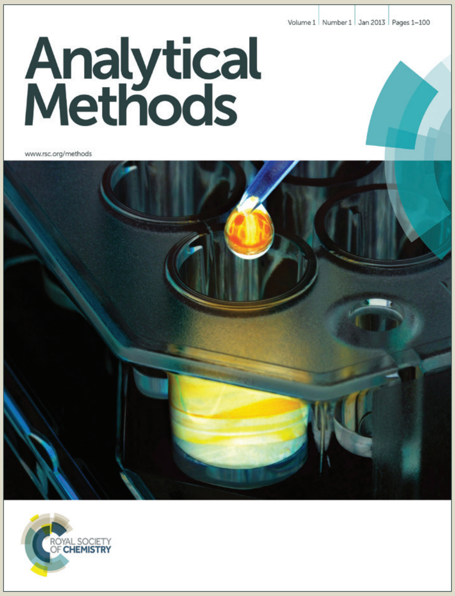

Volume 1 | Number 1 | Jan 2013 | Pages 1–100
Analytical Methods
www.rsc.org/methods
ROYAL SOCIETY OF CHEMISTRY

Multiple technical routes in the application of proteomics expression profile in French liver

Hong Jin[1,2]*, Yang Zhang[2]* , Liqi Xie[2], Huali Shen[2], Caiyun Fang[1], Haojie Lu[1,2], Mingxia Gao[1], Huizhi Fan[1] and Penyuan Yang[1,2]

1: Department of Chemistry, Fudan University, 220 Handan Road, Shanghai 200433, China

2: Institutes of Biomedical Sciences, Fudan University, 138 Yixueyuan Road，Shanghai, 200032, China

*: Co-first authors, equally contributed to the article

Corresponding author: pyyang@fudan.edu.cn, Tel.: 0086 2165642009; fax: 0086 21 65642009.

Abstract

Liver is the biggest digestive gland and the essential important organ in human body. The establishment of the proteome database of the human liver would provide useful information for human liver disease treatment. To maximized protein identification and compare corresponding analysis depth, multiple technical routes were applied in proteome profiling of French liver. Five strategies, including two 2DE methods and three multi-dimensional liquid chromatography (MDLC) methods were evaluated. By using above five routes, 1627 unique proteins were finally identified. All kinds of bioinformatics analyses focused on physicochemical properties and functional classification were used to illustrate the hallmark of the protein expression profile. Furthermore, the comparison of these data with the existing liver expression profile provided by UNIGENE and UNIPROT to investigate the identification efficiency of these dataset. The different technical methods were evaluated and compared to make a model in the organ proteome profile.

Keywords: French human liver, proteomics, 2DE, MALDI-TOF-TOF-MS, ESI

1. Introduction

The expression profile of human liver protein could serve as a reference to motivate biomarker and drug screening for liver disease, and preventive vaccines for future application [1]. Proteomic analysis is a powerful technological tool developed in this decade and widely used in the world. The tool can help to achieve the whole set of expression profile data for any human organ. Several proteomic techniques [2-5] have been utilized to establish human organ/tissue proteome expression profile. To improve our liver proteome dataset analytically, we need definitely to develop standard operating procedures (SOPs) to establish the liver protein database, in order to understand the mystery for liver in human beings, to draw the protein atlas for the biggest human organ. The vast number of proteins presented in liver requires various method for protein mixture separation before it is introduced into mass spectrometry. [6]

A common analytical methodology relying on the separation and analysis of an intact molecule to discern information about protein function or identity was named top down strategy. [7]The most classical top down separation strategy is two dimensional electrophoresis (2DE), which has been introduced more than three decades ago. [8-11] In this conventional separation technique, the proteins are in-gel separated by two different physicochemical principles, for example, by isoelectric point (pI) and molecular weight. The gel spots belong to certain proteins were excised and in-gel digested before identification by either MALDI-TOF-TOF-MS or ESI-MSMS. [12-15]Accordingly, the approximate size and pI of identified proteins could be speculated based on the position of gel spots on 2D gel. Also, some post translational modifications can be visualized on 2D gels where “trains” of spots indicated heterogeneity due to modifications such as phosphorylation or glycosylation. 2DE have been successfully applied in profiling hepatic proteome and comparative analysis of liver disease. [16,17] However, this technique has several disadvantages in detection of very acidic, basic, hydrophobic proteins and low-abundant proteins in the presence of high-abundant proteins.

MDLC based bottom-up strategies, which use peptide detection to infer protein presence, provide a promise alternation to 2DE, for its high-throughput separation ability of highly complex samples. [18] Nowadays, MDLC coupled tandem mass spectrometric analysis has been wildly applied in analyses of complete cell lysates, organisms, tissue extracts, subcellular fractions, and other subproteomes. [19] MDLC strategies are usually conducted in two workflows. One using protein fractionation and separation before protein digestion (Sort-then-break), the other directly separate protein digests without pre-fraction of proteins, commonly referred to as 'shotgun'. [20] In either route, large number of MSMS spectra can be collected in a few hours, which enables detection of more low abundant peptides, although with lower protein sequence coverage compared with 2DE. And the shortage of protein coverage will often lead to much more protein groups to interrupt accurate identification of the protein. In conclusion, MDLC based proteome profiling is complementary to 2DE based method.

To maximum protein identification and compare corresponding analysis depth of different technique routes for liver proteomes, five technique routes were used to analyze the protein expression profile of French human liver sample. Both 2DE and MDLC based strategies coupled with either MALDI-MS or ESI-MS were included. Through various bioinformatics methods, we can obtain global view of the identified proteins on function and pathways involved in liver.

2. Materials and methods

2.1 Preparation of protein samples from French human liver tissue specimens

The routine five technical routes (see Figure 1) were used in the expression profile of French human liver sample. Among them, two were mainly 2DE based technique. The other three were mainly shotgun with liquid chromatography technique. The MALDI-TOF-TOF-MS were used in first four routes. MASCOT was used as search engine for the results in all those five routes.

French human liver was provided by French Health Institute. Ten healthy donors were carefully selected according to the standard condition approved by the Institutional Revie w Board (IRB). The liver tissue was cut into small pieces and washed with glacial NaCl solution (0.9%) to remove blood and some possible contaminants. Then Human liver tissue is crushed with a hammer in a liquid nitrogen-cooled stainless tube for several times.

All experiments were performed in compliance with the relevant laws and institutional guidelines under Human Liver Proteome Project (HLPP). The institutional committee had approved all the experiments. The informed consent was obtained for any experimentation with human subjects.

2.2 The 2DE approach

Total protein and Sequential Extraction

The human liver tissue powder is transferred to a glass homogenizer and treated with lysis buffer which contains 9M Urea, 2% Chaps, 0.14% PMSF and 0.5% DTT for total extraction. Tissue sample was homogenized in ice bath. The resulting homogenate was swirled for at least 20 min and centrifuged for 30 min. at 12,000g, 4°C. The supernatant was collected. Protein concentration of sample was measured using bovine serum albumin (BSA) as standard by the Bradford assay [21]. Transfer the supernatant into several tubes after protein quantification and store at –80°C for the future use.

Proteins are treated with lysis buffer which contains 40 mM Tris for the first step. Then extracts are centrifuged at 16,000*g* for 30 min. at 4°C. Transfer the supernatant into a clean tube. Add the second lysis buffer which contains 8M Urea, 4% Chaps, and 0.5% DTT into the pellet. Centrifuge the solution at 16,000*g* for 30 min. at 4°C. Then transfer the supernatant into a clean tube. Add the third lysis buffer which contains 5M urea, 2M thiourea, 2% Chaps, 2% SB3-10 and 0.5% DTT into the pellet. Centrifuge the solution at 16,000*g* for 30 min. at 4°C. Transfer the supernatant into several tubes after protein quantification and store at –80°C for the future use.

The lysate of total, E1, E2, and E3 was separated with 2DE. The protein concentration of each part was determined by Bradford method. 2DE was performed using non-linear pH 3-10 IPG strips and linear pH 4-7 IPG strips (Amersham Pharmacia, Uppsala, Sweden).1mg and 2mg proteins were applied to 18cm and 24 cm strips correspondingly. Strips were rehydrated overnight at room temperature. IEF was performed using an IPGphor apparatus (Amersham Pharmacia). After rehydration, the strips were focused at voltages ranging from 100 V to 8000 V for a total of 64000V. After the first dimensional separation, IPG strips were reduced, alkylated and detergent exchanged in different equilibration solution for 15 min respectively. 18cm strips were used to run the electrophoresis. Second dimensional electrophoresis was performed. 24cm strips were embedded on top of the precast gel (Amersham Pharmacia, Sweden) with the same method before. Gels were stained with Coomassie brilliant blue staining. The stained gels were scanned using ImageScanner (Amersham Pharmacia) and analyzed with ImageMaster software (Amersham

Pharmacia).

Total Protein Extraction Followed RP-HPLC Separation

Human liver tissue is treated as mentioned. Then 0.2 g total protein is dissolved with lysis buffer which contains 8M Urea, 0.14% PMSF and 0.5% DTT for total protein. Use reversed phase column (300 Ǻ，5μ，4.6x 250mm）with binary gradient (A: 0.1%TFA in water; B: 0.1%TFA in ACN) to separate the whole protein into 10 fractions. After the salt peak, collect the fraction every 12 minutes with flow rate 0.8 ml/min. Collect the fractions with the same condition five times. Quantify the protein concentration and store at –80°C for the future use.

Liquid chromatography separation and fractions collection were performed by using a Shimadzu LC-2010A system with FRC-10A system and SCL-10A *VP* system. Protein sample was loaded onto the C18 reversed-phase column (250×4.6mm, 5µm, 300Å, Hypersil, elite HPLC, China). 0.1%TFA was used as buffer A, 100% acetonitrile + 0.1 % TFA were used as buffer B. Elution was conditioned as linear gradient and listed as following: 10 min of 100% buffer A, 20 min linear gradient from 0% to 30% buffer B, then 80 min linear gradient from 30% to 90% buffer B, 5 min of 90% buffer B and 2 min back to 0% buffer B. All fractions were collected every 3 minutes automatically. The chromatograms were monitored at 215 nm and all fractions were lyophilized and dissolved in 2DE buffer.

Pooled fractions were re-dissolved in a rehydration buffer as usual. Focusing was carried out in 18cm strip with a pH range of 3-10 (NL) at 20℃ as before. Rehydration procedure and the procedure transferred to the second dimension were the same as before.

2.3 In-gel digestion

Coomassie blue stained spots

Excised gel pieces were placed in 96-well plate and destained with 100 μL 50% ACN / 50 mM ammonium hydrogen carbonate for two times (15 min each time). After removal of solution, 100 μL of 100% ACN was added and remained for10 min. Then, ACN was discarded and the gel pieces were incubated at 37℃ for 10 min. in order to remove the excessive ACN. 3 μL 12.5 ng/μL (sequence grade, fresh diluted in 25 mM NH_4HCO_3; Promega, USA) trypsin solution was added to each of the dry gel pieces and incubated overnight at 37℃. Peptides were extracted by adding 60 μL of 0.1% TFA, 50% ACN to each well and remaining for 30min.The supernatants were removed to a fresh 96-well plate and the peptide extraction procedure was repeated. Then, combined the two extracts and dried the solution by using high purity nitrogen. 5mg/ml matrix (50% ACN/0.1%TFA) was added to the bottom of the dried 96-well plate and gently pipetted up and down several times to dissolve the extracted peptides. Matrix and sample were spotted onto a stainless steel MALDI target plate. The dried deposited fractions were analyzed by using a 4700 Proteomics Analyzer MALDI-TOF-TOF-MS (Applied Biosystems).

Silver stained spots

Excised gel pieces were placed in a 96-well plate and were de-stained with 15μL of fresh solution containing 15mM $K_3Fe(CN)_6$ and 50mM $Na_2S_2O_3$ (v/v 1:1) for 20min. Wash gel pieces twice with water by waiting for another 20min. After removal of solution, 80 μL of 100% ACN was added and remained for10 min. Then, ACN was discarded and the gel pieces were incubated at 37℃ for 10min. in order to remove the excessive ACN. All the consequent procedures were same as coomassie blue stained spots.

2.4 Shotgun route

The tissue specimens was suspended in the protein extraction buffer (8 M urea, 20 mM Tris-HCl, pH 8.5, 1 mM DTT, phosphatase inhibitors: 1 mM PMSF, 0.2 mM Na2VO3 and 1 mM NaF; and Protease Inhibitor Cocktail dissolved in the lysis buffer according to usage specification). Lysis was performed by homogenization on ice and aided by vortex of 30 min. The lysate was centrifuged at 26, 000g for 90 min at 40°C. The supernatant was collected as the global protein sample. Protein concentration was measured by a Bio-Rad assay using BSA as standard. The proteins were reduced with 10 mM DTT at 37°C for 1 h and then alkylated with 25 mM iodoacetamide for an additional 30 min at room temperature in the dark. After diluting the urea to 2 M with 25 mM ammonium bicarbonate (pH 8.5), a tryptic digestion was performed on one portion of the global protein sample overnight at 37°C.

Protein and peptide pre-fractionation by SEC

Size exclusion chromatography (SEC) was performed at the intact protein level as follows: 1.0 mg protein was injected onto a Shodex PROTEIN KW-803, 8 mm x 300 mm column (Tokyo, Japan) and the SEC separation was performed with an isocratic gradient produced by a Shimadzu LC-2010A (Kyoto, Japan) system at a rate of 0.3 mL/min consisting of 2 M urea, 25 mM NH_4HCO_3, pH 8.5, for the mobile phase. Parallel SEC separations of the same protein sample were performed. All eluted fractions were lyophilized. Then sequencing grade trypsin was added at a mass ratio of 1:50 and incubated at 37°C for 16 h. The peptide pre-fractionation was performed in a similar way for peptide separation, the mobile phase was changed into 25 mM NH_4HCO_3 (pH 8.5) and without any removal of urea or post-column digestion.

Strong cation-exchange chromatography-RPLC separation

LC experiments were performed with an Ulti-Mate™ nano scale LC system combined with a FAMOS microautosamplerand a Switchos valves from LC Packings (Amsterdam, The Netherlands). Strong cation-exchange chromatography SCX-RPLC systems were used to separate the samples for obtaining more proteins.

2.5 Mass spectrometry

2DE-MALDI-TOF-TOF-MS

Mass spectra were obtained in a mass range of 500-3200 Dalton by using laser (337 nm, 200 Hz) as ionization source. The instrument was used in reflector-positive mode with an acceleration voltage of 15 kV. Trypsin digested peptides of horse myoglobin was used as external mass standard to calibrate the instrument, and then default calibration was applied on the sample peptides. The TOF-TOF mass spectra were acquired by the Data Dependent Acquisition method with 6 strongest precursor ions selected automatically from one MS scan for MSMS analysis. Protein identification using combined raw data (PMF+ MS-MS) was performed with GPS software (Applied Biosystems; containing MASCOT search engine) against the non-redundant IPI human (version 3.25) protein sequence database. The mass tolerance was set as 100 ppm, and

MSMS tolerance was 0.6 Da for automatic data analysis. For partial data, we use different criterion to run MS.

Online 2DLC-MSMS

The procedure of separation was similar to the established comprehensive nano SCX-RPLC-MSMS system [11]. The sample loading and column equilibrium solvent of the SCX dimension: 95% water15% ACN 10.1% FA with flow rate 5 mL/min. 20mM ammonium acetate salt solutions of 0, 5, 10, 20, 30, 40, 60, 80, 150, 400, 600, 800, 1000, 1500 and 2000 mM was injected into the SCX column for a step gradient elution of peptides. A cartridge type trap column was used to pre-concentrate and desalt each peptides fraction eluted from SCX column prior to nano-RPLC separation. SCX fractions were separated on a nano-flow reversed phase column. A continuous gradient elution was performed with solvent A (0.1% FA , 5% ACN,95% water, v/v) and B (0.1% FA , 95% ACN, 5% water, v/v) at a flow rate of 250 nL/min. Quadrupole orthogonal TOF mass spectrometers were used to acquire ESI/MSMS spectra (QSTAR XL). Protein identification using combined raw data (PMF+ MS-MS) was performed with GPS software (Applied Biosystems; containing MASCOT search engine) against the non-redundant IPI human (version 3.25) protein sequence database. The mass tolerance was same as before. For partial data, we use different criterion to run MS. PMF was run at first. When the PMF score was greater than 100，this score would be taken as confidential result and then MSMS procedure was skipped. When the PMF score was smaller than 100, then six highest peaks among the PMF were picked and MSMS was run to determine if the score was above the confidential level.

2.6 Bioinformatics analysis

Analysis of chromosomal localization

The localization information of protein in chromosomes was provided by the IPI database. The information of the chromosome length was provided by Bioconductor (http://www.bioconductor.org). Chromosome plot was completed by the software which was provided by Integrative Genomics Viewer (IGV). [22]

Gene Ontology analysis

Gene Ontology item for each IPI protein was retrieved from GOA Database (Gene Ontology Annotation, http://www.ebi.ac.uk/GOA), and all the GO items were then slimed to parent GO items in order to summarize the protein number in the specific functional class. All the work was fulfilled by a home-made MATLAB program.

The reference database for liver

The cluster data collected from non-redundant clusters of genetic sources were provided by NCBI the United States UNIGENE (http://www.ncbi.nlm.nih.gov/unigene) database. UniGene Cluster includes a single gene sequences and related information represents for each individual, such as gene expression of tissue types, and map location information. The expression information of tissue types provided from UNIGENE can be used as the largest information source of liver transcriptome. The Knowledgebase provided from UNIPROT (http://www.uniprot.org) was by far the most important protein sequence, function, classification, cross-references and other information access center. The tissue expression of protein-specific information was also provided by UNIPROT. This can be used as a liver-background proteomics dataset.

3. Results and Discussion

The comprehensive proteomics is necessary and solid basis for panoramic display of cellular proteome and exploratory research of internal linkage of proteins. This can give us a global overview and can establish the reference dataset for disease treatment.

The routine five technical routes were used in the expression profile of French human liver project. Among them, two were 2DE based methods, the other three were mainly shotgun with liquid chromatography technique. The MALDI-TOF-TOF-MS were used in first four routes. MASCOT was used as search engine for the results in all those five routes.

The pooled fractions eluted from the column were subjected to 2DE by using IEF strip pH 3 -10 (fraction 3 and fraction 10) and pH 4 -7 (the rest of 6 fractions). Due to the enrichment effect, the protein patterns of 2DE were nearly as complex as the protein pattern of the un-fractionated sample. This was confirmed by the observation of the spot numbers of 2DE fraction 3-10.

In this route, ABI 4700 MAIDI-TOF-TOF MS was selected as the instrument for protein identification. Total 14629 spots picked from 2D-Gel were put on the 81 MALDI plates for sequential protein identification. For the research of former scientists employing the strip-2DE-technique, there was more than one protein in most spots of 2D-Gel for tissue expression profile or standard sample. In common situation, one spot on the gel corresponding to all identified proteins above the threshold value given by MASCOT, is defined as Rank_All, and the top score one is defined as Rank_1. However Rank_1 was chosen as protein candidate in most

2DE references. In order to review the confidential level of Rank_1 and Rank_All, a reversed database was used to evaluate the above two strategies.

Proteome Profiling of French Liver

The comprehensive proteomics is necessary and solid basis for panoramic display of cellular proteome and exploratory research of internal linkage of proteins. This can give us a global overview and can establish the reference dataset for disease treatment. Five technique routes were applied in proteome profile of French liver. As show in figure 1, route 1 and 2 were based on 2DE separation, excepted that route 2 incorporates 1D RPLC separation before to 2DE. The other three shotgun strategies take advantage of orthorgic SCX-RPLC separation. To maximum the identification, additional SEC separation was set before or after trypsin digestion in route 4 and 5. The MALDI-TOF-TOF-MS were used in first four routes, while ESI-Q-TOF was applied in route 5. MASCOT was used as search engine for all five routes.

Figure1. Global framework of proteome profiling for human liver.

Analysis of 2DE based strategy

In the identification of 2DE proteins, the precursor ion score distributions of all the target and decoy results from one gel spot (Rank_All) were almost the same. While for Rank_1 results (highest rank of protein score), there were significant differences in protein score, ion score, MSMS number between target and decoy results (Suppl. Fig 1). Therefore, it is reasonable to set the threshold of confident identification accordingly to Rank_1 results. A total of 132 MALDI plates together with 14629 spots were analyzed by MS in route 1. Since not all the spots on the MALDI plate have MSMS information, we divided the MALDI peaklists into two groups, with or without MSMS spectra for individual quality control, and set both 95% confidence. 4728 out of 5185 spots (91%) with MSMS spectra passed corresponding threshold, which belonged to 609 non-redundant proteins. While only 28% of PMF spectra were successfully identified, resulting in 393 non-redundant proteins (Suppl. Fig 2). So we can conclude that MSMS data play important role for improving the protein score and protein identification. Combining the two groups, 724 proteins were successfully identified from 14629 gel spots.

Although the 2DE routes provide high sequence coverage in protein identification, the proteome profiling efficiency was restricted by the high redundant protein identifications, since a single protein can migrate to multiple spots on 2D gels (20.21 spots per protein in route 1). To expand the dynamic range and reduce the redundancy, RPLC was used for protein fraction before 2DE in the second route. In the RPLC-2DE routes, 3955 spots spotting on 39 MALDI plates were analyzed, and 315 non-redundant proteins passed correspondent threshold defined in route 1 (Suppl. Fig 3, 4). The redundancy reduced nearly 50% after RPLC pre-fraction, from 20. 21 to 12.56 spots per protein. Although the sequence coverage of protein as well as successful identification ratio of MALDI spots were totally lower than the route 1, route 2 still can greatly increase the identification efficiency by means of less redundancy. The detail comparison information of route 1 and route 2 show in table 1.

Table 1. Protein identification efficiency between route 1 and route 2.

Analysis of Multidimensional LC based strategy

In the LC-MALDI strategy of route 3 and 4, the number of MSMS ions shows major effect to improve the protein score. The peak-lists of each fraction were merged together for database searching. The emPAI value dependent on the number of MSMS ions was used for quantification. Altogether, 299 non-redundant proteins were identified in the 21 fraction of route 3 (SCX-RPLC-4700). Unique peptide/spectrum was 0.85, and there are 4.2 peptides for each protein. The route 4 and 5 were both using 3DLC for sample fraction, 996 and 247 non-redundant proteins were identified by these two routes respectively, and 81.9% of route 3 identified proteins were detected in route 4 and 5. An extra of 767 and 138 proteins were identified by route 4 and 5 respectively. From the above results, we can conclude that 3DLC route can mostly cover 2DLC route and more proteins can be identified by 3DLC. Comparing with route 3 and 4, the 94 fraction of route 4 was analyzed by ESI coupled MS analyzer QSTAR XL. The number of MSMS ions has no significant correlation with protein score. And the value of unique peptide per spectrum is lower than that of in the LC MALDI route. Which means the identified MS graph contribution of unique peptides is relative lower. However, the number of identified protein was more than LC MALDI route. The detail comparison information of route 3, route 4 and route 5 show in table 1.

Table 2. Protein identification efficiency between route 3, route 4 and route 5.

Overlap of the 5 technique routes

There are total 1627 non-redundant proteins correspondent to 1030 genes in total five technical routes. 1075 proteins（66.1%）were identified by only one technical route. 552 (33.9%)

proteins were identified by both two technical routes or above (Figure 2A). There were 724 non-redundant proteins identified by route 1, while 315 non-redundant proteins identified by route 2, and 211 proteins were identified by both technical routes. Although compared with the first route, there are 211/724 = 29% similarity between first and second routes. In the second route, 104/（104+211）=33.0％ new proteins were provided compared to the first route 27%（3955/14629）. This means RPLC pre-fraction before 2DE is a good complementary method to the pure 2DE technical route (Figure 2C, a). There are only 93 proteins overlapped in the three shotgun routes which means the orthogonality of the experiments (Figure 2C, b). There are 828 proteins identified by top down route and 1154 proteins identified by shotgun routes. There are 355 proteins were identified by both routes, less than 50% overlap. This means these two experimental routes have strong complementary property with each other. (Figure 2C, c). The identified peptide number corresponding to each technique route are showed in Figure 2A, 2D. For quantification, the similarity of the curve of protein identification for both two top down routes was very high, and so did three shotgun routes. But the similarity between top down and shotgun was not so high. (Figure 2B).

Figure 2. Protein number comparison of the 5 technique routes.

Comparison of Physical and chemical property

The physical and chemical properties of identified peptides (redundant peptide) of each technical route were analyzed. The identified proteins are biased to the high molecular zone, low iso-electric points and high hydrophobicity (in the right side of next figure). Little difference of physical and chemical properties for identified proteins was observed between 2DE based routes and shotgun routes. While shotgun routes detected more basic proteins compared with 2DE based routes. Accordingly, the distributions of all identified proteins in isoelectric points shows significant different between the two routes. There were more hydrophilic high abundant proteins identified in the cytoplasm in the 2DE based routes. In the physics, chemical property analysis of each technique route, the results showed that more long peptide with high molecular weight in shotgun routes was identified. And there were more proteins identified with higher and wider iso-electric points in Route 5 due to the large numbers of identified peptides. There were more quantity of α-helix, β-sheet and β-turn in proteins identified in shotgun routes than in 2DE based routes. The contents of amino acid G, Q, R, S were relatively higher in the identified peptides in shotgun routes. Amino acid N was relatively more in the route 5. Amino acid R was relatively less in route 1. In conclusion, top down and shotgun route complement with each other for protein identification (Suppl. Fig 5).

Figure 3. The physical and chemical distribution of the identified proteins.

Over 90% identified proteins were liver proteins

To compare the coverage of different technical routes for the expressed protein in liver, a cross linked database for liver expressed proteins was constructed. For proteomics, UNIPROT database was used. This database only collects proteins of abundant annotation information validated by experiments. 1986 proteins expressed in liver were screened out according to the keyword Tissue_Specificity. In transcriptome, NCBI UNIGENE database was used. UNIGENE is a transcript expression library which is assembled EST into mRNA by sequence technique. Liver expressed mRNA can be screened out within this database according to keyword “Restriction Expression”. There were 19983 UNIGENE correspondent to 40170 proteins for liver expressed in UNIGENE (version 208). Combining UNIGENE and UNIPROT, liver protein database with 40913 proteins was established. The number of identified proteins in the fourth technical route is most. 84% of proteins are proved to be expressed in the liver. All other technical routes are proved above 90%. The proteins expressed in liver in UNIPROT by each technical route were only in the range of 12% to 17%. (Figure 4)

Figure 4. Comparison of identified proteins with known liver expressed proteins.

Functional annotation of liver proteins

There is not too much difference for the subcellular location in all five technical routes. Among them, the most number of identified proteins is cytoplasm, then membrane or mitochondrion, with lysosome at least. The subcellular location is similar between two 2DE based routes, so did the first two routes within shotgun. The distribution of Biological Process in all five technical routes is similar. The number of identified proteins is higher in Transport and metabolism. On the other hand, cell cycle, apoptosis, cell adhesion and cell motility which are related to the cell state are least. The tissue sample is come from adult liver, so the expression rate of this kind of proteins would be low (Figure 5).

Figure 5. Functional comparison of identified proteins in 5 technique routes.

24 human chromosomes were connected from begin to end according to their base position to composite the whole human genome. The width of chromosome position and length are directly proportional. When the identified proteins of all five technical routes were mapped to the genome, the bar graph shape of identified proteins was similar between two 2DE based routes, so did three routes within shotgun. Especially, the pattern of most abundant proteins on chromosomes in route 3 and route 4 are almost the same. Chromosome 1 has the most identified proteins in all technical routes, and then chromosome 12, 6 and 2 (Figure 6).

Figure 6. Chromosome analysis of identified proteins in 5 technique routes.

Generally, it was considered that the proteomics techniques were difficult for the identification of low abundant proteins due to some technical deficiencies. Therefore the coverage derived from proteome is always lower than transcriptome. In addition to technical reasons, it could also be caused by some biological reasons. In organisms, the expression occurred highly at the transcription level for many genes. However, the corresponding expressed protein abundance was relatively low on account of the intermediate degrading process. So this had some disadvantages for the identification at the protein level. On the contrast, this might also be owing to the slow degrading process and the high efficient translation rate, which could be easily identified at the protein level.

Therefore, to identify the disease-associated proteins is the key to study mechanisms of liver cancer. Since most important proteins which attend the signal transduction and biological process are low abundance, we need to have a very powerful technical method to identify those important low abundant proteins besides using classical method.

It is well known that about a half million human proteins are encoded by more than 20, 000 genes. It is proteins that execute the biological processes directly. Thus the state of an organism is essentially reflected in its proteome rather than genome. Liver cancer is one of the most fatal diseases worldwide and the mortality rate has been raised up to the second among malignancies in China. Different technical methods were used and compared to choose the best alterative one for different purpose. Our strategy can provide the objective technical method for the organ proteome.

4. Concluding Remarks

By using five different routes with different experimental strategies on whole protein expression profile in French human liver tissue, 1627 unique proteins correspondent to 1030 genes were finally identified, and all the identified proteins were analyzed with various bioinformatics method to observe the general function of French liver. The standard operating procedures (SOPs) had been tried to be set up to yield good model references for the others. The establishment of the proteome database of the French human liver will be a piece of very useful information for human liver disease understanding and treatment.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No: 20475011 and 20299030), the National Basic Research Priorities Program (2001CB5102, 2004CB520804), the National High Technology Development Program: 863-project (2002AA2Z2042 and 2002AA2Z2041), the National Key research Project (2004BA711A18) and Shanghai Key Project of Basic Science Research (04DZ14005 and 03DZ14024), Shanghai STDP fund (07JC14003)

References

[1] F. He, Chinese Human Liver Proteome Project: A Pathfinder of HUPO Human Liver Proteome Project, J. Proteome Res., 2010, 9 (1):1–2.

[2] A. Abbott, And now for the proteome, Nature, 2001, 409:747

[3] S. Fields, Proteomics in genomeland, Science, 2001, 291:1221.

[4] P. Akhilesh, M. Matthias, Proteomics to study genes and genomes. Nature, 2001, 405:837-846.

[5] W. Potter, S. Paul, Proteomics to study genes and genomes. Nature, 2001, 413:869-875.

[6] N.C VerBerkmoes, J.L Bundy, L.Hauser, et al, Integrating “Top-Down” and “Bottom-Up” Mass Spectrometric Approachesfor Proteomic Analysis of Shewanella oneidensis. Journal of Proteome Research, 2002; 1(3), 239-252.

[7] K.M. Millea, I.S. Krull, S. A. Cohen, et al, Integration of Multidimensional Chromatographic Protein Separations with a Combined “Top-Down” and “Bottom-Up” Proteomic Strategy. Journal of Proteome Research, 2006, 5 (1), 135-146.

[8] M. Adamczyk, J.C, Gebler, J. Wu, Selective analysis of phosphopeptide within a protein mixture by chemical modification, reversible biotinylation and mass spectrometry, Rapid Commun. Mass Spectrom. 15 (2001) 1481.

[9] C. Greenough, R.E. Jenkins, N.R. Kitteringham et al, A method for the rapid depletion of albumin and immunoglobulin from human plasma, Proteomics, 4 (2004) 3107.

[10] R. Pieper, C.L. Gatlin, A.J. Makusky et al, Multi-component immunoaffinity subtraction chromatography: An innovative step towards a comprehensive survey of the human plasma proteome, Proteomics, 2003, 3, 422.

[11] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg et al, IMAC enrichment protocol**,** Nat. Biotechnol. 20 (2002) 301.

[12] S.P. Gygi, B. Rist, R. Aebersold et al, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, Nat. Biotechnol. 17 (1999) 994.

[13] V. Badock, U. Steinhusen, K. Bommert et al, Pre-fractionation of protein samples for proteome analysis using reversed-phase high-performance liquid chromatography, Electrophoresis 22 (2001) 2856.

[14] G.V. Bergh, S. Clerens, F. Vandesande et al, Fluorescent two-dimensional difference gel electrophoresis and mass spectrometry identify age-related protein expression differences for the primary visual cortex of kitten and adult cat, Electrophoresis 24 (2003)1471.

[15] A. Butt, M.D. Davison, G.J. Smith et al, Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis.Proteomics 1 (2001) 42.

[16] W.Ying, Y. Jiang, F.He et al, A Dataset of Human Fetal Liver Proteome Identified by Subcellular Fractionation and Multiple Protein Separation and Identification Technology*S

Molecular & Cellular Proteomics, 5, (2006) 1703-1707.

[17] H.Shen, G.Cheng, P. Yang et al, Expressed proteome analysis of human hepatocellular carcinoma in nude mice (LCI-D20) with high metastasis potential, PROTEOMICS, Volume 6, Issue 2, pages 528–537, 2006

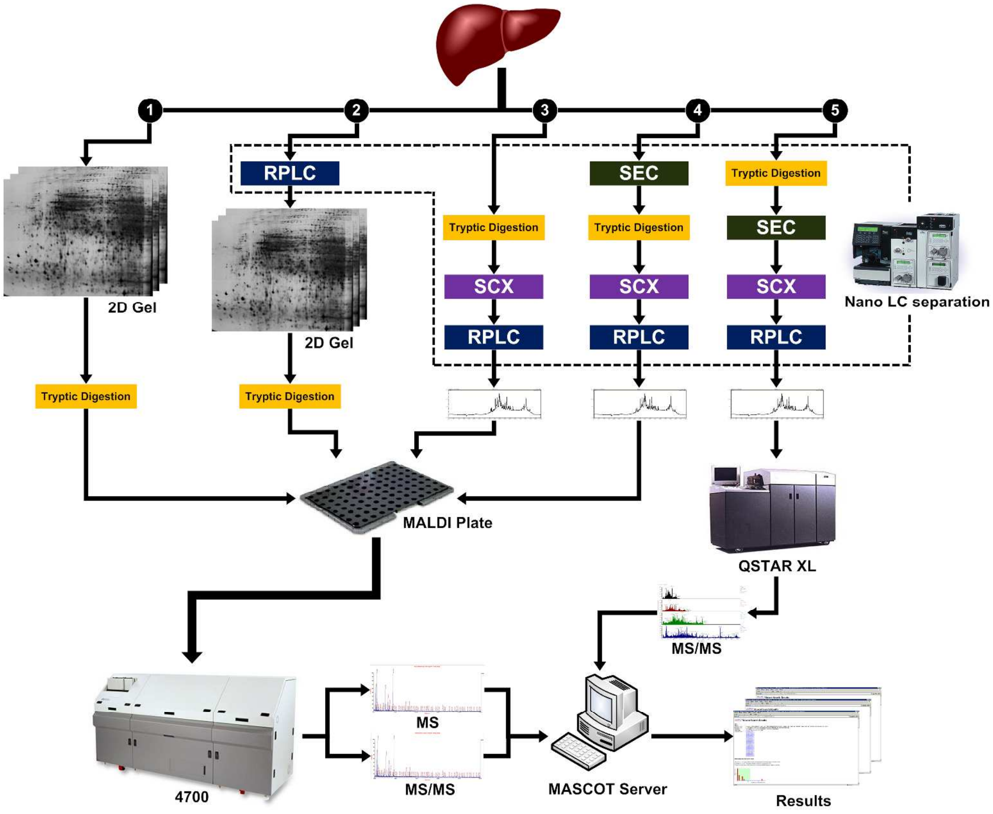
[18] B. Domon, R. Aebersold, Mass Spectrometry and Protein Analysis, Science Vol. 312 no. 5771 2006, 212-217.

[19] P. Hao, J. Qian, B. Dutta et al, Enhanced Separation and Characterization of Deamidated Peptides with RP-ERLIC-Based Multidimensional Chromatography Coupled with Tandem Mass Spectrometry, J. Proteome Res., 2012, 11 (3), 1804–1811

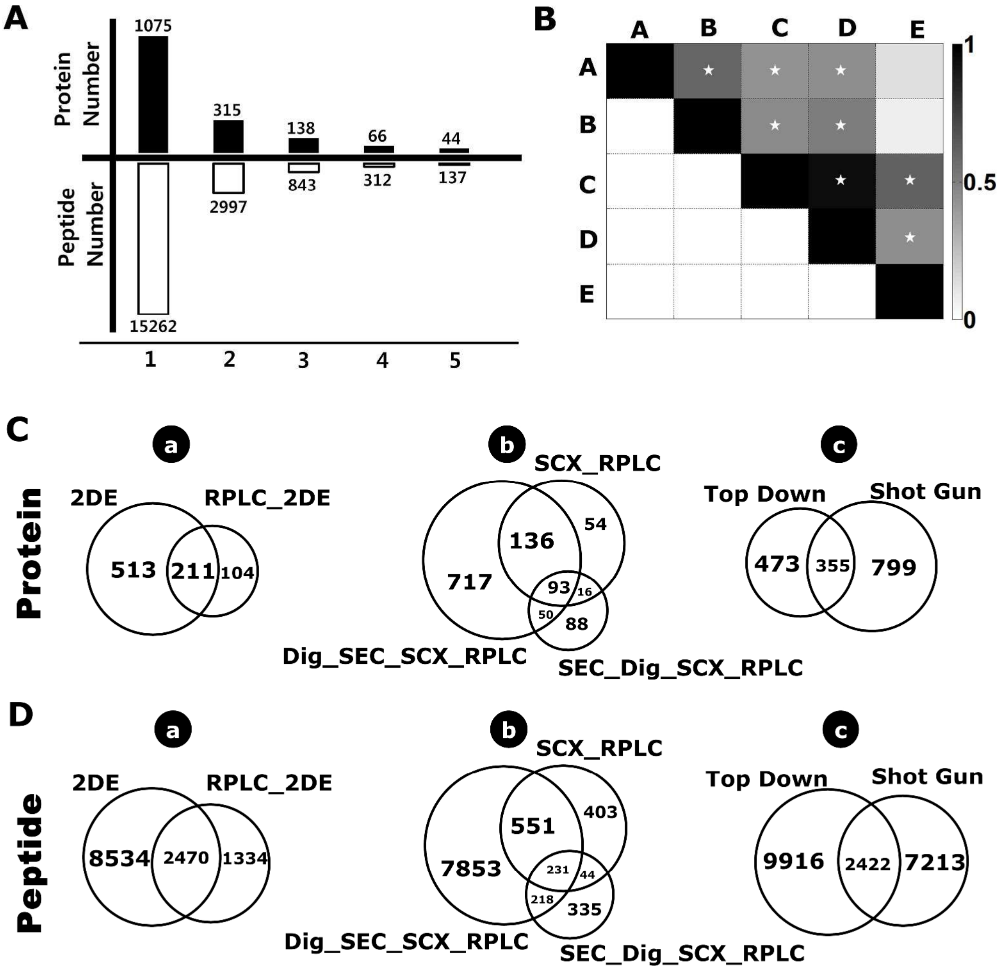
[20] X. Han, A. Aslanian and J. R. Yates, Mass spectrometry for proteomics, Current Opinion in Chemical Biology, 12 (5) 2008, 483–490.

[21] M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. (1976) 72, 248-254.

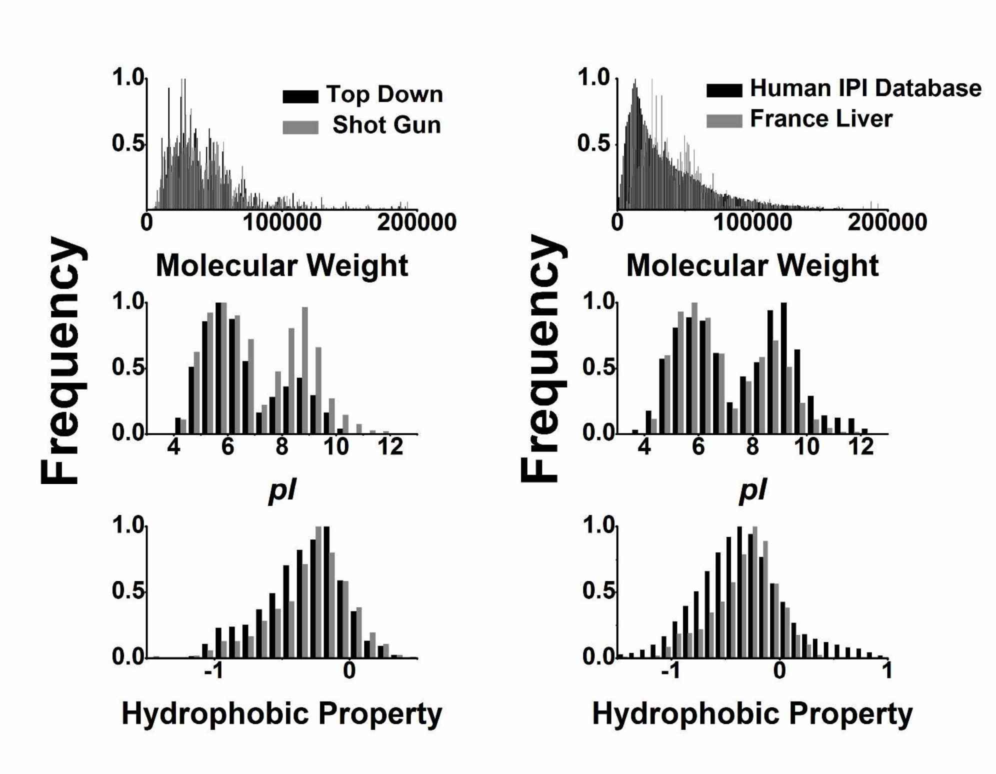
[22] H. Thorvaldsdóttir, J.T. Robinson and J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, *Brief Bioinform* (2013) 14 (2): 178-192



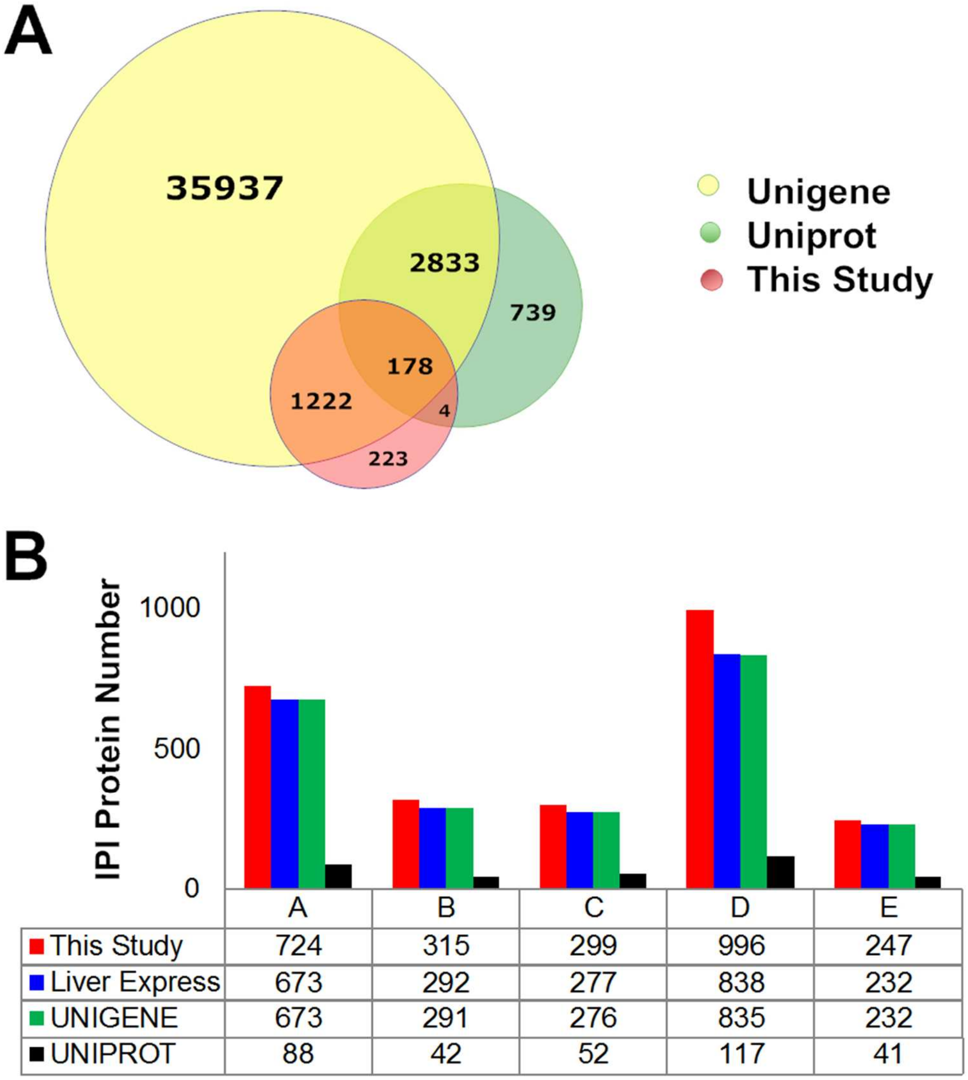
340x277mm (300 x 300 DPI)



514x499mm (300 x 300 DPI)

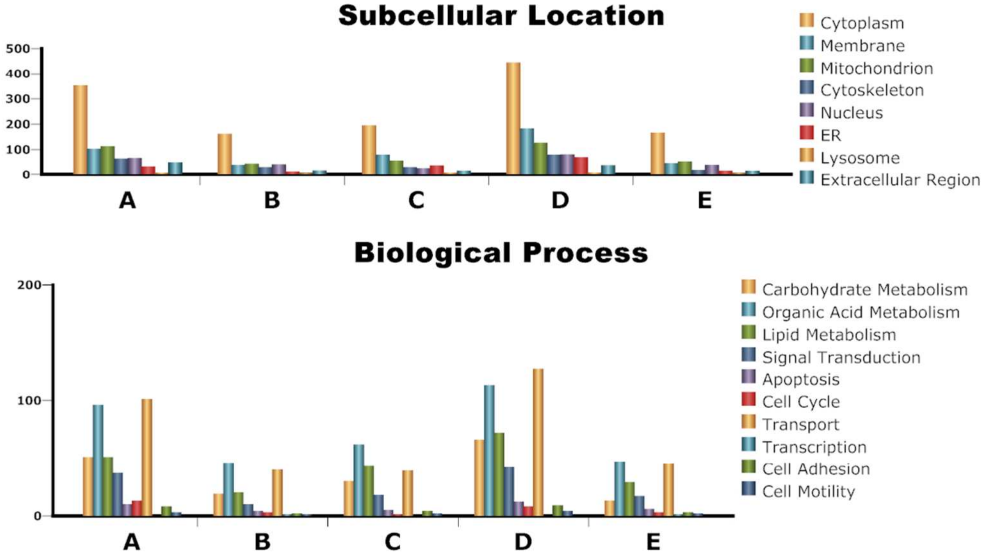


279x215mm (300 x 300 DPI)

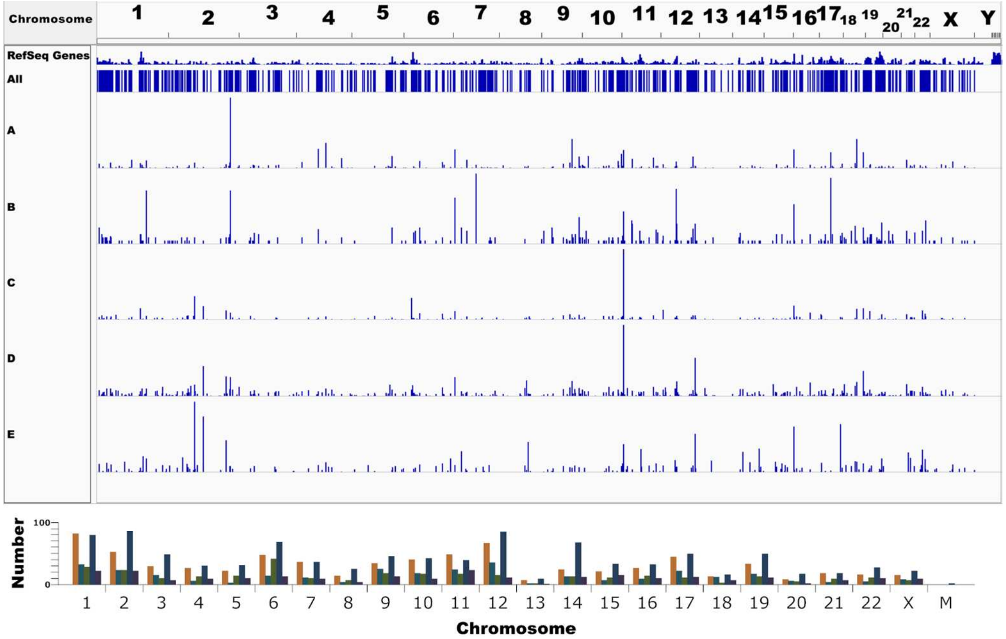


	A	B	C	D	E
🟥 This Study	724	315	299	996	247
🟦 Liver Express	673	292	277	838	232
🟩 UNIGENE	673	291	276	835	232
⬛ UNIPROT	88	42	52	117	41

82x91mm (300 x 300 DPI)



100x58mm (300 x 300 DPI)



143x90mm (300 x 300 DPI)

Table 1. Protein identification efficiency between route 1 and route 2.

	Route 1	**Route 2**
MALDI Spot Number	14629	3955
Success Spot Number	6438	988
Success Ratio	44%	25%
Protein Number	724	315
MALDI Spots per Protein	20.21	12.56
Peptide Number	15262	2997
Peptide Number per Protein	21.08	9.51

Table 2. Protein identification efficiency between route 3, route 4 and route 5.

	Route *3*	**Route *4***	**Route *5***
Fraction Number	21	94	16
Peptide Number	1470	14367	919
Unique Peptide Number	1252	9429	832
Protein Number	299	996	247
Unique Peptide Number per Protein	4.2	9.8	3.4
Unique Peptide Number / Spectrum	0.85	0.66	0.91