Supplementary Data

The study of metabonomics combined with diversity of intestinal flora in LDP intervention in kidney-yin deficiency hyperthyroidism rats

Shiyu Ma[a] Lan Shen[a,b*] Chen Meiwan[c] Xiao Lin[a,b*] Qiang Wang[a] Ruofei Du[a] Yi Feng[b]

([a] School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China;

[b] Engineering Research Center of Modern Preparation Technology of Traditional Chinese medicine of Ministry of Education, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China;

[c] State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao 999078, China.)

To whom correspondence should be addressed at

Lan Shen

[a] School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

[b] Engineering Research Center of Modern Preparation Technology of Traditional Chinese medicine of Ministry of Education, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

Tel/Fax: 86-21-51322211; E-mail: alansusu@sina.com

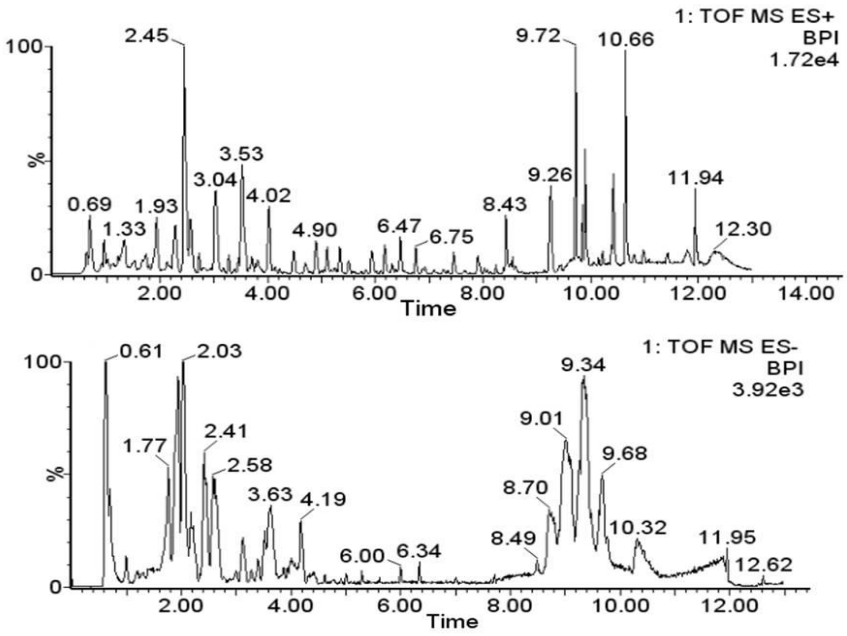

Fig. 1S The metabolic fingerprinting of normal rat urine in both positive and negative ion mode

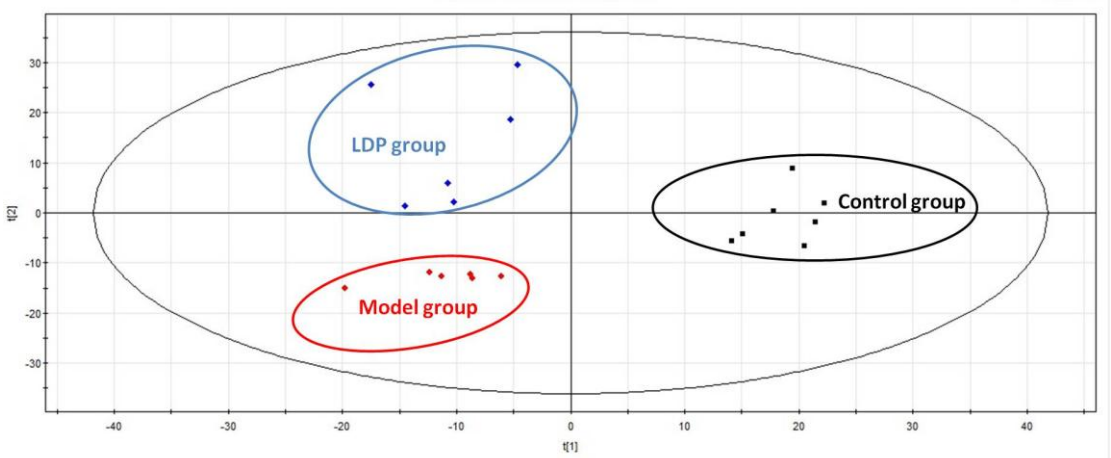

Fig. 2S The PLS-DA clustering score map of urine at positive ion for control group (n=7), LDP-treated group (n=6) and model group (n=6)

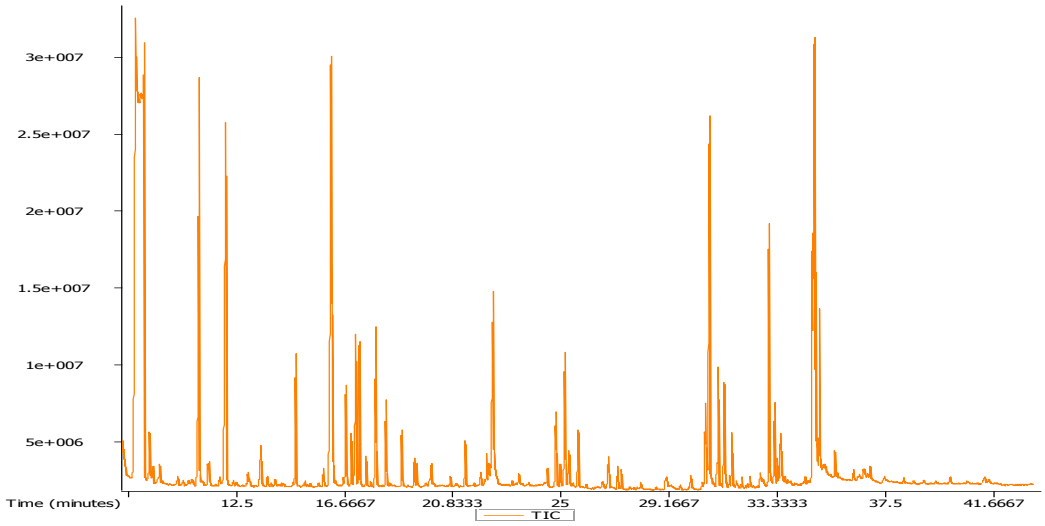

Fig. 3S The GC-MS fingerprinting of normal rat excrement

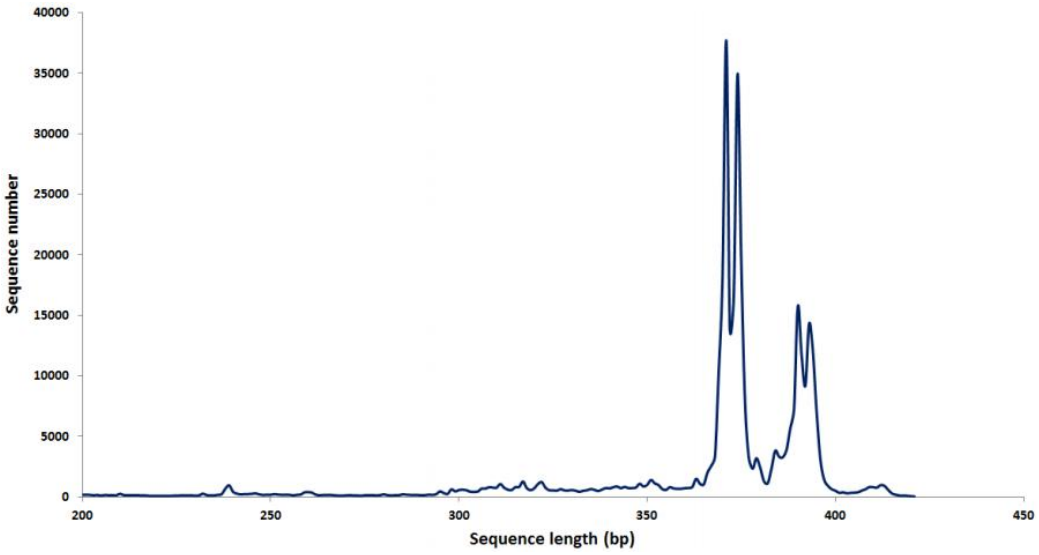

Fig. 4S The length distribution of mixed samples of 16S rDNA

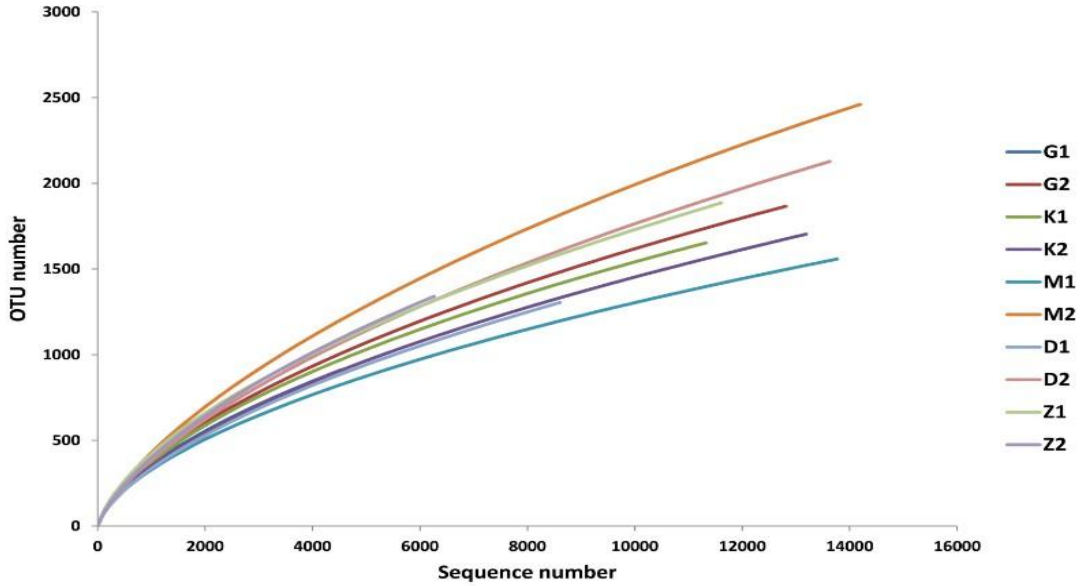

Fig. 5S Sample dilution curve in the case OUT= 0.03 of each groups rats

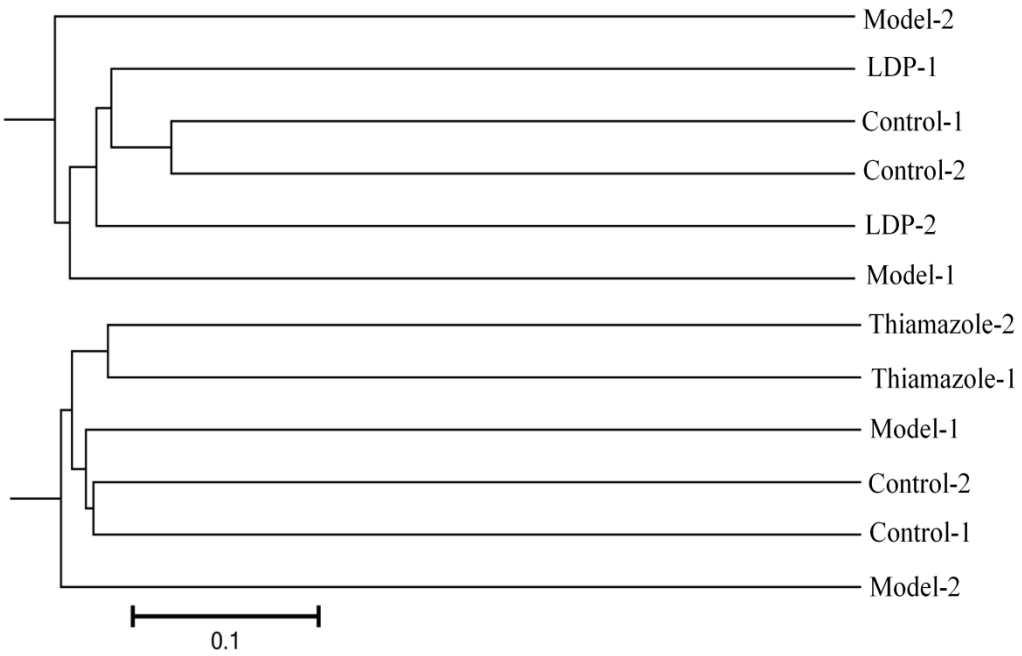

Fig. 6S The community structure similarity dendrogram of each sample in case of OUT = 0.03

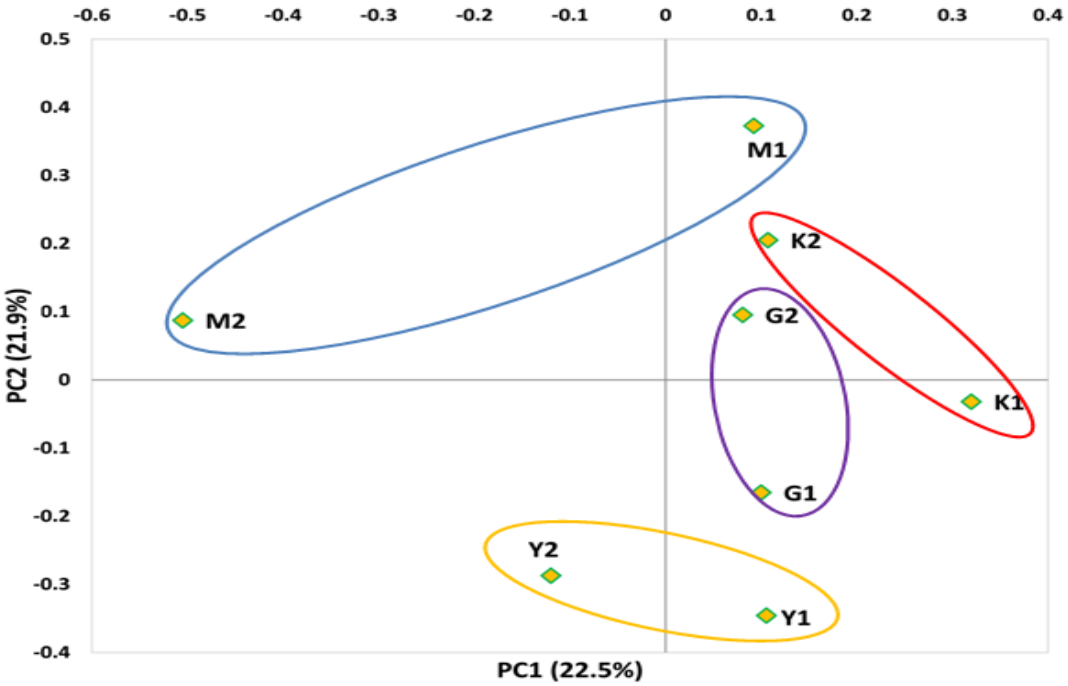

Fig.7S The PCoA analysis of each sample in the case of OUT = 0.03

Tab.1S The results of T_3, T_4, FT_3, FT_4 content test（X±SD）

Index	Mean			
	T_3	T_4	FT_3	FT_4
Control	0.54±0.01	44.7±12.48	1.53±0.41	16.48±3.96
Model	1.59±0.30**	143.8±39.44**	5.85±1.33**	48.42±7.84**
LDP-low	1.55±0.90	143.1±45.46	4.59±0.94	41.37±8.29
LDP-middle	1.18±0.22	124.2±15.61	3.41±0.56##	29.27±5.30##
LDP-high	1.26±0.81	146.4±30.35	3.38±0.80##	27.63±7.19##
Positive	1.04±0.23	149.5±27.07	3.10±0.33##	28.90±7.16##

Compare Model with Control, *mean $P<0.05$; **mean $P<0.01$;
Compare LDP group with Model, #mean $P<0.05$; ##mean $P<0.01$.

Tab. 2S Numbers of reads, unique reads and OTUs (97% similarity level) in the library of each sample, and coverage estimation of each library in groups.

Samples	**OTUs**	**ACE**	**Chao**	**Shannon**	**Coverage（%）**
Control	1198	3075.9098	2289.3584	4.864814	95.34
Model	1021	2305.6144	1780.9745	4.936413	96.45
LDP-low	903	2542.2812	1827.5391	4.697136	94.35
LDP-middle	1288	3438.1594	2458.8298	5.39530	94.28

				3	
LDP-high	1264	3324.3515	2354.7592	5.153814	94.96
Thiamazole tablets	1365	3600.5188	2632.1782	5.264399	93.80

Tab. 3S The read number, min length, max length and average length of each groups rats

Sample	**read number**	**min length**	**max length**	**average length**
Control-1	11321	200	418	367
Control-2	13187	200	420	368
Model-1	13764	200	420	368
Model-2	14194	200	418	363
LDP-low-1	8612	200	420	369
LDP -low-2	13628	200	419	383
LDP-middle-1	11610	200	420	365
LDP-middle-2	6261	202	417	368
LDP-high-1	4520	200	420	370
LDP-high-2	12812	200	420	367
Thiamazole-1	8116	200	417	367

Thiamazole-2	11550	200	418	367

1. Metabonomics of urinary

1.1 UPLC- TOF/MS analysis

Chromatographic separations were performed on an ACQUITYTM UPLC System (Waters Corporation, Milford, MA), equipped with a binary solvent delivery system and an autosampler. A BEH C18 column (2.1 mm×100 mm, 1.7 um, Waters Corporation, Milford, USA) was used. The column was maintained at 40 ◦C with a flow rate of 0.4 mL/min. The mobile phase was composed of water (Phase A) and acetonitrile containing 0.1% formic acid (Phase B) and the optimized gradient elution program of positive ions was set as followed: 0-1 min: 1-15% B; 1-5 min: 15-30% B;5-8 min: 30-80% B; 8-9 min: 80-100% B; 9-11 min: 100% B; 11-13 min: 100-1% B. The optimized gradient elution program of negative ions was set as followed: 0-1 min: 1-10% B; 1-5 min: 10-25% B;5-7 min: 25-50% B; 7-9 min: 50-100% B; 9-11 min: 100% B; 11-13 min: 100-1% B. MS spectrometry was carried out on a Waters Q-TOF Premier MS system (Waters Corp., Milford, MA) with an electrospray ionization source (ESI) operating in positive ion mode and negative ion mode. Nitrogen was used as the drying gas. The desolvation gas flow rate was 650 L/h and dessolvation temperature was maintained at 350 ◦C. Cone gas flow rate was maintained at 50 L/h and the source temperature was set at 110 ◦C. Capillary voltage and cone voltage were 3000 V and positive mode in 35 V(negative mode in 55V), respectively. The scan time and interscan delay were 0.28 s and 0.02 s, respectively. All analyses were obtained using an independent reference lock mass ion to ensure accuracy and reproducibility and leucine-enkephalin (m/z

566.2771 in positive mode, m/z 554.2615 in negative mode) was used as the lock mass at the concentration of 50 pg/μL and a flow rate of 10 mL/min. Data were acquired in centreid mode with a scan range from 50 to 1000 and a lock spray frequency of 10 s and averaged over 10 scans for correction.

1.2 Data collection and analysis

UPLC/MS spectra data were first processed by Markerlynx Applications Manager Version 4.1 (Waters, Manchester, UK), including the detection and retention time (R.T.) alignment of peaks in each chromatogram by Apex-Track-peak detection package incorporated in this software. The data were combined into a single matrix after aligning peaks with retention time-exact m/z pair and associated peak intensity. Then, ion intensities of each detected peak were normalized within each sample, to the sum of the intensities in that sample. Specifically, some parameters proposed [1-2] were used as parameters for data processing and were set as follows: retention time: 0-13.10 min, mass range: 50-1000 Da, mass tolerance: 0.05 Da, minimum intensity: 15% of the base peak intensity, maximum mass per retention time: 6 and retention time tolerance: 0.04 min. The processed data were then introduced to SIMCA-P version 11.5 (Umetrics, Umea, Sweden). Multivariate statistical analysis of the data: Principal Component Analysis (PCA) and Partial Least Square Discriminant Analysis (PLS-DA) were performed. PCA is an unsupervised multivariate statistical analysis method that transforms correlated variables of a dataset into a smaller number of independent variables, i.e., the principal components. PCA model was used as an overview to see the outliers, groups and trend of data and PLS-DA was then constructed to bring out specific variations between different treatment groups. Variable Importance in the

Projection (VIP) value of validated PLS-DA model is taken as the measurement index for peak selecting, meaning higher VIP value stands for greater contribution of a variable to the separation of different treatment groups and variables with VIP >1.0 are supposed to contribute significantly to the separation. T-test was also performed to select significantly differential peaks (p-value <0.05) according to their intensities.

2. Metabolites of fecal

2.1 GC TOF/MS analysis

GC TOF/MS analysis was performed on an Aligent 6890 gas chromatograph system equipped with a Pegasus III time-off light mass spectrometer (Pegasus HT, Leco Co., CA, USA). Chromatography was carried out using a DB-5MS capillary column 30 mm*250 mm i.d., 0.25 mm film thickness (Agilent J & W Scientific, Folsom, CA, USA) and helium was used as the carrier gas at the flow rate of 1 mL/min. The temperature of injector and the transfer line was set at 270 ◦C and 250 ◦C, respectively. And the oven temperature was initially set at 80 ◦C for 0.2 min and raised to 180 ◦C at the rate of 5 ◦C /min, then to 220 ◦C at the rate of 4 ◦C/ min and ramped to 280 ◦C at the rate of 20 ◦C /min1 and held for 10 min. The ion source temperature was 220 ◦C. The mass spectra of m/z 20-600 were acquired with electron impact ionization (70 eV) at the full scan mode. The detector voltage was set at 1450 V.

2.2 Data collection and analysis

2.2.1 Data extraction and pretreatment.

Raw GC/TOF MS data files were converted to CDF files by ChromaTOF software (v3.30, Leco Co., CA, USA) and CDF files were then processed in Matlab 7.0 (the

MathWorks, Inc, USA), including baseline correction, background reduction, smoothing, alignment, time-window setting and multivariate curve resolution. The resultant data were comprised of sample names, peak retention time and peak intensity. Some artificial peaks generated by noise, column bleed and by-products in the silylation procedure were removed manually from the dataset. Then, the three dimensional dataset was normalized to the area of the internal standard, mean centered and treated by unit variance scaling for further statistical analysis[3-6].

2.2.2 Metabolite identification

Metabolites were identified by importing the resolved mass spectra to mass spectra library search software National Institute of Standards and Technology (NIST) 08 library using NIST MS Search 2.0 software. Metabolites with similarity >600 were considered reliable (similarity 999 means a perfect match between the compound in the sample and the compound in the NIST library) and authentic reference standards were used to further validate identified metabolites including both retention time and mass spectrum match.

2.2.3 Multivariate analysis and potential biomarker selection.

The data consisting of sample names, peak indices (RT-m/z pair) and peak areas were introduced to SIMCA-P 12.0 software (Umetrics, Ume Sweden) for multivariate statistical analysis like Partial Least Square Discriminant Analysis (PLS-DA), which is a supervised pattern recognition if samples belong to different classes or groups, based on prior knowledge. By using PLS-DA, it is possible to visualize clustering and trends of all samples and efficiently detect the influential variables (i.e., metabolites) explaining the differences between samples, or groups of samples, by interpreting the variable weights. Variable

Importance in the Projection (VIP) value of PLS-DA model was employed as a screening parameter for biomarker selection and higher VIP value means greater contribution to the intergroup separation. Besides, p value of Students t-test based on peak area of compounds among different groups was taken as another measurement index for biomarker selecting. Only those with both VIP value > 1.0 and p value of t-test o 0.05 were considered influential in separating different treatment groups and would be chosen as potential biomarker candidates.

Reference

1. W.J. Cong, Q.L. Liang, L. Li, J. Shi, Q.F. Liu, Y. Feng, Y.M. Wang and G.A. Luo, Talanta,2012, 89,91-98.

2. Zhu Y, Feng Y, Shen L, Xu D, Wang B, Ruan K, Cong W, J Chromatogr B Analyt Technol Biomed Life Sci, 2013, 925, 110-116.

3. X. F. Gao, E. Pujos-Guillot and J. Se'be'dio, Anal. Chem., 2010, 82, 6447-6456.

4. A. Jiye, J. Trygg, J. Gullberg, A. I. Johansson, P. Jonsson, H. Antti, S. L. Marklund and T. Moritz, Anal. Chem., 2005,77, 8086-8094.

5. Y. M. Liu, T. L. Chen, Y. P. Qiu, Y. Cheng, Y. Cao, A. H. Zhao and W. Jia, Anal. Bioanal. Chem., 2011, 400, 1405-1417.

6. Yunyun Zhu, Wenjuan Cong, Lan Shen, Hai Wei, Yuan Wang, Lingyi Wang, Kefeng Ruan, Fei Wua and Yi Feng, Mol. BioSyst., 2014, 10, 304-312