Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1084 (2005) 214–221
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Strategy for metabonomics research based on high-performance liquid chromatography and liquid chromatography coupled with tandem mass spectrometry

Jun Yang, Guowang Xu*, Yufang Zheng, Hongwei Kong, Chang Wang, Xinjie Zhao, Tao Pang

National Chromatographic R&A Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 116011 Dalian, China

Available online 26 November 2004

Abstract

Metabonomics, the study of metabolites and their roles in various disease states, is a novel methodology arising from the post-genomics era. This methodology has been applied in many fields. Current metabonomic practice has relied on mass spectrometry (MS), gas chromatography–mass spectrometry (GC–MS), and nuclear magnetic resonance (NMR) to analyze metabolites. In this study, a strategy was developed for applying high-performance liquid chromatography (HPLC) and LC–MS–MS to metabonomics research. One of the key problems to be solved in this strategy is to match the peaks between the chromatograms. A peak alignment algorithm has been developed to match the chromatograms before the pattern recognition. As an application example, the strategy described above was applied to metabonomics research on liver diseases, and the false-positive result of live cancer diagnosis from the hepatocirrhosis and hepatitis diseases was effectively reduced to 7.40%. Based on the pattern recognition, several potential biomarkers were found and further identified by the following LC–MS–MS experiments. The structures of eight potential biomarkers were given for distinguishing the liver cancer from the hepatocirrhosis and hepatitis diseases.

Keywords: Metabonomics; Peak alignment; Liver; Cancer; Pattern recognition

1. Introduction

Metabonomics, defined as the quantitative measurement of the dynamic multi-parametric metabolic response of living systems to pathophysiological stimuli or genetic modification [1], has recently demonstrated enormous potentials in many fields such as plant genotype discrimination [2,3], toxicological mechanisms, disease processes, and drug discovery [4–10]. According to the destination, the metabonomics research could be divided into four approaches: metabolite target analysis, metabolite profiling, metabolomics and metabolic fingerprinting [11].

It is generally accepted that a single analytical technique could not provide sufficient visualization of the metabolome, and therefore, multiple technologies are needed for a comprehensive view. Accordingly, many analytical technologies have been enlisted to profile the metabonome [12–24]. Among the analytical technologies, NMR is the most popular method used in the metabonomics research [13–16,24], which could provide the high selectivity for all compounds. While being very effective, NMR has two significant disadvantages: poor sensitivity and resolution, which led to the masking of low abundance analytes by high concentration components. Another popular technology used in the metabonomics research is gas chromatography–mass spectrometry (GC–MS) [2], which could provide higher resolution and higher sensitivity compared to NMR. But the GC–MS could not analyze the non-volatile components of the metabolites in the bio-fluids and tissue directly. While most of the metabolites in these samples belong to this class, HPLC and/or LC–MS [22] has been employed to analyze the non-volatile components.

After the data was already acquired by various chromatographic techniques, the next step was to compare them using versatile chemometric methods to mine the useful

* Corresponding author. Tel.: +86 411 83693403; fax: +86 411 83693403.
E-mail address: dicp402@mail.dlptt.ln.cn (G. Xu).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.10.100

information behind them. But the retention time variation from chromatogram to chromatogram has been a significant impediment against the use of chemometric techniques in the analysis of chromatographic data due to the variations in the mobile phase composition, gradient reproducibility, temperature variations and column variability. To make use of the information of all visible peaks contained in the chromatograms, the first thing we should do is the peak matching between the chromatograms. In the past decades, dozens of algorithms have been developed for this purpose [25–32]. The representative work is the "COW" presented by Nielsen et al. [25]. Recently, Jonsson et al. [32] developed a strategy for identifying differences in large series of metabolomic sample analyzed by GC–MS, in which a time window based algorithm was employed to align the peaks in GC–MS. Compared with peak alignment in GC chromatogram, there is more difficult in HPLC chromatograms for its worse reproducibility.

In this study, a strategy was developed to apply HPLC-based techniques to metabonomics researches and its flow chart was given in Fig. 1. At first, the metabolome or its subset was analyzed by selective and sensitive liquid chromatography techniques. Then a peak alignment algorithm was developed to match the chromatograms before the pattern recognition method was applied. From the results of the pattern recognition, the variable loadings were obtained and potential biomarkers were further identified by the following LC–MS–MS experiments. To describe the strategy clearly, the diagnosis of liver diseases was used as an example. Based on the method developed, the false-positive result of live cancer diagnosis from other liver diseases was reduced effectively and potential biomarkers were identified.

2. Experimental

2.1. Sample collection and pretreatment

Spontaneous urine samples were collected from 50 healthy adults, 77 patients with liver infectious diseases (27 hepacirrhosis patients, 30 acute hepatitis patients, 20 chronic hepatitis patients) and 48 liver cancer patients. Their ages were 50.6 ± 16.2 (the age range was 20–85). All patients were from the First and Second Affiliated Hospitals of Dalian Medical University of China. All the diagnoses of these patients were confirmed by histopathology. Urine samples free of preservatives were collected and kept at $-20\,^{\circ}C$ until analysis.

For the analysis of urinary *cis*-diol metabolites, the samples were thawed at room temperature. The spontaneous urine added with internal standard 8-bromoguanosine (Br8G; Sigma, St. Louis, MO, USA) was extracted on a phenylboronic acid column as described elsewhere [33–35].

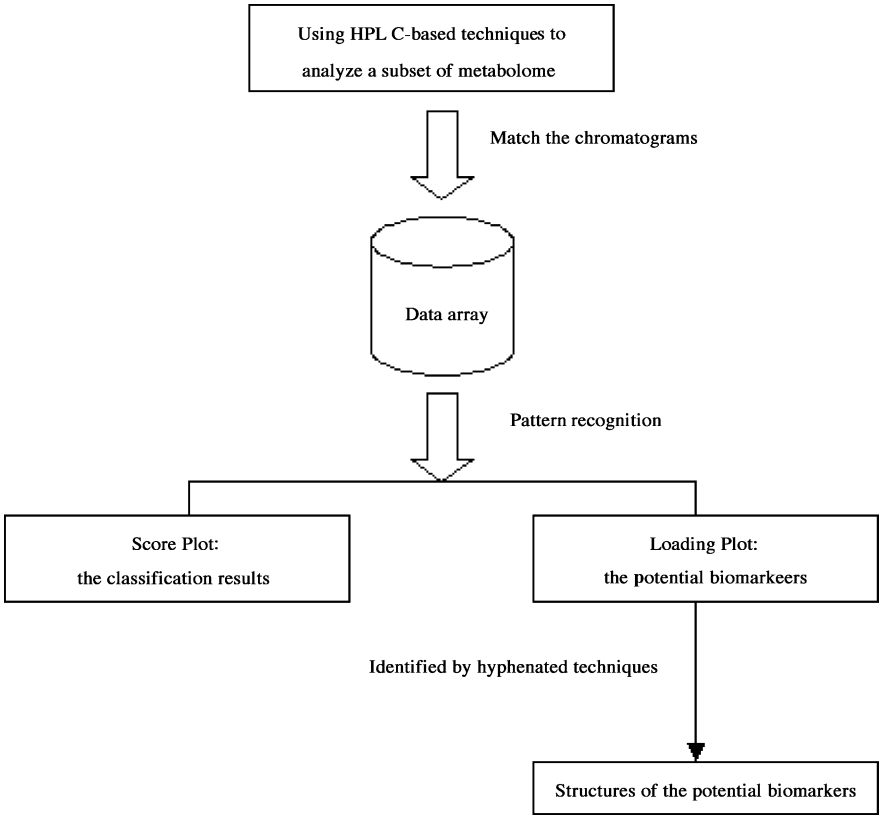

Fig. 1. Flow chart of the HPLC-based metabonomics research strategy.

Table 1
LC–MS–MS method parameters

Time (min)	%MeOH (the rest being 5 mM ammonium acetate, pH 4.5)
HPLC conditions:	
0	0
5	0
20	15
35	60
50	60
55	0
70	0
Turbo Ion Spray source conditions:	
IonSpray voltage (V)	−4200
Declustering potential (V)	−30
Entrance potential (V)	−10
Collision energy (eV)	−10
CUR (curtain gas) (p.s.i.)	25
GS1 (nebulizer gas) (p.s.i.)	30
GS2 (auxiliary gas) (p.s.i.)	40
TEM (temperature) (°C)	500
Ihe (interface heater)	ON
CAD (collision gas)	High

1 p.s.i. = 6894.76 Pa.

2.2. Reversed-phase (RP) HPLC analysis

The HPLC system consisted of three Shimadzu HPLC-10ATVP pumps, an autoinjector model SIL 10ADVP, an SPD-10AVP UV–vis detector, set at 254 nm and a SCL 10AVP interface (Shimadzu, Kyoto, Japan), a Hypersil ODS 5 μm C_{18} HPLC column (250 mm × 4.6 mm) (Elite, Dalian, China). The mobile phases and gradient were given in Table 1. And a typical chromatogram was shown in Fig. 2.

2.3. Analysis of HPLC data

The flow chart of the data processing was shown in Fig. 3. At first, the peak information of these samples was exported in CSV format using the Shimadzu Class-VP version 6.10 software. In order to compare these samples, their peak data were aligned to generate a two-dimensional array firstly, of which row was corresponding to the samples and the column was corresponding to the peak areas appearing in the chromatograms. After adjusted and normalized, the data were sent to perform principal component analysis (PCA) by a laboratory-made script written in MATLAB.

2.4. LC–MS–MS

The LC–MS–MS equipment consisted of: an HP 1100 series HPLC system (Agilent Technologies, Palo Alto, CA, USA) with a Hypersil ODS 5 μm C_{18} HPLC column (250 mm × 4.6 mm) (Elite), a Q-TRAP LC–MS–MS system from Applied Biosystems/MDS Sciex (USA) equipped with a turbo ionspary source. The Q-TRAP system from Applied Biosystems is an LC–MS–MS linear ion trap mass spectrometer. To get an unequivocal identification of the metabolites, information dependent acquisition (IDA) mode was employed to get the tandem mass spectra of the urinary metabolites. Using the IDA mode of the Q-TRAP system, the survey scans [like enhanced MS scan (EMS), constant neutral loss scan (CNL)] and dependent scans—enhanced product ion (EPI) are sequentially performed, and repeated in the entire duration of the HPLC analysis, thus offering maximum information from a single injection [36].

From the result of PCA, the potential biomarkers with the highest loadings were carefully selected for further identification by LC–MS–MS. The sample extracted on a phenylboronic acid column as described [33–35] was then analyzed by the LC–MS–MS system with a 3:1 splitting ratio before turbo ion spray interface. In the LC–MS–MS method, the HPLC parameters were the same as the ones described above. The other parameters are given in Table 1. The criteria of IDA were: the two most intensity peaks, which intensity exceeded 1×10^5 and mass range resided between 200 and 600 u, were

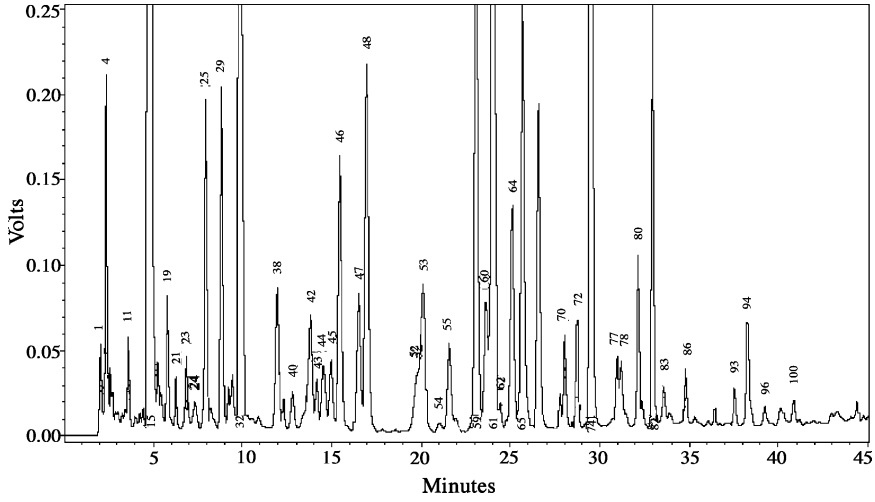

Fig. 2. Typical chromatogram of RP HPLC of *cis*-diol compounds in urine.

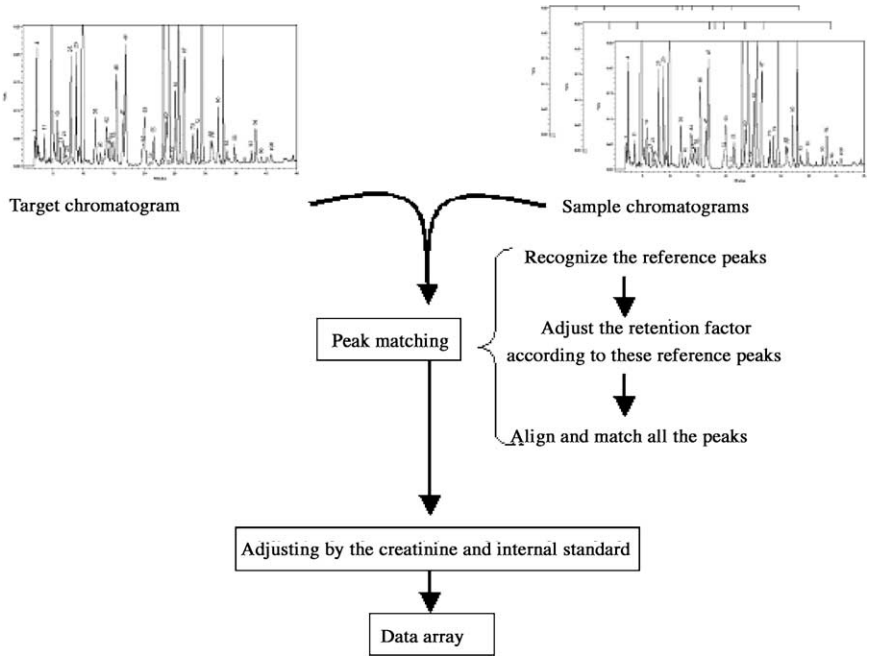

Fig. 3. Flow chart of the HPLC peak alignment and data normalization.

selected as the precursor ions of EPI. And the scan mass range of EPI was from 80 to 600 u.

3. Results and discussion

It can be known from Fig. 1 that the strategy for metabonomics research consists of the data collection, data analysis and biomarker identification. To describe the strategy clearly, the differentiation of liver cancer from other liver diseases using metabonomics method is used as an example.

Although it would be ideal to have information on the status of the whole metabolic complement of a cell, there might be instances when it would suffice to derive information on only a subset of the total metabolome. As shown in our previous researches [35,37–39], the compounds with *cis*-diol structure, mainly the metabolites of nucleic acids (nucleosides), were an important class of compounds related to the diagnoses of cancer. So this class of compounds was chosen to elucidate the strategy developed in this study. And the HPLC analytical method has been established in our previous study [37–39]. In following sections, we shall mainly describe a peak alignment method to match the chromatograms accurately, the pattern recognition method to classify the samples and identification of the potential biomarkers using the hyphenated techniques.

3.1. Peak alignment algorithm

The main intention of the algorithm is to define a series of prominent marker peaks firstly, which could be easily identified by selecting a wider retention window. The retention time was then divided into several zones, which were residing between the prominent peaks. While the prominent peaks' retention values have been assigned to a series of pre-defined values, all peaks' retention values were scaled according to these prominent reference peaks. After shrinking or enlarging the zones, the retention values of the peaks residing in these zones would be very stable in different chromatograms.

The alignment was carried out based on the adjusted retention value (arv) that was calculated according to Eq. (1):

$$\mathrm{arv}_i = \frac{\mathrm{rv}_i - \mathrm{RV}_j}{\mathrm{RV}_{j+1} - \mathrm{RV}_j}(\mathrm{ARV}_{j+1} - \mathrm{ARV}_j) + \mathrm{ARV}_j \qquad T_j < t_i \leq T_{j+1} \tag{1}$$

where arv_i is the adjusted retention value of the peak i (i is the peak No., equals to 1, 2, 3, . . ., 113); rv_i is the retention factor; t_i is the retention time of the peak; j ($j = 0 \ldots C$) is the number of reference peaks. In this study, C is equal to 6. ARV_j, RV_j and T_j are the reference peak's adjusted retention value, retention value and retention time, respectively. ARV_0, RV_0 and T_0 were all equal to 0.

As an example, two chromatograms (Fig. 4a) of 15 mixed standards run on two different days were chosen to test the algorithm. Two retention factors were employed for comparison: one was the peak retention factor (k), the other $\log k$. After the alignment, the results were shown in Fig. 4b using the mirror "spectrum", in which the bars were corresponding to the peak heights (y-coordinate) just like in mass spectrum, the x-coordinate was corresponding to corresponding arv values, the positive and negative parts corresponding to two chromatograms respectively. It can be observed that prior to the alignment, the retention time differences in two chro-

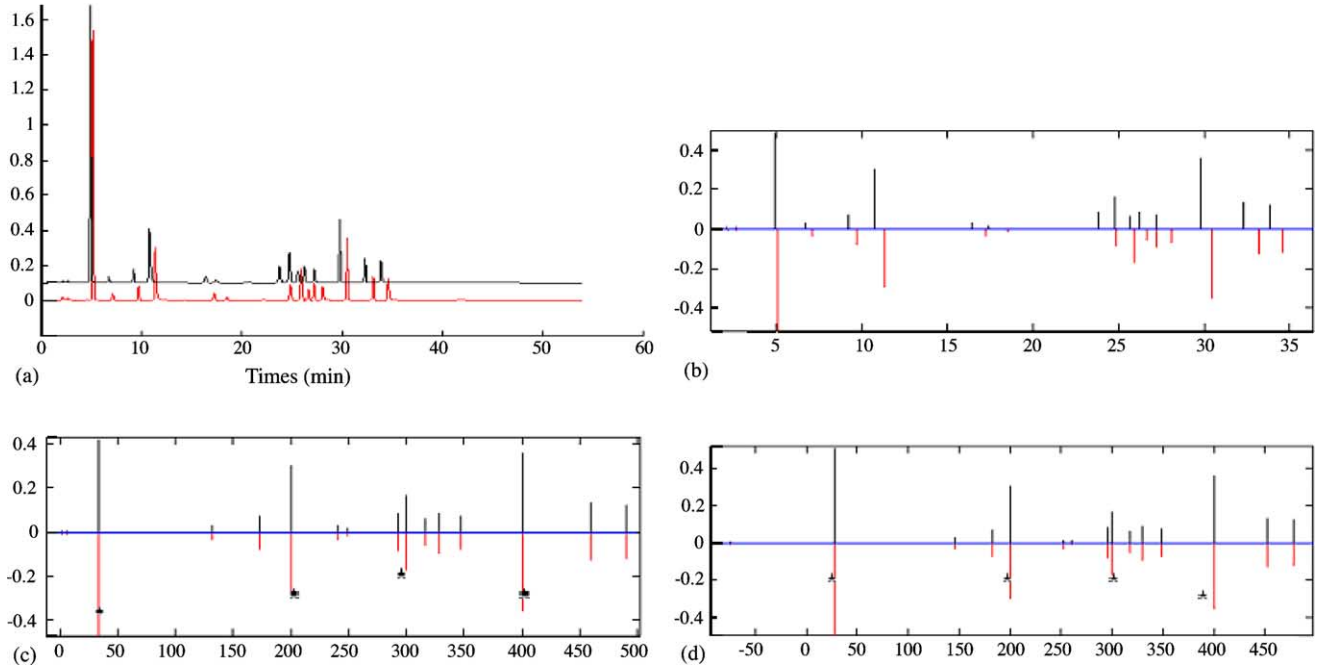

Fig. 4. Superimposed sections and "mirror spectrum" of the chromatographic trace from HPLC of a standards mixture running in different days. Up (Black): target chromatogram. Down (red): sample chromatogram. (a) Overlapped chromatograms prior to alignment; (b) "mirror spectrum" prior to alignment (using retention time as the retention value); (c) "mirror spectrum" after alignment (using k as retention value)**; and (d) "mirror spectrum" after alignment (using $\log k$ as retention value)**. The x-coordinate value of pane b–d was corresponding to the adjusted retention value calculated according to Eq. (1). *Note*: *The predefined reference peaks. **"k" was calculated according to the equation: $k' = (\text{retention\_time} - \text{void\_time})/\text{void\_time}$, where void_time was equaled to the time of the first fluctuation of the baseline. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

matograms were significant, but after the alignment, all of the peaks were satisfactorily matched (Fig. 4c and d). It should be pointed that although in Fig. 4c and d we cannot find the difference in using k and $\log k$ as the retention factor, for different practical samples different matching results will be obtained depending on the chromatographic conditions and samples.

3.2. Pattern recognition

After the peaks were aligned, area of each peak was normalized to the internal standard Br8G and corrected by the concentration of creatinine. Based on the peak detection software from the Shimadzu HPLC instrument, 113 peaks were found in the target chromatogram (Fig. 3). A number of peaks that were detected in the target chromatogram were apparently absent in the sample chromatograms. This primarily occurred because the comparison peaks had a S/N lower than 1000. Therefore, the true peak areas for these peaks were somewhere between zero and the detection limit. The value of 1×10^{-6} was given to such peaks. At last, all chromatograms were represented by the array, which size equals to the target chromatogram.

Then the PCA method could be performed to the data. The principal components are displayed as a set of 'scores' (t), which highlight clustering or outliers, and a set of 'loadings' (p), which highlight the influence of input variables on t. And the score plot and the loading plot were shown in Fig. 5. From the score plot, it could be observed that patients with hepatitis and patients with hepatocirrhosis clustered in one region, while liver cancer patients were located in a different cluster region. It illustrated that only 7.40% of the hepatocirrhosis patients and none of the hepatitis patients are classified as cancer. Contrast to the result of several target metabolites analysis described before, the false positive result was effectively eliminated [37–39].

3.3. Identification of the potential biomarkers

As mentioned above, the loading plot highlighted the most significant variables. The variables with the highest loading values were corresponding to the most significant those. In another word, the potential biomarkers were the furthest ones from the origin in the loading plot. According to the loading plot of the PCA, which was shown in Fig. 5b, the potential biomarkers were then identified by LC–MS–MS.

According to their retention time, the corresponding spectra were used to identify the potential biomarkers mentioned above. Because selectively absorbed on the phenylboronic acid affinity column, the potential biomarkers must be metabolites with *cis*-diol group, the nucleosides were the urinary major compounds in this class. Generally, the nucleosides consist of a sugar moiety, primarily ribose in nucleosides of RNA and denoted S, jointed to a base moiety,

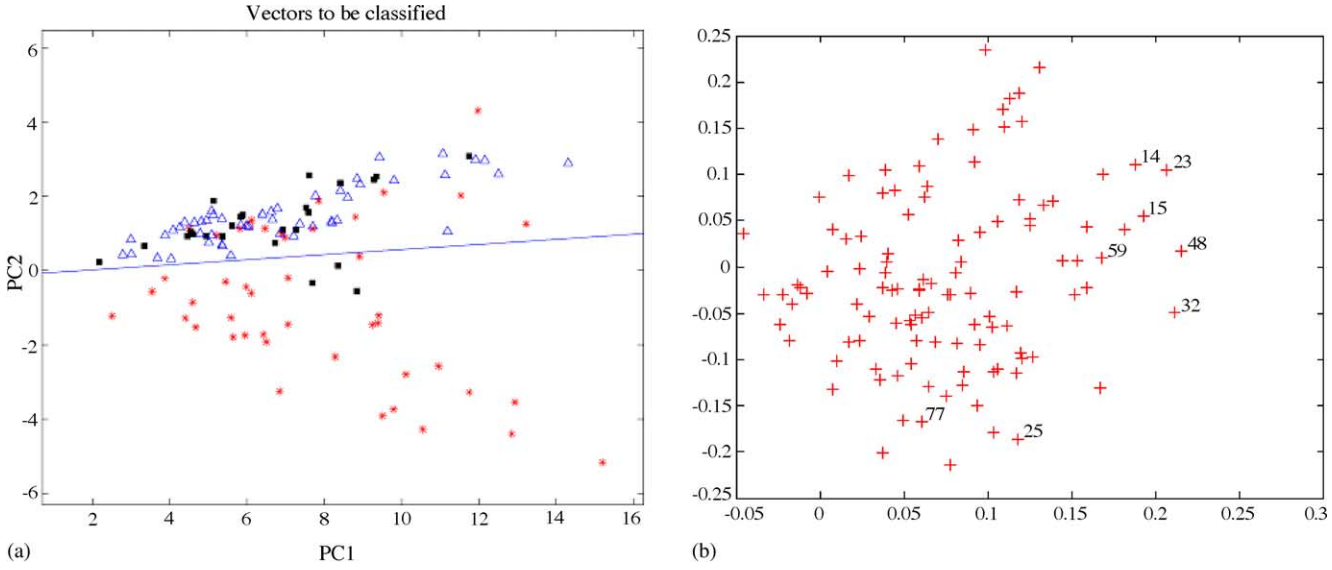

Fig. 5. (a) Score plot of the metabonomics data (*k* was used as the retention value in peak matching method). (*) liver cancer, (△) hepatitis, (■) hepatocirrhosis. (b) Loading plot of the data.

usually by a glycosidic bond, for example, adenine in the case of adenosine, and denoted B. The major peak in mass spectra obtained upon the fragmentation of the nucleosides molecular ion is usually base moiety ($[B-H]^-$) (with the exception of pseudouridine). In an ideal spectrum of the nucleoside, the loss of the water ($[M-H-18]^-$) and loss of CH_2O $[M-H-30]^-$) could be seen. And the de-protonated dimmer of the molecule was often found in the spectrum.

Fig. 6 gives an IDA view of a potential biomarker. From the EMS spectrum of the 15#, *m/z* 243.2 $[M-H]^-$, 110.1 $[B-H]^-$ and 153.1 $[B-H+43]^-$ could be found clearly. It is convinced that 15# was pseudouridine (Pseu). To exclude the confusion coming from the adhere peaks, the tandem spectrum of the $[M-H]^-$ was employed to corroborate the deduced structure. From the tandem spectrum, two more fragments (225.2 $[M-H–H_2O]^-$ and 183.2 $[M-H–2CH_2O]^-$) were found to confirm the structure. Because C–C bond residing between B and S, the most abundance peak was 153.1 $[B-H+43]^-$ other than 110.1 $[B-H]^-$ like other nucleosides. 15# was convinced to be Pseu. Similarly, other poten-

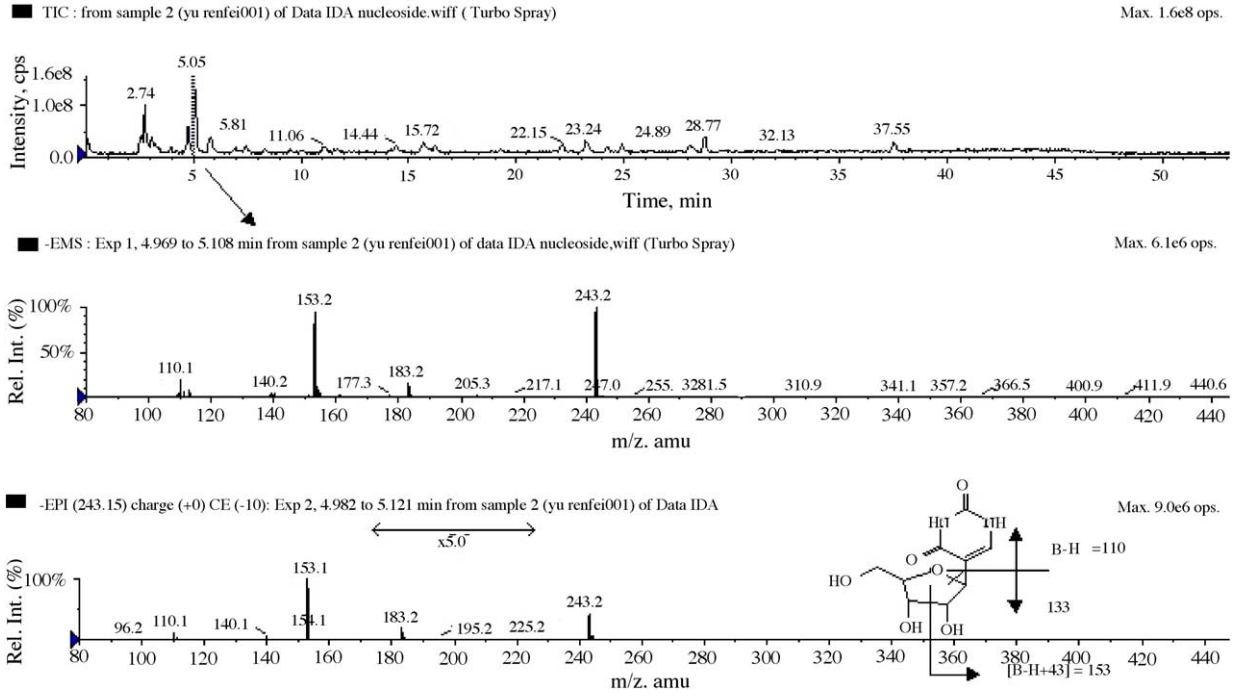

Fig. 6. IDA view corresponding to 15# peak. Top panel: TIC of the sample; middle panel: EMS of 15# peak; bottom panel: EPI of 243 ($[M-H]^-$).

Table 2
The possible identification result of the eight potential biomarkers

Peak No.	Potential structure	Major fragmentation	Confirmed by standards' retention time and spectra
14	Dhu	$[M-H]^-$, $[B-H+44]^-$, $[M-H-S]^-$	
15	Pseu	$[M-H]^-$, $[B-H+44]^-$, $[M-H-S]^-$	√
23	acp3U	$[M-H]^-$, $[M-H-S]^-$	
25	U	$[M-H]^-$, $[M-H-S]^-$	√
32	m1A	$[M-H]^-$, $[M-H-S]^-$	√
48	PCNR, m3U or m5U	$[M-H]^-$, $[M-H-S]^-$	
59	m1I	$[M-H]^-$, $[M-H-S]^-$	√
77	mcm5s2U or mcmo5U	$[M-H]^-$, $[M-H-S]^-$	

tial biomarkers' structures were also elucidated and the result was given in Table 2.

In order to validate the result given in Table 2, the available standards were injected to LC–MS–MS under the same conditions. The result was eventually confirmed by comparing the retention times and the spectra between samples and standards. And its interesting to find peak 48# was an overlapped peak comprised with two or three compounds: PCNR and mU (m3U or m5U) for two distinct spectra appeared in the front and tail of the peak.

4. Conclusions

In this study, we gave a strategy to apply HPLC-based techniques to the metabonomics research. And as an example, the diagnosis of liver diseases was used to test the strategy. At first, all compounds with *cis*-diol structure in the urine of the healthy volunteers and patients were determined using affinity column and RP HPLC. After all chromatograms have been obtained, in order to utilize all the information embedded in these chromatograms, a peak alignment algorithm has been developed to align the chromatograms and transfer the information into an array. Then PCA could be employed to mine the information. Additionally, the potential biomarkers were further identified by the LC–MS–MS. And it can be concluded the strategy could be helpful for us to apply the HPLC-based techniques, a powerful tool in separation, to the metabonomics researches.

On the other hand, the strategy mentioned above has also its limitation. The peak capacity of one-dimensional chromatography is not yet sufficient for metabonomics researches with several thousands of metabolites. In this case, the two-dimensional liquid chromatography-based techniques or other higher resolution techniques were needed. This will be our next work.

Acknowledgements

The studies have been supported by the high-tech R&D plan (2003AA223061), National Basic Research Program (2003CB716003) and the sociality commonweal project of State Ministry of Science and Technology of China, the Knowledge Innovation Program of the Chinese Academy of Sciences (K2003A16) and Liaoning province foundation of science and technology.

References

[1] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Nat. Rev. Drug Discov. 1 (2002) 153.
[2] J. Taylor, R.D. King, T. Altmann, O. Fiehn, Bioinformatics 18 (2002) S241.
[3] O. Fiehn, Plant Mol. Biol. 48 (2002) 155.
[4] N. Aranihar, B.K. Singh, G.W. Stockton, K.H. Ott, Biochem. Biophys. Res. Commun. 286 (2001) 150.
[5] D.B. Kell, R.M. Darby, J. Draper, Plant Physiol. 126 (2001) 943.
[6] K.H. Ott, N. Aranibar, B. Singh, G.W. Stockton, Phytochemistry 62 (2003) 971.
[7] D.B. Kell, R.D. King, TIBTECH 18 (2000) 93.
[8] D.J. Oliver, B. Nikolau, E.S. Wurtele, Metab. Eng. 4 (2002) 98.
[9] L.M. Raamsdonk, B. Teusink, D. Broadhurst, et al., Nat. Biotechnol. 19 (2001) 45.
[10] A.F. David, Trends Genet. 17 (2001) 680.
[11] O. Fiehn, Comp. Funct. Genomics 2 (2001) 155.
[12] S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Trends Biotechnol. 16 (1998) 373.
[13] E. Holmes, J.K. Nicholson, G. Tranter, Chem. Res. Toxicol. 14 (2001) 182.
[14] H.C. Keun, T.M.D. Ebbels, H. Antti, M.E. Bollard, O. Bechonert, G. Schlotterbeck, H. Senn, U. Niederhause, E. Holmes, J.C. Lindon, J.K. Nicholson, Chem. Res. Toxicol. 15 (2002) 1380.
[15] J.T. Brindle, H. Antti, E. Holmes, G. Tranter, J.K. Nicholson, H.W. Bethell, S. Clarker, P.M. Schofield, E. Mckilligin, D.E. Mosedale, D.J. Grainger, Nat. Med. 3 (2002) 1439.
[16] J.K. Nicholson, M.E. Bollard, J.C. Lindon, E. Holmes, Drug Discov. 1 (2002) 153.
[17] H. Tweeddale, L. Notley-McRobb, T. Ferenci, J. Bacteriol. 180 (1998) 5109.
[18] P.D. Fraser, M.E. Pinto, D.E. Holloway, P.M. Bramley, Plant J. 24 (2000) 551.
[19] B.R. Baggett, J.D. Cooper, E.T. Hogan, J. Carper, N.L. Paiva, J.T. Smith, Electrophoresis 23 (2002) 1642.
[20] K. Arlt, S. Brandt, J. Kehr, J. Chromatogr. A 926 (2001) 319.
[21] Y. Soga, Y. Ueno, H. Naraka, Y. Ohashi, M. Tomita, T. Nishioka, Anal. Chem. 74 (2002) 2233.
[22] D.V. Huhman, L.W. Sumner, Phytochemistry 59 (2002) 347.
[23] A. Aharoni, C.H. Ric de Vos, H.A. Verhoeven, C.A. Maliepaard, G. Kruppa, R. Bino, D. Goodenowe, OMICS: J. Interact. Biol. 6 (2002) 217.
[24] N.J. Bailey, P.D. Stanley, S.T. Hadfield, J.C. Lindon, J.K. Nicholson, Rapid Commun. Mass Spectrom. 14 (2000) 679.

[25] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[26] D. Bylund, R. Danielsson, G. Malmquist, K.E. Markides, J. Chromatogr. A 961 (2002) 237.
[27] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, J. Chromatogr. A 996 (2003) 141.
[28] G. Malmquist, R. Danielsson, J. Chromatogr. A 687 (1994) 71.
[29] M.E. Parrish, B.W. Good, F.S. Hsu, F.W. Hatch, D.M. Ennis, D.R. Douglas, J.H. Shelton, D.C. Watson, C.N. Relley, Anal. Chem. 53 (1981) 826.
[30] J.A. Pino, J.E. McMurry, P.C. Jurs, B.K. Lavine, A.M. Harper, Anal. Chem. 57 (1985) 295.
[31] D. Bylund, R. Danielsson, K.E. Markedes, J. Chromatogr. A 915 (2001) 43.
[32] P. Jonsson, J. Gullberg, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom, T. Moritz, Anal. Chem. 76 (2004) 1738.
[33] P.A. Limbach, P.F. Crain, J.A. McCloskey, Nucleic Acids Res. 22 (1994) 2183.
[34] C.W. Gehrke, K.C. Kuo, Chromatography and Modification of Nucleosides, Part C, Elsevier, Amsterdam, 1990.
[35] H.M. Liebich, G.W. Xu, C. Di Stefano, R. Lehmann, H.U. Haring, P.Z. Lu, Y. Zhang, Chromatographia 45 (1997) 396.
[36] J.W. Hager, J.C.Y. Leblanc, J. Chromatogr. A 1020 (2003) 3.
[37] G.W. Xu, H.M. Liebich, Am. Clin. Lab. 20 (2001) 22.
[38] Y.F. Zheng, G.W. Xu, D.Y. Liu, J.H. Xiong, P.D. Zhang, C. Zhang, Q. Yang, S. Lv, Electrophoresis 23 (2002) 4104.
[39] J. Yang, G.W. Xu, H.W. Kong, Y.F. Zheng, T. Pang, Q. Yang, J. Chromatogr. B 780 (2002) 27.