Analyst

Cite this: *Analyst*, 2012, **137**, 4703

www.rsc.org/analyst

**PAPER**

# Network generation enhances interpretation of proteomics data sets by a combination of two-dimensional polyacrylamide gel electrophoresis and matrix-assisted laser desorption/ionization-time of flight mass spectrometry†

**Xijun Wang,* Aihua Zhang,* Hui Sun,* Gelin Wu, Wenjun Sun and Guangli Yan**

***Received 1st July 2012, Accepted 15th August 2012***
**DOI: 10.1039/c2an35891c**



Recent advances in proteomic technologies have enabled us to create detailed protein–protein interaction maps in diseases. As the size of the interaction dataset increases, powerful computational methods are required in order to effectively interpret network models from large scale interactome data. In this study, we carried out comparative proteomics to construct and identify the proteins networks associated with hepatic injury (HI) which are largely unknown, as a case study. All proteins expressed were separated and identified by two-dimensional gel electrophoresis (2-DE) and matrix-assisted laser desorption/ionization time-of-flight-time-of-flight mass spectrometry (MALDI-TOF/TOF MS). Protein-interacting networks and pathways were mapped using STRING analysis program. We have performed for the first time a comprehensive profiling of changes in protein expression of HI rats, to uncover the networks altered by treated with $CCl_4$. Identification of fifteen spots (seven over-expressed and eight under-expressed) were established by MALDI-TOF/TOF MS. These proteins were subjected to functional pathway analysis using STRING software for better understanding of the biological context of the identified proteins. It suggested that modulation of multiple vital physiological pathways including DNA repair process, cell apoptosis, oxidation reduction, signal transduction, metabolic process, intracellular signaling cascade, regulation of biological processes, cell communication, regulation of cellular process, and molecular transport. In summary, the present study provides the first protein-interacting network maps and novel insights into the biological responses and potential pathways of HI. The generation of protein interaction networks clearly enhances the interpretation of proteomic data, particularly in respect of understanding molecular mechanisms of panel protein biomarkers.

## Introduction

The term proteomics encompasses the large-scale detection and analysis of proteins and their post-translational modifications. Driven by major improvements in mass spectrometric instrumentation, methodology, and data analysis, the proteomics field has burgeoned in recent years.[1] Proteomics provides important information that may not be inferable from indirect sources. It now has a range of sensitive and quantitative approaches for measuring protein structures and dynamics that promise to revolutionize our understanding of molecular mechanisms in human diseases.[2] However, mass-scale proteomics suffer from some pitfalls: protein and protein expression is not significant *per se*, but only if inserted in a detailed protein–protein interactions. This requires the development of a more robust and systematic tool to permit the automated construction and further analysis of molecular networks. Advent of network-based analysis methods can help in overcoming these problems but requires careful interpretation.[3] Beyond the identification of an integrated network of interest, further analysis of the network system is required to identify key target nodes that may represent novel therapeutic targets.

Disruption of protein–protein interactions can result in emergence of various diseases including tumor, neurodegenerative, cardiovascular, autoimmune, *etc.* Therefore, investigation of interacting protein networks is important in the diagnosis of diseases and in revealing the mechanism of their emergence and development, as well as the efficiency of different therapeutic approaches.[4] Traditional strategies for protein expression analysis have focused on identifying individual proteins that exhibit differences between two states of interest. Although useful, they fail to detect biological processes which are distributed across an entire network of proteins and subtle at the level of individual

*National TCM Key Lab of Serum Pharmacochemistry, Heilongjiang University of Chinese Medicine, and Key Pharmacometabolomics Platform of Chinese Medicines, Heping Road 24, Harbin 150040, China. E-mail: xijunwangls@126.com; aihua-zhang@163.com; phar_research@hotmail.com; Fax: +86-451-82110818; Tel: +86-451-82110818*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2an35891c

proteins. The rise of high-throughput technologies has led to the production of large amounts of biological data.[5] Proteomics technology has recently reached a point where it is possible to generate network inference approaches under different conditions.[6] The challenge no longer lies in obtaining protein expression profiles, but rather in interpreting the results to gain insights into biological mechanisms. Most of the successful studies building on the new approaches have focused, however, on using network-based tools to gain a better understanding of the relationship between the proteins implicated in a selected disease.[7]

The liver is the largest internal organ in the body that takes central roles in metabolic homeostasis and detoxification of various substances, as well as in the synthesis and storage of nutrients. To fulfill these complex tasks, thousands of biochemical reactions are going on in the liver to cope with a wide range of foods and environmental variations, which are densely interconnected into an intricate protein network. To date, the proteomic studies of liver have generated a list of identified proteins using methods like SDS-PAGE, 2-DE, LC-MS/MS, and SELDI-TOF.[8–10] Proteomic studies of HI disorders have the potential to identify protein biomarkers for diagnosis and disease monitoring. Protein–protein interaction networks are typical examples of complex biological systems that are difficult to understand from raw experimental data alone. Development and use of protein regulatory networks to mechanistically interpret this data is an important development in molecular biology, usually captured under the banner of systems biology. As a result, the repertoire of methods for the reconstruction of comprehensive and protein-specific maps of regulatory interactions, or interactomes, has also exploded in the past few years.[11–13]

Interaction networks provide a framework to visualize disease processes, but their complexity often makes their interpretation an overwhelming task.[14–16] With an ever increasing amount of available data on biological networks, modeling and understanding the structure of these large networks is an important problem with profound biological implications. Specifically, the identification of key target nodes within complex molecular networks remains a common objective in scientific research. Therefore, developing an effective computational method to uncover biological modules should be highly challenging and indispensable. In this tutorial, we carry out comparative proteomics from large scale proteomics data sets to create detailed protein–protein interaction maps in HI disease and aim to give an example for enriching biological discovery analysis. Using the molecular interaction database STRING, and taking as a starting point a list of proteins identified in a mass spectrometry-based proteomics experiment, we show how to build, visualize, and analyze a protein–protein interaction network.

## Materials and methods

### Reagents

Ultra-pure reagents for polyacrylamide gel preparation were obtained from Bio-Rad 18 cm immobilized pH gradient (IPG) strips (pH3-10), IPG buffer and dry strip cover fluid were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). 3-[(3-Cholamidopropyl)dimethylammonio]-propanesulfonate (CHAPS), glycine, ammonium persulphate (APS), TEMED, and trifluoroacetic acid (TFA) were obtained from Amresco (Solon, OH, USA). Dithiothreitol (DTT), PMSF, iodoacetamide, urea, thiourea, α-cyano-4-hydroxycinnaic acid (CCA), and TPCK-trypsin were purchased from Sigma Chemical (St. Louis, MO, USA). Distilled water was purchased from Watson's Food & Beverage Co., Ltd. (Guangzhou, China). The other reagents that were used in 2-DE were purchased from Bio-Rad.

### Animal sampling and handling

The animals were randomly assigned to control and HI groups of 8 rats in each group. The rats in the HI group were orally administrated 20% $CCl_4$ (1 ml per kg body weight) for 8 consecutive days (at 8:00 p.m.), and on day 9, the rats were anesthetized *via* an intraperitoneal injection of 1% pentobarbital sodium (0.15 ml per 100 g body weight). The rats in the control group were given distilled water for 8 consecutive days and anesthetized on day 9 as in the HI group. Blood (5 ml) was collected from the hepatic portal vein, and serum was separated *via* centrifugation at 4500 rpm for 5 min at 4 °C, flash frozen in liquid nitrogen and stored at −80 °C until the proteomics analyses were performed. All animal care and experimental procedures were performed in compliance with the policies on the care and use of animals of the Ethical Committee of Heilongjiang University of Chinese Medicine.

### Two-dimensional gel electrophoresis and image analysis

The plasma obtained was disrupted in a homogenization buffer containing 30 mM Tris, 7 M urea, 2 M thiourea, 4% CHAPS, and 40 mM DTT, after which the lysates were centrifuged at 12 000 rpm for 60 min at 4 °C. The supernatant was precipitated using the TCA sediment method to recover the total proteins, and the interfering components were removed using the 2-D Clean-Up Kit (GE Healthcare, Piscataway, NJ). The supernatant was removed, divided into several aliquots and stored at −70 °C until use. The protein concentration of each sample was measured using the Bradford method. A total of 350 mg of protein was loaded onto an 18 cm linear IPG strip (pH 3–10, Amersham Biosciences, Piscataway, NJ) for first-dimension isoelectric focusing (IEF). The strips were placed into a Protean IEF cell (Bio-Rad) and rehydrated at 50 V for 12 h, after which the proteins were separated based on their pI according to the following protocol: 250 V with linear climb for 60 min, 1000 V with rapid climb for 60 min, 10 000 V with linear climb for 5 h, and 10 000 V with rapid climb until a level of 60 000 V h was reached. After IEF, the IPG strips were equilibrated for 15 min in a buffer containing 50 mM Tris–HCl, pH 8.8, 30% glycerol, 7 M urea, 2% SDS, and 1% DTT; they were then further treated in a similar buffer (containing 4% iodoacetamide) for 15 min and directly applied onto 12% homogeneous SDS-PAGE gels for 6 h electrophoresis using a Protean II xi cell system (Bio-Rad). Moreover, two kinds of electrophoresis conditions that were suitable for the separation of proteins with higher molecular weight (10 mA per gel for 30 min followed by 30 mA per gel for 5.5 h) and for the separation of proteins with lower molecular weight (10 mA per gel for 30 min followed by 20 mA per gel for 8 h) were used. The gels were then silver-stained using Bio-Rad Silver Stain Plus kit reagents (Bio-Rad) according to the

 

manufacturer's instructions. The silver-stained gels were scanned using a GS-800 densitometer (Bio-Rad) and then analyzed using PDQuest software (Bio-Rad). Labscan Software (Applied Biosystems, Foster City, CA) was used for image analysis that included background abstraction, spot intensity calibration, spot detection, and matching. The intensity of each spot was quantified by calculating the spot volume after the gel image had been normalized. Each sample was processed in triplicate to ensure reproducibility, and the Student's *t*-test was used to evaluate the average change in protein abundance corresponding to each target spot across the gels. Protein spots with more than a 3-fold significant change (either increasing or decreasing) in density (Student's *t*-test; $p < 0.05$) were considered to be differentially expressed and were selected for further identification using MALDI-TOF/TOF MS.

### MALDI-TOF/TOF mass spectrometry and database analysis

Protein spots of interest with differential expression levels were excised from the 2-D gels with an EXQuest spot cutter (Bio-Rad), washed 3 times in distilled water, destained with 15 mM potassium ferricyanide in 50 mM sodium thiosulfate for 2 min at room temperature, dehydrated in acetonitrile (ACN) and finally dried in a centrifugal vaporizer. After a 30 min incubation, the gels were digested for more than 12 h at 37 °C. The peptides were then extracted twice using 0.1% TFA in 50% ACN dried under the protection of $N_2$. For MALDI-TOF/TOF MS, the peptides were mixed with 0.5 μl MALDI matrix (5 mg $ml^{-1}$ α-cyano-4-hydroxycinnamic acid diluted in 0.1% TFA and 50% ACN) and spotted onto the MALDI sample plates. The MS measurements were performed on a Proteomics Analyzer with delayed ion extraction (Applied Biosystems). The data were acquired in the delayed ion extraction mode using a 20 kV accelerating voltage and a 100 ns extraction delay time. Using the individual PMF spectra, peptides with a signal-to-noise ratio exceeding 20 that passed through a mass exclusion filter were subjected to fragmentation analysis. MS accuracy was internally calibrated using trypsin-digested peptides of horse myoglobin. MS/MS accuracy was calibrated against the MS/MS fragments of *m*/*z* (1373.60 and 1028.44), which is one of the peaks that is generated in the myoglobin peptide mass fingerprint (PMF). The parameters for peak matching were as follows: the minimum signal-to-noise ratio was 20, the mass tolerance was 0.2 Da, the minimum peak to match reference masses was 4, and the maximum outlier error was 50 ppm. Bioinformatics databases were searched for protein identification based on tryptic fragment sizes. The Mascot search engine (http://www.matrixscience.com) was initially used; we searched the theoretical peptide masses in the National Center for Biotechnology Information database (www.ncbi.nlm.nih.gov/; in the public domain, provided by the National Institutes of Health, Bethesda, MD) and SwissProt (http://www.expasy.org/; in the public domain, provided by the Swiss Institute of Bioinformatics, Geneva, Switzerland) using the assumption that peptides are monoisotopic and may be oxidized at methionine residues and that cysteine is completely modified by iodoacetamide. The protein score is $-10 \times \log(P)$, where $P$ is the probability that the observed match is a random event. The missed cleavage sites allowed were up to 1, and scores that were higher than 61 were considered significant.

### Proteomics analysis

2D-PAGE tests were performed. Protein spots with more than a 3-fold change in density with consistent increases or decreases were considered to be differentially expressed and were selected for further identification *via* a MALDI-TOF MS/MS analysis. All experiments were performed at least in triplicate to ensure reproducibility.

### Bioinformatics analysis of proteomic data

Identified proteins were further analysed using STRING, chosen as the source for protein–protein interactions, to statistically determine the functions and pathways most strongly associated with the protein list. STRING, a meta resource that aggregates most of the available information on protein–protein associations, scores and weights it, further allows exploring the network of natural drug-target relations, also in the context of associated binding proteins, combines a vast number of databases derived in different ways (*e.g.* experimentally determined interactions, protein neighbourhood data, or data acquired *via* text mining). Prior to upload and analysis using STRING, the mean ratio of each quantified protein in a group was calculated and the fold change between the groups calculated. A dataset containing protein identifiers and corresponding expression values was uploaded into the application. A fold change cut-off of 2.0 was set to identify molecules whose expression was significantly differentially regulated. The PANTHER classification system and KEGG (www.genome.jp/kegg/) pathway database were used to further examine the significantly expressed molecules.

## Results

### Identification of the differentially expressed proteins

To better understand differentially expressed proteins in HI rats, we investigated dynamic proteomic differences between the model and control groups. After analyzing the 2-DE gels, the peptides were extracted from each differentially expressed protein spot using in-gel trypsin digestion, and the proteins were identified using MS/MS. Representative 2-DE images for the control and model groups are shown in Fig. 1A and B. The

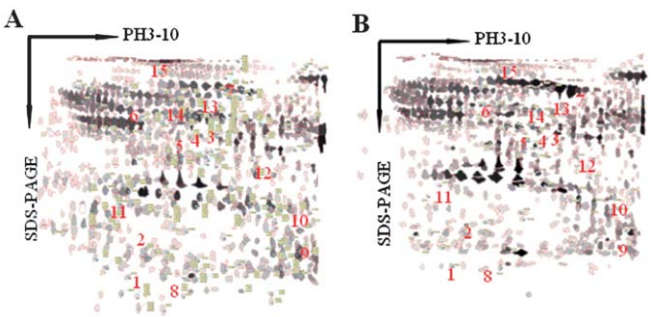


**Fig. 1** A representative two-dimensional electrophoresis (2-DE) representative proteomic maps of control (A) and HI (B) group. Numbers indicate proteins that were expressed differentially in the two groups. Protein spots 1–7 were upregulated, whereas the other spots were downregulated in the HI group compared to the control group.

results of the MS/MS analysis—including the protein score, coverage, number of identified peptides, and best ion score for each spot—are summarized in ESI Table 1.† The result of MALDI-TOF/TOF MS analysis of spot 4 and 8 is shown in Fig. 2 as an example. The resolution in 2-DE silver-stained gels resulted in approximately 1000 spots. Among all of the protein spots that were separated using 2-DE profiles with high resolution and reproducibility, 15 spots appeared to be significantly changed in percent volume as identified by PMF that were based on MALDI-TOF/TOF MS and database searching. We found these proteins (7 of which were over-expressed and 8 of which were under-expressed) that were expressed in the animal model. Relative expression level of the differentially expressed proteins is shown in Fig. 3. The over-expressed proteins in the HI group included Ig kappa chain C, zinc finger protein 407, serotransferrin, haptoglobin, macroglobulin, alpha-1-antitrypsin and complement factor B. The under-expressed proteins included fibrinogen alpha chain precursor, glyceraldehyde-3-phosphate dehydrogenase, albumin, 2,4-dienoyl-CoA reductase, transthyretin, α-1-inhibitor 3 precursor, vitamin D-binding protein and prothrombin. These differentially expressed proteins appear to be involved in metabolism, energy production, immunity, chaperoning, antioxidation and signal transduction.

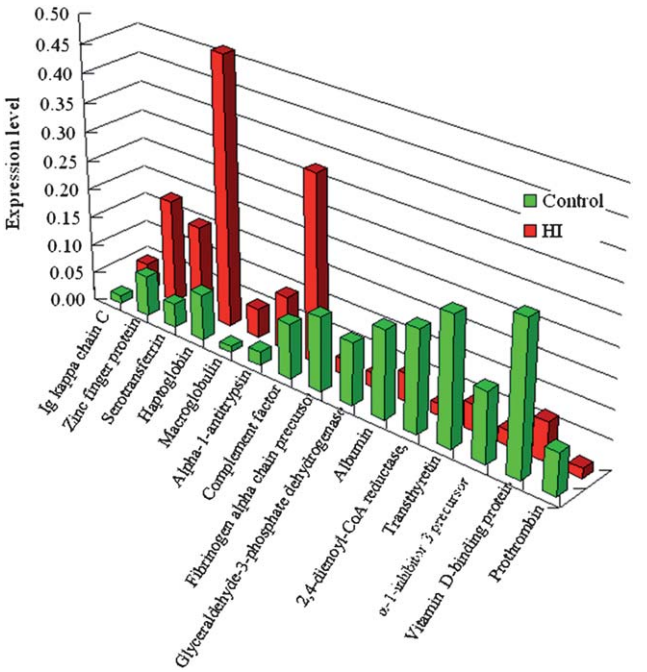


**Fig. 3** Relative expression level of the differentially expressed proteins obtained by MALDI-TOF/MS.

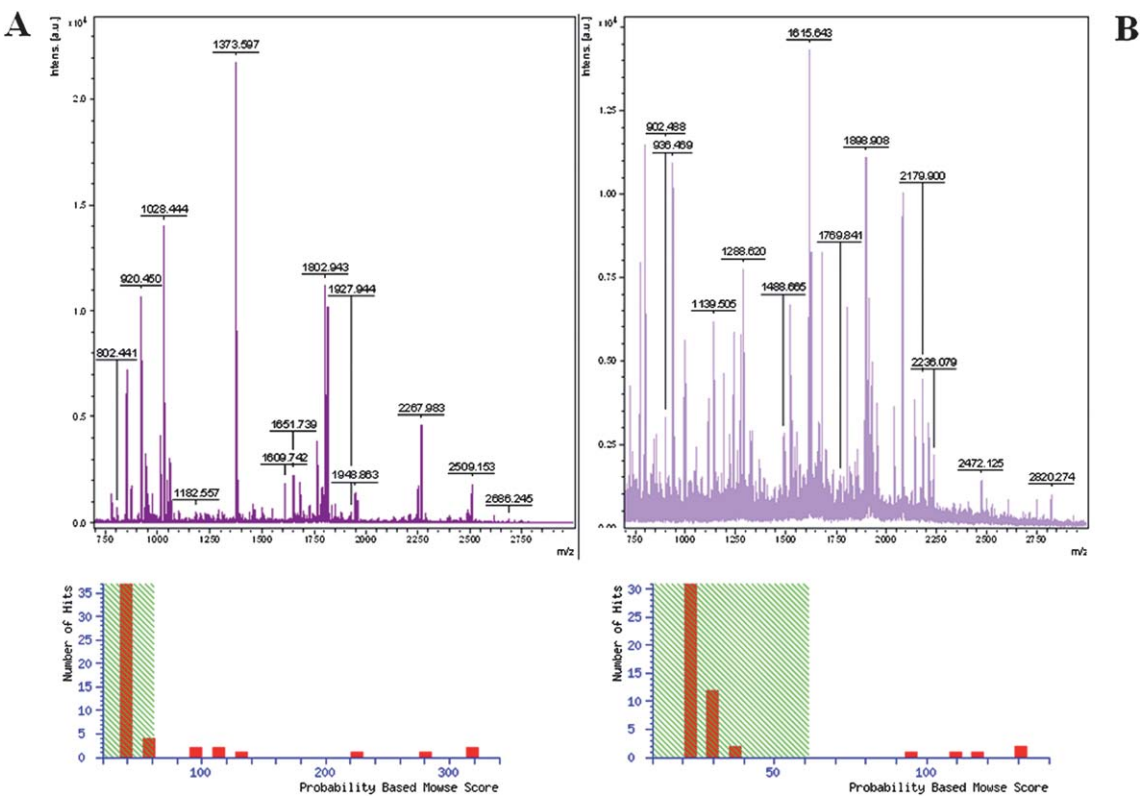


**Fig. 2** Peptide mass fingerprint spectrum of haptoglobin (spot 4) and fibrinogen alpha chain precursor (spot 8). The spot was in-gel digested with trypsin. After desalting, the peptide mixture was analyzed by MALDI-TOF-MS. Spectral masses (in mass per charge unit, *m*/*z*) obtained by MALDI-TOF/MS were analyzed by using bioinformatics through the Mascot and Profound search engines. The *x*-axis represents the mass-to-charge ratio (*m*/*z*), and the *y*-axis represents the relative abundance. All protein identifications are provided in ESI Table 1.† Below: the probability-based Mowse scores obtained using the Mascot search engine. Among predicted proteins with differential Mowse scores shown as multiple bars on the *x*-axis, only proteins with Mowse scores greater than 61 were considered significant, which were 318 and 131 ($p < 0.05$) for haptoglobin and fibrinogen alpha chain precursor, respectively.

 

### Functional analysis of proteins

We employed a proteomics strategy that involved combining 2-DE and MALDI-TOF/TOF MS to perform functional enrichment analysis, with a special focus on the changes in the HI biological process. From all the gene ontology (GO) categories covered by the differentially expressed proteins and interactions, the distinct regulatory patterns provides evidence that novel biomarkers are actively involved in multifunctional pathways that are likely essential for HI. The biological functions of these critical proteins can be sorted into five groups: (A) generation and degradation of the extracellular matrix, including fibrinogen alpha chain precursors and macroglobulin; (B) the regulation of transcription and translation, as provided by zinc finger proteins; (C) acute phase reaction and immunity protection, as provided by Ig kappa chain C, alpha-1-antitrypsin, α-1-inhibitor 3, prothrombin and vitamin D-binding protein; (D) oxygenation and cell apoptosis, to which 2,4-dienoyl-CoA reductase contributes; and (E) transport and metabolism, as provided by glyceraldehyde-3-phosphate dehydrogenase, haptoglobin, serotransferrin and transthyretin. The characteristic functions of these differentially expressed proteins were enriched within clusters that were based on biological processes such as "immunity", "cellular apoptosis", "transport", "signal transduction", "cell growth and proliferation" and "metabolism". For example, the modulation of several key members of the immunity cluster was revealed by proteome analysis, as highlighted by Ig kappa chain C, α-1-inhibitor 3 and prothrombin. As expected, currently available HI biomarkers rely on the measurement of substances that are key to the development, transport and metabolism of HI, and may be potential therapeutic targets for HI.

### Protein–protein interaction network visualization

To facilitate access to the protein–protein interaction data, STRING integrates information about interactions from metabolic pathways, crystal structures, binding experiments and drug–target relationships. Inferred information from phenotypic effects, text mining and chemical structure similarity is used to predict relations between disease and targets. As seen in Fig. 4 and 5, association network of differentially expressed proteins using STRING was constructed. Modes of action are shown in different colors. Networks are represented as graphs where the nodes are the entities and the edges their interactions. Differentially expressed proteins are represented as pill-shaped nodes, while proteins are shown as spheres. Nodes that are associated to each other are linked by an edge. Thicker lines represent stronger associations. Lines and, for directed edges, arrows of different colours stand for different edge types in the actions view: binding (blue), activation (green), inhibition (red), catalysis (magenta), same activity (cyan) and reaction (black).

### Analysis of interaction networks

Based on the proposed framework, we test our hypothesis that HI produces a disturbance in the protein expression, having a global effect in networks and essential biological pathways. To examine the protein functions and pathways of the differentially expressed proteins, the enriched GO terms were categorized for HI networks to identify the functional change. From all the GO categories covered by the differentially expressed proteins and their interactions, we have listed the ten most significant categories of the enriched GO terms in ESI Table 2.† These closely connected and coexpressed differentially expressed proteins are regarded as the signatures of the underlying targets. The significantly over represented categories indicated that this emergent new protein network was composed by proteins involved in DNA repair process, cell apoptosis, oxidation reduction, signal transduction, metabolic process, intracellular signaling cascade, regulation of biological process, cell communication, regulation of cellular process, and molecular transport.

## Discussion

Large-scale protein–protein interaction maps provide new insights into protein functions, pathways, molecular machines and functional protein modules.[17] Moreover, studies have indicated that protein–protein interaction networks are vital to understanding disease specificity.[18] Recent advances in high-throughput methods for taking global measurements of protein levels from biological samples have driven the field of systems biology.[19] A common application of such methods is to identify proteins that are likely to be involved in the disease process. However, this approach may overlook proteins that are important, but may not be the most highly differentially regulated, such as other upstream mediators of critical processes. Identification of target proteins for diseases represents a major objective of current biomedical research.[20] Analyzing the interaction network can be the key to understand how complex processes lead to diseases and has greatly contributed to used in identifying potential targets that are generally identified from the most highly differentially expressed proteins.[21]

In recent years, major advances in the various types of biological networks are being constructed including protein–protein interaction networks, and protein regulatory networks, *etc.*[22] Recent work by Yu *et al.*, for example, shifts the emphasis from high degree hubs to nodes that are 'bottlenecks' in the network.[23] Zhang and Chen further discovered that regulatory patterns for protein factor hubs changed substantially under different contexts.[24] Recent studies have also shown that house-keeping proteins and tissue-specific proteins have different topological properties in the human protein–protein interaction network.[25–28] Several studies of protein interaction networks have focused on detecting highly connected protein modules which generally correspond to meaningful biological units such as protein complexes and functional modules.[29] However, biological networks including protein–protein interaction networks are generally very sparse. Most methods only identify strongly connected subgraphs as modules, so only a few modules were detected. Furthermore, because these approaches heavily rely on the local topological connectivity, they ignore the impact of global organization of networks. High-throughput methods for obtaining global measurements of protein levels in biological samples have provided a large amount of data for identification of target proteins and proteins of interest. These targets may be mediators of functional processes involved in disease and therefore represent key points of control for disease.

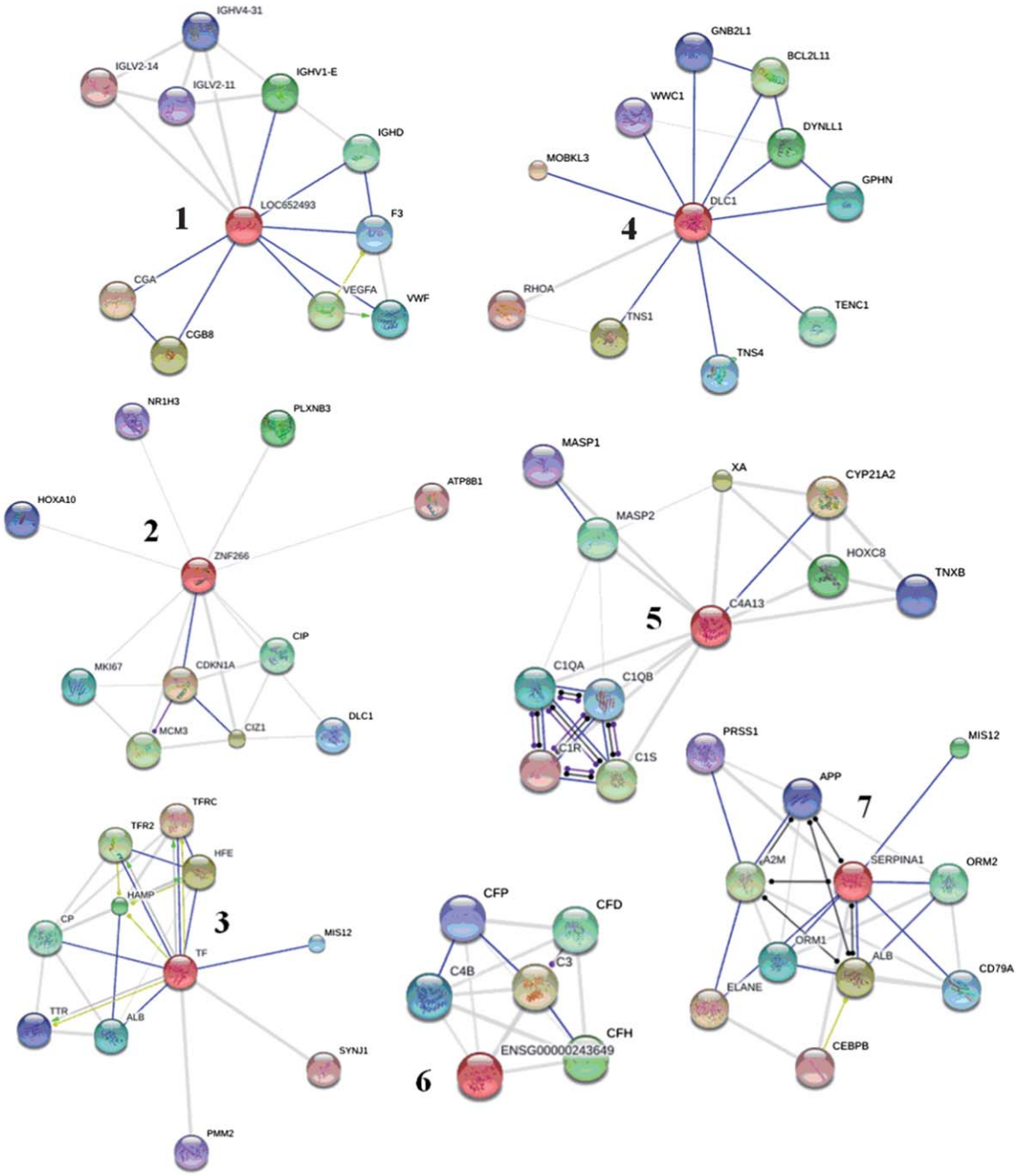


**Fig. 4** Protein interaction network generated with STRING and visualized with Cytoscape for proteins identified with increased expression. Modes of action are shown in different colors. Differentially expressed proteins are represented as pill-shaped nodes, while other proteins are shown as spheres. Nodes that are associated to each other are linked by an edge. Thicker lines represent stronger associations. Lines and, for directed edges, arrows of different colours stand for different edge types in the actions view: binding (blue), activation (green), inhibition (red), catalysis (magenta), same activity (cyan) and reaction (black).

A huge number of computational methods have been developed for detecting and analyzing the structure of biological networks.[30–32] One approach to gaining wider views on the roles of proteins in biological systems relies on network biology.[33,34] Known and inferred protein associations are used to build a network of proteins, thus establishing a map of all the associations in the organism and allowing deductions to be made as to the role of proteins that are poorly understood and annotated. Clearly, both proposed and demonstrated protein–protein associations could aid us in understanding the role of proteins. Therefore, we constructed a network of HI proteins using the protein–protein association data from STRING database, and analyzed these data. We have found 15 differentially expressed proteins identified by MALDI-TOF/TOF MS, including Ig kappa chain C, zinc finger protein 407, serotransferrin, haptoglobin, macroglobulin, alpha-1-antitrypsin and complement factor B, fibrinogen alpha chain precursor, glyceraldehyde-3-phosphate dehydrogenase, albumin, 2,4-dienoyl-CoA reductase,

 

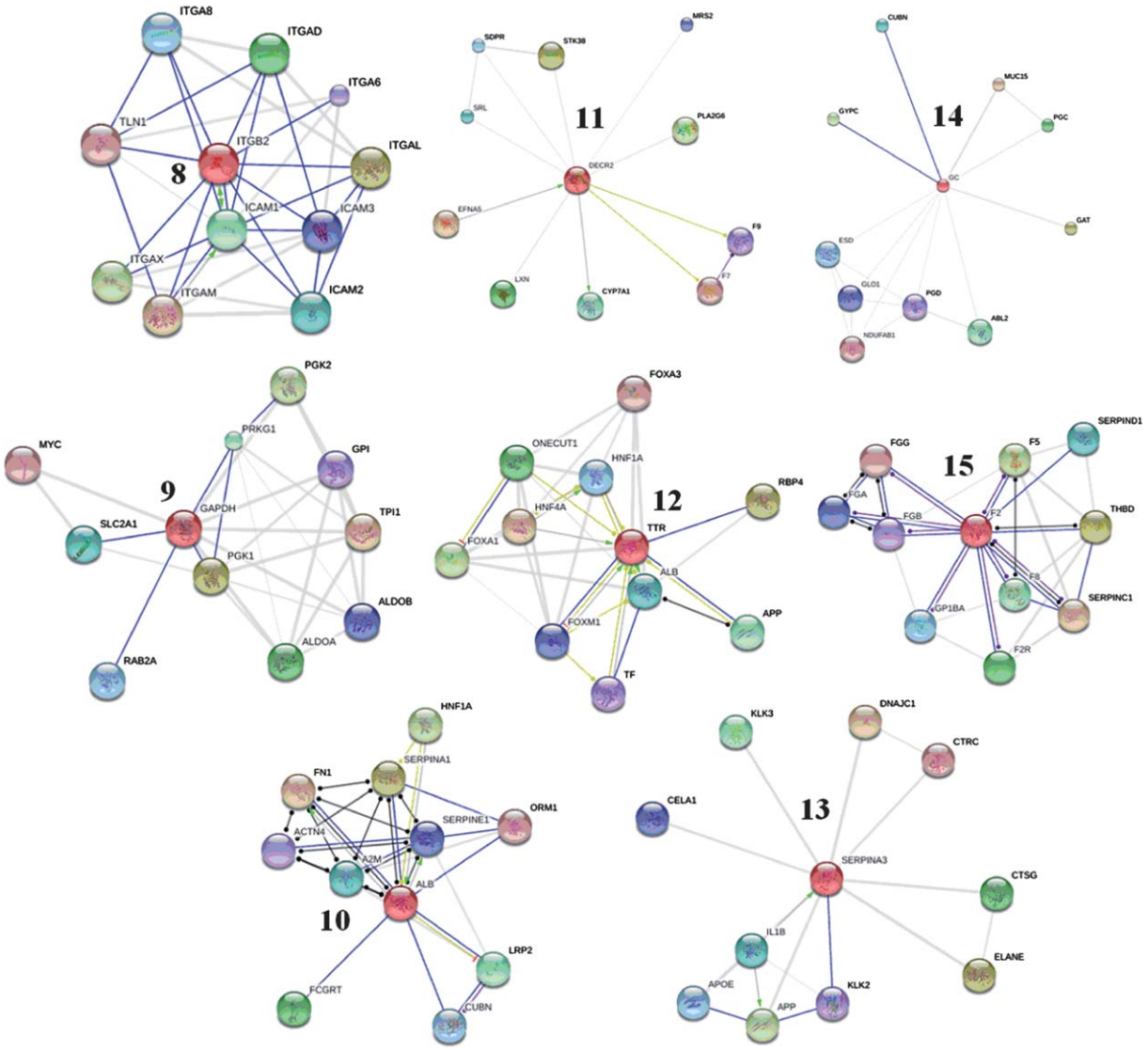


**Fig. 5** Protein network of the identified proteins constructed with the STRING software for enrichment analysis of proteins with decreased expression. Each node represents either a protein entity or a control mechanism of the interaction. The legend of the interaction network is summarized under the figure. Numbers show the identified proteins. The relations of binding, expression, and direct regulation were obviously displayed among these proteins.

transthyretin, α-1-inhibitor 3 precursor, vitamin D-binding protein and prothrombin. These proteins represent targets for further experimental investigation that will provide biological insight and potentially could be exploited for novel therapeutic approaches to liver disease. The characteristic functions of these differentially expressed proteins were enriched within clusters that were based on biological processes such as immunity, cellular apoptosis, transport, signal transduction, cell growth and proliferation and metabolism. As seen in Fig. 4 and 5, association network of differentially expressed proteins was constructed using STRING. From all the GO categories covered by the differentially expressed proteins and their interactions, we have listed the ten most significant categories of the enriched GO terms in ESI Table 2.† The significantly over represented categories indicated that this emergent new protein network was composed by proteins involved in DNA repair process, cell apoptosis, oxidation reduction, signal transduction, metabolic process, intracellular signaling cascade, regulation of biological process, cell communication, regulation of cellular process, and molecular transport. Global protein network analysis was then performed to obtain additional functional significance of the identified proteins and to guide for subsequent functional analysis. Our data may lead to a better understanding of molecular mechanisms of HI induced by $CCl_4$. We offer STRING as a broadly applicable approach for analysis of proteomic data which may help increase the current understanding of molecular networks in a variety of biological problems. This study analysed the disease mechanisms by considering the interactions of the proteins and employing topological analyses. Network-based target identification from proteomics data has the significance to compare this approach with an existing method for identification of important proteins from global proteomics data, differential regulation.

To study the complexity of disturbances in protein networks of HI, we constructed protein expression networks of HI and identified core pathways and therapeutic targets of HI by

network pharmacology. Sets of proteins associated with HI were extracted from the whole network to create differentially expressed protein-associated subnetworks. The characteristic functions of the differentially expressed proteins were based on biological processes including immunity, cellular apoptosis, transport, signal transduction, cell growth and proliferation and metabolism. The detection of these proteins with distinct regulatory patterns provides evidence that novel biomarkers are actively involved in multifunctional pathways. Novel protein targets and previously unrecognized members of some subsystems could be postulated; these insights help us to better understand the mechanisms underlying HI. The overall network was significantly enriched by proteins associated with signal transduction, cellular communication and regulatory processes. The identified targets were found to encompass a variety of biological processes mediated through complex networks. Furthermore, the subnetworks may also play an important role in discovering biomarkers. It has been shown that identified markers for class prediction are more reproducible if their identification is based on subnetworks rather than single proteins. Novel protein targets and the protein-associated subsystems provide new insights into the mechanisms underlying pathogenesis. The findings demonstrate that the network target-based methods are of importance for elucidating the inter-relationship between complex diseases. Generalization of the proposed method for identifying subnetwork markers used for class prediction will be the focus of future work.

## Conclusions

The advent of the 'omics' era in biology research has brought new challenges and requires the development of novel strategies to answer previously intractable questions. Disruption of the biological processes results in phenotypes that characterize each disease, with the phenotype depending on the influence of the affected biological processes on the larger biological network. Molecular interaction networks provide a framework to visualize biological processes, but their complexity often makes their interpretation an overwhelming task. Network pharmacology makes it possible to study larger and more intricate systems than before. Development of network pharmacology has enabled a detailed understanding of the underlying mechanism of disease. In this tutorial, we aim to give an illustrative example of the key aspects that any researcher needs to consider how to build, visualize, and analyze a protein–protein interaction network when working with high interaction data sets. The potential target identification problem can be solved effectively by our method, for extracting differentially regulated subnetworks from protein–protein interaction networks based on data from global quantification technologies, which may become an effective strategy for the underlying mechanism of disease. Knowledge of altered protein expression levels and perturbations to the interactome will assist in the development of targeted-network combination therapies with the potential to treat liver diseases. However the interpretation of high-throughput data is limited in its ability to provide a systems level understanding of changes to the proteome. This study, for the first time, incorporated protein databases to construct the protein–protein interaction network of abnormally expressed proteins in rat samples of HI disorder and the pathophysiology of complex diseases is characterized by the involvement of various biologic pathways. Therefore, identification and characterization of candidate proteins, and proteins involved in a given disease by the integrated use of proteomics disciplines will probably represent one of the milestones of future health care.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

This work was supported by grants from the Key Program of Natural Science Foundation of State (Grant no. 90709019), the National Specific Program on the Subject of Public Welfare (Grant no. 200807014), National Key Subject of Drug Innovation (Grant no. 2009ZX09502-005), and National Program on Key Basic Research Project of China (Grant no. 2005CB523406).

## References

1 D. Lavigne, L. Guerrier, V. Gueguen, J. B. Michel, E. Boschetti, O. Meilhac and D. Letourneur, Culture of human cells and synthesis of extracellular matrix on materials compatible with direct analysis by mass spectrometry, *Analyst*, 2010, **135**(3), 503–511.
2 C. Behrends, M. E. Sowa, S. P. Gygi and J. W. Harper, Network organization of the human autophagy system, *Nature*, 2010, **466**(7302), 68–76.
3 L. M. Stuart, J. Boulais, G. M. Charriere, E. J. Hennessy, S. Brunet, I. Jutras, G. Goyette, C. Rondeau, S. Letarte, H. Huang, P. Ye, F. Morales, C. Kocks, J. S. Bader, M. Desjardins and R. A. Ezekowitz, A systems biology analysis of the Drosophila phagosome, *Nature*, 2007, **445**(7123), 95–101.
4 C. Jørgensen, A. Sherman, G. I. Chen, A. Pasculescu, A. Poliakov, M. Hsiung, B. Larsen, D. G. Wilkinson, R. Linding and T. Pawson, Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells, *Science*, 2009, **326**(5959), 1502–1509.
5 A. Zhang, H. Sun, B. Yang and X. Wang, Predicting new molecular targets for rhein using network pharmacology, *BMC Syst. Biol.*, 2012, **6**, 20.
6 L. Bonetta, Protein–protein interactions: interactome under construction, *Nature*, 2010, **468**(7325), 851–854.
7 C. C. Liao, A. Mehta, N. J. Ward, S. Marsh, T. Arulampalam and J. D. Norton, Analysis of post-operative changes in serum protein expression profiles from colorectal cancer patients by MALDI-TOF mass spectrometry: a pilot methodological study, *World J. Surg. Oncol.*, 2010, **8**, 33.
8 S. C. Arruda, S. Barbosa Hde, R. A. Azevedo and M. A. Arruda, Two-dimensional difference gel electrophoresis applied for analytical proteomics: fundamentals and applications to the study of plant proteomics, *Analyst*, 2011, **136**(20), 4119–4126.
9 E. Alzahrani and K. Welham, Design and evaluation of synthetic silica-based monolithic materials in shrinkable tube for efficient protein extraction, *Analyst*, 2011, **136**(20), 4321–4327.
10 A. W. Jones and H. J. Cooper, Dissociation techniques in mass spectrometry-based proteomics, *Analyst*, 2011, **136**(17), 3419–2349.
11 R. F. Deighton, L. E. Kerr, D. M. Short, M. Allerhand, I. R. Whittle and J. McCulloch, Network generation enhances interpretation of proteomic data from induced apoptosis, *Proteomics*, 2010, **10**(6), 1307–1315.
12 F. Li, O. V. Glinskii, J. Zhou, L. S. Wilson, S. Barnes, D. C. Anthony and V. V. Glinsky, Identification and analysis of signaling networks potentially involved in breast carcinoma metastasis to the brain, *PLoS One*, 2011, **6**(7), e21977.
13 T. Simões, N. Charro, J. Blonder, D. Faria, F. M. Couto, K. C. Chan, T. Waybright, H. J. Isaaq, T. D. Veenstra and D. Penque, Molecular profiling of the human nasal epithelium: a proteomics approach, *J. Proteomics*, 2011, **75**(1), 56–69; P. A. Ortiz, M. E. Bruno,

 

T. Moore, S. Nesnow, W. Winnik and Y. Ge, Proteomic analysis of propiconazole responses in mouse liver: comparison of genomic and proteomic profiles., *J. Proteome Res.*, 2010, **9**(3), 1268–1278.
14 M. Heo, S. Maslov and E. Shakhnovich, Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**(10), 4258–4263.
15 X. Wang, B. Yang, A. Zhang, H. Sun and G. Yan, Potential drug targets on insomnia and intervention effects of jujuboside A through metabolic pathway analysis as revealed by UPLC/ESI-SYNAPT-HDMS coupled with pattern recognition approach, *J. Proteomics*, 2012, **75**(4), 1411–1427.
16 X. Wang, B. Yang, H. Sun and A. Zhang, Pattern recognition approaches and computational systems tools for ultra performance liquid chromatography-mass spectrometry-based comprehensive metabolomic profiling and pathways analysis of biological data sets, *Anal. Chem.*, 2012, **84**(1), 428–439.
17 Z. J. Li, N. Liu, Q. D. Han and Y. Y. Zhang, Proteome reference map and regulation network of neonatal rat cardiomyocyte, *Acta Pharmacol. Sin.*, 2011, **32**(9), 1116–1127.
18 G. C. Koh, P. Porras, B. Aranda, H. Hermjakob and S. E. Orchard, Analyzing protein–protein interaction networks, *J. Proteome Res.*, 2012, **11**(4), 2014–3201.
19 C. K. Fagerquist and O. Sultan, Induction and identification of disulfide-intact and disulfide-reduced β-subunit of shiga toxin 2 from *Escherichia coli* O157:H7 using MALDI-TOF-TOF-MS/MS and top-down proteomics, *Analyst*, 2011, **136**(8), 1739–1746.
20 N. Mischerikow, P. van Nierop, K. W. Li, H. G. Bernstein, A. B. Smit, A. J. Heck and A. F. Altelaar, Gaining efficiency by parallel quantification and identification of iTRAQ-labeled peptides using HCD and decision tree guided CID/ETD on an LTQ Orbitrap, *Analyst*, 2010, **135**(10), 2643–2652.
21 A. I. Lamond, M. Uhlen, S. Horning, A. Makarov, C. V. Robinson, L. Serrano, F. U. Hartl, W. Baumeister, A. K. Werenskiold, J. S. Andersen, O. Vorm, M. Linial, R. Aebersold and M. Mann, Advancing cell biology through proteomics in space and time (PROSPECTS), *Mol. Cell. Proteomics*, 2012, **11**(3), O112.017731.
22 B. Kholodenko, M. B. Yaffe and W. Kolch, Computational approaches for analyzing information flow in biological networks, *Sci. Signaling*, 2012, **5**(220), re1.
23 H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J. F. Rual, A. Dricot, A. Vazquez, R. R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A. S. de Smet, A. Motyl, M. E. Hudson, J. Park, X. Xin, M. E. Cusick, T. Moore, C. Boone, M. Snyder, F. P. Roth, A. L. Barabási, J. Tavernier, D. E. Hill and M. Vidal, High-quality binary protein interaction map of the yeast interactome network, *Science*, 2008, **322**(5898), 104–110.
24 F. Zhang and J. Y. Chen, Discovery of pathway biomarkers from coupled proteomics and systems biology methods, *BMC Genomics*, 2010, **11**(Suppl 2), S12.
25 C. E. Parker, T. W. Pearson, N. L. Anderson and C. H. Borchers, Mass-spectrometry-based clinical proteomics–a review and prospective, *Analyst*, 2010, **135**(8), 1830–1838.
26 E. Moreno-Gordaliza, B. Cañas, M. A. Palacios and M. M. Gómez-Gómez, Novel insights into the bottom-up mass spectrometry proteomics approach for the characterization of Pt-binding proteins: the insulin-cisplatin case study, *Analyst*, 2010, **135**(6), 1288–1298.
27 Z. J. Li, N. Liu, Q. D. Han and Y. Y. Zhang, Proteome reference map and regulation network of neonatal rat cardiomyocyte, *Acta Pharmacol.Sin.*, 2011, **32**(9), 1116–1127.
28 B. Shrestha, P. Nemes, J. Nazarian, Y. Hathout, E. P. Hoffman and A. Vertes, Direct analysis of lipids and small metabolites in mouse brain tissue by AP IR-MALDI and reactive LAESI mass spectrometry, *Analyst*, 2010, **135**(4), 751–758.
29 G. R. Gourley, L. Yang, L. Higgins, M. A. Riviere and L. L. David, Proteomic analysis of biopsied human colonic mucosa, *J. Pediatr. Gastroenterol. Nutr.*, 2010, **51**(1), 46–54.
30 B. Kholodenko, M. B. Yaffe and W. Kolch, Computational approaches for analyzing information flow in biological networks, *Sci. Signaling*, 2012, **5**(220), re1.
31 C. Tang, J. Iwahara and G. M. Clore, Visualization of transient encounter complexes in protein–protein association, *Nature*, 2006, **444**(7117), 383–386.
32 E. A. Franzosa and Y. Xia, Structural principles within the human-virus protein–protein interaction network, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**(26), 10538–10543.
33 J. Zhao, C. Geng, L. Tao, D. Zhang, Y. Jiang, K. Tang, R. Zhu, H. Yu, W. Zhang, F. He, Y. Li and Z. Cao, Reconstruction and analysis of human liver-specific metabolic network based on CNHLPP data, *J. Proteome Res.*, 2010, **9**(4), 1648–1658.
34 Q. Wu, Q. Zhao, Z. Liang, Y. Qu, L. Zhang and Y. Zhang, NSI and NSMT: usages of MS/MS fragment ion intensity for sensitive differential proteome detection and accurate protein fold change calculation in relative label-free proteome quantification, *Analyst*, 2012, **137**(13), 3146–3153.