Supporting Information

NMR for Mixture Analysis: Concentration Ordered Spectroscopy

Bin Yuan[†], Zhiming Zhou[†], Bin Jiang[†], Ghulam Mustafa Kamal[§], Xu Zhang[†], Conggang Li[†], Xin Zhou[†], Maili Liu[†]*

[†] State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, National Center for Magnetic Resonance in Wuhan, Wuhan Institute of Physics and Mathematics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences-Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430071, China; University of Chinese Academy of Sciences, Beijing 10049, China

[§] Department of Chemistry, Khwaja Fareed University of Engineering and Information Technology Rahim Yar Khan 64200, Pakistan.

* Corresponding author email: ml.liu@apm.ac.cn

Contents:

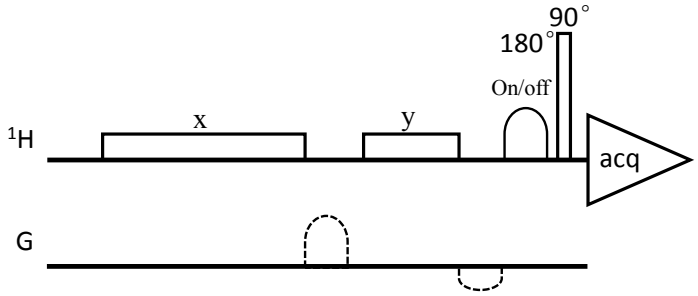

Figure S1. The modified Pre-SAT180 pulse sequence, with 90˚ phase shifted pre-saturation employed to further suppress residual water peak. Two experiments were performed, with the power of 180˚ adiabatic pulse on or off. The sum of two experiments leads to the final spectrum for quantitation.

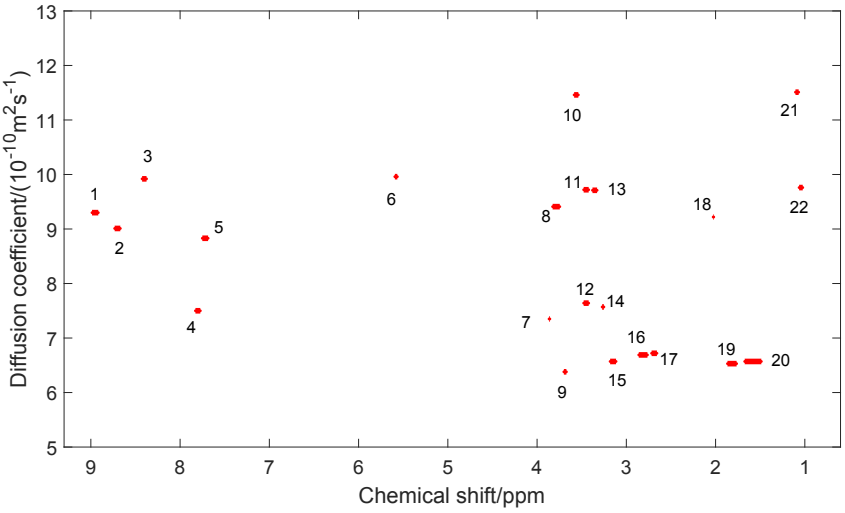

Figure S2. Plot of diffusion coefficients as a function of chemical shift (DOSY) for the peaks of the *Red Bull*® sample. The peak assignment: niacin (**1**, **2**, **3**, **5**), caffeine (**4**, **7**, **12**, **14**), acesulfame-K (**6**, **18**), arginine (**9**, **15**, **19**, **20**), ethanol (**10**, **21**), citrate (**16**, **17**), propylene glycol (**8**, **11**, **13**, **22**). The plot indicates that severe overlap of niacin, acesulfame-K and propylene glycol, and caffeine and arginine along the diffusion dimension. The large variation in the D values of the niacin is possibly due to low signal to noise ratio.

Table S1. NMR data for constructing CORDY of amino acids

Component	Input data								Output data				
	Peak ID	Chemical shift(δ)	Area	Line shape[a]	Possible H type	Diffusion coefficient (D, $\times 10^{10}$ m/s)			Assignment	Proton Number	Concentration (C. mM)		
						D-value	Mean	Deviation			C-value	Mean	Deviation
Glycine	14	3.564	0.842	s	CH3, CH2, CH	10.0	9.96	0.00	2-CH2	2	5.98	5.98	0.00
L-arginine	12	3.767	0.538	t	CH3, CH2, CH	6.31	6.41	0.06	2-CH	1	7.65	7.55	0.08
	18	3.232	1.051	t	CH2, CH	6.44			5-CH2	2	7.47		
	21	1.898	1.058	m	CH2, CH	6.42			3-CH2	2	7.52		
	22	1.673	1.063	m	CH2	6.45			4-CH2	2	7.55		
L-glutamate	11	3.787	3.865	t	CH3, CH2, CH	7.03	7.01	0.02	2-CH	1	54.9	52.8	1.8
	19	2.513	7.295	m	CH2, CH	7.00			4-CH2	2	51.8		
	20	2.128	7.286	m	CH2, CH	7.00			3-CH2	2	51.7		
L-histidine	1	8.642	0.696	s	CH	7.00	6.99	0.01	2'-CH	1	9.88	9.72	0.15
	4	7.376	0.690	s	CH	6.99			4'-CH	1	9.80		
	10	4.022	0.679	t	CH2, CH	6.98			2-CH	1	9.65		
	16	3.334	1.345	m	CH2, CH	6.99			3-CH2	2	9.56		
Threonine	8	4.254	0.346	m	CH2, CH	7.68	7.66	0.03	3-CH	1	4.91	4.85[b]	0.11
	13	3.593	0.346	d	CH3, CH2, CH	7.66			2-CH	1	4.92		
	23	1.316	0.998	d	CH3, CH2, CH	7.63			4-CH3	3	4.72		
Tryptophan	2	7.718	0.388	d	CH	6.33	6.34	0.03	4'-CH	1	5.52	5.44	0.17
	3	7.525	0.386	d	CH	6.34			7'-CH	1	5.48		
	5	7.307	0.387	s	CH	6.30			2'-CH	1	5.50		
	6	7.270	0.383	t	CH	6.36			6'-CH	1	5.44		
	7	7.187	0.393	t	CH	6.31			5'-CH	1	5.58		
	9	4.066	0.395	dd	CH2, CH	6.38			2-CH	1	5.60		
	15	3.475	0.376	dd	CH2, CH	6.33			3-CH2(1/2)	1	5.34		
	17	3.301	0.358	dd	CH2, CH	6.35			3-CH2(1/2)	1	5.09		

[a] Line-shape: s: singlet; d: doublet; dd: doublet-doublet; t: triplet; m: multiplet.

[b] used as internal reference.

Table S2. NMR data for constructing CORDY of Red Bull

Component	Input data								Output data				
	Peak ID	Chemical shift(δ)	Area	Line shape[a]	Possible H type	Diffusion coefficient (D, $\times 10^{10}$ m/s)			Assignment	Proton Number	Relative Concentration (C)		
						D-value	Mean	Deviation			C-value	Mean	Deviation
Acesulfame K	6	5.579	3.67	q	CH	9.96	9.59	0.52	5-CH	1	1.48	1.47	0.02
	18	2.023	10.8	d	CH3, CH2	9.22			6-CH3	3	1.46		
Arginine	9	3.685	70.4	t	CH3, CH2, CH	6.38	6.5	0.1	2-CH	1	28.5	27.2	1.0
	15	3.148	129.8	dd	CH2, CH	6.57			5-CH2	2	26.2		
	19	1.812	131.2	m	CH2, CH	6.53			3-CH2	2	26.5		
	20	1.595	135.5	m	CH2, CH	6.57			4-CH2	2	27.4		
Caffeine	4	7.804	15.0	d	CH	7.50	7.52	0.12	8-CH	1	6.08	6.02	0.05
	7	3.861	44.7	d	CH3, CH2, CH	7.35			7-CH3	3	6.02		
	12	3.441	44.7	s	CH3, CH2, CH	7.64			1-CH3	3	6.02		
	14	3.259	44.2	s	CH3, CH2, CH	7.57			3-CH3	3	5.95		
Citrate	16	2.810	943.6	d	CH3, CH2, CH	6.69	6.71	0.02	1,3-CH2 (1/2)	2	190.75	191.05	0.42
	17	2.687	946.6	d	CH3, CH2, CH	6.72			1,3-CH2 (1/2)	2	191.34		
Ethanol	10	3.561	11.7	q	CH2, CH	11.5	11.49	0.04	1-CH2	2	2.36	2.30	0.08
	21	1.085	16.7	d	CH3	11.5			2-CH3	3	2.24		
Niacin	1	8.943	2.46	d	CH	9.30	9.27	0.48	2-CH	1	0.99	1.00[b]	0.02
	2	8.703	2.46	dd	CH	9.01			6-CH	1	0.99		
	3	8.405	2.45	ddd	CH	9.92			4-CH	1	0.99		
	5	7.716	2.53	dd	CH	8.83			5-CH	1	1.02		
Propylene glycol	8	3.786	190.1	m	CH2, CH	9.41	9.7	0.2	2-CH	1	76.8	73.7	2.3
	11	3.448	180.4	dd	CH2, CH	9.72			1-CH2 (1/2)	1	73.0		
	13	3.348	181.8	dd	CH2, CH	9.71			1-CH2 (1/2)	1	73.5		
	22	1.044	530.4	d	CH3	9.76			3-CH3	3	71.5		

[a] Line-shape: s: singlet; d: doublet; dd: doublet-doublet; ddd: doublet-doublet-doulet; q; quadruplet; t: triplet; m: multiplet.

[b] used as internal reference.

Table S3. NMR data (CH region, δ6.3~δ9.4) for constructing CORDY of human urine

Component	Input data				Output data				
	Peak ID	Chemical shift(δ)	Area	Line shape[a]	Assignment	Proton Number	Relative concentration		
							C-value	Mean	Deviation
1,7-dimethylxanthine	24	7.870	11.7	s	8-CH	1	4.7		
1-methylhistidine	17	8.216	220.5	s	2-CH	1	88.4	93.8	7.6
1-methylhistidine	36	7.219	247.4	d	5-CH	1	99.2		
1-Methylnicotinamide	1	9.267	14.3	s	2-CH	1	5.7	5.6	0.2
1-Methylnicotinamide	3	8.956	13.5	d	6-CH	1	5.4		
1-Methylnicotinamide	4	8.883	14.1	d	4-CH	1	5.6		
4-aminohippurate	26	7.751	34.7	d	2,6-CH	2	6.9	7.3	0.5
4-aminohippurate	40	6.971	38.3	d	3,5-CH	2	7.7		
4-Hydroxyphenylacetate	43	6.851	29.5	d	3,5-CH	2	5.9		
formate	9	8.445	118.1	s	CH	1	47.3		
hippurate	25	7.823	262.9	d	2,6-CH	2	52.7	54.1	1.6
hippurate	30	7.624	134.0	t	4-CH	1	53.7		
hippurate	31	7.541	278.8	t	3,5-CH	2	55.9		
histidine	21	8.022	101.3	s	2-CH	1	40.6	35.5	7.3
histidine	38	7.133	75.7	s	5-CH	1	30.3		
Oxypurinol	19	8.178	46.2	s	3-CH	1	18.5		
tyrosine	37	7.182	75.7	d	2,6-CH	2	15.2	14.7	0.7
tyrosine	42	6.887	70.9	m	3,5-CH	2	14.2		
U1	28	7.693	74.1	d	CH	1	29.7	30.9	1.6
U1	32	7.490	79.8	d	CH	1	32.0		
U2	41	6.947	35.5	d	CH	1	14.2	13.3	1.4
U2	46	6.469	30.7	d	CH	1	12.3		
U3	11	8.324	36.8	d	CH	1	14.8	15.5	1.4
U3	23	7.957	85.4	dd	2CH	2	17.1		
U3	44	6.658	36.6	d	CH	1	14.7		
U4	20	8.036	15.0	d	CH	1	6.0	5.9	0.1
U4	45	6.547	14.5	d	CH	1	5.8		
	2	9.115	4.29	s	CH	1	1.72		
	5	8.667	2.49	s	CH	1	1.00[b]		
	6	8.639	6.63	d	CH	1	2.66		

	7	8.567	8.49	s	CH	1	3.40		
	8	8.532	23.5	dd	CH	1	9.43		
	10	8.376	4.98	s	CH	1	2.00		
	12	8.317	5.18	s	CH	1	2.08		
	13	8.287	2.91	s	CH	1	1.17		
	14	8.276	6.79	td	CH	1	2.72		
	15	8.259	3.59	s	CH	1	1.44		
	16	8.252	8.25	d	CH	1	3.31		
	18	8.203	43.2	s	CH	1	17.3		
	22	7.968	16.7	s	CH	1	6.69		
	27	7.720	24.5	d	CH	1	9.81		
	29	7.661	91.9	d	CH	1	36.9		
	33	7.457	190.5	t	CH	1	76.4		
	34	7.413	200.2	t	CH	1	80.3		
	35	7.263	188.2	td	CH	1	75.5		
	39	6.998	15.7	s	CH	1	6.30		
	47	6.425	3.54	d	CH	1	1.42		

[a] Line-shape: s: singlet; d: doublet; dd: doublet-doublet; t: triplet; dt: triplet-doublet; m: multiplet.

[b] used as internal reference.

NMR for Mixture Analysis: Concentration Ordered Spectroscopy

Bin Yuan,[a] Zhiming Zhou,[a] Bin Jiang,[a,b] Ghulam Mustafa Kamal,[c] Xu Zhang,[a,b,d] Conggang Li,[a,b,d] Xin Zhou,[a,b,d] Maili Liu,*,[a,b,d]

a. State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Center for Magnetic Resonance, Wuhan Institute of Physics and Mathematics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430071 (China)

b. University of Chinese Academy of Sciences, Beijing 100029, (China)

c. Department of Chemistry, Khwaja Fareed University of Engineering and Information Technology Rahim Yar Khan 64200 (Pakistan)

d. Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, 430074 (China).

*. Corresponding Author Email: Maili Liu, ml.liu@apm.ac.cn.

NMR Experiments

NMR experiments were performed for the analysis of the mixture of amino acids in the samples of *Red Bull*® and human urine by using Bruker Avance spectrometers (Bruker Biospin, Germany) operating at ^{1}H frequency of 599.81MHz (Oxford Magnet), 600.13MHz and 949.9MHz, respectively. All the spectrometers were equipped with a cryoprobe and the operating temperature was set at 25°C. All the quantitative analyses were performed with four dummy scans by default before acquisition. For the analysis of the model mixture of amino acids, quantitative 1D NMR data were recorded using the sequence as shown in Figure S1 with adiabatic pulse off, pre-acquisition delay of 14s to ensure magnetization fully relaxed, 2.3s for water peak pre-saturation and collection of 4 scans of 1.36s acquisition. 64k data points were acquired covering a spectral window of 12019.23 Hz, up to a total experimental time of 2.3 min. An exponential line-broadening of 0.3 Hz was applied to the data before Fourier Transformation (FT), with zero-filling to 128k. The peak area was integrated manually after careful phase and baseline correction. The overlapped peaks were deconvoluted using the module "line shapes" in Topspin 4.0. The chemical shifts were calibrated with reference to CH_3 group of threonine whose chemical shift was found to be as $\delta 1.316$ in HMDB database. As for the second sample i.e. *Red Bull*®, the quantitative 1D NMR data using the modified PreSAT180[1] (Figure S1) were collected with pre-acquisition delay of 12s, 3.42s for data acquisition, 4.5s for water peak pre-saturation and collection of 16 scans to ensure a reasonable signal to noise ratio. 64k data points were acquired covering a spectral window of 9590.79 Hz, up to a total experimental time of 7.5 min. The data were zero-filled to 128k before Fourier Transformation (FT). For human urine, the Bruker sequence ZGPR was used with 2s for pre-saturation, 2.15s for acquisition and 13s for further relaxation delay. 64k complex data points were acquired covering a spectral window of 15243.9 Hz, with total experimental time of 33 min. The data were zero filled to 256k before FT with 0.3 Hz line broadening. The ^{1}H spectrum of human urine is rather

complicated and thus deconvoluted automatically using the commercial software MestReNova. The integral values of selected multiplets can be easily calculated from the fitted peaks[2,3].

BPP-LED pulse sequence was used to measure the self-diffusion coefficients and for constructing DOSY conventionally, with a diffusion time of 80 ms, eddy current delay of 5 ms, gradient pulse duration of 2 ms, relaxation delay of 2 s, respectively. Eight gradient strengths linearly spaced between 1.49 and 47.03 Gauss/cm (for *Red Bull*®) and 1.51 and 47.84 Gauss/cm (for amino acids) were used with 32 (for *Red Bull*®) and or 8 (for amino acids) scans for each gradient. 64k (for *Red Bull*®) and or 32k (for amino acids) data points were acquired for each gradient value. An exponential weighted function (0.3 Hz) was applied to both datasets zero-filled by a factor of 2 before FT. The diffusion coefficients (D) are derived from the manually selected peaks using the module "T1T2" in Topspin 4.0.

Constructing CORDY

The first step for constructing CORDY is to prepare a working-table containing NMR peak identification (id), chemical shift, area (A_i), line-shape and potential spin (^{1}H) number (N_i). The self-diffusion coefficient may also be included as additional supplement. All this information can be obtained from quantitative and diffusion NMR experiments, general knowledge of NMR and used as input for the processing.

The next step, also the main part of the processing, is to define values of the N_{ij} for each peak (***i***) of component (***j***), respectively. It is recommended to start with a peak of aromatic (N_i=1) or methyl (N_i=3) region and to use the corresponding C_i (= A_i/N_i) as divisor applied to all peaks (Eq. S1a, S1b). Depending upon the properties the peaks with N=1, 2, or 3, $|\Delta| \leq 0.1$ (or 5%) can be considered as originating from the same group or component (***j***), and their mean concentration (C_j) can be derived (Eq. S1c).

$$N_{ij} = round\,(A_{ij}/C_i, 0), \tag{S1a}$$

$$\Delta_{ij} = A_{ij}/C_i - N_{ij}, \tag{S1b}$$

$$C_j = \frac{1}{m}\sum_{i=1}^{m}(A_{ij}/N_{ij}). \tag{S1c}$$

Where $round(x, 0)$ is an operation of rounding x to the nearest integer with deviation of Δ_{ij}. This process is repeated until all the peaks are grouped or assigned to different components.

Mathematically, the searching processes correspond to finding the greatest common divisors C_j for each component and minimizing the cost function f_{min} (Eq. S2):

$$f_{min} = \sum_{j=1}^{l}\left[\sum_{i=1}^{m}\left|A_{ij}/C_j - N_{ij}\right|\right] = \sum_{ij}|\Delta_{ij}| \tag{S2}$$

When tuning the peak between the groups to minimize f_{min}, one should pay attention to the requirements, such as N=1, 2, or 3, $|\Delta| \leq 0.1$ (or 5%) and peak properties. To speed up the searching process, one could sample out the grouped peaks from the working list and just search the remaining peaks. Once the peaks are grouped or assigned to different components, one can construct the CORDY, $S(\delta, C)$, according to the chemical shifts and the derived concentrations:

$$S(\delta,C) = \sum_{ij} \frac{S_{ij}(\delta)}{A_{ij}} N_{ij} \exp\left(\frac{-(C - C_{ij})^2}{2(C_j\Delta)^2}\right) \quad \text{(S3)}$$

$$C_{ij} = C_j\left(1 + \frac{A_{ij}}{C_j} - N_{ij}\right) \quad \text{(S4)}$$

Where $S(\delta)$ represents 1D NMR spectrum. The vertical scale of CORDY corresponds to concentration by default. It can be the absolute concentration if an internal or external reference is used, similar to conventional quantitative NMR. For simplicity, "$C_j\Delta$" is chosen as the error instead of standard deviation of peaks' concentration.

The algorithms are written in MATLAB and still being improved. The code is free for academic use upon request by email. The whole procedure is now semi-automatic and can be developed to be completely automatic in the future using scripts in commercial software (e.g. MestReNova NMR).

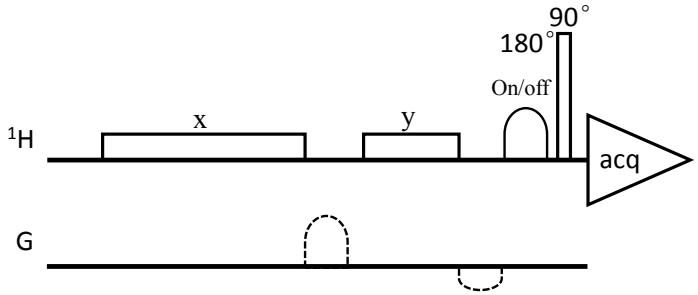

Figure S1. The modified Pre-SAT180 pulse sequence, with 90° phase shifted pre-saturation employed to further suppress residual water peak. Two experiments were performed, with the power of 180° adiabatic pulse on or off. The sum of two experiments leads to the final spectrum for quantitation.

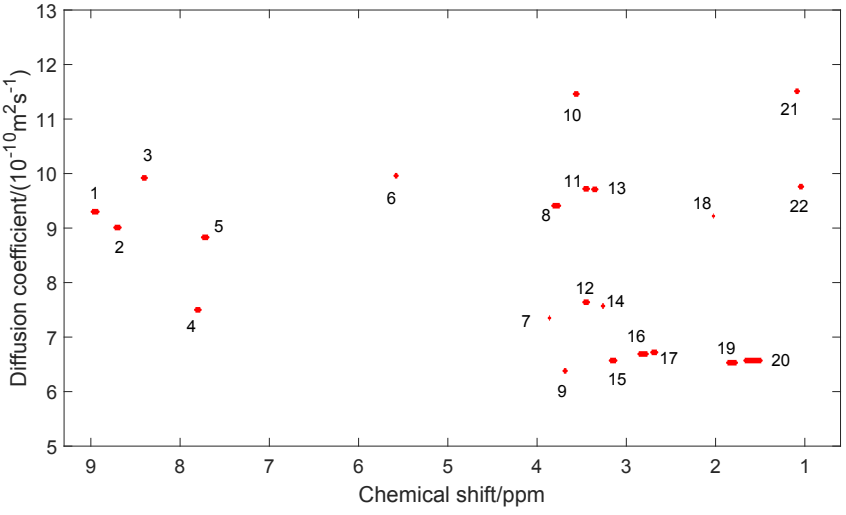

Figure S2. Plot of diffusion coefficients as a function of chemical shift (DOSY) for the peaks of the *Red Bull*® sample. The peak assignment: niacin (**1**, **2**, **3**, **5**), caffeine (**4**, **7**, **12**, **14**), acesulfame-K (**6**, **18**), arginine (**9**, **15**, **19**, **20**), ethanol (**10**, **21**), citrate (**16**, **17**), propylene glycol (**8**, **11**, **13**, **22**). The plot indicates that severe overlap of niacin, acesulfame-K and propylene glycol, and caffeine and arginine along the diffusion dimension. The large variation in the *D* values of the niacin is possibly due to low signal to noise ratio.

(1) Mo, H.; Raftery, D. *Journal of magnetic resonance* **2008**, *190*, 1-6.

(2) Bernstein, M. A.; Sykora, S.; Peng, C.; Barba, A.; Cobas, C. *Analytical Chemistry* **2013**, *85*, 5778-5786.

(3) Vadim Zorin, M. B., Felipe Seoane, Stanislav Sykora, Carlos Cobas. In *SMASH*: Italy, 2017.