q. If P has a characteristic peak at X_P, but Q has no spectral contribution in the spectral region investigated. Subtle changes in peak position and peak width can be revealed by the characteristic pattern of cross-peaks around (X_P, X_P) in the 2D asynchronous spectrum. The variation in the intensity of the peak can be deduced by the ASAP approach described in another paper of our

previous work [S1-2].

References

S1-1   J. Chen, Q. Bi, S.X. Liu, X. P. Li, Y. H. Liu, Y. J. Zhai, Y. Zhao, L.M. Yang, Y.Z. Xu, I. Noda, J.G. Wu, Double Asynchronous Orthogonal Sample Design Scheme for Probing Intermolecular Interactions, J. Phys. Chem. A., 2012, **116**, 10904–10916.

S1-2   X.P. Li, A.Q. He, K. Huang, H.Z. Liu, Y. Zhao, Y.J. Wei, Y.Z. Xu, I. Noda, J.G. Wu, Two-dimensional Asynchronous Spectrum with Auxiliary Cross Peaks in Probing Intermolecular Interactions, Rsc. Adv., 2015, **5**, 87739-87749.

Part 2: The detailed description of the model system

The equilibrium constant of Eq. 2 in the model system is set as 0.01 L/mol.

The spectrum of S possesses a characteristic peak that can be described by a Gaussian function, and the spectral function of S is present in Eq. S2-1.

$$f_S(x) = \varepsilon_S e^{-\frac{(\ln 2)(x - X_S)^2}{W_S^2}} \tag{S2-1}$$

where ε_S, X_S, and W_S, respectively are the molar absorptivity, peak position, and half-width at the half-height of the characteristic peak of S.

The spectrum of R possesses a characteristic peak that can be described by a Gaussian function, and the spectral function of R is present in Eq. S2-2.

$$f_R(x) = \varepsilon_R e^{-\frac{(\ln 2)(x - X_R)^2}{W_R^2}} \tag{S2-2}$$

where ε_R, X_R, and W_R, respectively are the molar absorptivity, peak position, and half-width at the half-height of the characteristic peak of R.

Table S2-1 The initial concentrations of S and T

	C_S (mol/L)	C_T (mol/L)
1	0	0.1
2	0.55	0.1
3	0.50	0.1
4	0.45	0.1
5	0.60	0.1

Table S2-2 Peak parameters of S and R in the model system

	Peak-position	Peak-width	Absorbance
S	305.6	35	1.0
R	306.6	30	1.0

The cross-peaks pattern in the first model system without noise

In the noisy free model system, 1D spectra constructed by the five solutions are shown in **Figure S2-1**. Then, we use the five 1D spectra to construct a 2D asynchronous spectrum (**Figure 1**). In this case, the characteristic pattern of cross-peaks is clear, we can find that the left and lower cross-peaks are stronger than the right and upper cross-peaks. Thus, we judge that **Figure S2-1** belongs to situation 3 in **Scheme 1**. In situation 3, which means the peak of S undergoes a redshift, and the width of the characteristic peaks decreases upon interacting with T. This result is in good agreement with the preset peak parameter of S and R listed in **Table**

S2-2.
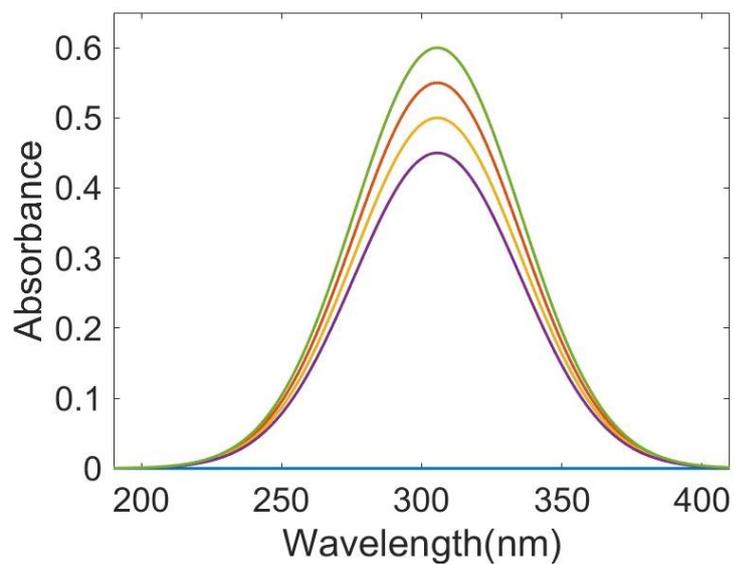
Figure S2-1 The 1D spectrum of the noisy free model system.
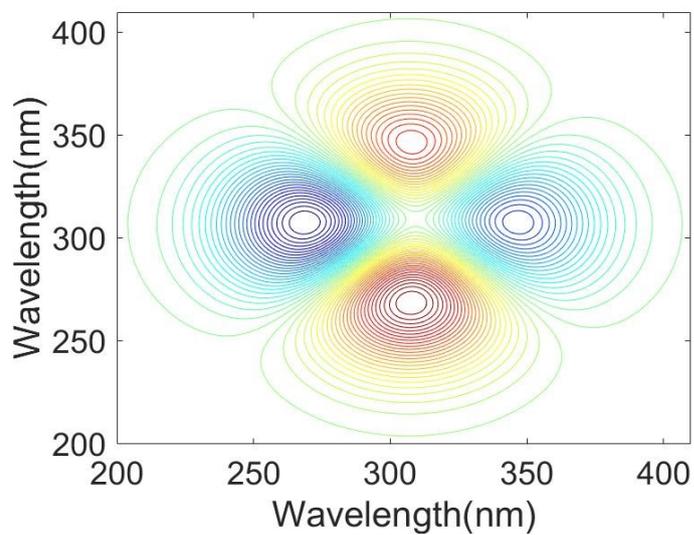
Figure S2-2 The 2D asynchronous spectrum of the noisy free model system.
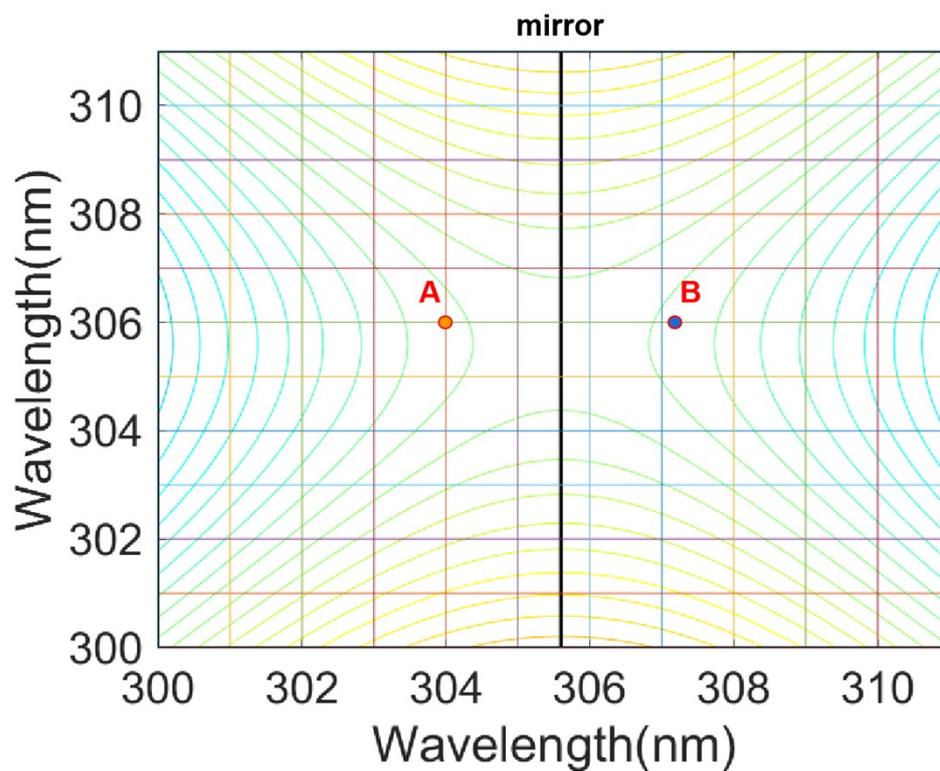
Figure S2-3 A schematic contour map of a magnified 2D asynchronous spectrum, where the cross peaks are symmetric with respect to a vertical mirror ($x = 305.6$ nm, the black bold vertical line). The intersection of grid lines is the discrete data point of the 2D asynchronous spectrum. Point A belongs to the array of discrete data points, while its mirror image (point B) does not belong to the array of discrete data points.

Part 3: A brief description of the Kolmogorov–Smirnov two-sample test

In this paper, the Kolmogorov-Smirnov two-sample test (denoted as K-S test hereafter) is used to verify whether the two sets of data come from the same distribution. This part provides a brief description of the Kolmogorov-Smirnov test, and a detailed description of the Kolmogorov-Smirnov test can be found in the literature [S3-1].

In the K-S test, the null hypothesis H_0 and the alternative hypothesis H_1 are defined as follows:

1) H_0 corresponds to the case where there is no significant difference between the two sets of data.

2) H_1 corresponds to the case where there is a significant difference between the two sets of data.

The criterion value α to reject the null hypothesis was set at 0.05.

Herein, we use the following two sets of data (X, Y) as an example to show how to use the K-S test to check whether the two sets of data come from the same distribution.

X = [2.91, -0.93, 1.17, 1.51, 0.5, 2.15, -0.94, -0.65, -0.8, -0.18, 0.23, -1.79, 0.98, -0.22, 0.17, -1.61, -0.65, -0.76, -0.04, -0.21, -1.54, 1.15, -0.5, 1.07, -0.96, 0.45, -1.38, -0.12, -0.57, -1.8, -0.89, -0.2, 0.97, 0.21, -0.61, 0.4, 0.2, 2, -0.31, -2.42, 0.81, 0.77, 1.75, -1.71, -1.26, 1.54, -0.9, 0.27, 0.57, 0.23];

Y = [0.81, -0.29, -0.46, 0.35, -0.86, -0.98, -1, 1.99, -1.23, 0.8, -0.26, 1.13, 0.47, 0, -1.2, 0.35, -2.02, 0.46, -0.17, -1.13, 0.18, 0.53, -0.02, 0.81, 1.22, -0.73, 1.59, -1.38, 0.94, 0.17, -0.33, 0.3, -0.68, -0.59, -1.53, 0.81, -1.01, -0.73, 0.05, -1.98, -2.2, 0.58, 1.26, 0.09, -0.47, -0.8, -0.12, 0.48, -0.59, -2.5, -0.74].

In this case, the sizes of both X and Y, which are respectively, denoted as m_X and m_Y, are 51.

The analysis is carried out via the following procedure:

1) Both X and Y are rearranged in ascending order. The sorted X and Y are denoted as X_1 and Y_1.

2) Empirical distribution functions for X and Y (denoted as $F_X(t)$, $F_Y(t)$, respectively), , were obtained via Eq. S3-1A and Eq. S3-1B.

$$F_X(t) = \begin{cases} 0 & t < X_1(1) \\ \dfrac{j}{m_X} & X_1(j) \le t < X_1(j+1), 1 \le j < m_X \\ 1 & t \ge X_1(m_X) \end{cases} \qquad \text{(S3-1A)}$$

$$F_Y(t) = \begin{cases} 0 & t < Y_1(1) \\ \dfrac{j}{m_Y} & Y_1(j) \le t < Y_1(j+1), 1 \le j < m_Y \\ 1 & t \ge Y_1(m_X) \end{cases} \qquad \text{(S3-1B)}$$

3) The value of D_{XY} is calculated via Eq. S3-2. In this case, the value of D_{XY} turns out to be 0.118.

$$D_{XY} = \max [F_X(t) - F_Y(t)] \qquad \text{(S3-2)}$$

4) Under the following conditions: 1) $m_X > 50$ or $m_Y > 50$; 2) The criterion value α is set as 0.05, the corresponding c_α, is calculated via Eq.

S3-3. The value of c_α is 0.269.

$$c_\alpha = 1.36 \sqrt{\frac{m_X + m_Y}{m_X m_Y}} = 0.269 \tag{S3-3}$$

5) The value of Asymp. Sig., which means asymptotic significance (denoted as P), can be calculated via Eq. S3-4.

$$P = 2e^{-\frac{2}{(1.36)^2}(\frac{D_{XY}}{c_\alpha})^2} \tag{S3-4}$$

If $P \geq 0.05$, we accept the null hypothesis (H_0);

If $P < 0.05$, we reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1).

For the data set X, Y, we have P=0.981>0.05, hence, X and Y come from the same distribution.

A brief process of the K-S test is shown in **Scheme S3-1**.
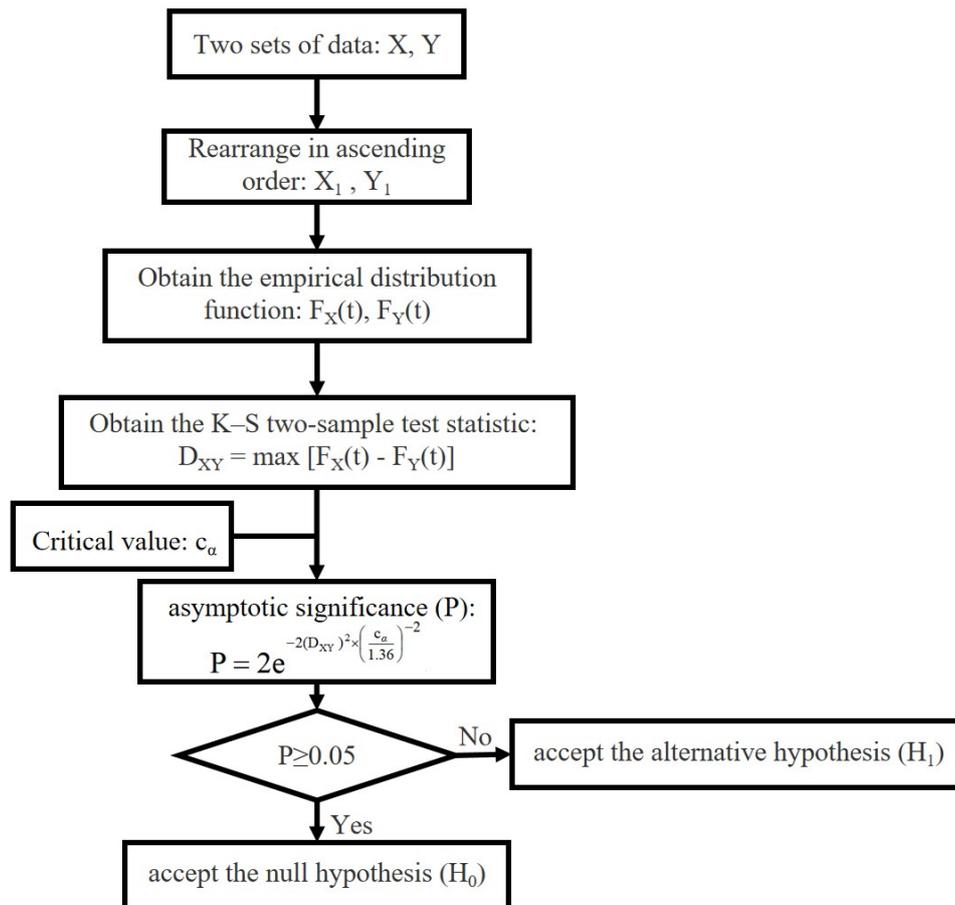
Scheme S3-1 Process of the K-S test.

References

S3-1   J.W. Pratt, J.D. Gibbons, Kolmogorov-Smirnov Two-Sample Tests. Concepts of Nonparametric Theory, 1981, 318–344.

Part 4: Analysis of the model system via the Kolmogorov-Smirnov two-sample test

In the 1D spectrum of S (**Figure S4-1**), a characteristic peak whose peak position is 305.6 nm can be observed. The K-S test is used to check whether the cross-peak around (305.6, 305.6) has a mirror symmetry with respect to $x = 305.6$ nm in **Figure 1**.

In the K-S test, two sets of data from the 2D asynchronous spectrum shown in **Figure 1** were generated:

The data of the first data set are selected randomly from the array of discrete data points within a rectangular region ((250:305.6)⊗(240:375), the region is marked by a red rectangle in **Figure S4-2**). The size of the first data set is 133. The data are $\{\Psi(x_A^1, y_A^1), \Psi(x_A^2, y_A^2), …, \Psi(x_A^{133}, y_A^{133})\}$.

The data of the second data set are selected from the array of the discrete data points within another rectangular region ((361.2:305.6)⊗(240:375), the region is marked by a blue rectangle in **Figure S4-2**). The size of the second data set is also 133.

To check whether mirror symmetry occurs or not, it is wise that the data points in the first data set and those in the second data set are mirror images of one another with respect to the mirror ($x = X_S = 305.6$ nm). However, X_S is not exactly located at any discrete data point of the 1D spectrum. Moreover, X_S is not exactly located at any middle point between

two adjacent discrete data points, either. Hence, the mirror images of the data points of the first data set cannot be within the array of discrete data points of the 2D asynchronous spectrum. Herein, the following rule is adopted in the selection of the data point of the second data set: For the ith data point of the first data set, whose coordinates are (x_A^i, y_A^i), the coordinates of the corresponding point in the second data set is (x_B^i, y_A^i). In this case, x_B^i is the abscissa of a discrete data point in the 1D spectrum, which is the closest to the value of $2X_S - x_A^i$ (the mirror image of x_A^i with respect to $x = X_S = 305.6$ nm.

Subsequently, the two data sets are subjected to the K-S test using the SPSS software. The calculation result shows that the value of D_{XY} is 0.632 and the corresponding value of Asymp. Sig. is 0.000. Therefore, we can reject H_0 and get a conclusion that there is no mirror symmetry with respect to $x = 305.6$ nm.
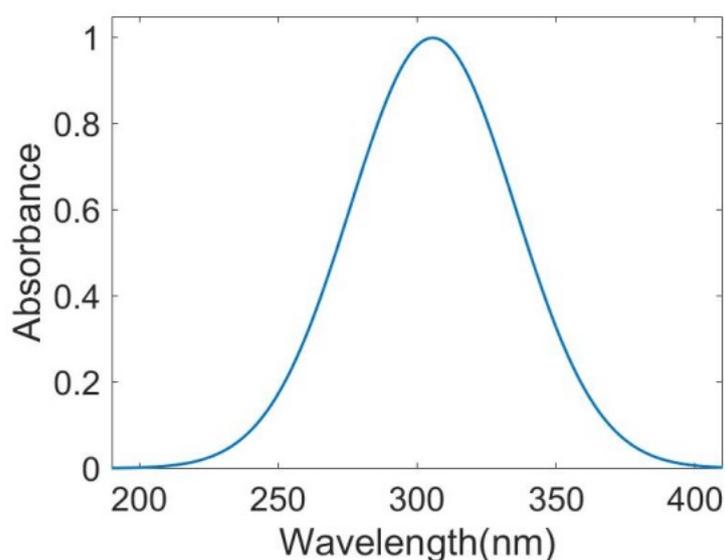
Figure S4-1 The 1D spectrum of S.
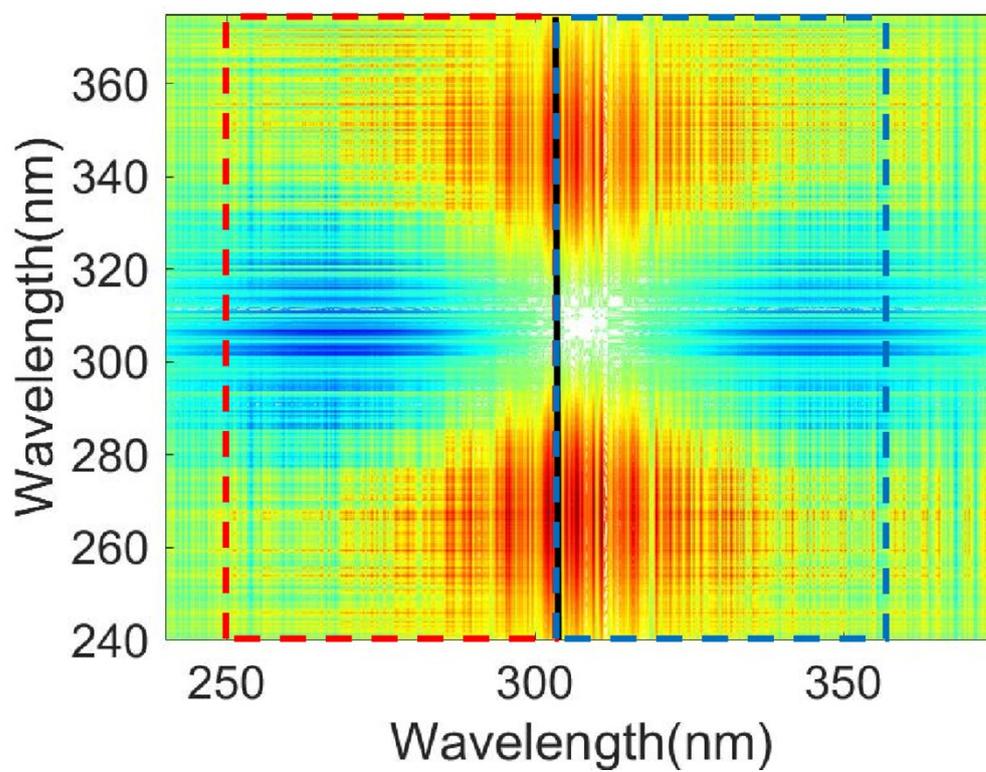
Figure S4-2 The 2D asynchronous spectrum of the model system.

Part 5: Detail description on the analysis of the N-methylpyrrolidone/artemisinin system

Reagents

Artemisinin with a purity of 98% was purchased from Aladdin. N-methylpyrrolidone, acetonitrile was of AR grade and obtained from Beijing Chemical Company.

Instrument

FTIR spectra were collected on a Thermo-Fischer Nicolet 6700 Fourier transform infrared spectrometer equipped with an attenuated reflection accessory. All the spectra were recorded at a resolution of 4 cm^{-1} and 32 scans were co-added. During the experiment, the FTIR spectrometer was purged by dry air to minimize the interference of water vapor.

Generation of 2D asynchronous spectra by using the DAOSD approach

In the experiment, both artemisinin and N-methylpyrrolidone were dissolved in acetonitrile. FTIR spectra of the acetonitrile solutions were recorded. Then a 2D asynchronous spectrum was generated by using the DAOSD approach.

The general procedures for the generation of the 2D asynchronous spectra by using the DAOSD approach are as follows:

1. A series of acetonitrile solutions with different concentrations of

artemisinin were prepared. Suitable concentration ranges of artemisinin were determined based on the fact that the relationship between the absorbance of the carbonyl band in FTIR spectra and the concentration of the artemisinin is linear. Similarly, suitable concentration ranges of NMP were obtained according to the fact that the absorbance of the amide band and concentrations of NMP are in a linear relationship. Based on the experimental result, suitable concentration ranges for artemisinin and NMP turn out to be 0-0.035 g/ml and 0-0.014 g/ml, respectively.

2. Based on the above result, a group of solutions containing both artemisinin and NMP was prepared. The concentrations of artemisinin and NMP are listed in **Table S5-1**. FTIR spectra of the group of solutions were recorded.

3. A 2D asynchronous spectrum ($\mathbf{\Psi_{NMP}}$) was generated from the set of 1D spectra obtained in step 2 based on a program written in our lab using the software of MATLAB (The Math Works, Inc.).

The Kolmogorov-Smirnov two-sample test in the N-methylpyrrolidone/artemisinin system

In the 1D spectrum of artemisinin **(Figure S5-1)**, a carbonyl band whose peak-position is 1739 cm^{-1} can be observed. The K-S test is used to check whether the cross-peak around (1739, 1739) has mirror symmetry with respect to $x = 1739$ cm^{-1} occurs or not in **Figure 5**.

In the K-S test, two sets of data from the 2D asynchronous spectrum shown in **Figure 5** were generated:

The data of the first data set are selected randomly from the array of discrete data points within a rectangular region $((1690:1739)\otimes(1675:1790)$, the region is marked by a red rectangle in **Figure S5-3**). The size of the first data set is 61. The data are $\{\Psi(x_C^1, y_C^1), \Psi(x_C^2, y_C^2), \ldots, \Psi(x_C^{61}, y_C^{61})\}$.

The data of the second data set are selected from the array of the discrete data points within another rectangular region $((1739:1788)\otimes(1675:1790)$, the region is marked by a blue rectangle in **Figure S5-3**). The size of the first data set is also 61.

To check whether mirror symmetry occurs or not, it is wise that the data points in the first data set and those in the second data set are mirror images of one another with respect to the mirror $(x = X_{artem} = 1739$ cm^{-1}). However, X_M is not exactly located at any discrete data point of the 1D spectrum. Moreover, X_{artem} is not exactly located at any middle point between two adjacent discrete data points, either. Hence, the mirror images of the data points of the first data set cannot be within the array of discrete data points of the 2D asynchronous spectrum. Herein, the following rule is adopted in the selection of the data point of the second data set: For the ith data point of the first data set, whose coordinates are (x_C^i, y_C^i), the coordinates of the corresponding point in the second data set is (x_D^i, y_C^i).

In this case, $x_D{}^i$ is the abscissa of a discrete data point in the 1D spectrum, which is the closest to the value of $2 X_{artem} - x_A{}^i$ (the mirror image of $x_C{}^i$ with respect to $x = X_{artem} = 1739$ cm^{-1}.

Subsequently, the two data sets are subjected to the K-S test using the SPSS software. The calculation result shows that the value of D_{XY} is 0.272 and the value of Asymp. Sig. is 0.022. Therefore, we can reject H_0 and get the conclusion that there is no mirror symmetry with respect to x=1739 cm^{-1}.

Table S5-1 The Initial Concentrations of solutions used to construct Ψ_{NMP}

	Group I (For the generation of Ψ_{NMP}) g/mL	
	C_{artem}	C_{NMP}
1	0	8.00×10^{-3}
2	1.001×10^{-2}	8.00×10^{-3}
3	1.555×10^{-2}	8.00×10^{-3}
4	2.004×10^{-2}	8.00×10^{-3}
5	2.503×10^{-2}	8.00×10^{-3}
6	3.009×10^{-2}	8.00×10^{-3}
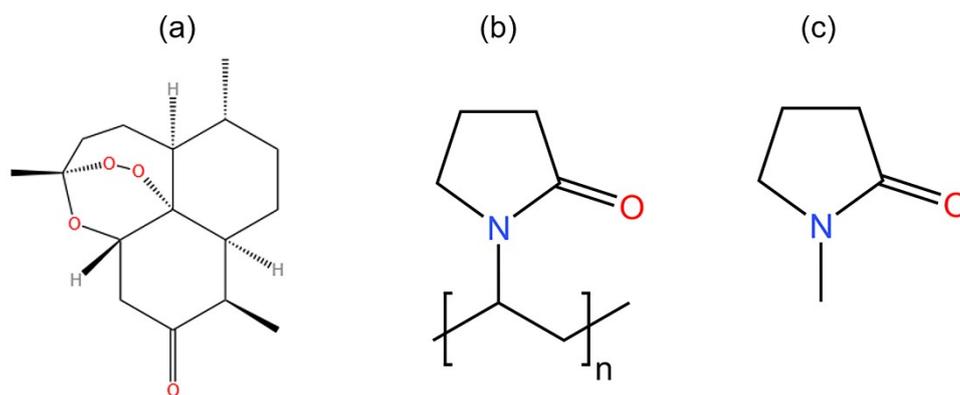
Scheme S5-1. Molecular structures of (a) artemisinin, (b) PVP, (c) NMP.
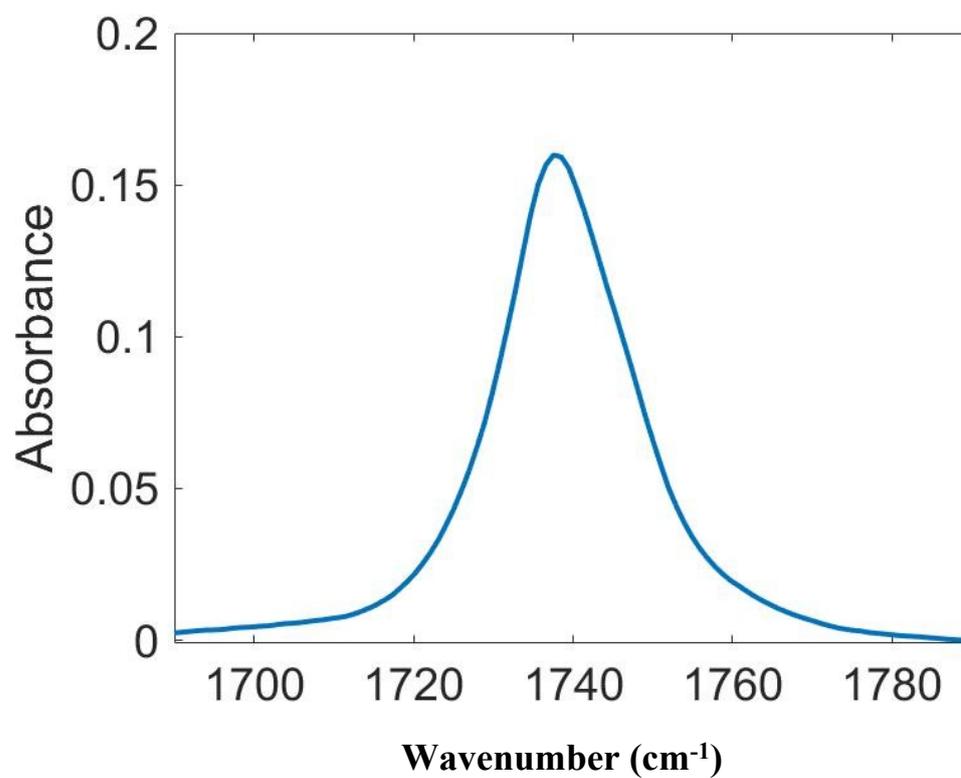
Figure S5-1 The FTIR spectrum in the carbonyl region of the artemisinin.
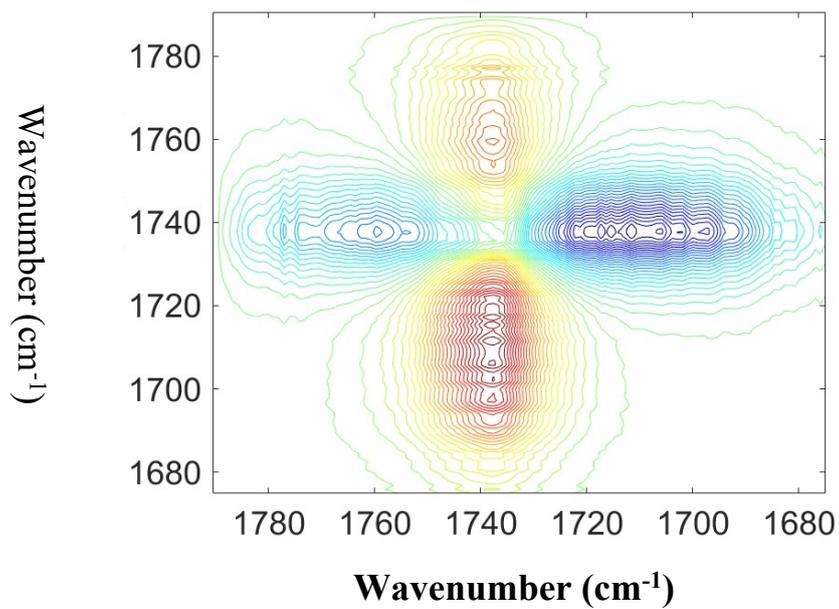
Figure S5-2 2D asynchronous spectra Ψ_{NMP} constructed by using the

DAOSD approach.
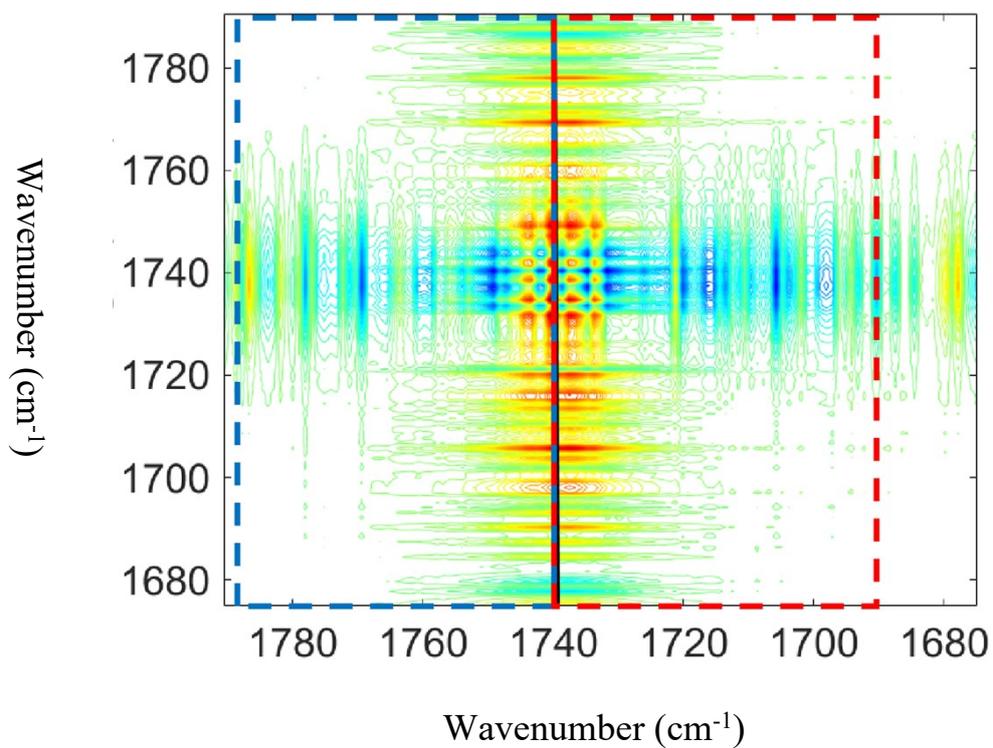
Figure S5-3 The 2D asynchronous spectrum of N-

methylpyrrolidone/artemisinin system.
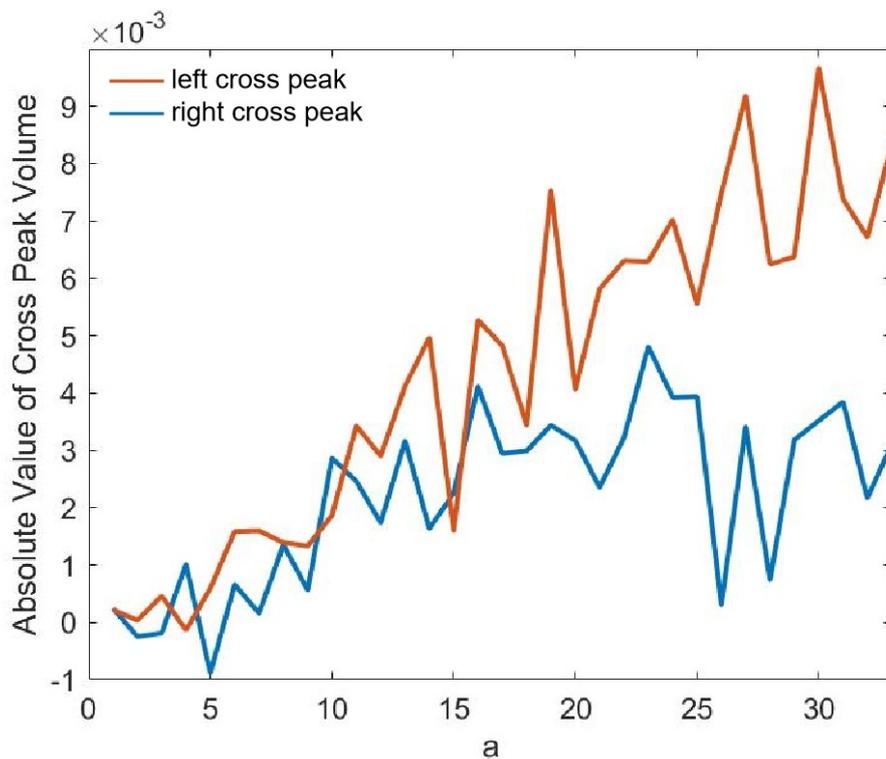
Figure S5-4 The absolute values of the volumes of the left and right

cross-peaks as a function of a.