Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Chromatography B, 826 (2005) 122–128
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Protein identification assisted by the prediction of retention time in liquid chromatography/tandem mass spectrometry

Yan Wang, Jie Zhang, Xue Gu, Xiang-Min Zhang *

Department of Chemistry & Research, Center of Proteome, Fudan University, No. 220, Handan Road, Shanghai 200433, China

Received 24 September 2004; accepted 28 August 2005
Available online 12 September 2005

Abstract

Two-dimensional liquid chromatography (2D-LC) coupled on-line with electrospray ionization tandem mass spectrometry (2D-LC-ESI-MS/MS) is a new platform for analysis and identification of proteome. Peptides are separated by 2D-LC and then performed MS/MS analysis by tandem MS/MS. The MS/MS data are searched against database for protein identification. In one 2D-LC-ESI-MS/MS run, we obtained not only the structural information of peptides directly from MS/MS, but also the retention time of peptides eluted from LC. Information on the chromatographic behavior of peptides can assist protein identification in the new platform for proteomics. The retention time of the matching peptides of the identified protein was predicted by the hydrophobic contribute of each amino acid on reversed-phase liquid chromatography (RPLC). By using this strategy proteins were identified by four types of information: peptide mass fingerprinting (PMF), sequence query, and MS/MS ions searched and the predicted retention time. This additional information obtained from LC could assist protein identification with no extra experimental cost.

Keywords: Proteomics; Protein identification; Retention time; Prediction; Tandem mass spectrometry

1. Introduction

Proteomics, the systematic study of the proteins expressed in a cell or tissue, is now the focus of many fields of scientists [1–3]. The first problem in conducting proteomics is the development of accurate, versatile and high-throughput protein identification strategies [4]. Mass spectrometry (MS) is one of the most useful methods for the protein identification of the complex mixture. But characterization of the complex mixture of peptides resulted from a total digest of complex protein mixture only by using MS is challenging. Liquid chromatography (LC) has been coupled with MS to improve the dynamic range and to reduce the complexity of sample introduced to the MS at any given time [5].

In one LC–MS run, we could obtain not only the information of mass-to-charge ratio (m/z) of peptide directly from MS, but also the retention time of peptide eluted from LC. The retention time of peptides in LC depends on the composition of peptides and the defined experiment (mobile phase composition, stationary phase, etc.). If the retention time of peptides couple with MS data, the confidence of peptide identification would be improved, and the number of protein identification would be increased. This additional information is significant and already available at no extra experimental cost. Palmblad et al. showed how to combine the information from chromatographic retention time with accurate mass measurement to improve protein identification by LC/MS [6,7].

Recently, two-dimension liquid chromatography (2D-LC) coupled on-line with electrospray ionization tandem MS (2D-LC-ESI-MS/MS) is a new platform for analysis and identification of proteome [8,9]. Tandem MS allows for the structural elucidation and the identification of analytes based upon molecular weight and the resulting fragment pattern [10], thus it can obtain enormous amount of information [11,12]. In this paper, the retention time for peptide from LC was used to assist the protein identification made by MS/MS. This additional information obtained from LC made the protein identification more reliable.

* Corresponding author. Tel.: +86 21 6564 3983; fax: +86 21 6564 1740.
E-mail address: xmzhang@fudan.edu.cn (X.-M. Zhang).

1570-0232/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2005.08.014

2. Experimental

2.1. Chemicals

Acteonitrile (ACN) was HPLC grade from Fisher Scientific (Fairlawn, NJ). The water used was MilliQ grade (Millipore, Bedford, MA, USA). Formic acid (FA) was HPLC grade obtained form Fluka. Urea, thiol urea, 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate(CHAPS), phynylmethylsulfonyl fluoride (PMSF), sequencing grade trypsin, dithiothreitol (DTT) were obtained from Sigma (St. Louis, MO). All chemicals used in making buffer solutions were analytical grade reagents.

2.2. Sample preparation

The liver cancer tissue of D20, which is a metastatic model of human hepatocellular carcinoma in nude mice with high metastatic potential, was obtained from *Liver Cancer Institute in Zhongshan Hospital, Fudan University*. The liver tissue was diced and washed with cold physiological saline solution (0.9% NaCl solution) to remove blood and other possible contaminants. The tissue was homogenized in lysis buffer consisting of 7 M urea, 2 M thiol urea, 2% (w/v) CHAPS, 50 mM DTT, 2% (w/v) SB3-10 and 1 mM PMSF using glass homogenization vessel in ice bath. The resulting homogenate was swirled for 20 min and centrifuged for 10 min at 15,000 × *g*. The supernatant was collected, fractionated in aliquots and stored at −80 °C. Protein concentration of sample was measured using bovine serum albumin (BSA) as standard by the Bradford assay [13]. The extraction was dilution to 2 M urea with 50 mM ammonium hydrogencarbonate and digested with trypsin (protein/trypsin 35:1, w/w) at 37 °C overnight.

2.3. Comprehensive nano SCX–RPLC–MS/MS system

All LC experiments were performed with an UltiMate™ nanoscale LC system combined with a FAMOS micro-autosampler and a SWITCHOS valves from LC Packings (Amsterdam, The Netherlands).

A FAMOS micro-autosampler with additional built-in six-port valves was used for sample injection. Elution of strong cation-exchange chromatography (SCX), sample preconcentration and desalting on the trapcolumn was performed with an auxiliary LC quaternary pump building in SWITCHOS that was operated isocratically at a flow-rate of 15 μL/min. The first dimension of SCX was separated on a 300 μm I.D. × 150 mm column (LC packings) packed with 5 μm, 90 Å, Bio-SCX stationary phase. The loading buffer of SCX was water with 2% acetonitrile and 0.1% formic acid, which was the same as the mobile phase A of RPLC. A step gradient approach was used, and salt gradient steps were 20 μL injections of 0 mM, 75 mM, 100 mM, 200 mM, 300 mM, 400 mM, 500 mM, 700 mM, and 1000 mM ammonium acetate.

Cartridge type trapcolumns (LC Packings) with a length of 1 mm and an I.D. of 300 μm were used to preconcentration, clean-up and desalt samples. The trapcolumn was filled with a 3 μm, 100 Å C18 PepMap™ stationary phase (LC packings). The loading solvent for sample preconcentration and clean-up and desalt consisted of an aqueous 2% acetonitrile/0.1% formic acid solution.

Samples were separated on a 15 cm × 75 μm I.D. nanoscale LC column (LC packings) packed with 3 μm, 100 Å, C18 PepMap™ stationary phase at a flow rate of 200 nL/min. Mobile phase A consisted of 0.1% formic acid in acetonitrile/water (2:98, v/v) and mobile phase B of 0.1% formic acid in acetonitrile/water (98:2, v/v). Following a 20 min isocratic elution with 5% B the column was developed using gradient conditions as follows: 5% B linearly increased to 50% B in 40 min, then increased up to 95% B in 10 min, further maintained at 95% B for 10 min for washing the column, and then ramp down to 5% B for equilibrium.

The retention time was expressed as the acetonitrile concentration in the solvent at the elution time. This was calculated by subtracting the gradient elapsed time from the peak elution time and then multiplying by the percentage of acetonitrile per minute in the linear gradient.

The nano RP column was linked directly to a nano-LC electrospray device. This device holds a PicoTip™ EMITTER Silica Tip™ needle (FS360-75-15-D-5, New Objective Inc., Woburn, MA, USA), which is a nano electrospray needle with a 15-μm tip.

2.4. Mass spectrometer

2D-LC–MS/MS analysis was performed on a QSTAR^{XL} ESI MS equipped with an orthogonal time-of-flight (TOF) mass analyzer (Applied Biosystems, USA). The instrument was operated in positive ion mode. The instrument was set to perform a MS survey scan of 1 s with *m*/*z* range of 400–2000. Information dependent acquisition (IDA) will be generated using the survey experiments conditions as a template. The collision gas was argon and the collision energy was kept at 51.2 V. Any peak with a threshold of 20 counts/s was automatically detected, and the top three precursors from the each MS survey scan were selected by the quadrupole for fragmentation. The instrument was calibrated prior to analysis using horse myoglobin.

2.5. Data processing and database searching

Peak list files were searched against a non-redundant protein database SWISS-PROT using the MASCOT search engine (http://www.matrixscience.com) [14]. Trypsin was selected as protein cleavage specificity. Both b-ions and y-ions were included in the database search. Carboxymethyl (C) was selected as a variable modification and no static modification was selected. The taxonomy was Homo sapiens. Peptide tolerance is ±0.8 Da, and MS/MS is ±0.5 Da.

3. Results and discussion

3.1. Accuracy of the predicted retention time of peptide

The retention of small peptides primarily depends on amino acid composition, and that retention can be predicted by sum-

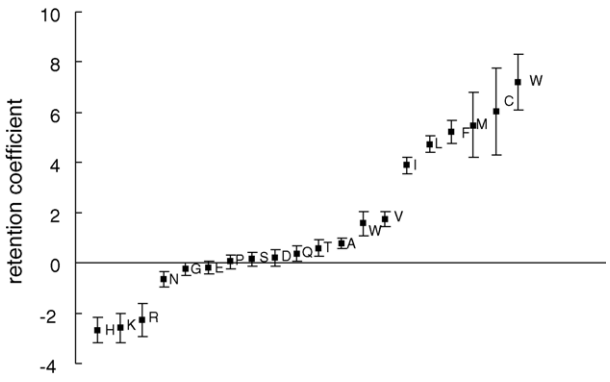

Fig. 1. Retention coefficient for the 20 amino acids.

ming the hydrophobic contribution to retention of each amino acid [15,16], followed as Eq. (1),

$$t_{\text{pre}} = \sum n_i R_{ci} + b_0 \quad (1)$$

where t_{pre} is the predicted retention time of peptide; R_{ci}, the retention coefficients for the ith amino acid; n_i, the number of each amino acid and b_0 is the intercept of the model. A multiple linear regression matrix approach was employed for solving the numerical value of the coefficients from the multivariate structure-retention dependence.

The chromatographic system developed was based upon separation of peptides in two dimensions with nano flow. To achieve maximal orthogonality, the first dimension was based on electrostatic interactions, and the second was based on hydrophobic interactions [17]. ESI–TOF MS/MS, as a detector for the on-line 2D separations, provided an additional dimension of separation (i.e. ions in m/z space).

One hundred and thirty-six peptides were identified by MS/MS. Fig. 1 shows retention coefficient for the 20 amino acids. Amino acids with aromatic or aliphatic side chains have a marked positive effect contribution to retention, while amino acids with neutral or polar side chains have no effect or a small negative contribution to retention.

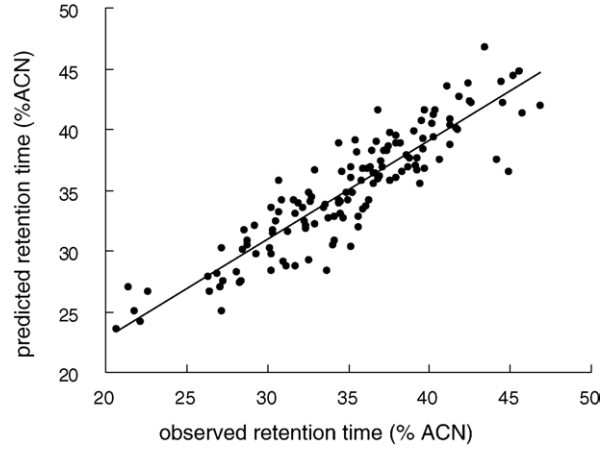

Fig. 2. Comparison of observed retention times of 136 peptides and their predicted retention times using Eq. (1). The correlation coefficients are 0.9011.

Fig. 2 shows the correlation between predicted and actual retention times for the 136 peptides tested by using Eq. (1). The correlation between observed and predicted retention times was 0.9011. The standard error of the prediction was 2.54 (% (v/v) ACN at elution).

3.2. LC assisted protein identification

Searches of 2D-LC–MS/MS data sets can produce an overwhelming volume of data. Many tools have been developed for high-throughput peptide identification for MS/MS such as SEQUEST, MASCOT and others. MASCOT, which incorporates probability-based scoring, integrates all three types of search, peptide mass fingerprint (PMF), sequence query, and MS/MS ions searched [18]. For more reasonable separating the good matches from the junk as efficiently as possible, MASCOT provides additional guidance in the form of a significance level. The score for an MS/MS match is based on the absolute probability (P) that the observed match between the experimental data and the database sequence is a random event. By default, the significance level is set at 5%. For this study, individual ions scores >24 indicate identity or extensive homology ($p < 0.05$) [18].

At the same time, in order to compare the observed and predicted retention time, $P(t_R)$ was used to show the one-tailed probability of the chi-squared distribution. In order to obtain more confidence, the significance level was set at $P(t_R) = 0.75$. If $P(t_R)$ is lower than 0.75, it means that there has the significant difference between the observed and predicted retention time.

3.2.1. Multiple peptides matching with one protein

In MS/MS ions search, confidence of protein identification comes largely from getting multiple matches of peptides from the same protein. If the ion scores of all the matched peptides from one protein are above the significance level, both peptides and protein can be identified confidently. However, if the ion scores of matched peptides are slightly below the significance level, even though the protein gets multiple hits, both peptides and the protein identifications are ambiguous. To solve this problem, we used peptide retention time from LC to assist peptides identification.

In Table 1, how information from LC could help identify the peptides and proteins were illustrated. For example, in Table 1, the first identified protein Q07244, had two matched peptides, IILDLISESPIK and GSYGDLGGPIITTQVTIPK, based on MASCOT (marked with underline). For the expectable mass 1339.98 and 1916.28, several different peptides had been found to match the mass. We selected the three peptides (if have) with top ion scores to further identify the peptide sequence by using predicting retention time in LC. The predicted retention time based on Eq. (1) and the observed time from TIC were listed in Table 1. The predicted retention time of peptide IILDLISESPIK, among the three peptides, was the closest to the observed retention time. $P(t_R)$ of the matched peptide IILDLISESPIK was 0.97, indicating there was only a 3% chance that the observed and predicted retention time of peptide IILDLISESPIK would be truly different. So there was no significant difference between the

Table 1
Protein identified by multiple matching peptides by MS/MS and retention time from LC

Number	Mass (exp)	Peptides	Ion score	Iden Pr	t_R predicted [ACN(%)]	t_R observed [ACN(%)]	P(t_R)
1	1339.98	IILDLISESPIK	11.1	Q07244	53.90	53.66	0.97
		LLNDEDPVVVTK	8.1		45.31		0.21
		EPDLRLENVQK	6.3		39.96		0.03
	1916.28	GSYGDLGGPIITTQVTIPK	17.2	Q07244	50.66	49.83	0.91
		VQGGALEDSQLVAGVAFKK	0.8		49.49		0.96
2	1527.93	DCGATWVVLGHSER	11.5	P00938	49.71	47.89	0.80
		EGAAECEAPGGPQGTR	1.2		39.77		0.20
	1602.09	VVLAYEPVWAIGTGK	19.3	P00938	53.04	52.34	0.92
		QSLTRHAVVHDPDK	3.0		35.37		0.00
3	940.72	NGIPEVASR	13.2		43.00	42.07	0.89
		IGIEIIKR	13.2	P10809	43.78		0.80
		IAQGSYFR	8.3		47.00		0.47
	1359.94	AEQGKFALEVAAK	14.8	P55786	42.88	46.76	0.55
		LLQSSEDWNAAK	12.5		48.09		0.85
		VGGTSDVEVNEKK	10.5	P10809	34.64		0.04
4	1292.93	LAILGIHNEVSK	18.8	P12814	46.28	45.28	0.88
		LSGPLISDFFAK	4.1		53.91		0.24
		SPLEVRLGAVPR	4.0		42.55		0.68
	1520.07	IKGEHPGLSIGDVAK	32.3	P09429	40.55	39.31	0.84
		AGTQIENIEEDFR	15.0	P12814	44.36		0.45
		QEVEENLNEVYR	3.3		39.90		0.92
5	1302.92	VLRCVCFFCSK	7.3	P24928	61.76	44.13	0.03
		KTKPIIFVSDR	3.4		41.91		0.73
		ALPNSGDETLMR	2.7		46.17		0.76
	1539.19	ILLAELEQLKGQGK	29.6	P08670	50.18	49.58	0.93
		LGGLMDPRQGVIER	13.9	P24928	48.15		0.84
		LLNLLADLVERDR	11.8		52.66		0.67

observed and predicted values for the matched peptide IILDLISESPIK. Even though the ion score of peptide IILDLISESPIK was only 11.1, lower than the significant level for mass, by comparing the predicted retention time with the observed retention time in LC, it was suggested that IILDLISESPIK was the peptide corresponding to the expectable mass 1339.98. Similarly, P(t_R) of the second matched peptide GSYGDLGGPIITTQVTIPK to protein Q07244, which was corresponding to the expectable mass 1916.28, was 0.91, indicating that there was also no significant difference between the observed and predicted values. So the identification of the second matched peptide GSYGDLGGPIITTQVTIPK to protein Q07244 by MASCOT was further affirmed by the information from LC. By using this method, the identification of protein Q07244 was more confident.

For the second protein, P00938, the two matched peptides identified by MASCOT had the closest the predicted retention time with the observed retention time from TIC. In addition, P(t_R) of the two matched peptides were both larger than 0.75. So the identification by MASCOT was further affirmed by the information from LC.

The third protein P10809 was identified by MASCOT with two peptides matched. For the first matching peptide, IGIEIIKR, the predicted retention time was much close to the observed retention time, and P(t_R) = 0.80. But for the second matching peptide, VGGTSDVEVNEKK, the predicted retention time of peptide was apparently different with the observed retention time. P(t_R) was only 0.04, lower than the significance level, indicating there was 96% chance that the observed and predicted retention time of peptide VGGTSDVEVNEKK would be truly different. In addition the ion score of VGGTSDVEVNEKK was the lowest among the three peptides fitting with the expectable mass 1359.94. So the peptide VGGTSDVEVNEKK identified by MASCOT was suspect. Therefore, protein P10809 could not be identified confidently with these two peptides.

The similar case happened to the forth and fifth protein listed in Table 1. So protein P12814 and P24928 could not be identified confidently with these two peptides provided by MASCOT.

3.2.2. *Only one peptide matched*

In MS/MS ions search, protein can also be identified by one peptide with high confidence only when this peptide is unique to the protein and its ion score is very high. However, if the ion score of the peptide is not very high, the protein identification might need other data to support. The retention time of peptide was a good proof. Table 2 lists some proteins identified only by one peptide, which was unique to proteins and its ion score was not very high. For example, No. 1 protein P04792 was identified only by peptide VSLDVNHFAPDELTVK with MASCOT (marked as underlined). The top three peptides for ion score (if have) fitting with the expectable mass 1783.18 were also shown in

Table 2
Protein identified by only one matching peptides by MS/MS and retention time from LC

Number	Mass	Peptides	Ion score	Iden Pr	t_R predicted [ACN(%)]	t_R observed [ACN(%)]	P(t_R)
1	1783.18	VSLDVNHFAPDELTVK	24.9	P04792	48.24	49.00	0.91
		EVIPHSRPYMASLQR	8.6		44.68		0.52
		RNAITMQPQNVQGLSK	3.9		45.20		0.57
2	1613.10	LVINGNPITIFQER	22.3	P04406	51.99	52.88	0.90
		NLRLPGSSDSPASASR	3.7		40.61		0.05
		SPRPRGDSAYHSQR	1.5		29.88		0.00
3	1381.95	GPSGGYRGSGGFQR	19.6	Q08211	36.37	48.63	0.04
		LRGSSLFMDTEK	7.5		48.32		0.96
		EHSDSNYTTQTT	0.6		36.15		0.04
4	1066.77	ESTLHLVLR	27.7	P02248	44.60	43.8	0.90
		LQEPPASAVR	5.0		40.21		0.57
		EPAAPVSIQR	4.6		39.51		0.49
5	1339.93	IILDLISESPIK	28.5	O07244	53.90	52.92	0.89
		VKGADINAEEAPK	5.9		36.53		0.01
		LVRQAGGGGGTGSPK	5.7		36.84		0.00
6	1252.88	GLTPSQIGVILR	27.1	Q02546	49.49	48.93	0.94
		AVKEALSAVLPR	8.8		43.75		0.43
		EKQTKPAEAPR	7.5		30.69		0.00
7	*1754.08*	VFDKDGNGYISAAELR	26.0	P02593	45.98	43.72	0.74
		EGTEDSALHGIEELKK	16.1		39.94		0.55
		GQGTICWVDCGDAESR	7.3		55.09		0.12
8	*1360.88*	AQFEGIVTDLIR	25.8	P38646	50.62	52.74	0.76
		KAIFMDCGIHAR	1.7		50.44		0.75
		EQLEEEEEEAK	1.7		36.58		0.00
9	1987.25	AIAELGIYPAVDPLDSTSR	*23.5*	*P06576*	53.05	51.3	0.81
10	1859.15	LDGLVETPTGYIESLPR	22.6	P55209	51.59	50.93	0.93
		ASLHALVGSPIIWGGEPR	5.4		53.59		0.72
		LVPESCPVENPEVPVPR	2.2		47.28		0.59
11	1591.98	KHPDASVNFSEFSK	26.0	P09429	39.22	40.96	0.78
		KHPDSSVNFAEFSK	3.5		39.22		0.78
12	1483.99	KYDAFLASESLIK	50.5	P53025	49.03	48.25	0.91
		KVYHYGDYVTLK	23.1	P17927	39.75		0.18
		AGYQSTLTRTECR	15.6		43.75		0.50
13	1360.90	AQFEGIVTDLIR	22.8	P38646	50.62	53.04	0.73
		KAIFMDCGIHAR	3.1		50.44		0.71
		GQCIKPLFGAVTK	2.0		49.82		0.65
14	1844.20	LAALNPESNTAGLDIFAK	24.1	O00299	54.30	51.7	0.72
		SPALLLSQLLPYMENR	3.5		59.29		0.32
		GDRGDPGPQGPPGLALGER	0.6		39.27		0.05
15	2168.39	GIVDQSQQAYQEAFEISKK	25.1	P29312	46.79	48.51	0.80
		KGIVDQSQQAYQEAFEISK	1.4		46.79		0.80
16	2495.71	IVSRPEELREDDVGTGAGLLEIK	23.2	P49368	50.68	48.27	0.73
		EIIEYYLRQLEEEGITFVPR	9.9		57.06		0.24
		VICAEEPYICKDFPETNNILK	6.9		61.36		0.09
17	1599.05	FHQLDIDDLQSIR	25.1	P16152	50.85	48.79	0.77
		NLTALGLNLVASGGTAK	2.9		51.50		0.70
18	1589.09	ELHINLIPNKQDR	26.4	P07900	42.34	44.72	0.71
		HQNIHSGEKPIVCK	1.0		38.76		0.34
		YSQVLANGLDNKLR	0.4		45.56		0.90
19	1237.74	DPWEPPREGR	17.4	P48681	37.02	48.81	0.05
		EHNRLATELR	13.8	P11678	37.15		0.05
		GLFPGGRHELR	10.1		40.51		0.19
20	952.70	LLLGATLPR	24.3	P00540	49.84	45.75	0.56
		IIIPEIQK	22.0	P55060	45.92		0.98
		LLLPGELAK	18.5		48.89		0.65

Table 2. The ion score of these three peptides was not very high. Base on Eq. (1), the matching peptide VSLDVNHFAPDELTVK with the predicted retention time of 48.24 min was the closest to the observed retention time of 49.00 min, and P(t_R) was 0.91. Although the ion score of the matching peptide was not very high, this match was affirmed further by the information from LC. So as a unique peptide to the protein P04792, the combined information on peptide VSLDVNHFAPDELTVK obtained from MS/MS and LC contributed to the identification of the protein P04792.

For proteins with number 2, 4, 5, 6, 8, 9, 10, 11, 15, 17, the cases were the same as the No 1. The predicted retention time provides another proof to the identification by MS/MS. So assisted by retention data from LC, the identification of these proteins by MS/MS was more reliable.

But for No. 3 protein Q08211 identified by MASCOT, the predicted retention time of the matching peptide, GPSGGYRGSGGFQR, was far away from the observed time from TIC. In addition P(t_R) was only 0.04, lower than the significance level, indicating there was 96% chance that the observed and predicted retention time of peptide GPSGGYRGSGGFQR would be truly different. So there was a significant difference between the observed and predicted values. While the predicted retention time of the second peptide fitting with the expectable mass 1381.95, LRGSSLFMDTEK was very close to the observed time from TIC. So the identification of peptide GPSGGYRGSGGFQR was ambiguous both from MS/MS and LC data, which meant protein Q08211 would not be identified only by this peptide. No. 7, 12, 13, 14, 16, 18,19 and 20 protein identification had the same problem, which the predicted retention times could not support the corresponding peptides identification by MS/MS, so the identified proteins by MS/MS were also suspect.

3.3. Protein identification by a four-step procedure

After a MS/MS ions searching, we have got a match, but is it the right match? Confidence in a PMF result often comes from having supporting evidence from other sources. For example, for 2D-PAGE–MS system, if the analyte originated from a spot at approximately 40 kDa on a 2D gel separation of yeast proteins, then the anticipated result of a peptide mass fingerprint is a 40 kDa yeast protein. If the top scoring protein fits this expectation, the search is deemed "successful". If the top scoring match is a 200 kDa protein from a different species, the initial reaction is likely to be that the search has "failed" [14]. By using online 2D-LC-ESI-MS/MS, when the ion score of MS/MS is much high, MS/MS is an unambiguous proof for the structural analysis. In our experiment, 113 proteins were identified confidently by MASCOT, which individual ions scores were all bigger than the significant level. But when the ion score of MS/MS was much low, the information from LC retention was an additional evidence to support the identification of protein by MS/MS. In this way proteins are identified by a four-step procedure, the first is PMF in MS, and the second is peptide sequence, and the third is ion score of MS/MS, the forth is retention time of LC. Such

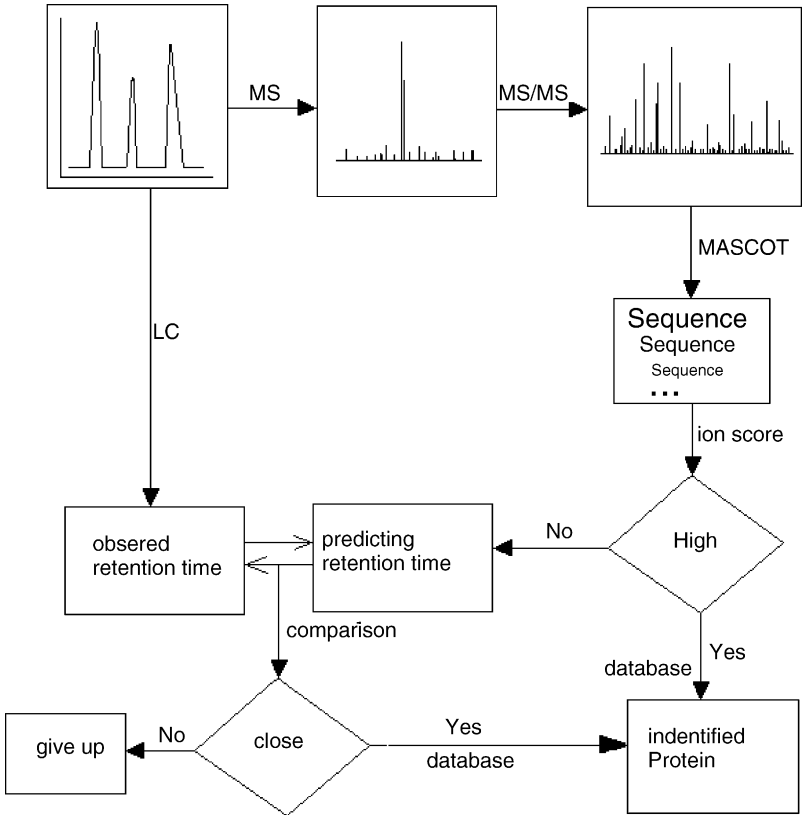

Fig. 3. Functional block diagram illustrating for high throughput protein identification.

an arrangement is illustrated schematically in Fig. 3. By using this strategy in our experiment, 13 kinds of proteins were identified further following the four-step procedure, and the amount of protein identification increased 11.5% relative to only by using MS/MS.

4. Conclusions

Peptide retention in RPLC depends mainly on the amino acid composition of peptides and can therefore be predicted by summing the relative hydrophobic contributions of each constitutive amino acid residue. Information of physicochemical properties of peptides in RPLC column can assist protein identification by tandem MS. The identification of protein is more reliable by using prediction method described here.

Acknowledgement

This work was supported by the National Basic Research Priorities Program (2001CB5102, 2004CB520804).

References

[1] P.A. Haynes, S.P. Gygi, D. Figeys, R. Aebersold, Electrophoresis 11 (1998) 1862.
[2] E. Jung, M. Heller, J.C. Sanchez, D.F. Hochstrasser, Electrophoresis 21 (2000) 3369.
[3] A. Pandey, A. Podtelejnikow, B. Blagoev, X. Bustelo, M. Mann, H. Lodish, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 179.
[4] A.J. Liska, A. Shevchenko, Trends Anal. Chem. 22 (2003) 291.
[5] G.J. Opiteck, J.W. Jorgenson, M.A. Moseley III, R.J. Anderegg, J. Microcolumn Sep. 10 (1998) 365.
[6] M. Palmblad, M. Ramström, K.E. Markides, P. Håkansson, J. Bergqiust, Anal. Chem. 74 (2002) 5826.
[7] M. Palmblad, M. Ramström, C.G. Bailey, S.L. McCutchen-Maloney, J. Bergquist, L.C. Zeller, J. Chromatogr. B 803 (2004) 131.
[8] H.J. Liu, S.J. Berger, A.B. Chakraborty, R.S. Plumb, S.A. Cohen, J. Chromatogr. B 782 (2002) 267.
[9] F.W. McLafferty, Int. J. Mass Spectrom. 212 (2002) 81.
[10] J.T. Wu, L. He, M.X. Li, S. Parus, D.M. Lubman, J. Am. Soc. Mass Spectrom. 8 (1997) 1237.
[11] D.F. Hunt, J.R. Yates III, J. Shabanowitz, S. Winston, C.R. Hauer, Proc. Natl. Acad. Sci. U.S.A. 83 (1996) 6233.
[12] K. Biemann, S.A. Martin, Mass Spectrom. Rev. 6 (1987) 1.
[13] M.M. Bradford, Anal. Biochem. 72 (1976) 248.
[14] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[15] D. Guo, C.T. Mant, A.K. Taneja, R.S. Hodges, J. Chromatogr. 359 (1986) 519.
[16] D. Guo, C.T. Mant, A.K. Taneja, J.M.R. Parker, R.S. Hodges, J. Chromatogr. 359 (1986) 499.
[17] J.C. Giddings, Anal. Chem. 56 (1984) 1258.
[18] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.